

تخمین دمای سطحی و عمقی خاک از داده‌های هواشناسی با استفاده از تکنیک‌های یادگیری ماشین در اقلیم فرا خشک

ابوالفضل بامری^۱، مهسا خالقی^{۲*}

(تاریخ دریافت: ۱۳۹۹/۱۱/۰۶ تاریخ پذیرش: ۱۴۰۰/۰۱/۲۸)

چکیده

برآورد دقیق دما در اعماق مختلف خاک در اندرکنش زمین و جو بسیار مهم است. در این مطالعه کاربرد شش مدل مختلف یادگیری ماشین شامل شبکه عصبی مصنوعی (ANN)، درخت تصمیم‌گیری (DT)، کیوبست (CB)، جنگل تصادفی (RF)، ماشین بردار پشتیبان (SVM) و رگرسیون خطی (LR) برای مدل‌سازی روزانه دمای خاک در شش عمق مختلف ۵، ۱۰، ۲۰، ۳۰، ۵۰ و ۱۰۰ سانتی متر در کرمان مورد بررسی قرار گرفت. مجموعه‌ای از داده‌های هواشناسی سهل الوصول شامل دمای حداکثر و حداقل، رطوبت نسبی، نقطه شبنم، تبخیر-تعرق و فشار جو به عنوان ورودی مدل‌ها استفاده شد. آنالیز درجه اهمیت و همبستگی برای متغیرهای ورودی بر اساس اطلاعات دوره آماری ۱۸ ساله انجام شد. با توجه به نتایج، عملکرد هر شش مدل بر اساس معیارهای ارزیابی ($R^2 > 0/86$ ، $RMSE > 2/8$ درجه سانتیگراد و $Bias > 0/14$ درجه سانتیگراد) در همه اعماق قابل قبول بود. با این حال، RF، ANN و SVM کارایی بسیار بالایی در تخمین دمای خاک ($R^2 > 0/97$) از خود نشان دادند. همچنین مدل DT و پس از آن LR عملکرد ضعیف‌تری نسبت به بقیه داشتند. بررسی درجه اهمیت متغیرها نشان داد که از بین پارامترهای ورودی، دمای حداکثر و حداقل دارای بیشترین تاثیر در پیش‌بینی دمای خاک در همه مدل‌ها داشت. در نهایت می‌توان با اطمینان ادعان داشت که مدل‌های یادگیری ماشین نظیر جنگل تصادفی، شبکه عصبی مصنوعی و بردار پشتیبان قابلیت تخمین دمای خاک سطحی و عمقی در اقلیم خشک را در شرایط نبود تجهیزات اندازه‌گیری دارند.

واژه‌های کلیدی: جنگل تصادفی، داده‌های اقلیمی، شبیه‌سازی دمای خاک، مدل‌های داده‌محور

بامری^۱، خالقی^۲ م. ۱۴۰۱. تخمین دمای سطحی و عمقی خاک از داده‌های هواشناسی با استفاده از تکنیک‌های یادگیری ماشین در اقلیم فرا خشک. تحقیقات کاربردی خاک. جلد ۱۰، شماره ۱. صفحه: ۵۴-۶۸.

۱- مربی گروه مهندسی علوم خاک، دانشکده آب و خاک، دانشگاه زابل

۲- دانش آموخته دکتری، گروه مهندسی آب، دانشکده آب و خاک، دانشگاه زابل

* پست الکترونیک: khaleghi.83@gmail.com

مقدمه

خروجی تاریخی که قبلاً مشاهده شده‌اند، ساختار سیستم را بیاموزند. سپس، آزمون بر روی سیستم آموزش دیده انجام می‌شود و میزان موفقیت مدل محاسبه می‌شود (Sattari *et al.*, 2020; Sihag *et al.*, 2020). مطالعه متعددی برای تخمین دمای خاک از پارامترهای هواشناسی از طریق روش‌های خطی یا غیرخطی، مدل‌های مبتنی بر یادگیری ماشین و سری زمانی انجام شده است (Mehdizadeh *et al.*, 2020; Kim & Singh 2014; Tabari *et al.*, 2011; Wu *et al.*, 2013). تکنیک‌های یادگیری ماشین نیز به طور گسترده‌ای برای تخمین داده‌های دمای خاک در سال‌های اخیر مورد توجه و استفاده قرار گرفته است. زادمهر و فرخیان (Zadmehr & Farrokhan, 2020) با استفاده از مدل‌های یادگیری ماشین سریع، شبکه عصبی مصنوعی و رگرسیون خطی چندگانه به برآورد دمای خاک از داده‌های هواشناسی پرداختند. آنها از داده‌های دمای هوا، سرعت باد، رطوبت نسبی هوا، فشار بخار، تبخیر و بارندگی به‌عنوان ورودی برای آموزش مدل‌ها استفاده کردند. نتایج این تحقیق نشان داد مدل‌های ماشین یادگیری حداکثر^۱ (ELM)، شبکه عصبی مصنوعی^۲ (ANN) و مدل‌های رگرسیون خطی چندگانه^۳ (MLR) عملکردی مطلوب در مدل‌سازی دمای خاک در تمام عمق‌ها داشتند. همچنین، تمامی مدل‌های مورد بررسی برآورد بهتری از دمای لایه سطحی خاک (عمق ۳۰ - ۰ سانتیمتری) داشته و با افزایش عمق دقت آن‌ها کاهش یافت. ستاری و همکاران (Sattari *et al.*, 2020) طی پژوهشی برآورد دمای خاک با پارامترهای هواشناسی با استفاده از سه مدل استخراج داده مبتنی بر درخت را در کشور ترکیه مورد مطالعه قرار دادند. با توجه به نتایج آن‌ها، هر سه روش درخت تصمیم‌گیری^۴ (DT)، درختان شیب دار تقویت شده^۵ (GBT) و ترکیب دو مدل مذکور (DT - GBT)، دمای خاک را در عمق ۵ سانتی‌متر با موفقیت تخمین زدند. با این حال، بهترین تخمین در عمق خاک ۱۰ و ۲۰ سانتی‌متر با مدل DT به دست آمد. میزان دقت مدل‌ها نیز با افزایش عمق خاک افزایش یافته است. همچنین در پیش‌بینی دمای خاک، مدت زمان تابش آفتاب و دمای هوا به عنوان مهم‌ترین عوامل تعیین‌کننده و بارش بی‌اهمیت‌ترین متغیر هواشناسی بود. فنگ

دمای خاک به عنوان یک عامل کلیدی و کنترل‌کننده در فرآیندها و خصوصیات فیزیکی، شیمیایی و بیولوژیکی خاک به شمار می‌آید که متعاقباً بر کمیت و کیفیت تولید محصولات کشاورزی تأثیرگذار است. دمای خاک، تبخیر-تعرق، جوانه‌زنی، تولید و تجزیه مواد مغذی، تهویه خاک، رشد گیاه، فعالیت ریشه و میکروارگانیسم‌های خاک را تحت تأثیر قرار می‌دهد. بسیاری از این فرآیندها در محدوده مشخصی از درجه حرارت به حداکثر مقدار خود رسیده و در درجه حرارت‌های بیشتر و کمتر از این محدوده، کاهش می‌یابند (Feng *et al.*, 2019; Zadmehr & Farrokhan, 2020). بنابراین، آگاهی در مورد دمای خاک و پیش‌بینی دمای خاک در اعماق مختلف به عنوان یک فاکتور اصلی تصمیم‌گیری برای بررسی تغییرات آینده در فرآیندهای مختلف نظیر هواشناسی سطحی و زیر سطحی، تبادل انرژی زمین-جو، سیستم‌های کشاورزی منعطف، طرح‌های بیابان‌زدایی، مدیریت منابع آب و خاک و ارزیابی ریسک زیست‌محیطی در نظر گرفته شده است (Samadianfard *et al.*, 2018; Bahmani *et al.*, 2018). علی‌رغم اهمیت و نیاز اساسی به داده‌های دمای خاک در زمینه‌های مختلف، دسترسی به آن داده‌ها در بسیاری از مناطق (به عنوان مثال کشورهای در حال توسعه) بسیار محدود است. مقادیر داده‌های دمای خاک باید توسط دماسنج‌هایی که در اعماق مختلف خاک نصب شده‌اند، اندازه‌گیری شود که این کار زمان‌بر و پرهزینه است (Mehdizadeh *et al.*, 2020). همچنین ایستگاه‌های هواشناسی و سینوپتیک کافی برای ثبت داده‌های دمای خاک در همه مناطق خصوصاً در مناطق کوهستانی و شیب‌های تند وجود ندارد. برای پر کردن این شکاف در داده‌ها، غالباً از داده‌های هواشناسی برای تخمین داده‌های دمای خاک استفاده می‌شود که ممکن است خیلی دقیق نباشد (Ozturk *et al.*, 2011). طی دهه‌های گذشته، بسیاری از محققان کارایی مدل‌های داده‌محور برای شبیه‌سازی روابط پیچیده ورودی و خروجی غیرخطی را گزارش کرده‌اند (Delbari *et al.*, 2019). مدل‌های مبتنی بر داده سعی می‌کنند با استفاده از داده‌های ورودی و

