

برآورد تبخیر و تعرق ماهانه گیاه مرجع توسط مدل رگرسیون درختی در نواحی مختلف آب و هوایی ایران

هادی وروانی^{۱*}، محمد امین مرادی و جواد وروانی

دانشجوی دکتری دانشگاه رازی؛ h.varvani@gmail.com

دانشجوی دکتری دانشگاه قزوین؛ m.a.moradi@ut.ac.ir

استادیار دانشکده کشاورزی دانشگاه آزاد اسلامی اراک؛ varvani_55@yahoo.com

چکیده

تبخیر و تعرق یکی از اجزای اصلی چرخه هیدرولوژی و تخمین نیاز آبیاری است. در سال های اخیر استفاده از سیستم های هوش مصنوعی برای برآورد پدیده های هیدرولوژی افزایش چشمگیری داشته است. از این رو این تحقیق با هدف امکان سنجی استفاده از مدل رگرسیون درختی در برآورد تبخیر و تعرق ماهانه گیاه مرجع (ET_o) و مقایسه آن با روش پنمن مانیت فائو در سه ناحیه اقلیمی ایران انجام شد. از مزایای مدل رگرسیون درختی نسبت به دیگر مدل های هوش مصنوعی از قبیل شبکه های عصبی، عدم نیاز به فرایند وقت گیر آزمون و خطا و نیز ارائه نتایج به صورت روابط ریاضی است. در مطالعه ی حاضر، متوسط ماهانه ی داده های حداقل و حداکثر دما، رطوبت نسبی، سرعت باد و تابش خورشیدی به عنوان داده های ورودی به مدل در نظر گرفته شده است. نتایج حاصله نشان داد که ضرایب تعیین مدل های تولید شده توسط مدل رگرسیون درختی در برآورد پارامتر تبخیر و تعرق مرجع برای نواحی گرمسیر، معتدل و سردسیر به ترتیب برابر است با ۰/۷۸، ۰/۸۰ و ۰/۸۹ که حاکی از توانایی بالای مدل مذکور است.

واژه های کلیدی: هوش مصنوعی، پنمن مانیت فائو

مقدمه

گیاهان زراعی، محاسبه ی تبخیر و تعرق گیاه مرجع نظیر تبخیر و تعرق چمن و یا تبخیر و تعرق یونجه از یک سطح استاندارد و سپس استفاده از ضرایب گیاهی می باشد. تبخیر و تعرق می تواند به طور مستقیم توسط لایسیمتر یا روش بیلان آب اندازه گیری و یا با داده های هواشناسی تخمین زده شود.

تبخیر و تعرق^۲ یکی از مهمترین اجزای چرخه ی آبشناسی می باشد و تخمین دقیق آن برای بسیاری از مطالعات نظیر بیلان آبی، طراحی و مدیریت سیستم های آبیاری، شبیه سازی تولید گیاهی و برنامه ریزی و مدیریت منابع آب ضروری است.

یکی از متداول ترین روش های تخمین تبخیر و تعرق

^۱ نویسنده مسئول، آدرس: کرمانشاه، دانشگاه رازی، دانشکده کشاورزی و منابع طبیعی، گروه مهندسی آب

* دریافت: اسفند ۱۳۹۰ و پذیرش: بهمن ۱۳۹۲

روش هارگریوز را در تخمین تبخیر و تعرق گیاه مرجع در ۱۲ ایستگاه هواشناسی خوزستان مقایسه نمود. تبخیر و تعرق گیاه مرجع با استفاده از روش پنمن مانیتث محاسبه و مبنای مقایسه‌ی دو روش هارگریوز و شبکه‌ی عصبی قرار گرفت. در این تحقیق نتیجه گرفته شد که شبکه‌های عصبی در برآورد تبخیر و تعرق گیاه مرجع نسبت به روش هارگریوز از دقت بیشتری برخوردارند.

تصمیم‌گیری درختی از جمله مدل‌های سلسله مراتبی^۱ است. بریمان و همکاران (۱۹۸۴) نشان دادند که رگرسیون و طبقه‌بندی درختی ابزار مفیدی برای بدست آوردن روابط بین متغیرهای هدف و پیش‌بینی کننده می‌باشد. ژو و همکاران (۲۰۰۵) با استفاده از درخت تصمیم به طبقه‌بندی نرم داده‌های ماهواره‌ای و سنجش از دور پرداختند و به نتایج قابل قبولی نیز در این زمینه دست یافتند.

از مدل رگرسیون درختی برای پیش‌بینی جریان برگشتی در شبکه‌های آبیاری و زهکشی نیز استفاده شده است. نتایج حاصله از تجزیه و تحلیل آماری نشان داد که ضریب همبستگی برای مزارع تک کشت و دوکشت بالاست که نشان‌دهنده تطابق خوب بین داده‌های شبیه‌سازی شده توسط مدل و داده‌های اندازه‌گیری شده بود (ماهان و همکاران ۲۰۰۸).

پژوهشگران دریافته‌اند که در پیش‌بینی بالا آمدن امواج بر روی سازه‌های سنگ‌چین، مدل درختی بسیار دقیق‌تر از فرمول‌های تجربی و منطق فازی عمل می‌کند (بنکدار و همکاران ۲۰۱۰).

