

Research Article

Estimating Soil Temperature at Various Climates Using Data-Driven Methods

Aynaz Vafaei¹, Erfan Abdi¹, Saeid Samadianfard^{2*}

Received: January 1, 2024

Accepted: June 8, 2024

Revised: June 6, 2024

Published online: September 22, 2024

1-M.Sc student, Dept. of Water Eng., Faculty of Agric., University of Tabriz, Iran

2-Assoc. Prof., Dept. of Water Eng., Faculty of Agric., University of Tabriz, Iran

*Corresponding Author, Email: s.samadian@tabrizu.ac.ir

Abstract

Background and Objectives

Soil temperature is one of the important factors in agriculture and hydrology, and its accurate measurement is very important to ensure the optimal growth and development of the plants. Soil temperature is a factor that affects many processes such as seed germination, soil moisture level, aeration, nitrification and availability of plant nutrients. Because the soil temperature data is measured in some synoptic stations, most of the data have limitations or are incomplete. However, choosing the best method to estimate soil temperature with other available meteorological data is an efficient approach in many fields. Soil temperature depends on several factors including color, slope, vegetation, density, humidity and amount of sunlight. Currently, some physical models are available that are intrinsically related to the state of soil heat flow and energy balance in underlying soils to estimate soil temperature. The importance of soil temperature in agricultural sciences and hydrology, on the one hand, and the existence of many difficulties in recording this vital parameter, have led researchers to seek a relationship between soil temperature and other parameters in order to be able to estimate soil temperature with optimal accuracy.

Methodology

In this research, daily soil temperature values were collected during the time period of 1990-2022 in Ramsar, Arak and Shiraz stations. On the other hand, the parameters of minimum temperature (T_{min}), maximum temperature (T_{max}), average temperature (T_m), maximum relative humidity (U_{max}), minimum relative humidity (U_{min}), average relative humidity (U_m), average wind speed (FFM) and Sunshine hours (SSHN) was considered as the input parameters and soil temperature (T_{soil}) as the target parameter. It is worth mentioning that the way of choosing different input compounds to estimate the value of soil temperature in the studied models is based on having a higher correlation with soil temperature based on the thermal map. Moreover, the ability of data-driven models of Gaussian process regression (GPR), support vector regression (SVR), M5P algorithm, linear regression (LR), and multilayer perceptron (MLP) neural network in estimating soil temperature was evaluated using different statistical parameters of correlation coefficient (R), root mean square error (RMSE), Nash Sutcliffe coefficient (NS), average absolute value of percentage error (MAPE) and Wilmot agreement index (WI).

Findings

The evaluation of five GPR, SVR, M5P, LR and MLP models for three stations of Arak, Ramsar and Shiraz shows that the 8th M5P scenario and the 8th LR scenario with lower root mean square error respectively (0.899 and 0.889) for Ramsar station, (0.958 and 0.949) for Arak station and (0.966 and 0.953) for Shiraz station have better performance than other studied models. Also, the evaluation of the impact of the input parameters in creating the scenario for the models shows that the parameters of relative humidity and air temperature had more important role than other input parameters. So that by adding parameters of relative humidity and air temperature, the accuracy of the model has increased. Therefore, these parameters are among the most key and important parameters of soil temperature.

Conclusion

The analysis and evaluation of soil characteristics has an important impact in the fields of hydrology, agriculture and climate. On the other hand, soil temperature has a direct relationship with the amount of moisture available to the plant, so that an increase in soil temperature can increase the transpiration rate of plants, and as a result, soil moisture decreases. Soil temperature is also an essential factor in agriculture because it determines whether plants can grow, and controls soil chemistry and biology and atmosphere-land gas exchange. Therefore, predicting soil temperature is very important for successful crop management and yield optimization. So, In this research, five data-driven methods of GPR, SVR, M5P, LR and MLP were used to predict soil temperature in Arak, Ramsar and Shiraz stations during the time period of 1990-2022. The obtained results were compared using statistical parameters and it was concluded that the 8th M5P scenario and the 8th LR scenario have shown the best performance in three stations with the lowest error compared to all scenarios. Therefore, the application of the mentioned models to predict the soil temperature has proper accuracy and is recommended for management and evaluation in terms of environmental and civil aspects.

Keywords: Estimation, Gaussian model regression, Meteorological parameters, Multilayer neural network, Support vector regression



مقاله پژوهشی

بر آورد دمای خاک در اقلیم‌های مختلف با استفاده از روش‌های داده‌محور

آیناز وفایی^۱، عرفان عبدی^۱، سعید صمدیان فرد^{۲*}

تاریخ پذیرش: ۱۴۰۳/۰۳/۱۹

تاریخ دریافت: ۱۴۰۲/۱۰/۱۱

تاریخ انتشار آنلاین: ۱۴۰۳/۰۷/۰۱

تاریخ ویرایش: ۱۴۰۳/۰۳/۱۷

۱- دانشجوی کارشناسی ارشد، گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز

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

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

چکیده

دمای خاک یکی از جنبه‌های مهم کشاورزی و هیدرولوژی است و اندازه‌گیری دقیق آن برای اطمینان از رشد و نمو مطلوب گیاه بسیار مهم است. دمای خاک عاملی است که بر بسیاری از فرآیندها مانند جوانه‌زنی بذر، میزان رطوبت خاک، تهویه، سرعت نیتریفیکاسیون و در دسترس بودن مواد مغذی گیاه تأثیر می‌گذارد. با توجه به این که داده‌های دمای خاک در بعضی از ایستگاه‌های سینوپتیک اندازه‌گیری می‌شود، اغلب داده‌ها دارای محدودیت و یا نواقصی هستند. با این حال انتخاب بهترین روش جهت پیش‌بینی و تخمین دمای خاک با سایر داده‌های هواشناسی موجود، رویکردی مؤثر و کارآمد در بسیاری از زمینه‌ها می‌باشد؛ لذا در مطالعه حاضر، توانایی مدل‌های داده محور رگرسیون فرایند گاوسی (GPR)، رگرسیون ماشین بردار پشتیبان (SVR)، الگوریتم M5P، رگرسیون خطی (LR) و شبکه عصبی پرسپترون چندلایه (MLP) در برآورد دمای خاک سه ایستگاه اراک، رامسر و شیراز طی دوره آماری ۳۲ ساله با استفاده از پنج معیار اعتبارسنجی مورد ارزیابی قرار گرفت. نتایج بدست آمده نشان داد که سناریو هشتم M5P و LR با داشتن جذر میانگین مربعات خطای کمتر به ترتیب «۰/۸۸۹ و ۰/۸۹۹» برای ایستگاه رامسر، «۰/۹۵۸ و ۰/۹۴۹» برای ایستگاه اراک و «۰/۹۶۶ و ۰/۹۵۳» برای ایستگاه شیراز، عملکرد بهتری نسبت به سایر مدل‌ها داشته‌است. همچنین پارامترهای رطوبت نسبی و دمای هوا از مؤثرترین پارامترهای هواشناسی مورد نیاز در برآورد دمای خاک شناخته شد، بطوری که افزودن این پارامترها باعث افزایش دقت مدل می‌شود.

واژه‌های کلیدی: پیش‌بینی، داده‌های هواشناسی، رگرسیون مدل گاوسی، رگرسیون ماشین بردار پشتیبان، شبکه عصبی چندلایه.

مقدمه

دمای خاک یکی از مهمترین شاخص‌های خصوصیات خاک است. دمای خاک و چگونگی تغییرات آن نسبت به زمان و مکان یکی از مهمترین عواملی است که نه تنها تبادل ماده و انرژی را در خاک تحت تأثیر قرار می‌دهد؛ بلکه میزان و جهت کلیه فرآیندهای فیزیکی خاک به صورت مستقیم یا غیرمستقیم وابسته به دمای خاک است (اسدی و همکاران ۲۰۱۴). این تأثیر مهمی بر حرارت خاک و خواص انتقال جرم و ساختار خاک و موادمغذی خاک دارد (بورو سن و همکاران ۲۰۰۷؛ لی و همکاران ۲۰۰۸؛ وو و همکاران ۲۰۱۰؛ علیزامیر و همکاران ۲۰۲۰؛ هاوو و همکاران ۲۰۲۳) و همچنین نقش مهمی در اکوسیستم ایفا می‌کند. دمای خاک متأثر از عوامل متعددی از جمله توپوگرافی، تابش خورشید، دمای هوا، توزیع بارش، میزان رطوبت خاک، نوع و ویژگی گرمایی خاک، همانند ظرفیت گرمایی، ضریب رسانایی گرمایی و گرمای ویژه می‌باشد. دمای خاک می‌تواند بر تجمع موادآلی خاک، واکنش‌های شیمیایی، تنفس خاک و تخریب موادآلی تأثیر بگذارد (سیفرید و همکاران ۲۰۱۰؛ بروکس و همکاران ۲۰۰۵؛ شیمیل و همکاران ۲۰۰۴؛ هوانگ و همکاران ۲۰۲۰؛ السید و همکاران ۲۰۲۳).

دمای خاک به چندین عنصر از جمله رنگ، شیب، پوشش گیاهی، تراکم، رطوبت و مقدار نور خورشید بستگی دارد. در حال حاضر برخی مدل‌های فیزیکی که به طور ذاتی به وضعیت جریان گرمای خاک و تعادل انرژی در خاک‌های زیرین برای تخمین دمای خاک استفاده می‌شوند در دسترس هستند (صمدیان فرد و همکاران ۲۰۱۸). مدل‌های فیزیکی اعمال شده برای پیش‌بینی دمای خاک بسیار پیچیده‌اند؛ زیرا آنها به طور ذاتی به معادلات ریاضی متکی بوده و نیاز به مجموعه‌ای از شرایط اولیه و مرزی در طیف وسیعی از مکان‌های فضایی به عنوان داده‌های اولیه مدل دارند که به دست آوردن آنها اگر غیرممکن نباشد، دشوار است و از طرفی

هزینه های محاسباتی بالایی دارند. چنین پیچیدگی‌هایی ممکن است در مکان‌های دور افتاده و یا کشورهایی که ابزارهای کارآمد در این زمینه ندارند امکان پذیر نباشد (ین و همکاران ۱۹۹۸).