^۱Decision Tree
^۲Gradient Boosted Tree

Extreme Learning Machine
Artificial Neural Network
Multiple linear regression models

در اغلب ایستگاه‌های هواشناسی دسترسی به این امکانات ممکن نیست، به کارگیری روش‌های مدل‌سازی جایگزین برای تخمین دمای خاک راهکاری نو و سودمند خواهد بود. تاکنون از مدل‌های مختلف داده‌محور برای برآورد دمای نیمرخ خاک با استفاده از داده‌های هواشناسی استفاده شده است. هرچند در اکثر مطالعات صورت گرفته یک یا دو روش مختلف مورد بررسی قرار گرفته و کاربرد همزمان شش مدل یادگیری ماشین (خصوصاً تکنیک جنگل تصادفی و کیوبست مدل) و مقایسه عملکرد آن‌ها در پیش‌بینی دمای سطحی و عمقی خاک مورد پژوهش قرار نگرفته است. لذا هدف از این تحقیق بررسی قابلیت شش مدل مختلف یادگیری ماشین شامل رگرسیون خطی^۴ (LR)، جنگل تصادفی (RF)، درخت تصمیم‌گیری (DT)، شبکه عصبی مصنوعی (ANN)، ماشین بردار پشتیبان^۵ (SVM) و کیوبست (CB) در پیش‌بینی دمای سطحی و عمقی خاک (نیمرخ خاک) بر اساس داده‌های سهل‌الوصول هواشناسی و مقایسه عملکرد آن‌ها در اقلیم فرا خشک است.

مواد و روش‌ها

منطقه مورد مطالعه

استان کرمان در جنوب شرقی ایران با در بر گرفتن بیش از ۱۱ درصد از وسعت ایران پهناورترین استان با حدود ۱۸۳۱۹۳ کیلومترمربع است. این استان از نظر موقعیت جغرافیایی بین ۵۵ درجه و ۲۵ دقیقه تا ۳۲ درجه عرض شمالی و ۲۶ درجه و ۵۳ دقیقه تا ۲۹ درجه و ۵۹ دقیقه طول شرقی از نصف‌النهار گرینویچ قرار گرفته است. استان کرمان یکی از مرتفع‌ترین استان‌های کشور در ناحیه کویری و کم‌آب ایران است. به طور کلی ناهمواری‌های استان کرمان را از نظر جغرافیایی و ویژگی‌های خاص منطقه‌ای می‌توان به دو دسته کوهستان‌های مرتفع و سرزمین‌های نسبتاً هموار و پست تقسیم‌بندی نمود. بر اساس روش‌های طبقه‌بندی معمول مانند روش‌های طبقه‌بندی کوپن و دومارتن، اقلیم غالب در استان کرمان، اقلیم فرا خشک (hyper-arid) می‌باشد.

توسعه مدل‌های یادگیری ماشین

و همکاران (Feng et al., 2019) به منظور برآورد دمای خاک در عمق‌های مختلف با استفاده از داده‌های هواشناسی در منطقه لاس پلاتو در چین از مدل‌های یادگیری ماشین حداکثر (ELM)، شبکه‌های عصبی رگرسیون تعمیم‌یافته^۱ (GRNN)، شبکه‌های عصبی انتشار^۲ (BPNN) و جنگل‌های تصادفی^۳ (RF) استفاده کردند. آنها به این نتیجه رسیدند که مدل ELM دارای عملکردی بهتر و همچنین سرعت محاسبه بهتری نسبت به BPNN، GRNN و همچنین مدل‌های RF در بازه‌های زمانی ساعتی است، بنابراین مدل ELM را برای تخمین دمای خاک در اعماق مختلف توصیه نمودند. طبری و همکاران (Tabari et al., 2015) در پژوهشی از شبکه عصبی مصنوعی (ANN) برای پیش‌بینی دمای کوتاه‌مدت خاک با استفاده از دمای هوا در ۶ عمق مختلف در دو ایستگاه مرطوب و خشک واقع در ایران استفاده کردند. نتایج حاکی از توان مطلوب ANN برای پیش‌بینی دمای کوتاه‌مدت خاک است. ذونعمت کرمانی (Zounemat-Kermani, 2013) دمای خاک را با استفاده از شبکه‌های عصبی مصنوعی در بازه‌های زمانی روزانه و هفتگی تخمین زد. سه پارامتر هواشناسی (دمای هوا، تابش و رطوبت نسبی) و دو متغیر هیدرولوژیکی (بارش و جریان) به عنوان ورودی در نظر گرفته شدند. مشاهده شده است که شبکه‌های عصبی مصنوعی در برآورد دمای خاک نسبت به روش‌های رگرسیون خطی چندگانه موفق‌ترند.

در مناطق خشک فراهم کردن محیط مناسب جهت رشد گیاه از اهمیت بسیار زیادی برخوردار است. افزایش دمای خاک در مناطق خشک می‌تواند با افزایش میزان تبخیر و کاهش میزان رطوبت و درصد کربن آلی باعث کاهش کیفیت خاک و افزایش حساسیت خاک نسبت به تنش‌های محیطی همچون تنش شوری و آبی شود. از سویی دیگر پارامتر دمای خاک تاثیر زیادی بر رشد و فعالیت میکروارگانیسم‌های خاک داشته و از این طریق بر رشد گیاه تاثیرگذار می‌باشد. لذا دستیابی به اطلاعات دقیقی از دمای خاک در اعماق مختلف به مدیریت پایدار کشاورزی کمک شایانی خواهد نمود. اما چون کسب این اطلاعات نیازمند تجهیزات هزینه بر هواشناسی بوده که

^۴Linear Regression
^۵Support Vector Machine
^۶Cubist Model

^۱General regression neural network
^۲Back Propagation Neural Network
^۳Random Forest

متغیرهای مستقل، هنگامی که تعداد زیادی متغیر مستقل موجود باشد.

مدل ماشین بردار پشتیبان (SVM): ماشین بردار پشتیبان، تکنولوژی جدید یادگیری ماشین است که روش آماری غیرپارامتریک و نیمه نظارت شده می‌باشد (Rezaei *et al.*, 2018). این روش در سال ۱۹۹۲ میلادی توسط بوسر و همکاران بر پایه تئوری یادگیری آماری معرفی شد. آنها طی سال‌های بعد تئوری ابرصفحه بهینه را به عنوان طبقه‌بندی کننده خطی ارائه داده و طبقه‌بندی کننده‌های غیرخطی را با کمک توابع کرنل معرفی نمودند. اصول اساسی آن چه که امروز به عنوان ماشین بردار پشتیبان شناخته می‌شود، نتایج کارهای بوسر و همکاران است و در نهایت گسترش ماشین بردار پشتیبان بر اساس رگرسیون نیز در سال ۱۹۹۵ توسط وپنیک به نتیجه رسید (Samadianfard & Panahi, 2018). ویژگی اصلی این روش، توانایی بالا در استفاده از نمونه‌های آموزشی کمتر و در عین حال رسیدن به دقت بالاتر در مقایسه با سایر روش‌ها است (Rezaei *et al.*, 2018). در حقیقت مدل‌های ماشین‌های بردار پشتیبان به دو گروه عمده الف) مدل طبقه‌بندی ماشین بردار پشتیبان و ب) مدل رگرسیون بردار پشتیبان تقسیم‌بندی می‌شوند. از مدل طبقه‌بندی ماشین بردار پشتیبان جهت حل مسائل طبقه‌بندی داده‌هایی که در کلاس‌های مختلف قرار می‌گیرند، استفاده می‌گردد و مدل رگرسیون بردار پشتیبان در حل مسائل پیش‌بینی داده‌های عددی کاربرد دارد (Samadianfard & Panahi, 2018).