مدل‌های درختی و شبکه عصبی مصنوعی در مدل‌سازی روابط عمق-دبی را با روابط (منحنی‌های) سنتی مقایسه شدند که نتایج حاکی از برتری دو روش تصمیم‌گیری هوشمند درختی و شبکه عصبی بودند، با این تفاوت که در جریان‌های با دبی بالا، مدل درختی پیش-

با این وجود امکان استفاده از لایسمتر برای اندازه‌گیری تبخیر و تعرق همواره میسر نیست، زیرا این روش زمان-گیر و نیازمند عملیات و برنامه‌ریزی دقیق می‌باشد؛ از این رو روش‌های غیرمستقیم بر پایه‌ی داده‌های هواشناسی برای تخمین تبخیر و تعرق گیاه مرجع استفاده می‌شود. این روش‌ها شامل معادلات تجربی و یا روش‌های پیچیده‌ای مبتنی بر فرآیندهای فیزیکی می‌باشند. روش‌های ترکیبی تغییرات تبخیر را با شار تابش خالص و خصوصیات انتقال آئرودینامیک یک سطح طبیعی مرتبط می‌نمایند (آلن و همکاران ۱۹۹۸).

چیو و همکاران (۱۹۹۵) در کشور استرالیا، نشان دادند که روش پنمن مانیتث فائو یکی از دقیق‌ترین روش‌های تخمین تبخیر و تعرق می‌باشد و می‌توان از آن برای اقلیم‌های مختلف استفاده نمود.

تبخیر و تعرق فرایندی پیچیده و غیر خطی است زیرا به عوامل متعدد اقلیمی نظیر دما، رطوبت، سرعت باد، تابش، نوع و مرحله‌ی رشد گیاه و غیره وابسته است. در چند سال اخیر استفاده از روش‌های حل هوش مصنوعی از قبیل شبکه عصبی و درختان تصمیم‌گیری، به‌عنوان راه‌کاری دقیق و سریع در بسیاری از زمینه‌ها از جمله برآورد نیاز آبی گیاهان رایج شده است. ساختار درخت تصمیم در یادگیری ماشین، یک مدل پیش‌بینی کننده می‌باشد که حقایق مشاهده شده در مورد یک پدیده را به استنتاج‌هایی در مورد مقدار هدف آن پدیده ارتباط می‌دهد. تکنیک یادگیری ماشین برای استنتاج یک درخت تصمیم از داده‌ها، یادگیری درخت تصمیم نامیده می‌شود که یکی از رایج‌ترین روش‌های داده‌کاوی است.

زانتی و همکاران (۲۰۰۷) قابلیت شبکه‌های عصبی مصنوعی را برای تخمین تبخیر و تعرق گیاه مرجع در یکی از ایالات آرژانتین با استفاده از حداکثر و حداقل دمای هوا مورد بررسی قرار دادند. برای آموزش شبکه از یک سری داده‌های هواشناسی روزانه (از سپتامبر ۱۹۹۶ تا اوت ۲۰۰۲) منطقه‌ی مورد مطالعه استفاده شد. رحیمی-خوب (۲۰۰۸) عملکرد شبکه‌های عصبی مصنوعی و

معادله (۱)

$$ET_s = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}$$

که در آن،

ET_0 : تبخیر و تعرق گیاه مرجع (میلی‌متر بر روز)

R_n : تابش خالص ورودی به سطح گیاه (مگاژول بر

مترمربع بر روز)

G : شار گرمای خاک (مگاژول بر مترمربع بر روز)

T : میانگین روزانه‌ی دمای هوا در ارتفاع دو متری (درجه

سانتیگراد)

U_2 : میانگین روزانه‌ی سرعت باد در ارتفاع دو متری (متر

بر ثانیه)

e_s : فشار بخار اشباع (کیلو پاسکال)

e_a : فشار بخار واقعی (کیلو پاسکال)

$e_s - e_a$: کمبود فشار بخار اشباع (کیلو پاسکال)

Δ : شیب منحنی فشار بخار (کیلو پاسکال بر درجه‌ی

سانتیگراد) و

γ : ضریب ثابت سایکرومتری (کیلو پاسکال بر درجه‌ی

سانتیگراد) می‌باشند.

ساختار درخت تصمیم

یک درخت معمولاً از ریشه^۱، شاخه^۲، گره^۳ و برگ^۴

تشکیل شده است. درخت تصمیم هم به‌طور مشابه از گره

ها که با دایره نشان داده می‌شوند و شاخه‌ها که نشان‌دهنده

اتصال بین گره‌ها می‌باشند، تشکیل شده است (شکل ۱).

درخت تصمیم به‌منظور سادگی در رسم، معمولاً از بالا به

پایین کشیده می‌شود به‌طوری که ریشه (گره اول) را ریشه

بینی‌های بسیار دقیق‌تری ارائه داد (باچاریا و همکاران ۲۰۰۳).

در تحقیق حاضر مقادیر تبخیر و تعرق محاسبه شده

ماهانه به روش پنمن مانتیث فائو با استفاده از داده‌های

ماهانه مربوط به مراکز استانی ایران (که بر اساس تحلیل

عاملی به ۳ ناحیه اقلیمی سرد، معتدل و گرم تقسیم بندی

می‌شود) با بهره‌گیری از مدل رگرسیون درختی مدل-

سازی شده است.

مواد و روش‌ها

محاسبه ET_0

در مطالعه‌ی حاضر، داده‌های اقلیمی متوسط ماهانه‌ی

حداقل و حداکثر دما، میانگین رطوبت نسبی، سرعت باد

و ساعات آفتابی مربوط به کلیه ایستگاه‌های سینوپتیک

کشور که دارای داده‌های آماری بالای ۳۰ سال مربوط به

دوره‌ی آماری تا سال ۲۰۰۵ بودند برای محاسبه‌ی تبخیر و

تعرق متوسط ماهانه‌ی گیاه مرجع مورد استفاده قرار

گرفته است.

گروه‌بندی مراکز استانی بر اساس تحلیل خوشه‌ای

در تحقیق حاضر به کمک نرم‌افزار SPSS 16.0 و با

استفاده از ۹ پارامتر هواشناسی جدول ۱، بر اساس تحلیل

خوشه‌ای، ۳۰ استان مطالعاتی گروه‌بندی شدند. برای این

کار ابتدا بایستی داده‌ها استاندارد شوند (۶). لذا پس از

استاندارد کردن داده‌ها به روش Z-SCORE، گروه‌بندی

با روش طبقاتی تجمعی و محاسبه فاصله اقلیدسی به

روش ورد انجام شد که به تدریج استان‌های مشابه طی

فرایند جمع‌آوری در یک گروه با هم ادغام شدند.