پیش‌بینی دمای خاک نقش مهمی در علوم زمین داشته و به پیش‌بینی و درک واکنش‌های اکوسیستم به تغییرات اقلیم کمک می‌کند (لی و همکاران ۲۰۲۲). اهمیت دمای خاک در علوم کشاورزی و هیدرولوژی از یک سو و وجود دشواری‌های فراوان در ثبت این پارامتر حیاتی پژوهشگران را بر آن داشته‌است تا به دنبال یافتن رابطه‌ای بین دمای خاک و پارامترهای دیگر باشند تا بتوانند دمای خاک را با دقت مطلوب برآورد و پیش‌بینی کنند (اسدی و همکاران ۲۰۱۴). در حال حاضر تحقیقات متعددی در این زمینه انجام گرفته‌است. میهالاکاکو در آتن و دوبلین برای برآورد تغییرات دمای خاک از تشخیص هوشمند و شبکه‌عصبی مصنوعی با دمای هوا، رطوبت نسبی و تابش خورشیدی به عنوان متغیرهای ورودی استفاده نمود (میهالاکاکو ۲۰۰۲). مایکل و گیلچریست با ارائه مدل چندجمله‌ای درجه چهارم موفق به محاسبه کمینه و بیشینه دمای خاک در هر روز از سال در ۱۸ منطقه از انگلستان شده اند (مایکل و گیلچریست ۱۹۸۳). کارانزا و همکاران از روش جنگل تصادفی RF برای پیش‌بینی رطوبت خاک منطقه ریشه بر روی یک حوضه کوچک کشاورزی با استفاده از اندازه‌گیری‌های درجا استفاده کردند (کارانزا و همکاران ۲۰۲۰). ژنگ و همکاران از تصاویر ماهواره‌ای، داده‌های تاریخی رطوبت خاک و روش RF برای پیش‌بینی دمای خاک روزانه استفاده کردند. آنها نشان دادند که مدل RF برای دمای خاک سطحی و عمیق در اوکلاهما می‌تواند بسیار دقیق باشد (ژنگ و همکاران ۲۰۱۹). فنگ و همکاران دمای خاک نیم ساعته را از داده‌های اتمسفر (دمای هوا، سرعت باد، رطوبت نسبی، تابش خورشیدی و کسری فشار بخار) بر اساس چندین روش یادگیری ماشین، از جمله مدل RF پیش‌بینی کردند. همچنین آنها

اشاره کردند که مدل RF می‌تواند به طور دقیق دمای خاک را در هر عمق خاک توصیف کند (فنگ و همکاران ۲۰۱۹). گیل و همکاران با اعمال مدل رگرسیون ماشین بردار پشتیبان SVR برای پیش‌بینی رطوبت خاک به صورت روزانه با استفاده از داده‌های هواشناسی و دمای خاک گذشته، نشان‌دادند که مدل SVR عملکرد پیش‌بینی بهتری نسبت به مدل‌های ساده شبکه‌عصبی مصنوعی ANN دارد (گیل و همکاران ۲۰۰۶). دلبری و همکاران عملکرد مدل SVR در پیش‌بینی دمای خاک روزانه در ایران را آزمایش کردند و نتیجه گرفتند که یک جزء تناوب می‌تواند به طور قابل توجهی کارایی پیش‌بینی را بهبود بخشد (دلبری و همکاران ۲۰۱۹). لی و همکاران برای پیش‌بینی دما و رطوبت خاک روزانه ایستگاه مورد نظر، از مدل‌های حافظه‌ی کوتاه‌مدت طولانی^۱ (LSTM)، ماشین بردار پشتیبان^۲ (SVR)، جنگل تصادفی، شبکه‌الاستیک^۳ (Enet) و تلفیق دیگر مدل‌ها مانند LSTM_soil، LSTM_NC استفاده کردند (لی و همکاران ۲۰۲۲). در کشور نیز تحقیقاتی در خصوص دمای خاک صورت گرفته‌است. مؤذن زاده و همکاران مدل SVR را در ایستگاه هواشناسی مراغه در شمال غرب ایران برای پیش‌بینی دمای خاک منطقه ریشه ارزیابی کردند. آنها نشان‌دادند که SVR و MLR هر دو می‌توانند به پیش‌بینی خوبی دست یابند، همچنین عملکرد مدل SVR در لایه سطحی بهتر از مدل MLR در لایه‌های عمیق‌تر است (مؤذن زاده و همکاران ۲۰۱۹). ثنایی و همکاران با استفاده از تحلیل سری‌های فوری و دوره آماری ۱۲ ساله به برآورد دمای خاک روزانه و سالیانه در عمق‌های مختلف در ۶ ایستگاه کردستان اقدام نمودند. در این تحلیل آنها تنها با استفاده از دمای هوا (۲ متری) و بدون در نظر گرفتن دیگر پارامترهای هواشناسی (تابش،

ساعات آفتابی، رطوبت هوا، ابرناکی) روابط همبستگی خود را جهت برآورد دمای خاک پیشنهاد نمودند (ثنایی و همکاران ۲۰۰۸). جعفری گلستانی و همکاران نیز با استفاده از روابط رگرسیونی، معادلات تجربی را جهت تخمین دمای خاک ایستگاه ساری در عمق‌های ۷ و ۲۰ سانتی‌متری پیشنهاد کردند (جعفری گلستانی و همکاران ۲۰۰۷). سبزی پرور و همکاران با مطالعه روی داده‌های هواشناسی ۳ ایستگاه سینوپتیک زاهدان، تهران و رامسر که دارای اقلیم متفاوتی می‌باشند، نتیجه گرفتند که دقت روش استنتاج تطبیقی عصبی-فازی چهار درصد بیشتر از روش آماری رگرسیون در پیش‌بینی دمای خاک در اعماق مختلف بوده‌است (سبزی پرور و همکاران ۲۰۱۰). اسدی و همکاران با استفاده از آمار و اطلاعات موجود در ایستگاه هواشناسی سینوپتیک گرگان و همچنین الگوریتم‌های مدلسازی داده‌کاوی شامل روش شبکه‌عصبی مصنوعی و نزدیک‌ترین k-همسایگی و روش رگرسیون چندمتغیره، روابط موجود بین داده‌های دمای هوا و رطوبت نسبی با دمای خاک در اعماق مختلف جستجو کرده و بهترین روش را معرفی کردند (اسدی و همکاران ۲۰۱۴). محمدی برای تخمین دمای خاک در عمق ۱۰ سانتی‌متری برای ایستگاه تبریز، از پارامترهای دمای متوسط هوا، ساعات آفتابی، سرعت باد و رطوبت نسبی استفاده کرد. او مدل‌های شبکه‌عصبی مصنوعی ANN و رگرسیون ماشین بردار پشتیبان SVR مورد بررسی قرارداد و نتایج نشان‌داد مدل SVR عملکرد بهتری نسبت به ANN داشته‌است (محمدی ۲۰۲۲). همانطور که به اهمیت اندازه‌گیری دمای خاک اشاره شد، ارزیابی و پیش‌بینی دقیق دمای خاک در برنامه‌ریزی کشاورزی و مدیریت آب، به کشاورزان اجازه می‌دهد زمان‌بندی فعالیت‌های مزرعه‌ای و

¹ Long short-term memory² Support Vector Machine³ Extra net

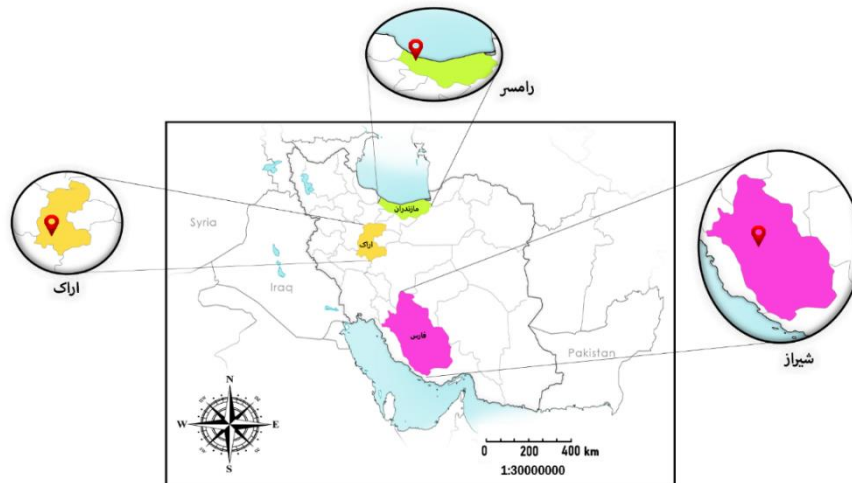
برنامه‌ریزی زمان کاشت را بهینه کنند و دستورالعمل‌های مربوطه را به بهترین شکل انجام دهند. در تحقیق حاضر برای برآورد دمای خاک در سه ایستگاه با اقلیم های مختلف رامسر (معتدل و مرطوب)، اراک (گرم و خشک) و شیراز (گرم و نیمه خشک) که تاکنون مورد مطالعه واقع نشده است از پنج مدل رگرسیون خطی، رگرسیون ماشین بردار پشتیبان، رگرسیون فرایند گاوسی، M5P و شبکه عصبی پرسپترون چندلایه که بر مبنای یادگیری ماشین و داده محور هستند استفاده شد و همچنین به بررسی مدل های ذکر شده از جهت انعطاف پذیری و سازگاری آن ها با داده های سه ایستگاه انتخابی پرداخته شد.

مواد و روش ها

منطقه مورد مطالعه

در این مطالعه سه ایستگاه از مناطق مختلف ایران با اقلیم‌های مختلف که بر اساس روش دو مارتن انتخاب

شد، مورد بررسی قرار گرفت. موقعیت جغرافیایی مناطق به این صورت می‌باشد که شهر رامسر بین ۳۶ درجه و ۳۲ دقیقه تا ۳۶ درجه و ۵۹ دقیقه عرض شمالی و ۵۰ درجه و ۲۰ دقیقه تا ۵۰ درجه و ۴۷ دقیقه طول شرقی از نصف‌النهار گرینویچ با ۲۱ متر ارتفاع از سطح دریا با آب و هوایی معتدل و مرطوب، شهر اراک روی مدار ۳۴ درجه و ۰۵ دقیقه و ۳۰ ثانیه در نیم کره شمالی از خط استوا قرار دارد. همچنین این شهر روی نصف‌النهار ۴۹ درجه و ۴۱ دقیقه و ۳۰ ثانیه طول شرقی از نصف‌النهار گرینویچ با ۱/۷۲ متر ارتفاع از سطح دریا با زمستان‌های سرد و مرطوب و تابستان‌های گرم و خشک و موقعیت شهر شیراز ۲۹ درجه و ۳۶ دقیقه شمالی و ۵۲ درجه و ۳۲ دقیقه با ۱/۵ متر ارتفاع از سطح دریا با آب و هوای گرم و نیمه خشک است. شکل ۱ موقعیت شهرهای مورد مطالعه را نشان می‌دهد.