مدل درختان تصمیم‌گیری (DT): در مدل‌سازی پدیده‌ها، وجود شرایط محیطی ممکن است باعث شود تا استفاده از یک رابطه کلی نتایج خوبی را به همراه نداشته باشد. بر این اساس معمولاً برای حل یک مسئله پیچیده، آن را به چند مسئله کوچکتر و ساده‌تر تقسیم نموده و سپس جواب‌های بدست آمده را با هم ترکیب می‌کنند. همین ایده ساده در مدل‌های درخت تصمیم‌گیری مورد استفاده قرار می‌گیرد. درخت تصمیم یکی از روش‌های داده‌کاوی و ابزارهای قوی و متداول برای دسته‌بندی (classification tree) و پیش‌بینی (regression tree) می‌باشد، که برخلاف شبکه‌های عصبی به تولید قانون می‌پردازد. بدین منظور، فضا یا محدوده مقادیر داده‌های ورودی به چند زیربازه تقسیم و برای هر زیربازه یک الگوی

در این تحقیق، روش‌های یادگیری ماشین در محیط نرم افزار R اجرا شده است. R یک زبان برنامه‌نویسی و محیط نرم‌افزاری برای محاسبات آماری و تحلیل داده است که حاوی محدوده گسترده‌ای از تکنیک‌های آماری از جمله: مدل‌سازی خطی و غیرخطی، آزمون‌های کلاسیک آماری، تحلیل سری‌های زمانی، رده‌بندی، خوشه‌بندی و قابلیت‌های گرافیکی می‌باشد. در حقیقت، روش‌های یادگیری ماشین برای آموزش و پیش‌بینی روابط پیچیده بین دمای خاک و پارامترهای هواشناسی موثر در میزان دمای خاک استفاده شد. هدف از ماشین‌های یادگیری پیدا کردن یک عملکرد تقریباً مفید است که زمینه پیش‌بینی ارتباط بین متغیرهای ورودی و نتایج موردنظر را نشان دهد (Mirakzehi *et al.*, 2017).

مدل شبکه عصبی مصنوعی (ANN): معماری ANN از یک ساختار غیرخطی تشکیل شده است که توسط یک لایه ورودی، یک یا چند لایه پنهان و یک لایه خروجی تشکیل شده است. هر لایه تعداد سلول‌های عصبی متفاوتی دارد که با وزن مشخصی به هر نورون در لایه بعدی مرتبط است (Zhu *et al.*, 2018). تعداد نورون‌های موجود در لایه ورودی برابر با تعداد متغیرهای ورودی بوده، در حالی که از یک نورون (برابر با تعداد متغیرهای خروجی) در لایه خروجی استفاده شده است. تاکنون، تعداد لایه‌های پنهان و تعداد نورون‌های موجود در هر لایه پنهان هنوز با قواعد صریح مشخص نشده است، که معمولاً بسته به نظر طراح شبکه و از آزمایش‌های تجربی همراه با روش آزمایش و خطا بدست می‌آید. در تحقیق حاضر یک لایه پنهان منفرد اتخاذ شد، زیرا یک لایه پنهان به طور کلی برای مقابله با اکثر مشکلات کافی بوده و همزمان از شلوغ شدن بیش از حد اجتناب می‌کند.

مدل رگرسیون خطی (LR): مدل رگرسیون خطی ساده یا چندگانه مدلی است که رابطه بین یک متغیر پاسخ و یک متغیر پیش‌بینی کننده (رگرسیون ساده خطی) یا بیش از یک متغیر پیش‌بینی کننده (رگرسیون خطی چندگانه) را با استفاده از پارامترهای وارد شده بصورت خطی و تخمین زده شده با روش حداقل مربعات نشان می‌دهد (Noi *et al.*, 2017). عوامل مهم در ایجاد یک رابطه رگرسیونی عبارتند از: ۱) انتخاب مناسب‌ترین تعداد متغیر مستقل، ۲) پی بردن به نوع همبستگی بین متغیرهای مستقل و وابسته و ۳) انتخاب موثرترین

مربوطه برای محاسبه مقدار پیش‌بینی شده استفاده می‌شود (Zhou *et al.*, 2019; Noi *et al.*, 2017). محبوبیت مدل کیوبیست به دلیل توانایی آن در استخراج روابط غیرخطی در داده‌های مورد مطالعه است، اما مسئله پیش‌بینی‌های محدودی را که برای سایر مدل‌های درخت رگرسیون رخ می‌دهد ندارد (Malone, 2013). مدل کیوبیست از ترکیب خطی دو مدل استفاده می‌کند. برای تفسیر آسانتر مدل کیوبیست، می‌توان آن را به عنوان الگویی در نظر گرفت که شامل بسیاری از قوانین است، که در آن هر قانون به یک مدل خطی چند متغیره مربوطه پیوست می‌شود. در مواردی که شرایط این قانون مطابقت دارد، مدل مربوط به محاسبه یک مقدار پیش‌بینی شده مربوطه اختصاص می‌یابد. (Zhou *et al.*, 2019).

گردآوری داده‌ها، آمار توصیفی و اجرای مدل‌ها

اندازه‌گیری دمای خاک با استفاده از دماسنج‌های مخصوص خاک استاندارد و مورد تایید WMO در اعماق ۵، ۱۰، ۲۰، ۳۰، ۵۰ و ۱۰۰ سانتی‌متر در ایستگاه سینوپتیک شهر کرمان انجام شد (IRIMO 2007). داده‌های هواشناسی مورد استفاده برای مدل‌سازی میانگین دمای روزانه خاک شامل حداکثر و حداقل دمای روزانه هوا (Tmax و Tmin)، تبخیر-تعرق (ET)، تابش خورشید (Rs)، رطوبت نسبی (RH)، تشعشعات (RAD)، نقطه شبنم (DEW) و فشار جو کاهش یافته به سطح دریا (QFF) بود. داده‌های دمای خاک و سایر داده‌های هواشناسی به صورت مقادیر میانگین روزانه و برای بازه زمانی ۱۸ ساله (۱۳۷۲-۱۳۹۰) در نظر گرفته شد که تعداد کل داده‌ها برای هر پارامتر با حذف داده‌های پرت و در نظر نگرفتن داده‌های مفقود برابر با ۶۳۷۲ بوده است. برای هر گونه مطالعه یا محاسبه آماری لازم است که داده‌های آزمایشی و یا میدانی که به صورت توده‌ای از اعداد خام هستند به شکل خاصی منظم گردند. تنظیم داده‌های عددی در جدول‌ها و ترسیم نمودار آنها، از اولین مراحل تجزیه آماری می‌باشد و این داده‌ها زمانی که مرتب شوند حاوی اطلاعات مهم و مفیدی می‌باشند. وضعیت توزیع داده‌ها برای کارهای آماری و از جمله مدل‌سازی با تکنیک‌های داده‌محور از اهمیت قابل توجهی برخوردار می‌باشد. وضعیت توزیع داده‌ها توسط ویژگی‌هایی مورد

یادگیری، معادله یا مدل ساده استخراج می‌شود (Hasani *et al.*, 2018). این فرآیند به طور بازگشتی در هر زیر مجموعه مشتق شده تکرار می‌شود. هنگامی که زیر مجموعه در یک گره تمام مقادیر یکسان متغیر هدف را به دست می‌آورد، یا وقتی تقسیم دیگر نتایج پیش‌بینی‌ها را بهبود نمی‌بخشد، به نتیجه می‌رسد (Granata *et al.*, 2017). معیارهای متفاوتی جهت ایجاد شاخه و تولید درخت تصمیم وجود دارد، اما از آنجا که پژوهش حاضر به استفاده از درخت تصمیم رگرسیونی پرداخته، معیار مورد استفاده در این مدل که انحراف حداقل مربعات (Least Squares Deviation or LSD) نام دارد.

مدل جنگل تصادفی (RF): جنگل تصادفی نوعی الگوریتم مجموعه طبقه‌بندی شده است که توسط برایمن (۲۰۰۱) پیشنهاد شده است. جوهر الگوریتم RF بهبود الگوریتم درخت تصمیم است (Ma *et al.*, 2016). به عبارتی، روش جنگل تصادفی یک روش یادگیری مبتنی بر دسته‌ای از درخت‌های تصمیم است. جنگل تصادفی می‌تواند الگوهای پیچیده را یاد بگیرد و ارتباط غیرخطی بین متغیرهای توضیحی و متغیرهای وابسته را در نظر بگیرد (Norouzi *et al.*, 2017). این مدل می‌تواند تعداد زیادی از متغیرهای ورودی را کنترل کند. در ضمن، این مدل حاوی روش مناسبی برای تخمین داده‌های از دست رفته است و می‌تواند با داده‌های از دست رفته، دقت را حفظ کند (Ma *et al.*, 2016; Norouzi *et al.*, 2017). روش RF نسبت به روش‌های دیگر دارای مزایایی مانند دقت پیش‌بینی بالا، توانایی در یادگیری روابط غیرخطی، توانایی بالا در تعیین متغیرهای مهم در پیش‌بینی و ماهیت غیر پارامتری می‌باشد (Norouzi *et al.*, 2017).