معادله فائو پنمن مانتیث

در این تحقیق با استفاده از معادله فائو پنمن مانتیث

(معادله ۱) که یکی از رایج‌ترین معادلات ترکیبی است،

مقادیر تبخیر و تعرق پتانسیل (ET_0) یا تبخیر و تعرق

استاندارد محاسبه گردید. به علت عدم وجود داده‌های

دقیق لایسیمتری در طولانی مدت، و بر طبق پیشنهاد

نشریه ۵۶ فائو، از این روش به عنوان روش استاندارد

جهت ارزیابی سایر روش‌ها استفاده شده است.

1 . Root

2 . Branch

3 . Node

4 . Leaf

برخلاف دیگر روش‌های هوش مصنوعی از قبیل شبکه‌ی عصبی مصنوعی، برای آموزش به مدل نیازی به فرایند وقت‌گیر آزمون و خطا نیست. فرایند تشکیل دادن رگرسیون درختی شامل ۵ مرحله است:

مرحله آغازگری^۴ (مقدار دهی اولیه)

در این مرحله متغیر پیش‌بینی کننده یا به عبارتی دیگر گره والد انتخاب می‌شود، که این فرایند با استفاده از الگوریتم ID3، که بخشی از مدل رگرسیون درختی است، صورت می‌گیرد.

ساختن درخت^۵: این مرحله با تقسیم‌شدن گره والد به دو گره فرزند شروع می‌شود. در هر گره والد تمام موضوعات و انشعابات ممکن ارزیابی شده و سرانجام بهترین انشعاب، که در آن انحراف معیار بین داده‌ها به حداقل می‌رسد انتخاب می‌شود؛ که برای این منظور از فرمول زیر استفاده می‌شود:

که در معادله مذکور، **معادله (۲)**

$$\Delta(s, t) = DN(s, t) - [f_L DN(s_1, t_L) + f_R DN(s_2, t_R)]$$

$$DN(s, t) = \sum_{x_i, y_i \in t} (y_i(t) - \bar{y}(t))^2$$

$$\bar{y}(t) = \frac{1}{n(t)} \sum_{x_i, y_i \in t} y_i(t)$$

که در آن،

S نشان دهنده ی شماره انشعاب یک متغیر

t گره والد

DN(s,t) انحراف گره والد t

DN(S₁,t_L) و DN(S₂,t_R) به ترتیب انحراف گره

فرزند سمت چپ و راست می باشند

f_L و f_R نیز به ترتیب کسری از گره والد هستند که به-

ترتیب در گره های t_L و t_R شرکت کرده اند

می‌گویند) در بالا قرار گیرد. انتهای یک زنجیره ریشه-شاخه-گره را یک برگ می‌نامند. از هر یک از گره‌ها دو شاخه دیگر می‌توانند منشعب شوند. هر گره مربوط به یک خصوصیت معین (پارامتر ورودی مستقل) است و شاخه‌ها به معنای بازه‌ای از مقادیر هستند. عمل انشعاب توسط یکی از متغیرهای پیش‌بینی کننده انجام می‌پذیرد، و بازه‌های انشعاب طوری انتخاب می‌شوند که مجموع مجذور انحراف از میانگین داده‌های هر گره را حداقل کنند.

هنگامی که خروجی یک درخت، یک مجموعه گسسته از یک مجموعه مقادیر ممکن است؛ به آن طبقه-بندی درختی^۱ گفته می‌شود (مثلا مونث یا مذکر، برنده یا بازنده). هنگامی که بتوان خروجی درخت را یک عدد حقیقی در نظر گرفت آنرا، رگرسیون درختی^۲ می‌نامند و یا به عبارت دیگر اگر متغیرهای ما عددی^۳ باشند از رگرسیون درختی، واگر مطلق و قیاسی باشند از طبقه-بندی درختی استفاده می‌شود. فرایند انشعاب در هر گره بارها تکرار می‌شود تا به گره پایانی یا همان برگ برسد که در برگ مجموع مجذور انحراف از میانگین داده‌ها به حداقل می‌رسد، با این کار درخت بزرگی به وجود خواهد آمد.

مزایای مدل درختی

به دلایل زیر کار با مدل درختی پیشنهاد می‌شود: این مدل به‌طور مستقیم با متغیرهای پیش‌بینی کننده مرتبط می‌باشد بنابراین نتایج مدل برای فهمیدن و شبیه‌سازی آسان هستند.

درخت‌های تصمیم‌گیری غیر پارامتریک بوده و هیچ دخالتی از سوی کاربر بر روی آن‌ها صورت نمی‌گیرد. خروجی مدل از دقت بالایی برخوردار است که می‌توان آنرا با سایر مدل‌ها مقایسه کرد.

1 . Classification Tree

2 . Regression Tree

3 . Numerical

4 . Initialization

5 . Tree Building

توابع عملکرد

برای ارزیابی عملکرد مدل رگرسیون درختی در پیش بینی تبخیر و تعرق از پارامترهای جذر میانگین مربعات خطا (RMSE)، شاخص پراکندگی (SA) و پارامتر انحراف (Bias) استفاده می شود که از روابط زیر قابل محاسبه هستند.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}}$$

$$SI = \sqrt{\frac{\sum ((x_i - \bar{x}) - (y_i - \bar{y}))^2}{n}}$$

$$Bias = \bar{y} - \bar{x}$$

که در آن x_i مقدار تبخیر تعرق بدست آمده از روش پنمن مانیتث فائو، \bar{x} میانگین مقادیر واقعی، y_i مقدار تخمینی بدست آمده از مدل رگرسیون درختی، \bar{y} میانگین مقادیر تخمینی مدل و n نیز در معادله بالا تعداد مشاهدات می باشد. از ضریب نکویی برازش (R^2) نیز برای مقایسه داده های پیش بینی شده توسط مدل و مقادیر محاسبه شده توسط روش پنمن مانیتث فائو استفاده شده است.