شکل ۱- موقعیت جغرافیایی ایستگاه‌های مورد مطالعه.

مجموعه داده ها

در تحقیق حاضر مقادیر دمای خاک روزانه در طی دوره آماری ۳۲ ساله در سال‌های ۱۹۹۰-۲۰۲۲ در ایستگاه‌های رامسر، اراک و شیراز که هر کدام از ایستگاه‌ها از نظر اقلیم تفاوت محسوسی دارند، جمع آوری گردید. از طرفی پارامترهای حداقل دما (T_{min}).

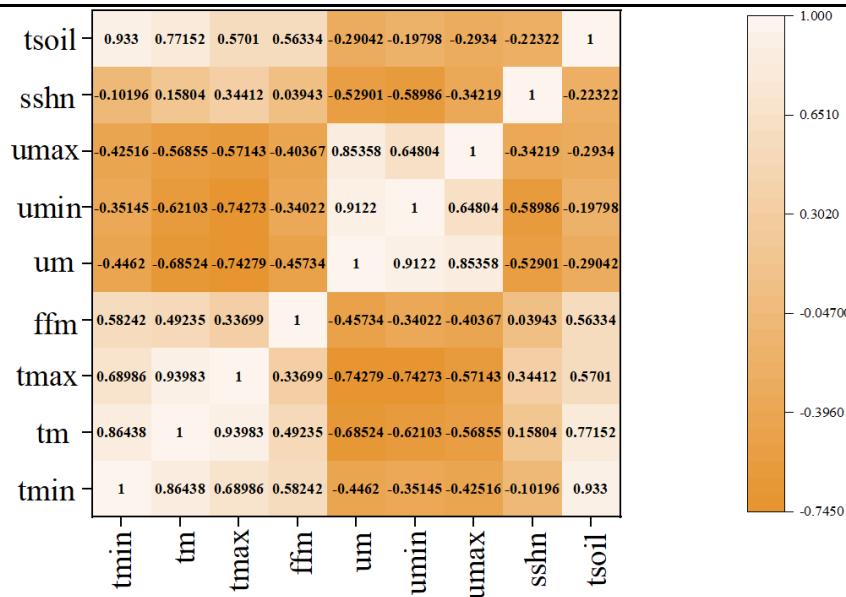
حداکثر دما (T_{max})، میانگین دما (T_m)، حداکثر رطوبت نسبی (U_{max})، حداقل رطوبت نسبی (U_{min})، میانگین رطوبت نسبی (U_m)، میانگین سرعت باد (FFM) و ساعت آفتابی (SSHN) به عنوان پارامترهای ورودی و دمای خاک (T_{soil}) به عنوان پارامتر هدف در نظر گرفته شد. علاوه بر این، مشخصات آماری پارامترهای ذکر شده در

بالاتر با دمای خاک بر مبنای نقشه حرارتی ارائه شده در شکل های ۲ الی ۴ صورت گرفته است.

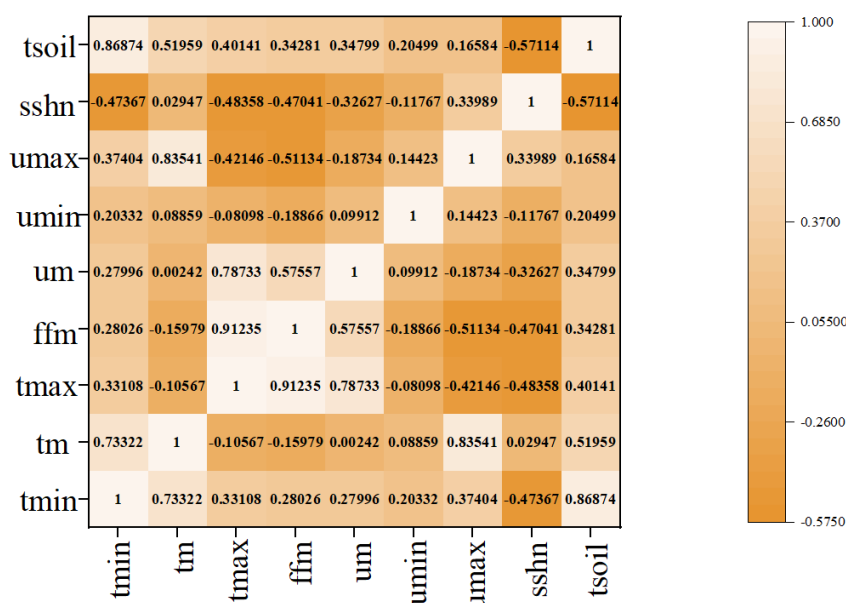
جدول ۱ ارائه شده است. شایان ذکر است که نحوه انتخاب ترکیبات مختلف ورودی، برای تخمین مقدار دمای خاک در مدل های مورد مطالعه بر اساس داشتن همبستگی

جدول ۱- مشخصات آماری پارامترهای مورد استفاده.

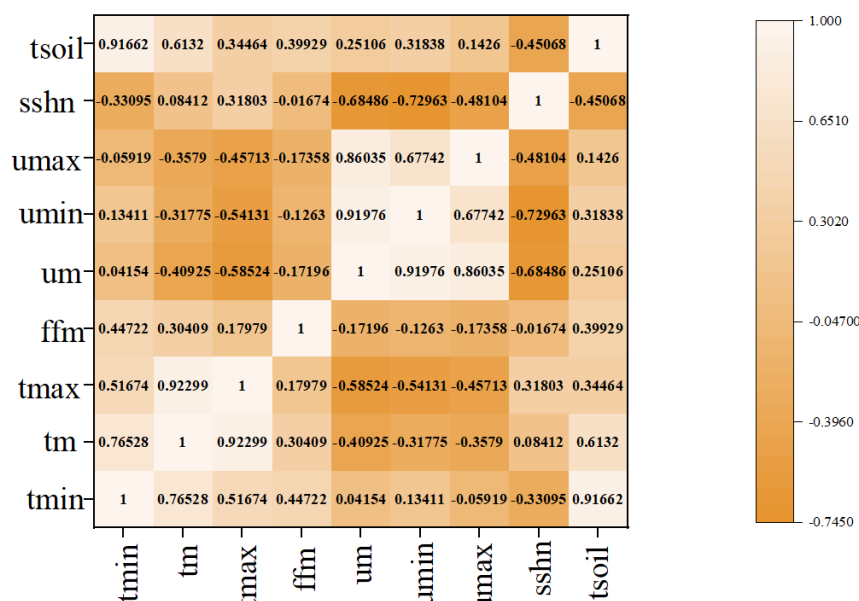
SSHN ($\text{kw m}^{-2}\text{day}^{-1}$)	$U_{\max}$ (%)	$U_{\min}$ (%)	U_m (%)	FFM (km h^{-1})	$T_{\max}$ ($^{\circ}\text{C}$)	T_{mean} ($^{\circ}\text{C}$)	$T_{\min}$ ($^{\circ}\text{C}$)	Statistics	Station
۶/۶۱	۷۸/۶	۳۶/۱	۵۶/۸۷	۲/۰۱	۱۱/۷۱	۵/۲۴	-۰/۲۵	Mean	اراک
۰	۲۵	۴	۱۴/۱۲۵	۰	-۸/۲	-۱۳/۶	-۲۲	Minimum	
۱۱/۳	۱۰۰	۹۸	۹۹	۷/۷۵	۲۷/۸	۲۱	۱۹	Maximum	
۳/۳۲	۱/۷۵	۱۸/۵	۱۷/۲۶	۱/۴۷	۵/۸۴	۵/۱۸	۵/۱۱	STDEV	
۳/۳۳	۹/۸۹	۷۳/۲۲	۸۴/۳۱	۱/۸۳	۱۲/۵۷	۹/۳۸	۶/۷۶	Mean	رامسر
۰	۶۹	۷	۴۶/۸۷۵	۰	۰/۶	-۱/۱	-۳/۶	Minimum	
۱۰/۸	۱۰۰	۹۷	۹۹	۸/۴	۳۱/۸	۲۰/۵	۱۵	Maximum	
۳/۴۲	۴/۷۶	۱۳/۷	۷/۷۶۷	۱/۱۵	۳/۵	۲/۴۷	۲/۸۸	STDEV	
۷/۹۱	۷/۶۵	۲۵/۲۷	۴۷/۳۲	۱/۶۹	۱۷/۶۴	۱۰/۳۱	۳	Mean	شیراز
۰	۲۶	۲	۱۲/۶۲۵	۰	۲	-۰/۴	-۶	Minimum	
۱۱/۶	۱۰۰	۹۷	۹۸/۲۵	۶/۷۵	۳۰/۸	۲۱/۳	۱۶/۴	Maximum	
۳/۰۷	۱/۳۳	۱۶/۲	۱۶/۸	۱/۰۷	۴/۶۳	۴	۴/۳۶	STDEV	



شکل ۲- نقشه حرارتی بر اساس همبستگی پارامترهای مورد مطالعه برای ایستگاه اراک.



شکل ۳- نقشه حرارتی بر اساس همبستگی پارامترهای مورد مطالعه برای ایستگاه رامسر.



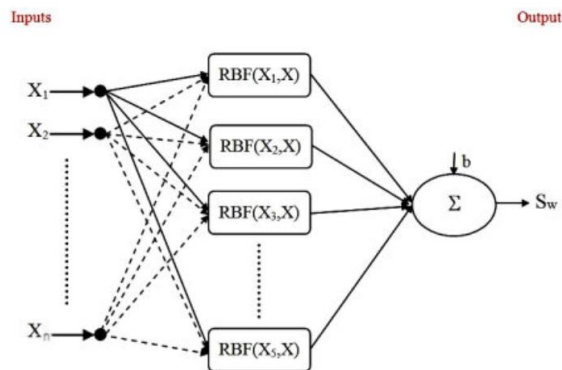
شکل ۴- نقشه حرارتی بر اساس همبستگی پارامترهای مورد مطالعه برای ایستگاه شیراز.