مدل کیوبیست (CB): مدل کیوبیست بر اساس درخت مدل M5 کوئینلان پیشنهاد شد (Quinlan, 1993). در مدل رگرسیون کیوبیست مفهومی، درخت رشد می‌کند و برگ انتهایی نقطه حاوی یک مدل رگرسیون خطی برای پیش‌بینی است. روش کیوبیست مجموعه‌ای از قوانین "اگر- پس از- پس از" را ایجاد می‌کند. هر قانون دارای یک مدل خطی چند متغیره مرتبط است. تا زمانی که مجموعه متغیرها شرایط قاعده را برآورده کند، از مدل

حداکثر، رطوبت، تشعشعات و ... است، به طوری که تا حد امکان بتوان از فقدان تجهیزات مخصوص اندازه‌گیری دمای خاک در ایستگاه‌های هواشناسی صرف‌نظر کرد. برای این منظور در این تحقیق، از ۶ روش یادگیری ماشین برای پیش‌بینی دمای سطحی و عمقی خاک استفاده گردید. مدل‌های بکار برده شده شامل: رگرسیون خطی، ماشین بردار پشتیبان، مدل کیوبیست، شبکه عصبی مصنوعی، درخت تصمیم‌گیری و جنگل تصادفی می‌باشد. به منظور ارزیابی مدل‌های بکار برده شده برای شبیه‌سازی دمای نیم‌رخ خاک، از چند شاخص آماری از جمله خطای میانگین مربعات ریشه (RMSE)، ضریب تبیین (R^2) و اریب (Bias) استفاده شد که به ترتیب در روابط زیر تعریف شده است (Omidvar et al., 2014):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - P_i)^2}{n}} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (4)$$

$$Bias = \bar{P} - \bar{O} \quad (5)$$

که در فرمول‌های مذکور: n تعداد کل نقاط مورد بررسی، \bar{O} میانگین مقادیر مشاهده شده، \bar{P} میانگین مقادیر پیش‌بینی شده، O_i مقدار مشاهده شده در نقطه i ام و P_i مقدار پیش‌بینی شده با استفاده از مدل در این نقطه است. جهت آنالیز، مدل‌سازی و همچنین تهیه دیاگرام همبستگی در این پژوهش از نرم افزار Rstudio استفاده شد.

نتایج و بحث

توصیف آماری خصوصیات اقلیمی منطقه مورد مطالعه مربوط به دوره آماری مورد بررسی در جدول ۱ خلاصه شده است. براساس نتایج گزارش شده در جدول ۱، پارامتر فشار جو کاهش یافته به سطح دریا (QFF) دارای تغییرپذیری کم، پارامترهای دمای حداکثر، تشعشعات و دمای خاک در تمامی عمق‌ها دارای تغییرپذیری متوسط و پارامترهای دمای حداقل، رطوبت، تبخیر-تعرق و نقطه شبنم دارای تغییرپذیری زیاد هستند. دلبری و همکاران (Delbari et al., 2019) و ستاری و همکاران (Sattari et al., 2017) ضریب تغییرات بالا برای رطوبت، نقطه شبنم،

بررسی قرار می‌گیرند که شامل: میانگین، میانه، حداقل، حداکثر، ضریب تغییرات و انحراف معیار می‌باشند. ضریب تغییرات (پراکندگی) شاخصی است که برای اندازه‌گیری توزیع پراکندگی داده‌های آماری و همچنین مقایسه پراکندگی دو یا چند صفت (متغیر) به کار می‌رود. به جز ضریب تغییرات، شاخص‌های پراکندگی دیگر، به مقیاس اندازه‌گیری داده‌ها وابستگی دارند در نتیجه برای مقایسه معیار پراکندگی بین دو یا چند جامعه مناسب نیستند، ضریب تغییرات معیاری از تغییرپذیری نسبی است که از تقسیم انحراف معیار به میانگین طبق فرمول ذیل بدست می‌آید (Bameri et al., 2015):

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2} \quad (1)$$

$$CV = \frac{SD}{\bar{x}} \times 100 \quad (2)$$

در فرمول‌های ارائه شده، SD انحراف معیار، N تعداد کل نقاط مورد بررسی، X_i مقادیر نمونه در نقطه i ام، \bar{X} میانگین نمونه‌ها و CV ضریب تغییرات براساس درصد می‌باشد. طبق طبقه‌بندی Wilding و همکاران (۱۹۸۳) اگر ضریب تغییرات داده‌ها کمتر از ۱۵ درصد باشد، داده‌ها دارای تغییرپذیری کم هستند. اگر مقدار این ویژگی بین ۱۵ تا ۳۵ درصد باشد، کلاس تغییرپذیری متوسط و در صورتی که مقدار این آماره بیشتر از ۳۵ درصد باشد، حاکی از تغییرپذیری زیاد داده‌ها است.

در طی فرآیند اجرای مدل‌ها، کل داده‌ها به دو دسته تقسیم‌بندی شده‌اند: داده‌های مرحله آموزش (Training data) و داده‌های مرحله آزمایش (Testing data)، به طوریکه در این تحقیق ۸۰ درصد از کل داده‌ها به آموزش مدل اختصاص یافت و ۲۰ درصد باقیمانده به عنوان داده‌های آزمون به مدل معرفی گردیده است. انتخاب داده‌های آموزش و آزمون به صورت تصادفی و توسط نرم‌افزار انجام شده و بدیهی است که داده‌های یک مرحله در مرحله دیگر مورد استفاده قرار نگرفته است.

ارزیابی عملکرد مدل

در مطالعه حاضر، هدف اصلی پیش‌بینی دمای خاک در اعماق مختلف با استفاده از تعداد محدود پارامترهای ورودی سهل‌الوصول هواشناسی نظیر دمای حداقل و

تبخیر و دمای حداقل را در تحقیقات خود گزارش کرده-
اند. صمدیان فرد و همکاران، (Samadianfard *et al.*, 2018) نیز ضریب تغییرات کم برای فشار جو و ضریب تغییرات متوسط برای تشعشعات و دمای خاک در تمامی عمق‌ها بیان داشته‌اند، هرچند متغیر دمای خاک با افزایش عمق خاک تغییرپذیری کمتری از خود نشان داد، که همسو با نتایج تحقیق حاضر می‌باشد.

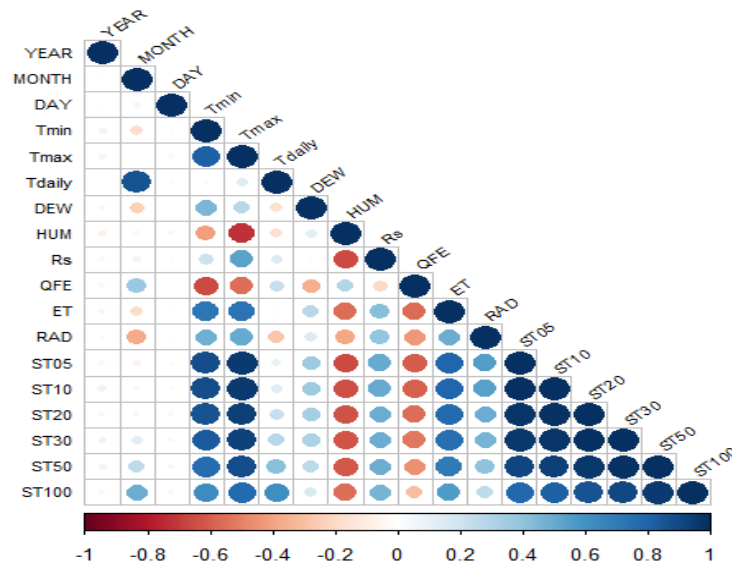
جدول ۱- توصیف آماری خصوصیات اقلیمی منطقه مورد مطالعه در دوره آماری ۱۸ ساله

Table 1. Statistical description of climatic characteristics of the study area in the statistical period of 18 years

Variables	Min	Max	Mean	Median	SD	Range	CV(%)
Tmin (°C)	-8.6	26.6	10.3	10.8	6.7	35.2	65.1
Tmax (°C)	3.0	42.0	28.8	30.4	7.0	39.0	24.4
Tdaily (°C)	-3.0	48.0	26.3	29.0	11.7	51.0	44.4
DEW (°C)	-21.9	14.6	-1.7	-1.3	5.0	36.4	301.0
HUM (%)	8.9	97.0	27.0	22.9	13.6	88.1	50.4
Rs (calcm ⁻²)	0.0	13.5	9.8	10.5	2.8	13.5	28.7
QFF (mbar)	814.6	835.5	824.2	824.1	3.5	20.9	0.4
ET (mm day ⁻¹)	0.0	70.4	9.2	9.7	4.1	70.4	44.6
RAD (W m ⁻²)	0.0	414.6	96.5	101.4	29.9	414.6	31.0
ST05 (°C)	2.7	44.9	25.8	27.8	8.6	42.3	33.4
ST10 (°C)	5.3	40.6	25.3	27.1	8.0	35.3	31.7
ST20 (°C)	6.2	35.5	23.9	25.7	7.0	29.3	29.2
ST30(°C)	7.3	43.8	24.0	25.8	6.7	36.4	28.0
ST50 (°C)	8.7	33.1	24.1	25.7	5.9	24.4	24.4
ST100 (°C)	11.0	30.7	23.4	24.7	4.9	19.7	21.1

بررسی همبستگی بین متغیرهای مورد بررسی نشان می‌دهد که بیشترین همبستگی بین دمای خاک با دمای حداقل و حداکثر و سپس تبخیر-تعرق و تشعشعات وجود دارد (شکل ۱). با توجه به شکل ۱، این همبستگی مثبت با افزایش عمق کاهش می‌یابد به طوری که بیشترین همبستگی دمای خاک در لایه سطحی مشاهده می‌شود. همبستگی بالا و رابطه مستقیم دمای حداکثر و حداقل با دمای خاک و نیز کاهش این همبستگی با افزایش عمق در تحقیقات پیشین نیز گزارش شده است (Delbari *et al.*, 2019).