داده های متوسط ماهانه دمای حداقل، دمای حداکثر، رطوبت نسبی، سرعت باد و ساعات آفتابی برای ایستگاه-های سینوپتیک مختلف با مراجعه به سازمان هواشناسی کل کشور تا سال ۲۰۰۵ بدست آمد. داده های ورودی به مدل عبارتند از:

متوسط حداقل دمای ماهیانه

متوسط حداکثر دمای ماهیانه

متوسط ماهیانه سرعت باد در ارتفاع ۲ متری از سطح

زمین

متوسط ماهیانه ساعات آفتابی؛ و خروجی که از مدل می-گیریم درحقیقت همان متوسط ماهیانه تبخیر و تعرق گیاه مرجع می باشد.

ابتدا با استفاده از نرم افزار XLMIner و روش تقسیم بندی استاندارد، به تقسیم بندی داده ها با اختصاص ۵۰ درصد از

Y_i مقدار جوابی است که به متغیر پیش بینی کننده X_i

نسبت داده می شود

و \bar{Y} میانگین حسابی جواب متغیرها در گره والد می باشد. بهترین انشعاب، انشعابی است که بیشترین مقدار $\Delta(S, t)$ را داشته باشد. با تکرار فرایند بالا برای هر انشعاب، درخت بزرگی شکل می گیرد که به درخت حداکثر^۱ معروف است که شاخه ها و گره های زیادی دارد و کار با آن سخت می باشد؛ بنابراین برای رسیدن به یک درخت بهینه و کارآمد، باید شاخه های اضافی را هرس کرد.

هرس کردن درخت^۲

دو روش هرس وجود دارد: الف- هرس قبل از شکل گیری درخت حداکثر^۳ ب- هرس بعد از شکل گیری درخت حداکثر^۴

در روش اول فرایند هرس اجازه نمی دهد شاخه های اضافی تولید شوند و درخت فقط تا جایی توسعه می یابد که یک میزان دقت خاصی که توسط کاربر تعیین می شود در نتایج مدل حاصل شود؛ ولی در روش دوم ابتدا درخت حداکثر تشکیل می شود و سپس فرایند هرس تا جایی که دقت مدل از مقدار تعیین شده توسط کاربر پایین تر نیاید انجام می گیرد.

انتخاب درخت بهینه^۵

درخت بهینه بر اساس حداقل کردن خطای پیش بینی انتخاب می شود که دو روش برای محاسبه خطای پیش بینی وجود دارد: الف- آزمون دستگاه مستقل ب- آزمون صحت سنجی روش اول هنگامی مورد استفاده قرار می گیرد که تعداد داده های زیادی داشته باشیم در غیر این صورت از روش دوم استفاده می شود.

1. Maximal Tree
2. Tree Pruning
3. Pre-Pruning
4. Past-Pruning
5. Optimal Tree Selection

نیمه خشک دارای ضریب تعیین برابر با ۰.۸۹۱ در مورد داده‌های آزمون می‌باشد که این مقدار در نواحی گرم و خشک برابر است با ۰.۷۸۶ که نشان می‌دهد این مدل در نواحی سردنیمه خشک دارای دقت بالاتری نسبت به نواحی گرم و خشک می‌باشد. منفی بودن پارامتر انحراف در مورد مناطق گرمسیر و سرد سیر در مورد داده‌های آزمون بیانگر این موضوع است که این مدل مقدار تبخیر و تعرق را دست پایین برآورد می‌کند و مثبت بودن این پارامتر در نواحی مرطوب نیز نشان‌دهنده‌ی این موضوع است که این مدل تبخیر تعرق را دست بالا برآورد می‌کند. شاخص پراکندگی در مورد داده‌های تست برای نواحی گرم و خشک از بقیه نواحی بیشتر بوده و نشان‌دهنده‌ی این موضوع است که پراکندگی داده‌های پیش‌بینی شده توسط مدل در این نواحی بیشتر است. اشکال ۳، ۴ و ۵ نمای کلی درخت بهینه، تولید شده توسط نرم افزار XLMIner، را در نواحی مختلف آب و هوایی نشان می‌دهند، درخت بهینه براساس حداقل کردن خطای پیش‌بینی با استفاده از آزمون صحت‌سنجی شکل می‌گیرد. اعداد نوشته شده بر روی شاخه‌ها نشان دهنده‌ی تعداد داده‌هایی است که در بازه آن شاخه قرار گرفته‌اند. قوانین حاکم بردخت تصمیم بهینه در نواحی گرم و خشک به صورت زیر می‌باشد:

اگر مقدار تابش کمتر از ۷.۰۸۳ باشد مقدار تبخیر و تعرق متوسط ماهیانه برابر است با ۱.۸۷ میلیمتر در روز.
اگر مقدار تابش مابین ۷.۰۸۳ و ۸.۸۷ باشد مقدار تبخیر و تعرق متوسط ماهیانه برابر است با ۳.۶۸۸ میلیمتر در روز.
اگر مقدار تابش بیش از ۸.۸۷ و سرعت باد نیز کمتر از ۱.۷۵ متر بر ثانیه باشد مقدار تبخیر و تعرق متوسط ماهیانه برابر است با ۵.۷۴ میلیمتر در روز.
اگر مقدار تابش بیش از ۸.۸۷ و سرعت باد نیز بیشتر از ۱.۷۵ متر بر ثانیه و رطوبت نسبی بیش از ۳۱.۵ درصد باشد مقدار تبخیر و تعرق متوسط ماهیانه برابر است با ۶.۷۵ میلیمتر در روز.