رگرسیون خطی (LR)

رگرسیون خطی^۴ یک روش آماری است که برای مدل سازی رابطه بین یک متغیر وابسته و یک یا چند متغیر مستقل استفاده می شود. رگرسیون خطی یک تحلیل پیش بینی ساده و رایج است که ضرایب معادله خطی را تخمین می زند و به بهترین وجه مقدار متغیر

وابسته را پیش بینی می کند. تحلیل رگرسیون این امکان را فراهم می آورد تا تغییرات متغیر وابسته را از طریق متغیرهای مستقل پیش بینی و سهم هر یک از متغیرهای مستقل را در تبیین متغیر وابسته تعیین کند. رگرسیون رابطه نزدیکی با ضریب همبستگی دارد بدین معنا که برای انجام رگرسیون، باید ضریب همبستگی را محاسبه

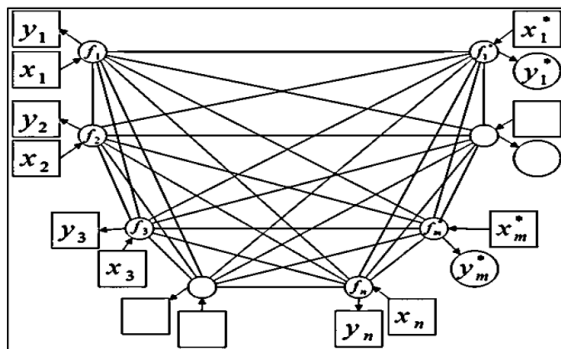
⁴ Linear Regression



شکل ۵- مدل تصویری رگرسیون بردار پشتیبان.

رگرسیون فرآیند گاوسی (GPR)

رگرسیون فرایند گاوسی مجموعه‌ای از متغیرهای تصادفی است که تعداد دلخواه محدودی از آنها با توزیع‌های گاوسی ادغام شده‌اند. رگرسیون فرآیند گاوسی یک رویکرد ناپارامتریک برای رگرسیون است که در حوزه یادگیری ماشین موج ایجاد می‌کند. مدل‌های فرایند گاوسی به دلیل دانش قبلی درباره وابستگی‌های تابعی و داده‌ها، برای تعمیم به هیچ فرایند اعتبارسنجی نیاز ندارند و مدل‌های رگرسیون فرایند گاوسی قادر به درک توزیع پیش‌بینی متناظر با ورودی آزمون هستند (پال و دسوال ۲۰۱۰). فرایندهای گاوسی به صورت مجموعه‌ای از متغیرهای تصادفی تعریف می‌شود، هر تعداد محدودی از آنها دارای توزیع گاوسی چند متغیره توأمی است. تصویری از مدل گاوسی در شکل ۶ آورده شده‌است.



شکل ۶- مدل تصویری رگرسیون فرآیند گاوسی.

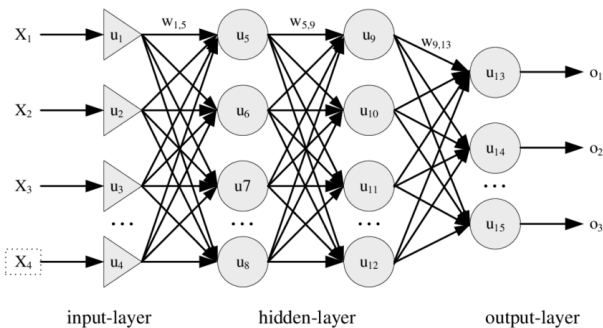
الگوریتم M5P

کرد. رگرسیون خطی متناسب با یک خط یا سطح مستقیم است که اختلاف بین مقادیر خروجی پیش‌بینی شده و واقعی را به حداقل می‌رساند. هدف از انجام تحلیل رگرسیون، شناسایی مدل خطی این رابطه است. اگر برای شناسایی و پیش‌بینی متغیر وابسته فقط از یک متغیر مستقل استفاده شود مدل را رگرسیون خطی ساده می‌گویند (رابط و همکاران ۲۰۲۱).

رگرسیون ماشین بردار پشتیبان (SVR)

در سال‌های اخیر، ابزار پیشرفته‌ای که در رابطه با هوش مصنوعی کاربردهای زیادی در روش یادگیری ماشین دارد، رگرسیون ماشین بردار پشتیبان می‌باشد (فراستی و همکاران ۲۰۲۱). رگرسیون ماشین بردار پشتیبان یکی از روش‌های یادگیری است که بر مبنای تئوری یادگیری آماری در سال ۱۹۶۳ میلادی معرفی شده است (بوسر و همکاران ۱۹۹۲). SVR یک الگوریتم یادگیری ماشینی قدرتمند است که برای طبقه‌بندی خطی و غیرخطی، رگرسیون و حتی کارهای تشخیص داده‌های پرت استفاده می‌شود. رگرسیون ماشین بردار پشتیبان یک سیستم یادگیری کارآمد بر مبنای تئوری بهینه‌سازی مقید است که از اصل استقرای کمینه‌سازی خطای ساختاری استفاده می‌کند و به یک جواب بهینه کلی منجر می‌شود (شکل ۵). SVR ها در کاربردهای مختلف سازگار و کارآمد هستند زیرا می‌توانند داده‌های با تعداد زیاد و روابط غیرخطی را مدیریت کنند. SVR یکی از روش‌های یادگیری با ناظر است که از آن برای طبقه‌بندی و رگرسیون استفاده می‌شود. همچنین روشی برای طبقه‌بندی دوتایی در فضای ویژگی‌های دلخواه است و از این روش مناسب برای مسائل پیش‌بینی بشمار می‌رود (فراستی و همکاران ۲۰۲۱).

پرسپترون چندلایه، نوعی شبکه عصبی مصنوعی است که برای یادگیری نظارت شده استفاده می‌شود. یکی از پایه‌ای‌ترین مدل‌های عصبی موجود، مدل پرسپترون چندلایه است که عملکرد انتقالی مغز انسان را شبیه‌سازی می‌کند و از شبکه‌های عصبی مصنوعی پیشخور محسوب می‌شوند. این یک شبکه عصبی کاملاً متصل است که از سه نوع لایه تشکیل شده است: لایه ورودی، لایه خروجی و یک یا چند لایه پنهان. MLP یک مدل غیرخطی است که می‌تواند روابط پیچیده بین ورودی‌ها و خروجی‌ها را یاد بگیرد. MLP با استفاده از پس انتشار آموزش داده می‌شود، که الگوریتمی است که وزن اتصالات بین نورون‌ها را تنظیم می‌کند تا خطا بین خروجی پیش‌بینی شده و خروجی واقعی را به حداقل برساند (آبیرامی و چیترا ۲۰۲۰). توابع فعال‌سازی رایج مورد استفاده در MLP ها شامل تابع سیگموئید، تابع ReLU و تابع مماس هذلولی هستند. در شکل ۸ مدل تصویری مربوط به این مدل آورده شده است.

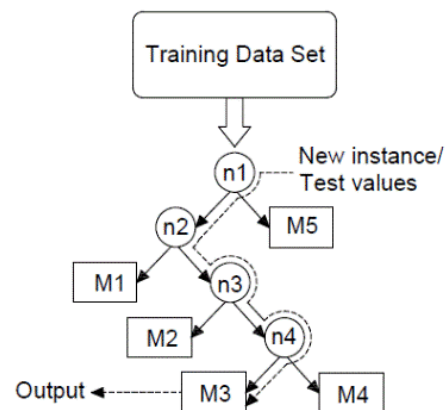


شکل ۸- مدل تصویری شبکه عصبی پرسپترون چندلایه (واهیونگورو و همکاران ۲۰۱۳).

سناریوهای مدل‌سازی و معیارهای ارزیابی

در مطالعه حاضر هشت سناریو مختلف از متغیرهای پیش‌بینی کننده پارامترهای حداقل دما (T_{min})، حداکثر دما (T_{max})، میانگین دما (T_m)، حداکثر رطوبت نسبی (U_{max})، حداقل رطوبت نسبی (U_{min})، میانگین رطوبت نسبی (U_m)، میانگین سرعت باد (FFM) و ساعت آفتابی (SSHN)، به عنوان ورودی مدل‌ها در جدول ۲

M5P یک الگوریتم درخت تصمیم است که برای مسائل رگرسیون استفاده می‌شود. مجموعه‌ای از نمونه‌های آموزشی وجود دارد که هر نمونه آموزشی با مجموعه‌ای از ویژگی‌ها مشخص می‌شود که مقادیر ورودی هستند و دارای هدف متناظر که همان مقدار خروجی است (کوینلان ۱۹۹۲). الگوریتم M5P یک درخت تصمیم‌گیری مرسوم را با یک مدل رگرسیون خطی ترکیب می‌کند. M5P با استفاده از یک مدل رگرسیون خطی چند متغیره، درخت‌های طبقه‌بندی و رگرسیون ایجاد می‌کند که می‌تواند تغییرات را در یک زیرفضای خاص به حداقل برساند (امیکسیز و همکاران ۲۰۱۸). الگوریتم M5P توابع رگرسیون خطی را در گره‌های پایانه اختصاص می‌دهد و با طبقه‌بندی یا تقسیم کل فضای داده به چندین زیرفضا، یک مدل رگرسیون خطی چند متغیره را برای هر زیرفضا برازش می‌دهد. روش درختی M5P به جای کلاس‌های گسسته، با مسائل کلاس پیوسته سروکار دارد و می‌تواند وظایفی را با ابعاد بسیار بالا انجام دهد (شکل ۷) (سیهاگ و همکاران ۲۰۱۹).



شکل ۷- مدل تصویری الگوریتم M5P (امیکسیز و همکاران

۲۰۱۸).

شبکه عصبی پرسپترون چند لایه (MLP)

شبکه‌های عصبی مدل‌های محاسباتی هستند که قادرند رابطه میان ورودی‌ها و خروجی‌های یک سیستم فیزیکی را توسط شبکه‌ای از گره‌ها که همگی به هم متصل هستند تعیین نمایند (قاسمی و همکاران ۲۰۲۱).

$$NS = 1 - \left[\frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \right] \quad [3]$$

$$WI = 1 - \left[\frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \right] \quad [4]$$

$$MAPE = \frac{\sum_{i=1}^n \left(\frac{O_i - P_i}{O_i} \right)}{n} \times 100 \quad [5]$$

در روابط، پارامترهای O_i و P_i به ترتیب مقادیر

پیش‌بینی و مشاهداتی می‌باشند.

ایجاد شد. در این پژوهش جهت ارزیابی عملکرد سناریوهای مختلف تعریف شده برای هر یک از روش‌های مورد مطالعه از پارامترهای آماری ضریب همبستگی (R)، جذر میانگین مربعات خطا ($RMSE$)، ضریب نش ساتکلیف (NS)، میانگین قدر مطلق درصد خطا ($MAPE$) و شاخص توافق ویلموت (WI)، مطابق روابط زیر، استفاده شده است.

$$RMSE = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \quad [1]$$

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

جدول ۲- ترکیب های مختلف ورودی، برای تخمین دمای خاک در مدل مورد مطالعه.