بررسی همبستگی بین متغیرهای مورد بررسی نشان می‌دهد که بیشترین همبستگی بین دمای خاک با دمای حداقل و حداکثر و سپس تبخیر-تعرق و تشعشعات وجود دارد (شکل ۱). با توجه به شکل ۱، این همبستگی مثبت با افزایش عمق کاهش می‌یابد به طوری که بیشترین همبستگی دمای خاک در لایه سطحی مشاهده می‌شود. همبستگی بالا و رابطه مستقیم دمای حداکثر و حداقل با دمای خاک و نیز کاهش این همبستگی با افزایش عمق در تحقیقات پیشین نیز گزارش شده است (Delbari *et al.*, 2019).



شکل ۱- دیاگرام همبستگی بین متغیرهای اقلیمی و دمای خاک

Figure 1. Diagram of the correlation between climatic variables and soil temperature

جنگل تصادفی در همه اعماق مورد بررسی و بر اساس هر سه معیار ارزیابی بهترین عملکرد را از خود نشان داده و پس از آن مدل شبکه عصبی کارایی بالایی در تخمین دمای خاک داشت. همچنین بر اساس مقادیر معیارهای ارزیابی بدست آمده، روش درخت تصمیم‌گیری و پس از آن روش رگرسیون خطی عملکرد ضعیف‌تری را نسبت به چهار مدل دیگر از خود نشان دادند. اگرچه این دو مدل به نوبه خود پیش‌بینی‌های قابل‌قبولی از دمای خاک داشتند که در شرایط عدم امکان استفاده از دیگر مدل‌ها همچنان در تخمین دمای خاک می‌توانند کاربردی باشند. شکل (۲) نمودار یک به یک مقادیر پیش‌بینی شده و مقادیر واقعی دمای خاک را در سطح و عمق خاک (۵ و ۱۰۰ سانتی‌متر) را برای برترین مدل نشان می‌دهد. دلبری و همکاران (Delbari et al., 2019) نیز در تحقیق خود کارایی بهتر مدل بردار پشتیبان را نسبت به رگرسیون خطی در پیش‌بینی دمای خاک خصوصا در لایه‌های عمقی اذعان داشتند. ستاری و همکاران عملکرد بالای مدل درخت تصمیم‌گیری را در تخمین دمای خاک سطحی در مطالعه خود گزارش کردند.

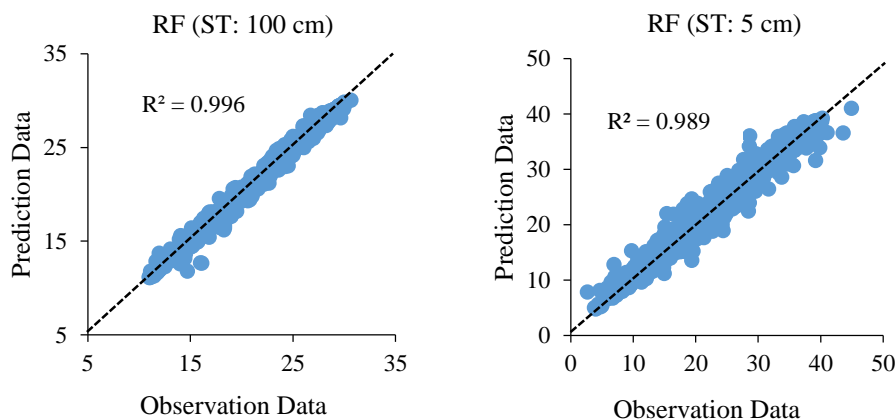
ارزیابی عملکرد مدل‌های یادگیری ماشین

ارزیابی آماری مدل‌های مورد مطالعه در جدول (۲ و ۳) نشان داده شده است. مقایسه الگوریتم‌های مختلف یادگیری ماشین برای همه اعماق خاک بیانگر قابلیت بالای روش‌های یادگیری ماشین در پیش‌بینی دمای خاک می‌باشد. هر چند بر اساس شاخص‌های آماری محاسبه شده برخی از روش‌ها قابلیت پیش‌بینی بالاتری از خود نشان دادند. مقادیر ضریب رگرسیون بالاتر از ۰/۹۷ هم در مرحله آموزش و تست و هم در کل، برای روش‌های جنگل تصادفی، شبکه عصبی مصنوعی، ماشین بردار پشتیبان و کیوبیست نشان‌دهنده کارایی بالای این روش‌ها در پیش‌بینی دمای خاک در اعماق سطحی صفر تا ۲۰ سانتی‌متر (جدول ۲) و نیز دمای عمقی خاک در اعماق ۳۰ تا ۱۰۰ سانتی‌متر (جدول ۳) می‌باشد. در حقیقت این ضرایب رگرسیون انطباق بسیار مطلوب مقادیر مشاهداتی و مقادیر پیش‌بینی شده توسط این مدل‌ها را بیان می‌کند. عملکرد قابل قبول این مدل‌ها در تحقیقات دیگر پژوهشگران نیز گزارش شده است (Fathololoumi et al., 2020; Delbari et al., 2019; Samadianfard et al., 2018; Sattari et al., 2017). در بین این مدل‌ها، مدل

جدول ۲- ارزیابی مدل‌های مختلف ماشین‌های یادگیری در پیش‌بینی دمای خاک سطحی (۰-۲۰ سانتی‌متر)

Table 2. Evaluation of different machine learning models in predicting surface soil temperature (0-20 cm)

Method	Data set	Bias (°C)			RMSE (°C)			R ²		
		5 cm	10 cm	20 cm	5 cm	10 cm	20 cm	5 cm	10 cm	20 cm
LR	train	0.000	0.000	0.000	1.817	1.689	1.675	0.956	0.956	0.941
	test	0.001	-0.113	0.086	1.772	1.619	1.765	0.956	0.959	0.939
	all	0.000	-0.034	0.026	1.803	1.669	1.703	0.956	0.957	0.940
ANN	train	0.000	0.000	0.000	1.244	1.062	0.964	0.980	0.983	0.981
	test	0.001	-0.041	0.000	1.273	1.054	1.036	0.977	0.982	0.979
	all	0.000	-0.012	0.000	1.253	1.060	0.986	0.979	0.983	0.980
DT	train	0.000	0.000	0.000	2.625	2.421	2.407	0.909	0.909	0.879
	test	0.062	-0.134	0.035	2.715	2.348	2.497	0.897	0.913	0.877
	all	0.019	-0.040	0.011	2.652	2.400	2.434	0.906	0.910	0.878
CB	train	0.015	0.023	0.002	1.318	1.079	1.002	0.977	0.982	0.979
	test	0.028	-0.021	0.012	1.361	1.091	1.079	0.974	0.981	0.977
	all	0.019	0.010	0.005	1.331	1.083	1.026	0.976	0.982	0.978
RF	train	-0.001	-0.001	-0.002	0.612	0.509	0.461	0.995	0.996	0.996
	test	-0.048	-0.054	0.001	1.394	1.156	1.050	0.973	0.979	0.978
	all	-0.015	-0.017	-0.001	0.919	0.763	0.693	0.989	0.991	0.990
SVM	train	0.010	-0.017	-0.005	1.189	0.909	0.850	0.981	0.987	0.985
	test	-0.033	-0.069	0.014	1.459	1.151	1.092	0.970	0.979	0.976
	all	-0.003	-0.033	0.001	1.276	0.988	0.929	0.978	0.985	0.982



شکل ۲- نمودار پراکنش دمای خاک پیش‌بینی شده توسط مدل RF در مقابل مقادیر واقعی دمای خاک در عمق‌های ۵ و ۱۰۰ سانتی-متر

Figure 2. Scatterplot of predicted soil temperature by RF model versus observation data of soil temperature in depth of 5 and 100 cm

(Sattari *et al.*, 2020). دقت مدل‌های مختلف یادگیری ماشین برای پیش‌بینی دمای خاک در تمامی اعماق بر اساس سه شاخص‌های آماری محاسبه شده، در شکل (۳)