داده‌ها به داده‌های آموزشی، ۳۰ درصد به داده‌های صحت‌سنجی و ۲۰ درصد به داده‌های آزمون پرداخته شد.

نتایج و بحث

شکل ۱ نتایج دندروگرام گروه‌بندی مراکز استانی از تحلیل خوشه‌ای را نشان می‌دهد. مطابق با خط تقاطع دندروگرام، ۳۰ مرکز استانی به ۴ گروه طبقه‌بندی شدند. طبق جدول ۲ مشاهده می‌شود که روش به کار رفته در تحلیل خوشه‌ای توانسته است به خوبی ۳۰ مرکز استانی کشور را با توجه به ۹ پارامتر اقلیمی در ۴ گروه طبقه‌بندی کند. با توجه به آنچه که از آمار و اطلاعات هواشناسی کشور دیده می‌شود می‌توان پذیرفت که معیار تحلیل خوشه‌ای فوق از دقت قابل‌قبولی برخوردار است. این داده‌ها در سه مرحله - آموزشی^۱ - صحت‌سنجی^۲ (ارزیابی) - آزمون^۳ با داده‌های واقعی مورد مقایسه قرار گرفتند که کل نتایج در جدول زیر آمده است با توجه به اینکه در صورت قرار دادن استان‌های مرطوب کم بارش و پربارش در گروه‌های متفاوت تعداد داده‌هایی که در این دو دسته قرار می‌گرفتند بسیار کم بود و قابلیت کاربرد مدل درختی را پایین می‌آورد بنابراین نواحی اقلیمی مرطوب پر بارش و کم بارش را در یک دسته مورد تجزیه و تحلیل قرار دادیم:

همانطور که از جدول بالا نیز پیداست ضریب نکویی برازش (ضریب تعیین) R^2 در مورد داده‌های آموزشی بالا است که این به خاطر تعداد داده‌های زیادی است که در اجرای مدل در این مرحله استفاده شده است، پایین‌تر بودن جذر میانگین مربعات خطا نیز به همین صورت توجیه می‌شود. البته ملاک سنجش عملکرد مدل، ضرایب نکویی برازش، RMSE، شاخص پراکندگی و پارامتر انحراف برای داده‌های صحت‌سنجی و آزمون می‌باشد. مقادیر پیش‌بینی شده توسط مدل در نواحی سرد و

1. Training

2. Validation

3. Test

نتیجه گیری

در تحقیق حاضر برای نخستین بار مدل رگرسیون درختی در تخمین تبخیر و تعرق پتانسیل متوسط ماهیانه در نواحی مختلف آب و هوایی کشور مورد استفاده قرار گرفت. با محاسبه ET_0 به روش پنمن مانیتث فائو، که تنها روش مورد تایید فائو در غیاب داده‌های لایسیمتری می‌باشد، و در نظر گرفتن این مقادیر محاسبه شده به عنوان مقادیر هدف در آموزش مدل، مقادیر ضریب نکویی برآزش، جذر میانگین مربعات خطا، شاخص پراکندگی و پارامتر انحراف برای عملکرد مدل رگرسیون درختی در این نواحی محاسبه شد. نتایج حاکی از آن بود که مدل رگرسیون درختی مقدار تبخیر تعرق متوسط ماهیانه را در نواحی گرم و خشک و سرد و نیمه خشک دست پایین و در نواحی مرطوب دست بالا برآورد می‌کند. با بررسی-های صورت گرفته مشخص شد که با استفاده از مدل رگرسیون درختی مقادیر ET_0 ماهیانه را با استفاده از ۴ پارامتر متوسط حداقل دمای ماهیانه متوسط حداکثر دمای ماهیانه، متوسط ماهیانه سرعت باد در ارتفاع ۲ متری، و متوسط ماهیانه ساعات آفتابی می‌توان با دقت قابل قبولی تخمین زد. علاوه بر این دقت مدل در نواحی سرد و نیمه خشک بیشتر از نواحی گرم و خشک می‌باشد.

اگر مقدار تابش بیش از ۸.۸۷ و سرعت باد نیز بیشتر از ۱.۷۵ متر بر ثانیه و رطوبت نسبی کمتر از ۳۱.۵ درصد و متوسط دمای حداقل ماهیانه نیز کمتر از ۲۳.۶۵ درجه سانتیگراد باشد مقدار متوسط تبخیر و تعرق ماهیانه برابر است با ۷.۳۹ میلیمتر در روز.

اگر مقدار تابش بیش از ۸.۸۷ و سرعت باد نیز بیشتر از ۱.۷۵ متر بر ثانیه و رطوبت نسبی کمتر از ۳۱.۵ درصد و متوسط دمای حداقل ماهیانه نیز بیشتر از ۲۳.۶۵ درجه سانتیگراد باشد مقدار متوسط تبخیر و تعرق ماهیانه برابر است با ۹.۲۹ میلیمتر در روز. نمودارهای ۶، ۷ و ۸ مقادیر تبخیر و تعرق واقعی را براساس تبخیر و تعرق پیش‌بینی شده نشان می‌دهند که در مورد داده‌های آزمون رسم شده-اند. همان طور که از نمودارهای زیر نیز پیداست مقادیر پیش‌بینی شده توسط مدل رگرسیون درختی در نواحی سرد تطابق خوبی با داده‌های محاسبه شده دارد. و کمترین تطابق نیز مربوط به داده‌های نواحی گرم و خشک می‌باشد