سناریو	ورودی	خروجی
1	T min	T-SOIL
2	T min	T-SOIL
3	T min	T-SOIL
4	T min	T-SOIL
5	T min	T-SOIL
6	T min	T-SOIL
7	T min	T-SOIL
8	T min	T-SOIL

نتایج و بحث

در این بخش، توانایی مدل‌های LR، SVR، GPR، M5P و MLP بر روی هشت مجموعه داده برای پیش‌بینی مقادیر دمای خاک مورد آزمون قرار گرفته است. میزان ارتباط

بین مقادیر اندازه‌گیری شده و پیش‌بینی شده دمای خاک برای هشت سناریو در جدول‌های ۳ و ۴ از نظر ضریب همبستگی، جذر میانگین مربعات خطا، ضریب نش ساتکلیف و شاخص توافق ویلموت، طی مراحل اعتبارسنجی گزارش شده است.

جدول ۳- پارامترهای ارزیابی مدل‌های مورد مطالعه در دوره آزمون ایستگاه‌های شیراز و اراک.

مدل	شیراز					اراک				
	WI	NS	MAPE	RMSE	R	WI	NS	MAPE	RMSE	R
GPR1	۰/۹۵۵	۰/۸۳۸	۱/۵۹۴	۲/۰۲۴	۰/۹۳	۰/۲۶۷	-۱۱۴/۸۱	۳/۶۵۹	۴/۴۱	۰/۹۲۹
GPR2	۰/۹۵۷	۰/۸۴۵	۱/۵۴۸	۱/۹۸۱	۰/۹۳۳	۰/۶۹۳	-۱/۵۰۵	۲/۹۵۵	۳/۶۳۳	۰/۶۳۳
GPR3	۰/۹۶۳	۰/۸۶۳	۱/۴۱۸	۱/۸۶۱	۰/۹۳۸	۰/۷۹	-۰/۳۴۸	۲/۶۰۱	۳/۲۳۹	۰/۷۱۹
GPR4	۰/۹۶۱	۰/۸۵۸	۱/۴۴	۱/۸۹۵	۰/۹۳۶	۰/۷۹۱	-۰/۳۴۷	۲/۵۸۳	۳/۲۲۵	۰/۷۲۳
GPR5	۰/۹۶۵	۰/۸۶۹	۱/۳۸۶	۱/۸۲۱	۰/۹۴۳	۰/۸۹۴	۰/۵۱۷	۲/۰۶۷	۲/۶۴۵	۰/۸۱۹
GPR6	۰/۹۶۴	۰/۸۶۹	۱/۳۹۲	۱/۸۲۳	۰/۹۴۳	۰/۸۹۱	۰/۵۰۴	۲/۰۸۷	۲/۶۷۴	۰/۸۱۵
GPR7	۰/۹۶۵	۰/۸۷	۱/۳۸۸	۱/۸۱۲	۰/۹۴۳	۰/۸۹۱	۰/۵۱	۲/۱۰۵	۲/۶۸۸	۰/۸۱۲
GPR8	۰/۹۶۳	۰/۸۶۸	۱/۴۰۴	۱/۸۲۸	۰/۹۴۲	۰/۹۳۴	۰/۷۱	۱/۶۶۵	۲/۱۶	۰/۸۸۷
SVR1	۰/۹۵۹	۰/۸۵۲	۱/۵۲۳	۱/۹۳۴	۰/۹۳۳	۰/۹۶۲	۰/۸۴۶	۱/۲۴۶	۱/۷۱۷	۰/۹۲۹

۰/۹۵۶	۰/۸۴۳	۱/۵۸۴	۱/۹۹	۰/۹۲۸	۰/۹۶۷	۰/۸۷	۱/۱۸۱	۱/۶۵۶	۰/۹۳۵	SVR2
۰/۹۶۱	۰/۸۶	۱/۴۷۲	۱/۸۸	۰/۹۳۶	۰/۹۶۸	۰/۸۷۷	۱/۱۶۴	۱/۶۲۸	۰/۹۳۸	SVR3
۰/۹۶	۰/۸۵۶	۱/۴۹۶	۱/۹۱۱	۰/۹۳۶	۰/۹۶۸	۰/۸۷۸	۱/۱۶۱	۱/۶۲۶	۰/۹۳۸	SVR4
۰/۹۶۵	۰/۸۷۱	۱/۴۱	۱/۸۰۷	۰/۹۴۵	۰/۹۷۶	۰/۹۰۴	۰/۹۶۹	۱/۴۲۷	۰/۹۵۲	SVR5
۰/۹۶۵	۰/۸۷۲	۱/۴۰۱	۱/۸۰۱	۰/۹۴۵	۰/۹۷۵	۰/۹۰۴	۰/۹۷	۱/۴۲۸	۰/۹۵۲	SVR6
۰/۹۶۴	۰/۸۷	۱/۴۱۲	۱/۸۱۲	۰/۹۴۵	۰/۹۷۶	۰/۹۰۴	۰/۹۵۹	۱/۴۱۹	۰/۹۵۳	SVR7
۰/۹۶۵	۰/۸۷	۱/۴۰۴	۱/۸۱	۰/۹۴۶	۰/۹۷۶	۰/۹۰۴	۰/۹۵۵	۱/۴۱۸	۰/۹۵۳	SVR8
۰/۹۵۷	۰/۸۴۲	۱/۵۸۳	۲	۰/۹۳۳	۰/۹۶۲	۰/۸۴۴	۱/۲۶۵	۱/۷۳	۰/۹۲۹	M5P1
۰/۹۵۹	۰/۸۴۷	۱/۵۳	۱/۹۶۶	۰/۹۳۵	۰/۹۶۷	۰/۸۶۸	۱/۱۶۵	۱/۶۴	۰/۹۳۵	M5P2
۰/۹۶۵	۰/۸۶۶	۱/۴۰۳	۱/۸۴۲	۰/۹۴۱	۰/۹۶۷	۰/۸۶۷	۱/۱۷۳	۱/۶۳۱	۰/۹۳۶	M5P3
۰/۹۶۴	۰/۸۶۳	۱/۴۱۵	۱/۸۵۸	۰/۹۴۱	۰/۹۶۹	۰/۸۷۶	۱/۱۶۱	۱/۶	۰/۹۳۹	M5P4
۰/۹۶۷	۰/۸۷۴	۱/۳۶۹	۱/۷۸۵	۰/۹۴۸	۰/۹۷۶	۰/۹۰۱	۰/۹۵۲	۱/۴۰۲	۰/۹۵۲	M5P5
۰/۹۶۲	۰/۸۷۴	۱/۳۷۱	۱/۷۸۴	۰/۹۴۳	۰/۹۷۷	۰/۹۰۱	۰/۹۵۲	۱/۴۰۲	۰/۹۵۳	M5P6
۰/۹۶۷	۰/۸۷۴	۱/۳۶۹	۱/۷۸۱	۰/۹۴۶	۰/۹۷۶	۰/۹۰۲	۰/۹۴۶	۱/۳۹۶	۰/۹۵۳	M5P7
۰/۹۶۷	۰/۸۷۴	۱/۳۶۹	۱/۷۷۱	۰/۹۵۸	۰/۹۷۶	۰/۹۰۲	۰/۹۴۶	۱/۳۹۶	۰/۹۶۶	M5P8
۰/۹۵۷	۰/۸۲	۱/۵۸۳	۲	۰/۹۳۳	۰/۹۶۲	۰/۸۴۴	۱/۲۶۵	۱/۷۳	۰/۹۲۹	LR1
۰/۹۵۷	۰/۸۲	۱/۵۸۳	۲	۰/۹۳۳	۰/۹۶۷	۰/۸۶۸	۱/۱۶۵	۱/۶۴	۰/۹۳۵	LR2
۰/۹۶۵	۰/۸۵۷	۱/۴۱۱	۱/۸۵۲	۰/۹۴۱	۰/۹۶۹	۰/۸۷۶	۱/۱۴۴	۱/۶۰۳	۰/۹۳۹	LR3
۰/۹۶۴	۰/۸۵۶	۱/۴۲	۱/۸۶۳	۰/۹۴	۰/۹۷۶	۰/۹۰۲	۰/۹۴۶	۱/۳۹۶	۰/۹۵۳	LR4
۰/۹۶۷	۰/۸۶۶	۱/۳۶۹	۱/۷۸۵	۰/۹۴۸	۰/۹۷۶	۰/۹۰۱	۰/۹۵۲	۱/۴۰۲	۰/۹۵۳	LR5
۰/۹۶۷	۰/۸۶۶	۱/۳۷۱	۱/۷۸۴	۰/۹۴۴	۰/۹۷۶	۰/۹۰۱	۰/۹۵۲	۱/۴۰۲	۰/۹۵۳	LR6
۰/۹۶۷	۰/۸۶۶	۱/۳۶۹	۱/۷۸۱	۰/۹۴۸	۰/۹۷۶	۰/۹۰۲	۰/۹۴۶	۱/۳۹۶	۰/۹۵۳	LR7
۰/۹۶۷	۰/۸۶۶	۱/۳۶۹	۱/۷۸۱	۰/۹۴۹	۰/۹۷۷	۰/۹۰۲	۰/۹۴۶	۱/۳۹۲	۰/۹۵۳	LR8
۰/۹۵۹	۰/۸۲۷	۱/۵۰۲	۱/۹۲۱	۰/۹۳۱	۰/۹۵۳	۰/۷۸	۱/۳۲۵	۱/۸۲۵	۰/۹۲۷	MLP1
۰/۹۶	۰/۸۲۵	۱/۴۸۴	۱/۹۰۳	۰/۹۳۲	۰/۹۵۷	۰/۸۰۸	۱/۳۳۶	۱/۷۷۵	۰/۹۳	MLP2
۰/۹۶۹	۰/۸۷۳	۱/۲۴۷	۱/۷۳۲	۰/۹۴۱	۰/۹۶۱	۰/۸۳۳	۱/۲۵۸	۱/۷۰۲	۰/۹۳۱	MLP3
۰/۹۶۷	۰/۸۶۴	۱/۲۹۳	۱/۷۷۳	۰/۹۴۲	۰/۹۶۱	۰/۸۳۳	۱/۲۵۸	۱/۶۹۸	۰/۹۳۱	MLP4
۰/۹۷۱	۰/۸۷۸	۱/۱۹	۱/۶۶	۰/۹۴۷	۰/۹۷۱	۰/۸۷۸	۱/۰۵	۱/۴۹۹	۰/۹۵۲	MLP5
۰/۹۶۴	۰/۸۴۹	۱/۳۷۹	۱/۸۴۹	۰/۹۴۹	۰/۹۶۸	۰/۸۶۷	۱/۱۱۸	۱/۵۷۴	۰/۹۵۱	MLP6
۰/۹۶۶	۰/۸۵۸	۱/۳۰۹	۱/۷۹	۰/۹۴۸	۰/۹۶۷	۰/۸۵۸	۱/۱۵۶	۱/۶۱۳	۰/۹۴۹	MLP7
۰/۹۶۳	۰/۸۴۴	۱/۳۹۳	۱/۸۷۵	۰/۹۴۸	۰/۹۵۱	۰/۷۷۹	۱/۴۴۶	۱/۹۰۵	۰/۹۴۳	MLP8