علت تفاوت نتایج این تحقیق با نتایج حاضر را می‌توان تفاوت در اقلیم منطقه مورد بررسی و نیز تفاوت در متغیرهای هواشناسی به عنوان ورودی مدل دانست

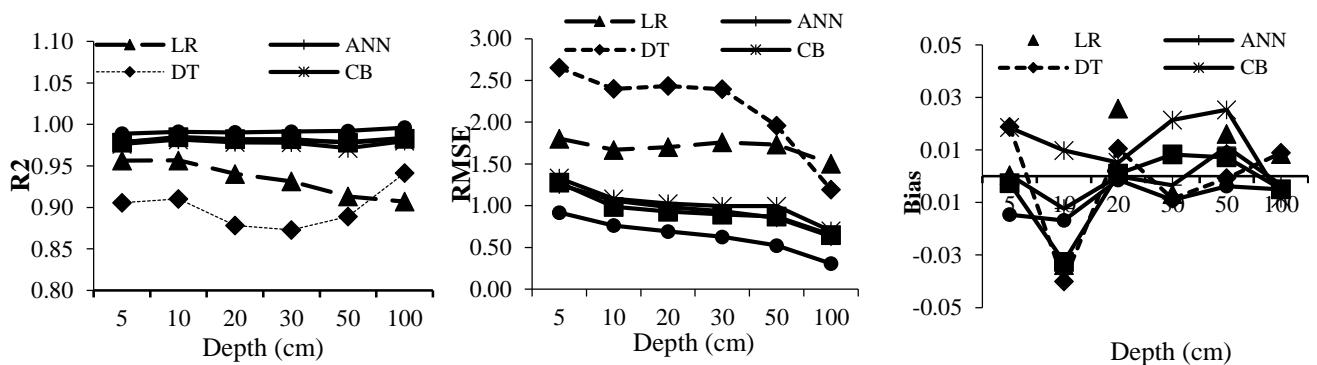
جدول ۳- ارزیابی مدل‌های مختلف ماشین‌های یادگیری در پیش‌بینی دمای خاک عمقی (۳۰-۱۰۰ سانتی‌متر)

Table 3. Evaluation of different machine learning models in predicting deep soil temperature (30-100 cm)

Method	Data set	Bias (°C)			RMSE (°C)			R ²		
		30 cm	50 cm	100 cm	30 cm	50 cm	100 cm	30 cm	50 cm	100 cm
LR	Train	0.000	0.000	0.000	1.733	1.748	1.490	0.934	0.912	0.909
	Test	-0.023	0.054	0.028	1.818	1.694	1.536	0.925	0.915	0.902
	All	-0.007	0.016	0.008	1.759	1.732	1.504	0.931	0.913	0.907
ANN	train	0.000	0.000	0.000	0.904	0.882	0.615	0.982	0.978	0.985
	test	-0.012	0.037	-0.011	0.999	0.788	0.665	0.977	0.982	0.982
	all	-0.003	0.011	-0.003	0.933	0.855	0.630	0.981	0.979	0.984
DT	train	0.000	0.000	0.000	2.384	1.972	1.186	0.875	0.889	0.942
	test	-0.029	-0.003	0.030	2.423	1.931	1.205	0.867	0.890	0.940
	all	-0.009	-0.001	0.009	2.396	1.960	1.192	0.873	0.889	0.942
CB	train	0.021	0.014	0.000	0.968	1.021	0.692	0.979	0.970	0.980
	test	0.022	0.052	-0.028	1.057	0.931	0.713	0.975	0.974	0.979
	all	0.021	0.025	-0.008	0.996	0.995	0.698	0.978	0.971	0.980
RF	train	-0.004	-0.005	-0.002	0.384	0.377	0.190	0.997	0.996	0.999
	test	-0.021	-0.001	-0.013	0.984	0.761	0.480	0.978	0.983	0.990
	all	-0.009	-0.004	-0.005	0.627	0.522	0.307	0.991	0.992	0.996
SVM	train	0.007	0.006	-0.004	0.806	0.825	0.595	0.986	0.981	0.986
	test	0.010	0.010	-0.007	1.075	0.970	0.763	0.974	0.972	0.976
	all	0.008	0.007	-0.005	0.895	0.871	0.650	0.982	0.978	0.983

بیش‌برآورد کردن دما توسط مدل‌ها می‌باشد. هر چند مقادیر این کم برآورد و یا بیش‌برآورد در اکثر مدل‌ها ناچیز و نزدیک به صفر بوده و موجب کاهش دقت پیش‌بینی نشده است. همچنین، همانطور که از شکل‌ها برمی‌آید برای همه‌ی لایه‌های عمقی خاک، مدل‌های رگرسیون خطی و درخت تصمیم ضعیف‌ترین عملکرد را نسبت به دیگر مدل‌های مورد ارزیابی داشتند.

به وضوح قابل مقایسه می‌باشد. علاوه بر شاخص آماری ضریب رگرسیون، نزدیک بودن مقادیر دو شاخص RMSE و Bias به صفر نیز بیان کننده عملکرد بهتر چهار مدل جنگل تصادفی، شبکه عصبی مصنوعی، ماشین بردار پشتیبان و کیوبیست در مقایسه با روش‌های رگرسیون خطی و درخت تصمیم می‌باشد. همچنین مقادیر منفی Bias نشان‌دهنده این می‌باشد که مدل‌ها تمایل به کم برآورد کردن دما داشته و مقادیر مثبت آن نشان‌دهنده