جدول ۱. پارامترهای مورد بررسی ایستگاه‌های سینوپتیک در مقیاس میانگین سالانه

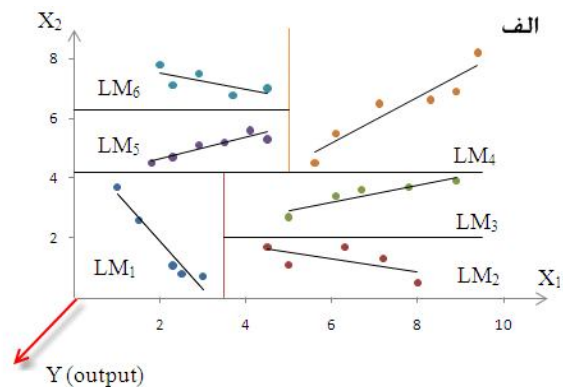
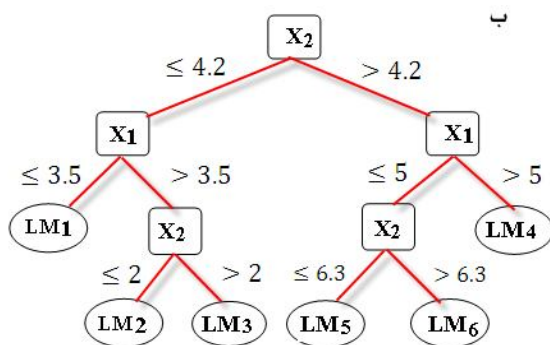
کد	مرکز استان	حداقل دما (C°)	حداکثر دما (C°)	دمای متوسط (C°)	دمای نقطه شبنم (C°)	درصد رطوبت نسبی	بارش (mm)	سرعت باد (m/s)	تعداد ساعات آفتابی	تبخیر و تعرق (mm/day)
۱	ارومیه	۵/۴	۱۷/۶	۱۱/۵	۲/۷	۶۰	۳۴۱	۲/۷	۲۸۲۹	۲/۸۳
۲	تبریز	۶/۹	۱۸	۱۲/۵	۱/۴	۵۴	۲۸۹	۶	۲۷۹۴	۳/۶۴
۳	اردبیل	۲/۸	۱۵/۳	۹	۳/۳	۷۱	۳۰۴	۷/۵	۲۴۵۴	۳/۲۱
۴	رشت	۱۱/۳	۲۰/۶	۱۵/۹	۱۲/۶	۸۱	۱۳۵۹	۲/۵	۱۶۲۰	۲/۲۵
۵	ساری	۱۳/۵	۲۲/۲	۱۷/۹	۱۳/۱	۷۸	۷۸۹	۳/۸	۱۹۸۷	۲/۴۳
۶	گرگان	۱۲/۷	۲۲/۸	۱۷/۸	۱۱/۴	۷۰	۶۰۱	۲/۱	۲۱۰۷	۲/۶۴
۷	بجنورد	۶/۷	۱۹/۷	۱۳/۸	۳/۲	۵۸	۲۵۷	۴/۶	۲۷۱۰	۳/۲۴
۸	مشهد	۷/۱	۲۱/۱	۱۴/۱	۳/۱	۵۵	۲۵۵	۴/۱	۲۸۹۲	۳/۴۶
۹	سنندج	۵/۵	۲۱/۴	۱۳/۴	-/۳	۴۷	۴۵۸	۳/۹	۲۸۷۳	۳/۵۲
۱۰	زنجان	۴	۱۸	۱۱	-/۸	۵۴	۳۱۳	۳/۷	۲۸۴۳	۳/۰۶
۱۱	قزوین	۶/۹	۲۱/۲	۱۴	۲/۳	۵۱	۳۱۶	۳/۹	۲۹۵۵	۳/۵
۱۲	تهران	۱۱/۹	۲۲/۷	۱۷/۳	۱/۵	۴۱	۲۳۲	۵/۲	۳۰۳۰	۴/۲۵
۱۳	سمنان	۱۲/۴	۲۳/۸	۱۸/۱	۳	۴۱	۱۴۰	۲/۷	۳۰۲۶	۳/۶۴
۱۴	کرمانشاه	۵/۹	۲۲/۶	۱۴/۳	-/۷	۴۷	۴۴۵	۴/۹	۲۹۰۷	۳/۸۶
۱۵	همدان	۳/۳	۱۹/۲	۱۱/۳	-/۸	۵۳	۳۱۷	۳/۲	۲۹۴۵	۳/۵۱
۱۶	اراک	۷	۲۰/۷	۱۳/۷	.	۴۶	۳۴۱	۲/۹	۲۹۷۳	۳/۳۳
۱۷	قم	۱۰/۲	۲۵/۹	۱۸	۱/۹	۴۱	۱۵۱	۳/۹	۳۱۵۲	۴/۲۱
۱۸	ایلام	۱۱/۵	۲۲/۳	۱۶/۹	-/۳	۴۰	۶۱۶	۴/۲	۳۱۱۴	۳/۸۳
۱۹	خرم آباد	۹/۱	۲۵/۲	۱۷/۲	۳/۲	۴۶	۵۰۹	۳/۲	۲۹۷۶	۳/۷۱
۲۰	اصفهان	۹/۱	۲۳/۴	۱۶/۲	-/۱	۴۰	۱۲۲	۴	۳۲۷۴	۴/۰۲
۲۱	یزد	۱۱/۷	۲۶/۵	۱۹/۱	-/۸	۳۱	۶۰	۵/۱	۳۲۲۳	۴/۸۴
۲۲	بیرجند	۸/۴	۲۴/۵	۱۶/۵	-/۸	۳۶	۱۷۰	۵/۱	۳۲۲۵	۴/۵۶
۲۳	اهواز	۱۷/۶	۳۲/۹	۲۵/۳	۹/۷	۴۳	۲۱۳	۵	۳۰۴۹	۵/۵
۲۴	شهرکرد	۳/۴	۲۰/۲	۱۱/۸	-/۶	۴۶	۳۲۷	۲/۳	۳۱۶۰	۳/۱۳
۲۵	یاسوج	۸	۲۲/۳	۱۵/۲	-/۲	۴۴	۸۶۵	۲/۶	۳۱۹۰	۳/۳۵
۲۶	بوشهر	۱۹/۵	۲۹/۷	۲۴/۶	۱۶/۹	۶۵	۲۷۹	۵/۹	۳۱۱۰	۴/۶۷
۲۷	شیراز	۹/۸	۲۵/۶	۱۷/۷	۱/۸	۴۱	۳۴۶	۴/۵	۳۳۵۸	۴/۴۴
۲۸	کرمان	۶/۹	۲۴/۷	۱۵/۸	-/۶	۳۲	۱۵۲	۶/۱	۳۱۷۴	۴/۸۵
۲۹	زاهدان	۱۰/۱	۲۶/۸	۱۸/۴	-/۵	۳۳	۹۱	۶/۵	۳۲۵۸	۴/۸۵
۳۰	بندرعباس	۲۱/۸	۳۲/۱	۲۷	۱۹/۳	۶۵	۱۸۲	۵/۶	۳۱۰۱	۴/۸۶