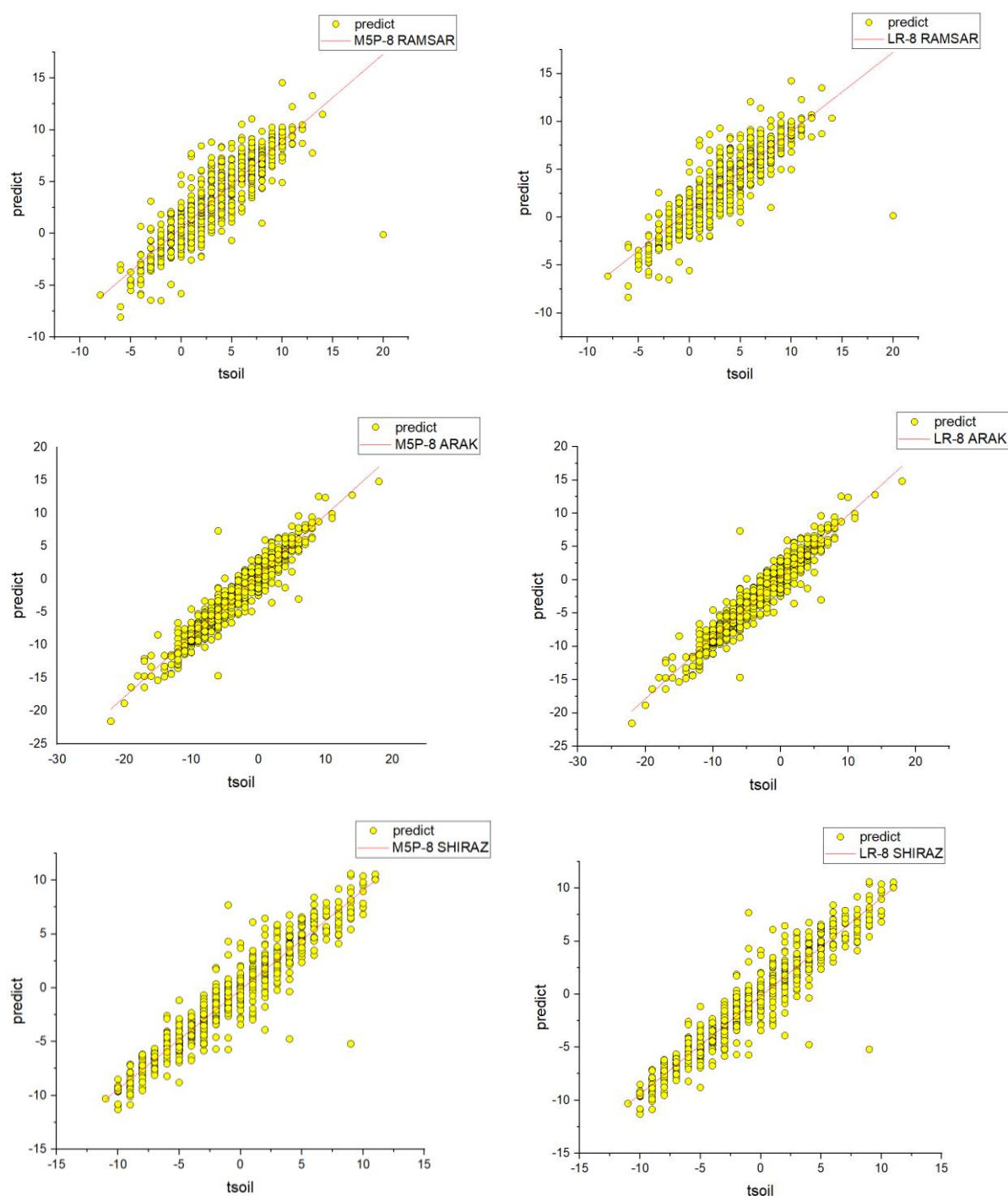
جدول ۴- پارامترهای ارزیابی مدل‌های مورد مطالعه در دوره آزمون ایستگاه رامسر.

مدل	رامسر				
	WI	NS	MAPE	RMSE	R
GPR1	۰/۹۱۹	۰/۷۴۹	۱/۵۹۳	۲/۰۰۹	۰/۸۶۷
GPR2	۰/۹۳۶	۰/۷۸	۱/۳۶۱	۱/۸۷۸	۰/۸۸۴
GPR3	۰/۹۳۸	۰/۷۸۱	۱/۳۴۱	۱/۸۷۴	۰/۸۸۵
GPR4	۰/۹۳۸	۰/۷۸۱	۱/۳۴۹	۱/۸۷۷	۰/۸۸۵
GPR5	۰/۹۳۷	۰/۷۷۷	۱/۳۵۵	۱/۸۹۱	۰/۸۸۳
GPR6	۰/۹۳۶	۰/۷۷۳	۱/۳۶۱	۱/۹۱۱	۰/۸۸۱
GPR7	۰/۹۳۸	۰/۷۷۹	۱/۳۳۴	۱/۸۸۵	۰/۸۸۵
GPR8	۰/۹۳۸	۰/۷۸۱	۱/۳۲۵	۱/۸۷۷	۰/۸۸۶
SVR1	۰/۹۱۸	۰/۷۴۴	۱/۶۱۳	۲/۰۲۸	۰/۸۶۹
SVR2	۰/۹۲۷	۰/۷۷۲	۱/۴۹۲	۱/۹۱۴	۰/۸۸۵
SVR3	۰/۹۳۲	۰/۷۷۶	۱/۴۵۴	۱/۸۹۸	۰/۸۸۲

۰/۹۲۹	۰/۷۶۹	۱/۴۷۱	۱/۹۲۵	۰/۸۷۸	SVR4
۰/۹۳	۰/۷۷۱	۱/۴۶۴	۱/۹۱۷	۰/۸۸	SVR5
۰/۹۳	۰/۷۶۷	۱/۴۵۳	۱/۹۳۲	۰/۸۷۶	SVR6
۰/۹۳۳	۰/۷۷۴	۱/۴۱۶	۱/۹۰۷	۰/۸۸	SVR7
۰/۹۳۷	۰/۷۸	۱/۳۷۸	۱/۸۷۸	۰/۸۸۴	SVR8
۰/۹۲۴	۰/۷۵۴	۱/۵۶	۱/۹۸۷	۰/۸۶۹	M5P1
۰/۹۳۵	۰/۷۷	۱/۳۶۹	۱/۹۲	۰/۸۷۹	M5P2
۰/۹۳۶	۰/۷۷۳	۱/۳۶۱	۱/۹۱	۰/۸۸	M5P3
۰/۹۳۶	۰/۷۷۵	۱/۳۵۵	۱/۹۰۲	۰/۸۸۱	M5P4
۰/۹۳۷	۰/۷۷۵	۱/۳۵۴	۱/۹۰۲	۰/۸۸۱	M5P5
۰/۹۳۷	۰/۷۷۵	۱/۳۵۴	۱/۹۰۲	۰/۸۸۱	M5P6
۰/۹۳۹	۰/۷۷۹	۱/۳۲۷	۱/۸۸۴	۰/۸۸۵	M5P7
۰/۹۴۱	۰/۷۸۷	۱/۳۲۱	۱/۸۴۳	۰/۸۹۹	M5P8
۰/۹۲۴	۰/۶۵۹	۱/۵۶	۱/۹۸۷	۰/۸۶۹	LR1
۰/۹۳۶	۰/۷۳	۱/۳۶۷	۱/۸۹۷	۰/۸۸۱	LR2
۰/۹۳۷	۰/۷۳۵	۱/۳۶۲	۱/۸۹	۰/۸۸۲	LR3
۰/۹۳۷	۰/۷۳۹	۱/۳۵۵	۱/۸۸۴	۰/۸۸۳	LR4
۰/۹۳۸	۰/۷۴	۱/۳۴۶	۱/۸۷۹	۰/۸۸۴	LR5
۰/۹۳۸	۰/۷۴۱	۱/۳۴۴	۱/۸۸۹	۰/۸۸۳	LR6
۰/۹۳۸	۰/۷۴۳	۱/۳۳۷	۱/۸۷۹	۰/۸۸۴	LR7
۰/۹۴۱	۰/۷۵۷	۱/۳۲۲	۱/۸۵۵	۰/۸۸۹	LR8
۰/۹۲۴	۰/۶۶۸	۱/۵۶۵	۲/۰۴۲	۰/۸۷۲	MLP1
۰/۹۱	۰/۶۱۴	۱/۸۰۶	۲/۳۱۵	۰/۸۸۳	MLP2
۰/۹۲۲	۰/۶۸۳	۱/۶۷	۲/۱۸۹	۰/۸۸۹	MLP3
۰/۹۲۱	۰/۶۸	۱/۶۹۸	۲/۲۱۳	۰/۸۸۹	MLP4
۰/۹۲۳	۰/۷۱	۱/۶۹	۲/۲۴۴	۰/۸۸۷	MLP5
۰/۹۳	۰/۷۴۱	۱/۵۸۵	۲/۱۵۴	۰/۸۸۶	MLP6
۰/۹۳۱	۰/۷۵	۱/۵۶۸	۲/۱۳۸	۰/۸۸۷	MLP7
۰/۹۲۵	۰/۷۲۹	۱/۷۰۴	۲/۲۶۳	۰/۸۹	MLP8