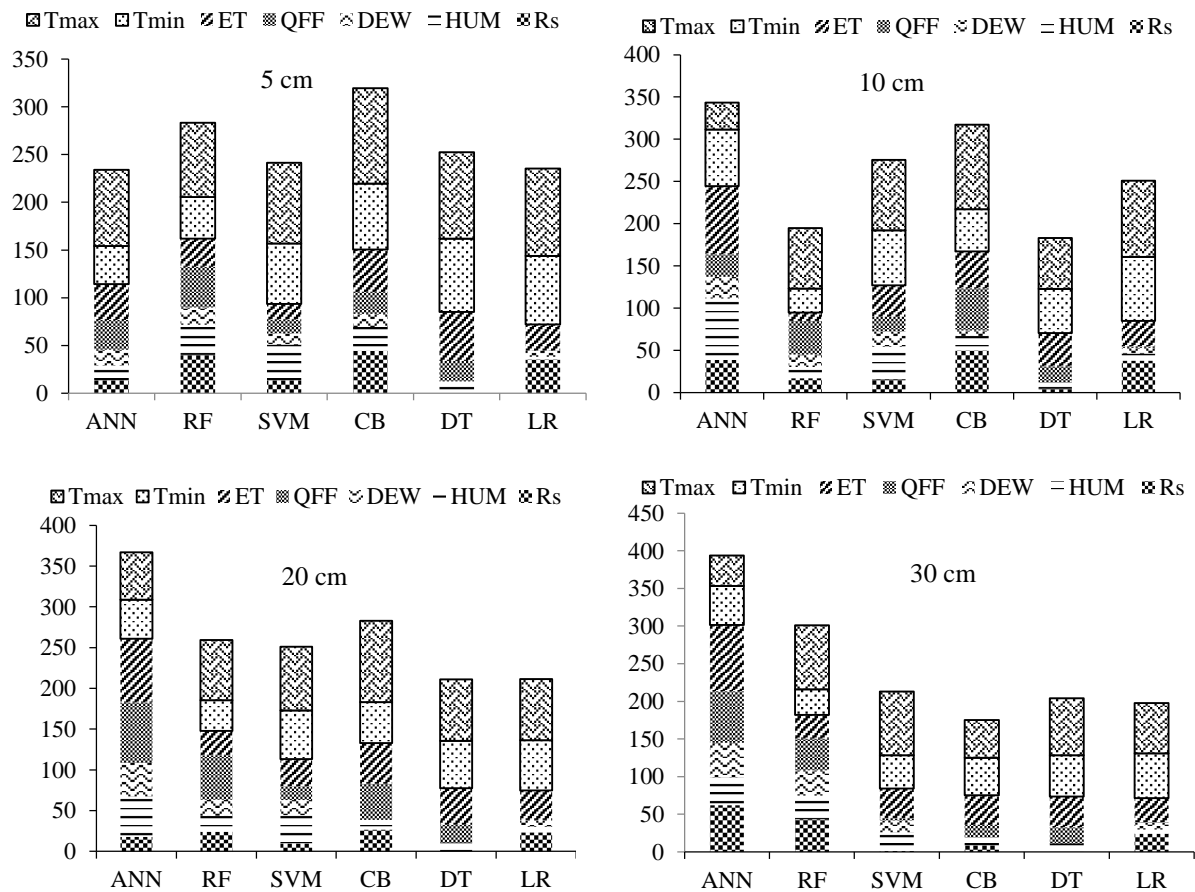


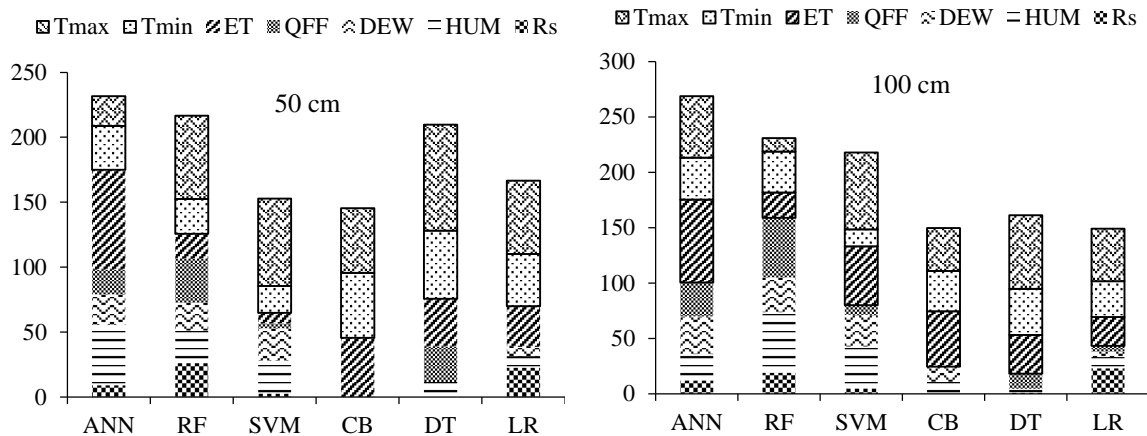
شکل ۳- مقایسه دقت مدل‌های مختلف یادگیری ماشین بر اساس شاخص آماری

Figure 3. Comparing the accuracy of different models of machine learning based on statistical index

اهمیت متغیرها نشان می‌دهد که متغیر دمای حداکثر و پس از آن دمای حداقل و تبخیر-تعرق بیشترین تاثیر را در بیان تغییرات خروجی یعنی دمای خاک در هر ۶ مدل به کار برده شده نشان داد. ستاری و همکاران (Sattari *et al.*, 2020) نیز دمای هوا و ساعات آفتابی را مهم‌ترین پارامتر در پیش‌بینی دمای خاک عنوان کردند. از دقت در نمودارها چنین برمی‌آید که در مدل شبکه عصبی مصنوعی و جنگل تصادفی تمامی متغیرها دارای تاثیر قابل توجه در پیش‌بینی دمای خاک هستند. همچنین در همه مدل‌ها به جز مدل درخت تصمیم و رگرسیون خطی، همه متغیرهای ورودی هرچند کم در تغییر دمای خاک تاثیر داشتند. متغیر دمای نقطه شبنم در برخی مدل‌ها دارای تاثیر نسبی و در برخی مدل‌ها تاثیر ناچیزی از خود نشان داد.

همچنین با توجه به نمودارها در مدل‌ها (بجز مدل‌های رگرسیون خطی و درخت تصمیم) با افزایش عمق دقت پیش‌بینی افزایش یا کاهش قابل‌ملاحظه‌ای مشاهده نشد. این در حالیست که در تحقیقات دلبری و همکاران (Delbari *et al.*, 2019) و ستاری و همکاران (Sattari *et al.*, 2017) گزارش شده است که با افزایش عمق دقت مدل‌ها کاهش می‌یابد. این مغایرت احتمالا به تفاوت در مدل‌های مورد بررسی و پارامترهای ورودی مدل برمی‌گردد. زیرا کاربرد صرف از پارامترهای به شدت تاثیرگذار در دمای سطح خاک به عنوان متغیر ورودی برای پیش‌بینی، سبب عملکرد ضعیفتر مدل‌ها در تخمین دمای عمقی خاک خواهد شد. اهمیت متغیرهای ورودی (دمای حداکثر و حداقل، تشعشعات، رطوبت، نقطه شبنم، فشار جو و ...) در پیش‌بینی دمای خاک در اعماق مختلف در مدل‌های مختلف یادگیری ماشین مورد بررسی قرار گرفت (شکل ۴). نتایج





شکل ۴- درجه اهمیت متغیرهای ورودی در مدل‌های مختلف یادگیری ماشین

Figure 4. The importance of input variables in different models of machine learning

یک منطقه فرا خشک مورد بررسی قرار گرفت. نتایج بدست آمده نشان داد که همه مدل‌ها توانایی خوبی در پیش‌بینی دمای خاک از خود نشان دادند. هرچند برخی مدل‌ها قابلیت بالاتری در این زمینه داشته و برخی مدل‌ها عملکردشان به طور نسبی ضعیف‌تر بود. دو مدل RF و ANN بر اساس معیارهای ارزیابی آماری بهترین کارکرد را در شبیه‌سازی دمای خاک داشتند. پس از این دو مدل روش‌های SVM و CB نیز دارای عملکرد قابل قبولی بودند. دقت روش‌های مذکور توسط محققان دیگر نیز در پیش‌بینی دمای خاک در اقلیم‌های مختلف مورد تایید قرار گرفته است. در این بین روش درخت تصمیم و رگرسیون خطی در ارزیابی عملکرد در سطح پایین‌تری نسبت به دیگر روش‌ها قرار گرفتند. این تفاوت در عملکرد به فرآیندهای مختلف شبیه‌سازی و روابط ریاضی پیچیده و غیرخطی متفاوت در مدل‌های پیش‌بینی برمی‌گردد. به طور کلی از نتایج این تحقیق چنین برمی‌آید که برخی از تکنیک‌های یادگیری ماشین نظیر جنگل تصادفی، شبکه عصبی مصنوعی و ماشین بردار پشتیبان با درجه اطمینان بالا قابلیت جایگزینی با روش‌های عملی تعیین دمای خاک را دارا هستند. این مدل‌ها به عنوان یک ابزار جانبی واجد شرایط می‌توانند برای تولید دمای دقیق خاک استفاده و سبب صرفه‌جویی در هزینه و زمان گردند.

علت برخی تفاوت‌ها در اهمیت پارامترها در مدل‌های مختلف احتمالاً به فرآیند پیچیده شبیه‌سازی و روابط ریاضی پیچیده و غیرخطی بین متغیرهای مستقل و وابسته در مدل‌های مختلف پیش‌بینی مربوط می‌شود (Ajamzadeh et al., 2017). به طور کلی نتایج تعیین اهمیت متغیرهای ورودی در همه مدل‌ها نشان داد که پارامتر دمای حداکثر و حداقل و تبخیر-تعرق دارای اهمیت بسیار زیادی در تعیین دمای خاک در تمامی اعماق می‌باشند. متغیر دمای نقطه شبنم و پس از آن فشار جو از درجه اهمیت نسبی کمتری برخوردار بوده و مدل‌ها به خوبی توانسته‌اند این موضوع را بیان کنند. این نتایج با مشاهدات میدانی (شکل ۳) مطابقت داشته و این انطباق عملکرد خوب مدل‌ها را در تعیین پارامتر تاثیرگذار نشان می‌دهد.

نتیجه‌گیری کلی

در تحقیق حاضر به ارزیابی و مقایسه شش مدل داده‌محور به منظور پیش‌بینی دمای خاک تا عمق ۱۰۰ سانتی‌متری از داده‌های سهل‌الوصول هواشناسی پرداخته شد. شش تکنیک یادگیری ماشین شامل رگرسیون خطی (LR)، درخت تصمیم‌گیری (DT)، ماشین بردار پشتیبان (SVM)، جنگل تصادفی (RF)، کیوبست (CB) و شبکه عصبی مصنوعی (ANN) برای شبیه‌سازی دمای خاک در