جدول ۲. گروه‌بندی اقلیمی مراکز استانی با استفاده از تحلیل خوشه‌ای

گروه	مرکز استان	اقلیم
۱	بیرجند، کرمان، یزد، زاهدان، قم، اصفهان، شیراز، تهران، سمنان	گرم و خشک
۲	مشهد، قزوین، بجنورد، سنندج، کرمانشاه، تبریز، اراک، شهرکرد، زنجان، همدان، ارومیه، ایلام، خرم‌آباد، یاسوج، اردبیل	سرد و نیمه‌خشک و نیمه مرطوب
۳	بوشهر، بندرعباس، اهواز	مرطوب کم بارش
۴	رشت، ساری، گرگان	مرطوب پر بارش

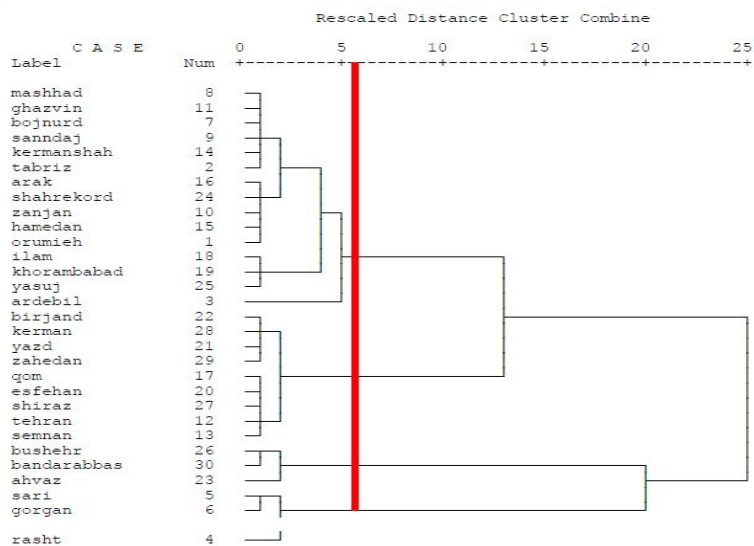
جدول ۳. نتایج کلی مدل

نوع داده	مرحله	شاخص پراکندگی	RMSE	پارامتر انحراف	R^2
گرم و خشک	آموزش	۰.۱۸	۰/۹۷۰	۰	۰/۸۵۶
	صحت	۰.۲۹	۱/۳۶۸	۰.۰۸۸	۰/۵۸۷
	سنجی	۰.۳۰۷	۱/۵۹۹	-۰.۳۱	۰/۷۸۶
معتدل	آموزش	۰.۱۴	۰/۶۱۶	۰	۰/۹۱۹
	صحت	۰.۱۸	۰/۸۵۹	-۰.۱	۰/۸۳۰
	سنجی	۰.۲۱	۰/۹۹۲	۰.۱۳	۰/۸
سرد و نیمه خشک	آموزش	۰.۱	۰/۵۳۲	۰	۰/۹۵
	صحت	۰.۱۵	۰/۸۲۷	-۰.۰۷	۰/۹۰۱
	سنجی	۰.۲۵	۰/۷۸۶	-۰.۰۹۸	۰/۸۹۱