ارزیابی پنج مدل SVR، GPR، M5P، LR و MLP برای سه ایستگاه اراک، رامسر و شیراز نشان می‌دهد که سناریو هشتم M5P و سناریو هشتم LR با داشتن جذر میانگین مربعات خطای کمتر به ترتیب «۰/۸۹۹ و ۰/۸۸۹» برای ایستگاه رامسر، «۰/۹۵۸ و ۰/۹۴۹» برای ایستگاه اراک و «۰/۹۶۶ و ۰/۹۵۳» برای ایستگاه شیراز، عملکرد بهتری نسبت به سایر مدل‌ها دارد؛ همچنین، ارزیابی تاثیر پارامترهای ورودی در ایجاد سناریو برای مدل‌ها نشان

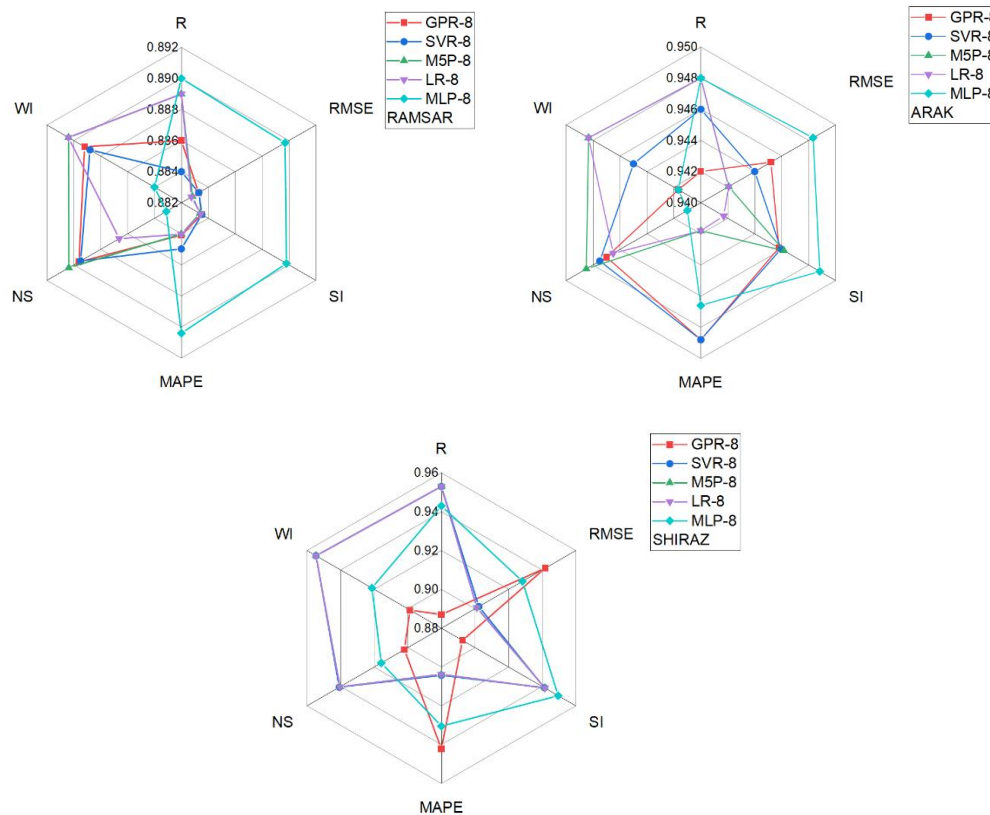
می‌دهد که پارامترهای رطوبت نسبی و دمای هوا نقش مهم‌تری نسبت به سایر پارامترهای ورودی داشته‌است. به طوری که با افزودن پارامترهای رطوبت نسبی و دمای هوا، دقت مدل افزایش پیدا کرده است؛ لذا این پارامترها از جمله کلیدی‌ترین و مهم‌ترین پارامترهای دمای خاک می‌باشند. در ادامه نمودار پراکنش مقادیر دمای خاک برای بهترین مدل‌های ذکر شده در شکل ۹ ارائه شده است.



شکل ۹- نمودار پراکنش مقادیر دمای خاک.

سناریوها هم از نتایج قابل قبولی برخوردار می‌باشند. میزان تغییرات ریشه میانگین مربعات خطا و ضریب همبستگی برای مدل M5P، ($RMSE > 1/294$) و ($R > 0/869$) بوده و برای مدل LR، ($RMSE > 1/396$) و ($R > 0/869$) می‌باشد.

همچنین، جهت بررسی و ارزیابی بهتر مدل‌ها، نمودار راداری شاخص‌های عملکرد برای سناریوی هشتم در ایستگاه در شکل ۱۰ نمایش داده شده است. نتایج حاکی از آن است که مدل‌های M5P و LR علاوه بر اینکه در سناریو هشتم بهترین عملکرد را داشته‌اند، در بقیه



شکل ۱۰- نمودار راداری برای بهترین شاخص‌های عملکرد.

هیدرولوژیکی (بارش و رواناب) دمای خاک را به صورت روزانه و هفتگی پیش بینی کرد. سپس مدل‌های استفاده شده با معیارهای ارزیابی RMSE، MAE و ضریب نش ساتکلیف مورد ارزیابی قرار گرفت و در آخر شبکه عصبی با کمترین خطا بهترین مدل انتخاب شد. داس و همکاران (۲۰۲۳)، از پنج مدل درخت تصمیم، جنگل تصادفی، گرادیان درخت تقویت کننده، ترکیبی از درخت تصمیم و درخت تقویت کننده گرادیان، و ترکیبی از جنگل تصادفی و درخت تقویت کننده گرادیان برای استخراج پیوندهای قوی بین عوامل مختلف هواشناسی و خاک در عمق‌های ۳۰ و ۵۰ سانتی‌متری استفاده کردند. و در آخر با ارزیابی نتایج، مدل ترکیبی جنگل تصادفی و درخت تقویت کننده گرادیان با خطای کمتر ($RMSE < 1/6$) و همبستگی بالا ($R < 0/950$) نسبت به بقیه مدل ها دقت بهتری داشتند. سبزی پرور و همکاران (۲۰۱۰)، از رگرسیون چند متغیره با بکارگیری سامانه استنتاج

مزیت اصلی مدل‌های یادگیری ماشینی این است که آنها توسط مکانیسم‌های داخلی ناشناخته و پیوندهایی که در داخل داده‌ها تعبیه شده‌اند، هدایت می‌شوند به جای اینکه لزوماً مکانیسم فیزیکی پیچیده را در نظر بگیرند (ژنگ و لی، ۲۰۱۹). اشمیت و همکاران (۲۰۲۰)، دقت پیش‌بینی و قابلیت مدل‌های یادگیری ماشین جنگل تصادفی، شبکه عصبی مصنوعی و رگرسیون خطی در استنتاج هیدرولوژیکی را بررسی و نشان‌دادند که مدل‌های یادگیری ماشین دقت پیش‌بینی بالایی دارند اما قابلیت تفسیر نسبتاً پایینی از مکانیسم‌های درونی دارند. همچنین با استفاده از ضریب همبستگی مشخص شد که مدل داده محور جنگل تصادفی از دیگر مدل‌ها عملکرد بهتری دارد. زونمت کرمانی (۲۰۱۳)، از شبکه‌های عصبی و رگرسیون خطی چندگانه برای پیش‌بینی دمای خاک در وایومینگ آمریکا سه عامل هواشناسی (دمای هوا، تابش خالص و رطوبت نسبی) و دو عامل

نتیجه گیری کلی

امروزه تحلیل و ارزیابی مشخصات خاک تأثیر مهمی در زمینه‌های هیدرولوژی و کشاورزی و اقلیم دارد. از طرفی دمای خاک با میزان رطوبت در دسترس گیاه رابطه مستقیم دارد به طوری که افزایش دمای خاک می‌تواند باعث افزایش نرخ تعرق گیاهان شود و در نتیجه رطوبت خاک کاهش پیدا کند. دمای خاک نیز یک عامل ضروری در کشاورزی است زیرا تعیین می‌کند که آیا گیاهان می‌توانند رشد کنند یا خیر، و شیمی و بیولوژی زمین و تبادل گاز اتمسفر-زمین را کنترل می‌کند؛ بنابراین، پیش‌بینی دمای خاک برای مدیریت موفق محصول و بهینه‌سازی عملکرد بسیار مهم است. در این پژوهش از پنج روش داده محور LR, M5P, SVR, GPR و MLP جهت پیش‌بینی دمای خاک در ایستگاه اراک، رامسر و شیراز در طی دوره آماری ۲۰۲۲ - ۱۹۹۰ برآورد گردید. نتایج بدست آمده با استفاده از پارامترهای آماری مورد مقایسه، حداقل دما (T_{min})، حداکثر دما (T_{max})، میانگین دما (T_m)، حداکثر رطوبت نسبی (U_{max})، حداقل رطوبت نسبی (U_{min})، میانگین رطوبت نسبی (U_m)، میانگین سرعت باد (FFM) و ساعت آفتابی (SSHN)، نتیجه گرفته شد که سناریو هشتم M5P و سناریو هشتم LR بهترین عملکرد را در سه ایستگاه با داشتن کمترین خطا نسبت به تمامی سناریوها نشان داده‌اند؛ بنابراین، کاربرد مدل‌های مذکور برای پیش‌بینی دمای خاک دقت مناسبی داشته و جهت مدیریت و ارزیابی از لحاظ جنبه‌های محیط زیستی و عمرانی توصیه می‌گردد. به طور کلی، مدل M5P یک مدل قدرتمند برای مدل‌سازی و عملیات پیش‌بینی است که مزایای متعددی را نسبت به سایر مدل‌های رگرسیون ارائه می‌دهد. برای بهبود ساختار مدل‌های داده محور، تحقیقات آینده نیاز به تمرکز بر استفاده از مدل‌های هیبریدی و الگوریتم‌های فرا ابتکاری و روش‌های ترکیبی مشابه برای پیش‌بینی متغیرهای اضافی (مانند متغیرهای تاخیری، بارش، نفوذ