References

- Ajamzadeh A., Mollaeinia M.R., and Ghandahari Gh. 2017. Comparison of artificial intelligence methods in predicting daily Time Series of minimum and maximum temperature and precipitation in tangab dam station (Fars Province). *Geographical Space*, 59: 205-228. (In Persian)
- Bahmani F., Piri-Sahragard H., and Piri J. 2018. Application of artificial intelligence methods to estimate soil daily temperature in arid and semi-arid climates. *Iranian Journal of Range and Desert Research*, 26 (1): 201-213. (In Persian)
- Bameri A., Khormali F., Kiani F., and Dehghani AA. 2015. Spatial variability of soil organic carbon in different hillslope positions in Toshan area, Golestan Province, Iran: geostatistical approaches. *Journal of Mountain Science*, 12(6). DOI: 10.1007/s11629-014-3213-z.
- Delbari M., Sharifazari S., and Mohammadi E. 2019. Modeling daily soil temperature over diverse climate conditions in Iran—a comparison of multiple linear regression and support vector regression techniques. *Theoretical and Applied Climatology*, 135(3-4): 991- 1001.
- Fan A.W., and Liu W. 2003. Simulation of the daily change of soil temperature under different conditions. *Heat Transfer—Asian Research*, 32(6):533–544.
- Fatholouloumi S., Vaezi A.R., Alavipanah S.K., Montzka C., Ghorbani A., and Biswas A. 2020. Soil temperature modeling using machine learning techniques. *Desert*, 25(2): 185-199.
- Feng Y., Cui N., Hao W., Gao L., and Gong D. 2019. Estimation of soil temperature from meteorological data using different machine learning models. *Geoderma*, 338: 67–77.
- Granata F., Papirio S., Esposito G., Gargano R., and Marinis G. 2017. Machine learning algorithms for the forecasting of wastewater quality indicators. *Water*. 9(105), doi:10.3390/w9020105.
- Hasani Z., Mirabbasi-Najafabadi R., and Ghasemi A.R. 2018. Prediction of groundwater quality of Khanmirza plain using decision tree method. *Hydrogeology*, 3(1): 99-110. (In Persian)
- IRIMO (2007) I.R. of Iran meteorological organization, data center. Official home page: <http://www.irimo.ir/eng/>.
- Kim S., and Singh V.P. 2014. Modeling daily soil temperature using data driven models and spatial distribution. *Theoretical and Applied Climatology*, 118(3):465–479.
- Ma W., Tan K., and Du P. 2016. Predicting soil heavy metal based on random forest model. IGARSS: 4331-4334. 978-1-5090-3332-4/16/\$31.00 ©2016 IEEE.
- Malone B. 2013. Use R for Digital Soil Mapping. Soil Security Laboratory. The University of Sydney. PP: 217.
- Mehdizadeh S., Ahmadi F., and Kozekalani Sales A. 2020b. Modelling daily soil temperature at different depths via the classical and hybrid models. *Meteorological Applications*, 27:e1941. DOI: 10.1002/met.1941.
- Mehdizadeh S., Fathian F., Sadegh Safari M.J., and Khosravi A. 2020a. Developing novel hybrid models for estimation of daily soil temperature at various depths. *Soil & Tillage Research*, 197(104513): 1-12.
- Mirakzehi Kh., Shahriari A., Pahlavan-Rad M.R., and Bameri A. 2017. Application of random forest method for predicting soil classes in low relief lands (Case study: Hirmand county). *Journal of Water and Soil Conservation*, 24(1): 67-84. (In Persian)
- Noi P.T., Degener J., and Kappas M. 2017. Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data. *Remote Sensing*. 9(398). doi:10.3390/rs9050398.
- Norouzi H., Nadiri A.A., Asghari Mogaddam A., and Gharekhani M. 2017. Prediction of transmissivity of malikan plain aquifer using random forest method. *Water and Soil Science*, 27(2):61-75. (In Persian)
- Omidvar K., Shafie Sh., Taghizadeh Z., and Alipour M. 2014. Evaluating the efficiency of the decision tree model in predicting rainfall in Kermanshah synoptic station. *Journal of Applied Researches in Geographical Sciences*, 34: 89-110. (In Persian)
- Ozturk M., Salman O., and Koc M. 2011. Artificial neural network model for estimating the soil temperature. *Canadian Journal of Soil Science*, 91(4): 551-562.
- Quinlan R. 1993. Combining instance based and model based learning. In Proceedings of the Tenth International Conference on Machine Learning. Amherst. MA. USA. 27–29. pp. 236–243.

- Rezaei M., Sameni A., and Fallah-Shamsi S.R. 2018. Advanced machine learning methods for wind erosion monitoring in southern Iran. *Journal of Environmental Erosion Research*, Vol: 29(8:1): 39-58. (In Persian)
- Samadianfard S., and Panahi S. 2018. Estimating daily reference evapotranspiration using data mining methods of support vector regression and M5 model tree. *Journal of Watershed Management Research*, 10(18):157-167. (In Persian)
- Samadianfard S., Ghorbani M.A., and Mohammadi B. 2018. Forecasting soil temperature at multiple-depth with a hybrid artificial neural network model coupled hybrid firefly optimizer algorithm. *Information Processing in Agriculture*, 5: 465-476.
- Sattari M.T., Avram A., Apaydin H., and Matei O. 2020. Soil temperature estimation with meteorological parameters by using tree-based hybrid data mining models. *Mathematics*, 8, 1407. doi:10.3390/math8091407.
- Sattari M.T., Dodangeh E., and Abraham J. 2017. Estimation of daily soil temperature via data mining techniques in semi-arid climate conditions. *Earth Sciences Research Journal*, 21(2): 85 – 93.
- Sihag P., Esmailbeiki F., Singh B., and Pandhiani S.M. 2020. Model-based soil temperature estimation using climatic parameters: The case of Azerbaijan Province, Iran. *Geology, Ecology, and Landscapes*, 4(3): 203-215.
- Tabari H., Hosseinzadeh-Talaei P., and Willems P. 2015. Short - term forecasting of soil temperature using artificial neural network. *Meteorological Applications*, 22(3), 576-585.
- Tabari H., Sabziparvar AA., and Ahmadi M. 2011. Comparison of artificial neural network and multivariate linear regression methods for estimation of daily soil temperature in an arid region. *Meteorology and Atmospheric Physics Journal*, 110(3):135-142.
- Wilding, L.P., Smeck, N. E., and Hall, G.F. 1983. Pedogenesis and Soil taxonomy. I. Concepts and Interactions. Elsevier Publishing Company, 303p.
- Wu W., Tang X.P., Guo N.J., Yang C., Liu H.B., and Shang Y.F. 2013. Spatiotemporal modeling of monthly soil temperature using artificial neural networks. *Theoretical and Applied Climatology*, 113(3-4):481-494.
- Zadmehr H., and Farrokhian-Firouzi A. 2020. Estimating soil temperature from meteorological data using extreme learning machine, artificial neural network and multiple linear regression models. *Iranian Journal of Soil and Water Research*, 51(4): 895-906. (In Persian)
- Zhou J., Li E., Wei H., Li Ch., Qiao Q., and Jahed-Armaghani, D. 2019. Random forests and cubist algorithms for predicting shear strengths of rockfill materials. *Applied Sciences*. 9(1621), doi:10.3390/app9081621.
- Zhu X., Wu G. Coulon F., Wu L., and Chen D. 2018. Correlating asphaltene dimerization with its molecular structure by potential of mean force calculation and data mining. *Energy Fuel*. 32:5779-5788.
- Zounemat-Kermani M. 2013. Hydrometeorological parameters in prediction of soil temperature by means of artificial neural network: case study in Wyoming. *Journal of Hydrologic Engineering*, 18: 707-718.

Estimation of Surface and Depth Soil Temperature from Meteorological Data Using Machine Learning Techniques in Hyper Arid Climate

Abolfazl Bameri¹, Mahsa Khaleghi^{2*}

(Received: January 2021 Accepted: April 2021)

Abstract

Accurate estimation of temperature at various soil depths is crucial for land-atmosphere interactions. In this study, the application of six different machine learning models including artificial neural network (ANN), decision tree (DT), cubist (CB), random forest (RF), support vector machine (SVM) and linear regression (LR) for modeling of daily soil temperature was studied at six different depths of 5, 10, 20, 30, 50 and 100 cm in Kerman. A set of accessible meteorological data including maximum and minimum temperatures, relative humidity, dew point, evapotranspiration and atmospheric pressure were used as input to the models. The degree of importance and correlational analysis was performed for the input variables based on the data of the 18-year statistical period. According to the results, the performance of all six models based on evaluation criteria ($R^2 > 0.86$, RMSE < 2.8 °C and Bias < 0.14 °C) was acceptable at all depths. However, RF, ANN and SVM showed very high efficiency in estimating soil temperature ($R^2 < 0.97$). Also, the DT model and then the LR model performed lower than the others. Examination of the importance of variables showed that among the input parameters, maximum and minimum temperature had the greatest effect on predicting soil temperature in all models. Finally, it can be safely acknowledged that machine learning models such as random forest, artificial neural network and support vector machine have the ability to estimate surface and depth soil temperatures in arid climates in the absence of measuring equipment. A set of meteorological data including maximum and minimum temperature, relative humidity, dew point, evapotranspiration and atmospheric pressure were used as input to the models.

Key words: Random Forest, Climatic Data, Soil Temperature Simulation, Data-driven Models

Bameri A., and Khaleghi M. 2022. Estimation of surface and depth soil temperature from meteorological data using machine learning techniques in hyper arid climate. *Applied Soil Research*, 10(1): 54-68.

1-lecturer, Department of Soil Science, Faculty of Water and Soil, Zabol University, Iran.

2- M.Sc. Graduated, Department of water engineering, Faculty of Water and Soil, Zabol University, Iran.

* Corresponding Author Email: khaleghi.83@gmail.com