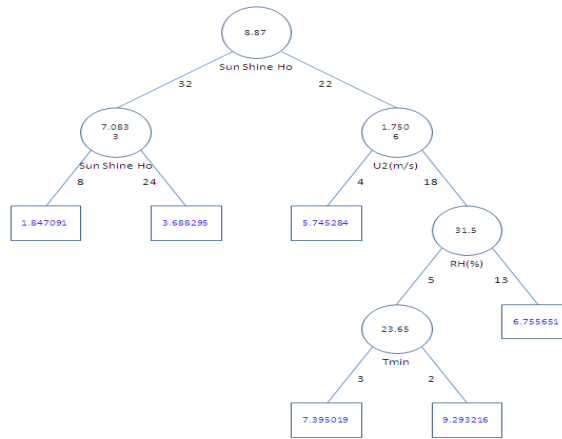


شکل ۱. شمای کلی شبکه درختی

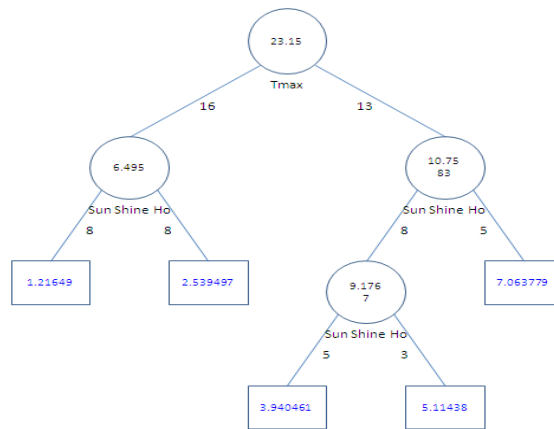
Dendrogram using Ward Method



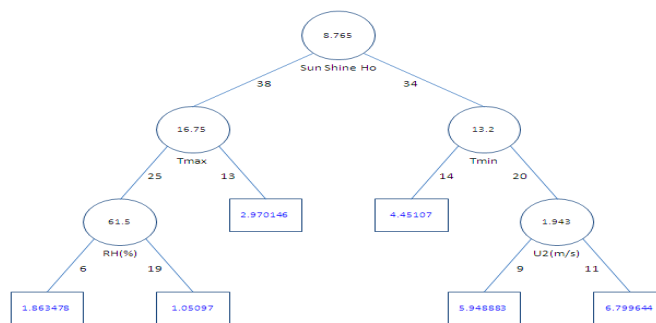
شکل ۲. دندروگرام مراکز استانی از نظر پارامترهای اقلیمی به روش ورد



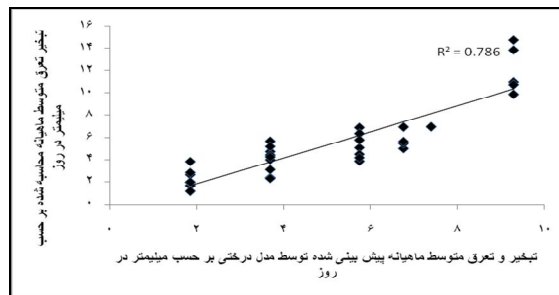
شکل ۳. درخت بهینه در نواحی گرم و خشک



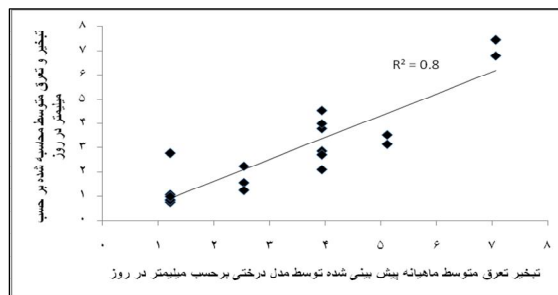
شکل ۴. درخت بهینه در نواحی معتدل



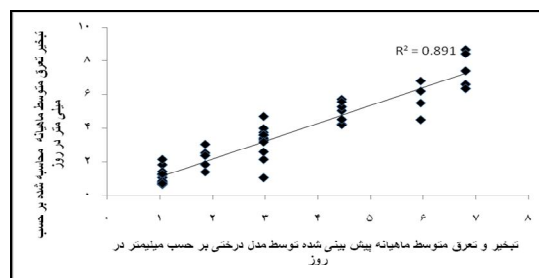
شکل ۵. درخت بهینه در نواحی سرد



شکل ۶. مقایسه داده های پیش بینی شده و داده های واقعی نواحی گرم و خشک



شکل ۷. مقایسه داده های پیش بینی شده و داده های واقعی نواحی مرطوب



شکل ۸. مقایسه داده های پیش بینی شده و داده های واقعی نواحی سردنیمه خشک

فهرست منابع

۱. حیدری، ح.، علیجانی، ب.، ۱۳۷۸، طبقه‌بندی اقلیمی ایران با استفاده از تکنیک های آماری چند متغیره، مجله پژوهش های جغرافیایی شماره ۳۷ : ۱۴-۲۶
2. Allen, R. G., Pereira, L. S., Raes, D. and Smith, M. 1998. Crop evapotranspiration: Guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper, 56(1,2): 300.
3. Beriman, L., Friedman, J., Olshen, R. and Stone, Ch. 1984. Classification and regression trees. Chapman & Hall/CRC.
4. Bhattacharya, B. and Solomatine, D.P. 2003. Neural networks and M5 model trees in modeling water level-discharge relationship for an Indian river. In European Symposium on Artificial Neural Networks, Bruges (Belgium), 23-25 April; 407-412.
5. Bonakdar, L. and Etemad-Shahidi, A. 2010. Predicting Wave run-up on Rubble-Mound Structures using M5 Model Tree. Journal of Ocean Engineering, 10(16): 57-69.
6. Chiew, FHS. , Kamaladassa, NN., Malano, HM. and MacMahon, TA. 1995. Penman-Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. Agricultural Water Management, 28: 9-21.
7. Mohan, S. and Vijayalakshmi, D.P. 2009. Prediction of irrigation return flows through a hierarchical modeling approach, Agricultural Water Management, 96(2): 233-246

8. Mokhtari, M. H., Ibrahim B., Mokhtari H., Zahedi G., Sheikhattar L., Movahed M. A., 2013. Neural network and multiple linear regression for estimating surface albedo from ASTER visible and near-infrared spectral bands. *Earth Interactions*, 17: 1–20.
9. Rahimikhoob, A., 2008. Comparative Study of Hargreaves's and Artificial Neural Network's Methodologies in Estimating Reference Evapotranspiration in a Semiarid Environment. *Agricultural Water Management*, 26:253-259
10. Xu, M., Pakorn, W., Pramod, K.V. and Manoj K.A. 2005. Decision tree regression for soft classification of remote sensing data, *Remote Sensing of Environment*, 97(3): 322-336
11. Zanetti, SS., Sousa, EF., Oliveira, VPS., Almeida, FT. and Bernardo, S. 2007. Estimating evapotranspiration using artificial neural network and minimum climatological data. *ASCE Journal of Irrigation and Drainage Engineering*, 133(2): 83-8