تطبیقی عصبی-فازی و پارامترهای حداکثر دما، حداقل دما، میانگین دمای روزانه، رطوبت نسبی، ساعت آفتابی و سرعت باد برای تخمین دمای خاک در هشت ایستگاه منتخب ایران با متغیرهای اقلیم و هواشناسی استفاده کردند. نتایج، RMSE را از ۲/۰۶ تا ۲/۷۵ (درجه سلسیوس) نشان داد و همچنین ملاحظه شد که بر خلاف مناطق خشک، در اقلیم مرطوب با افزایش عمق خاک دقت پیش‌بینی تا ۱۸ درصد افزایش مییابد، در حالیکه در ایستگاه‌های گرم و خشک دقت پیش‌بینی دمای خاک در لایه‌های کم عمق تا ۱۰ درصد بیشتر از لایه‌های عمیق مییابد. از طرفی در تحقیق دیگر سبزی پرور و همکاران (۲۰۱۰)، مطالعه‌ای را روی دمای خاک در ایستگاه‌های زاهدان و رامسر با استفاده از رگرسیون چند متغیره انجام دادند. در بهترین و بدترین حالت ضریب همبستگی به ترتیب ۰/۹۴ و ۰/۶۴ بود. بیلگلی (۲۰۱۰)، جهت پیش‌بینی دمای خاک روش‌های رگرسیون و شبکه عصبی مصنوعی را مقایسه کرد. نتایج، برتری عملکرد شبکه عصبی با دقت ۰/۹۹۸ (درجه سلسیوس) نسبت به روش رگرسیون را نشان داد. سانیخانی و همکاران (۲۰۱۸)، برآورد دمای خاک ماهانه با استفاده از مدل‌های ماشین یادگیری حداکثر، شبکه عصبی مصنوعی و درخت مدل M5 اجرا کرده، و نتایج به‌دست‌آمده نشان داد که مدل ELM به‌عنوان دقیق‌ترین مدل جهت برآورد دمای خاک در مقایسه با سایر مدل‌ها می‌باشد. با توجه به مطالعات انجام شده در زمینه پیش-بینی دمای خاک اکثر روش‌های مورد استفاده خطای بالای ۱/۵ درجه سانی گراد را دارند ولی در مقایسه با تحقیق حاضر، حداکثر خطای بهترین مدل‌ها که M5P و LR بوده ۰/۹۶ سانتی گراد می‌باشد که بیان‌گر دقت قابل قبول می‌باشد. همانطور که بیان شد مدل‌هایی که بر پایه یادگیری ماشین هستند از دقت و توانایی بالایی جهت پیش‌بینی و برآورد دمای خاک برخوردار می‌باشند که این امر موجب تسریع اندازه‌گیری و اقدام صحیح برای تصمیم‌گیری در این زمینه می‌شود.

مقادیر از دست رفته، نقاط پرت، نویز، تکراری بودن، ناسازگاری یا ناقص بودن قرار گیرد. برای اطمینان از کیفیت و در دسترس بودن داده ها، پاکسازی، اعتبار سنجی و یکپارچه سازی داده ها، و همچنین روش های جمع آوری و ذخیره سازی داده ها که متناسب با اهداف و زمینه باشد، ضروری است.

پذیری خاک) دارد، مدل های داده محور با وجود مزایایی که دارند، محدودیت ها و چالش هایی دارند. یکی از محدودیت ها تکیه بر روابط خطی است که می تواند منجر به پیش بینی های نادرست شود، به ویژه در سناریوهای طراحی پیچیده با تعاملات غیرخطی. کیفیت و در دسترس بودن داده ها می تواند تحت تأثیر مسائل مختلفی مانند

منابع مورد استفاده

- Abirami S and Chitra P, 2020. Energy-efficient edge based real-time healthcare support system. *Advances in Computers* 117(1): 339-368.
- Alizamir M, Kis O, Ahmed AN, Mert C, Fai CM, Kim S, Kin N and El-Shafie A, 2020. Advanced machine learning model for better prediction accuracy of soil temperature at different depths. *PLoS One* 15(4), 0231055.
- Asadi L, Hezarjaribi A, Ghorbani K, Zakernia M and Agha Shariatmadari Z, 2014. Estimating soil temperature using modern methods of data analysis. *Iranian Journal of Irrigation and Drainage* 8(1): 145-152 (in Persian with English abstract).
- Bilgili M, 2010. Prediction of soil temperature using regression and artificial neural network models. *Meteorology and Atmospheric Physics* 110: 59-70.
- Børresen MH, Barnes DL and Rike AG, 2007. Repeated freeze–thaw cycles and their effects on mineralization of hexadecane and phenanthrene in cold climate soils. *Cold Regions Science and Technology* 49(3): 215-225.
- Boser BE, Guyon IM and Vapnik VN, 1992. A training algorithm for optimal margin classifiers. Pp.144-152. In: Haussler D, (Ed.), 5th Annual ACM Workshop on COLT, Pittsburgh.
- Brooks PD, McKnight D and Elder K, 2005. Carbon limitation of soil respiration under winter snowpacks: potential feedbacks between growing season and winter carbon fluxes. *Global Change Biology* 11(2): 231-238.
- Carranza C, Nolet C, Pezij M and van der Ploeg M, 2021. Root zone soil moisture estimation with Random Forest. *Journal of Hydrology* 593, 125840.
- Das LC, Zhang Z and Crabbe MJC, 2023. Optimization of data-driven soil temperature forecast-The first model in Bangladesh. *Applied Sciences* 13(23): 12616.
- 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.
- Emeksiz C and Demir G, 2018. An investigation of the effect of meteorological parameters on wind speed estimation using bagging algorithm. *International Journal of Intelligent Systems and Applications in Engineering* 6(4): 311-321.
- Elsayed S, Gupta M, Chaudhary G, Taneja S, Gaur H, Gad M, Eid MH, Kovács A, Péter S, Gaagai A and Schmidhalter U, 2023. Interpretation the influence of hydrometeorological variables on soil temperature prediction using the potential of deep learning model. *Knowledge-Based Engineering and Sciences* 4(1): 55-77.
- Farasat M, Seyedian M and Daab K, 2021. Evaporation modeling of free surface water using SVM and LSSVM models. *Journal of Irrigation and Water Engineering* 11(3): 272-288 (in Persian with English abstract).
- 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 (2019): 67–77.
- Gill MK, Asefa T, Kemblowski MW and McKee M, 2006. Soil moisture prediction using support vector machines 1. *JAWRA Journal of the American Water Resources Association* 42(4): 1033-1046.
- Hao L, Yang C and Li X, 2023. Prediction of soil temperature field in panax notoginseng plough layer based on PSO-LSTM neural network. 6th International Symposium on Autonomous Systems (ISAS), Nanjing, China. IEEE.

- Huang R, Huang JX, Zhang C, Wen ZHUO, Chen YY, Zhu DH, Qingling WU and Mansaray LR, 2020. Soil temperature estimation at different depths, using remotely-sensed data. *Journal of Integrative Agriculture* 19(1): 277-290.
- Li HJ, Yan JX, Yue XF and Wang MB, 2008. Significance of soil temperature and moisture for soil respiration in a Chinese mountain area. *Agricultural and Forest Meteorology* 148(3): 490-503.
- Li Q, Zhu Y, Shangguan W, Wang X, Li L and Yu F, 2022. An attention-aware LSTM model for soil moisture and soil temperature prediction. *Geoderma* 409, 115651.
- Meikle RW and Gilchrist AJ, (1983). A mathematical method for estimation of soil temperatures in England and Scotland. *Agricultural Meteorology* 30(3): 221-225.
- Mihalakakou G, 2002. On estimating soil surface temperature profiles. *Energy and Buil* 34: 251-259.
- Moazenazadeh R and Mohammadi B, 2019. Assessment of bio-inspired metaheuristic optimisation algorithms for estimating soil temperature. *Geoderma* 353: 152-171.
- Mohammadi B, 2022. Use of new methods to determine the inputs effective in estimating soil temperature. *Nivar* 46(118): 25-36.
- Pal M and Deswal S, 2010. Modelling pile capacity using Gaussian process regression. *Computers and Geotechnics* 37: 942-947.
- Quinlan JR, 1992. Learning with continuous classes. Pp.343-348. 5th Australian Joint Conference on Artificial Intelligence, Hobart, Australia.
- Rabet A, Dastranj A, Asadi S and Asadi Nalivan O, 2021. Determination of groundwater potential using artificial neural network, random forest, support vector machine and linear regression models (Case study: Lake Urmia watershed). *Ecohydrology* 7(4): 1047-1060 (in Persian with English abstract).
- Sabziparvar AA, Tabari H and Aeni A, 2010. Estimation of mean daily soil temperature by means of meteorological data in some selected climates of Iran. *JWSS-Isfahan University of Technology* 14(52): 125-138 (in Persian with English abstract).
- Sabziparvar AA, Zare Abyaneh H and Bayat Varkeshi M, 2010. A model comparison between predicted soil temperatures using ANFIS model and regression methods in three different climates. *Journal of Water and Soil* 24(2): 274-285 (in Persian with English abstract).
- Samadianfard S, Ghorbani MA 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(4): 465-476.
- Sanikhani H, Deo RC, Yaseen ZM, Eray O and Kisi O, 2018. Non-tuned data intelligent model for soil temperature estimation: A new approach. *Geoderma* 330: 52-64.
- Schimel JP, Bilbrough C and Welker JM, 2004. Increased snow depth affects microbial activity and nitrogen mineralization in two Arctic tundra communities. *Soil Biology and Biochemistry* 36(2): 217-227.
- Schmidt L, Heße F, Attinger S and Kumar R, 2020. Challenges in applying machine learning models for hydrological inference: A case study for flooding events across Germany. *Water Resources Research* 56(5): 025924.
- Seyfried M, Harris R, Marks D and Jacob B, 2001. Geographic database, Reynolds Creek Experimental Watershed, Idaho, United States. *Water Resources Research* 37(11): 2825-2829.
- Sihag P, Mohsenzadeh Karimi S and Angelaki A, 2019. Random forest, M5P and regression analysis to estimate the field unsaturated hydraulic conductivity. *Applied Water Science* 9: 1-9.
- Wahyunggoro O, Permanasari AE and Chamsudin A, 2013. Utilization of neural network for disease forecasting. 59th ISI World Statistics Congress, Hong Kong.
- Wu X, Yao Z, Bröggemann N, Shen ZY, Wolf B, Dannenmann M, Zheng X and Butterbach-Bahl K, 2010. Effects of soil moisture and temperature on CO₂ and CH₄ soil-atmosphere exchange of various land use/cover types in a semi-arid grassland in Inner Mongolia, China. *Soil Biology and Biochemistry* 42(5): 773-787.
- Yin X and Arp PA, 1993. Predicting forest soil temperatures from monthly air temperature and precipitation records. *Canadian Journal of Forest Research* 23(12): 2521-2536.
- Zeng L, Hu S, Xiang D, Zhang X, Li D, Li L and Zhang T, 2019. Multilayer soil moisture mapping at a regional scale from multisource data via a machine learning method. *Remote Sensing* 11 (3): 284.
- Zhang Z and Li J, 2019. *Big Data Mining for Climate Change*, Elsevier, Amsterdam, The Netherlands.
- 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(6): 707-718.