JOURNAL OF SOUTHERN COMMUNICATION ENGINEERING
ISLAMIC AZAD UNIVERSITY BUSHEHR BRANCH
E-ISSN: 2980-9231
https://jce.bushehr.iau.ir
https://doi.org/10.30495/jce.2023.1996606.1224

Vol. 13/ No. 52/Summer 2024

Research Article

Optimizing Solar Radiation Prediction Based on The Internet of Things Platform in Photovoltaic Power Plant

Neda Ashrafi Khozani, Instructor[1] | Maryam Mahmoudi, Instructor[2*] | Shabnam Nasr Esfahani, Assistant Professor[3]

[1]Department of Computer Engineering, Meymeh Branch, Islamic Azad University, Meymeh, Iran, anahid.shaabani@gmail.com

[2]Department of Computer Engineering, Meymeh Branch, Islamic Azad University, Meymeh, Iran, Mahmoudi.m174@iau.ac.ir

[3]Department of Electrical Engineering, Meymeh Branch, Islamic Azad University, Meymeh, Iran, nasr.sh.2010@gmail.com

Correspondence
Maryam Mahmoudi, Instructor of Computer Engineering, Meymeh Branch, Islamic Azad University, Meymeh, Iran, Mahmoudi.m174@iau.ac.ir

Received: 16 September 2023
Revised: 15 December 2023
Accepted: 22 December 2023

Abstract

The solar radiation value parameter is one of the most important parameters in determining the output power value of photovoltaic panels. Accurate prediction of this parameter is crucial for dispatching and load management planning. Managers and designers encounter economic and managerial challenges due to the uncertainty and difficulty in predicting solar radiation levels. This research introduces a highly accurate prediction method utilizing tree-based methods, enhanced by meta-heuristic algorithms to boost performance. The proposed method emphasizes preventing overfitting and ensuring high reliability for use in Internet of Things systems. Meta-heuristic algorithms are utilized for optimizing tree-based methods, as well as for feature and instance selection. Employing meta-heuristic methods as the main innovation in this research not only optimizes machine learning model settings but also mitigates the impact of noise, outliers, and ineffective inputs, thereby enhancing the final output quality. Utilizing an innovative fitness function in model optimization enhances prediction accuracy and adaptability to real photovoltaic power plant environments. The final outcome is a strong model that has a score of 0.95 with the R-square criterion and is optimal model.

Keywords: Internet of Things, Decision Tree, Machine Learning, Bat Algorithm, Photovoltaic Power Plants.

Highlights

- Accurate prediction of the amount of solar radiation as an important parameter in determining the amount of output power of photovoltaic panels.
- Optimizing tree-based models with meta-heuristic algorithms for prediction of solar radiation parameter.
- Balance the accuracy, complexity, and suitability of the final model for IoT device implementation.
- In the end, a strong model that has a score of 0.95 with the R-square criterion was obtained in this research.
- The final model can be implemented in the environment of power plants based on the Internet of Things.

Citation: N. Ashrafi Khozani, M. Mahmoudi, and Sh. Nasr Esfahani, "Optimizing Solar Radiation Prediction Based on The Internet of Things Platform in Photovoltaic Power Plant," *Journal of Southern Communication Engineering*, vol. 13, no. 52, pp. 33–44, 2024, doi: 10.30495/jce.2023.1996606.1224, [in Persian].

مقاله پژوهشی

بهینه‌سازی پیش‌بینی تابش خورشیدی بر اساس بستر اینترنت اشیاء در نیروگاه فتوولتائیک

ندا اشرفی خوزانی[۱] | مریم محمودی*[۲] | شبنم نصر اصفهانی[۳]

[۱] دانشکده مهندسی کامپیوتر، واحد میمه، دانشگاه آزاد اسلامی، میمه، ایران، anahid.shaabani@gmail.com

[۲] دانشکده مهندسی کامپیوتر، واحد میمه، دانشگاه آزاد اسلامی، میمه، ایران، Mahmoudi.m174@iau.ac.ir

[۳] دانشکده مهندسی برق، دانشگاه آزاد اسلامی، واحد میمه، میمه، ایران
nasr.sh.2010@gmail.com

نویسنده مسئول
*مریم محمودی، مربی، دانشکده کامپیوتر، واحدمیمه، دانشگاه آزاد اسلامی، میمه، ایران،
Mahmoudi.m174@iau.ac.ir

تاریخ دریافت: ۲۵ شهریور ۱۴۰۲
تاریخ بازنگری: ۲۴ آذر ۱۴۰۲
تاریخ پذیرش: ۱ دی ۱۴۰۲

چکیده:

پارامتر مقدار تابش خورشیدی یکی از مهم‌ترین پارامترها در تعیین مقدار توان خروجی پنل‌های فتوولتائیک است. پیش‌بینی دقیق این پارامتر برای برنامه‌ریزی در واحدهای دیسپچینگ(Dispatching) و مدیریت بار از اهمیت ویژه‌ای برخوردار است. عدم قطعیت در میزان تابش خورشیدی و سختی پیش‌بینی آن، طراحان را با چالش‌های اقتصادی و مدیریتی مواجه می‌کند. در این پژوهش یک روش پیش‌بینی با دقت بالا با استفاده از روش‌های مبتنی بر درخت و بهبود عملکرد این روش‌ها به کمک الگوریتم‌های فرا ابتکاری ارائه می‌شود. تأکید اصلی در روش پیشنهادی عدم بیش-برازش و قابلیت اتکای بالا و قابلیت به‌کارگیری در سیستم‌های اینترنت اشیاء است. بنابراین، الگوریتم‌های فراابتکاری در بهینه‌سازی روش‌های مبتنی بر درخت و همینطور در انتخاب ویژگی و انتخاب نمونه‌ها نیز دخیل شده‌اند. لذا استفاده از روش‌های فراابتکاری به عنوان جنبه‌ی نوآوری اصلی این پژوهش، نه تنها استفاده برای به دست آوردن تنظیمات بهینه‌ی مدل‌های یادگیری ماشین بلکه در کاهش اثر نویزها، داده‌های پرت(Outlier) و ورودی‌های کم‌اثر نیز به بهبود کیفیت خروجی نهایی کمک کرده است. به علاوه مناسب‌سازی نتایج پیش‌بینی برای استفاده عملی در محیط نیروگاه‌های فتوولتائیک موضوع پراهمیتی است. این موضوع که از طریق تابع برازش نوآورانه این پژوهش در بهینه‌سازی مدل‌ها انجام پذیرفته است، باعث می‌شود که خروجی نهایی علاوه بر دقت بالا از نظر سهولت پیاده‌سازی در محیط‌های واقعی نیروگاه‌های فتوولتائیک نیز بهینه باشد. خروجی نهایی، یک مدل قوی است که با معیار مربع-R دارای امتیاز۰/۹۵ است و از نظر سادگی تا حد زیادی بهینه است.

کلید واژه‌ها: الگوریتم خفاش، اینترنت اشیاء، درخت تصمیم، نیروگاه‌های فتوولتائیک، یادگیری ماشین

https://doi.org/10.30495/jce.2023.1996606.1224

۱-مقدمه

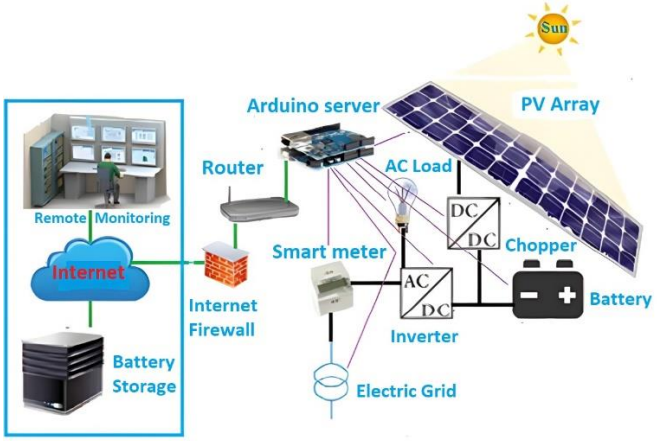
به طور کلی، سیستم‌های فتوولتائیک به سیستم‌هایی اطلاق می‌شود که از نور خورشید برای تولید برق استفاده می‌کنند. این سیستم‌ها شامل مجموعه‌ای از ماژول‌های فتوولتائیک(سلول‌های خورشیدی) هستند که نور خورشید را به انرژی الکتریکی

تبدیل می‌کنند. استفاده از سیستم‌های فتوولتائیک[1] به عنوان یکی از انواع منابع تولید انرژی پاک، در سال‌های اخیر به طور چشمگیری در بسیاری از کشورها افزایش یافته است. اما این سیستم‌ها دارای عدم قطعیت در تولید انرژی هستند که می‌تواند برای شبکه‌های برق مشکلاتی ایجاد کند. برای مثال، افزایش تولید انرژی در سیستم‌های فتوولتائیک در روزهای آفتابی باعث افزایش ولتاژ در شبکه می‌شود و در صورت عدم مدیریت مناسب می‌تواند باعث اتلاف انرژی، خرابی تجهیزات و حتی ایجاد نقص در شبکه شود. بنابراین، برای جلوگیری از این مشکلات، مدیریت دقیق تولید انرژی سیستم‌های فتوولتائیک و پیش‌بینی دقیق عدم قطعیت در تولید انرژی این سیستم‌ها برای شبکه‌های برق بسیار مهم و ضروری است.

اینترنت اشیاء[2] امروزه نقشی حیاتی در جنبه‌های مختلف زندگی انسان‌ها از کشاورزی و صنعت تا پزشکی و شهرهای هوشمند بازی می‌کند. به صورت خلاصه اینترنت اشیاء فناوری نوینی است که در آن برای هر موجودیتی(اعم از انسان و یا اشیاء) قابلیت دریافت و ارسال داده از طریق شبکه‌ی اینترنت یا اینترانت، فراهم می‌گردد. اینترنت اشیاء در واقع زیرساختی از شبکه پویای خود تنظیم شونده با پروتکل‌های ارتباطی استلندارد و یکپارچه است که در آن اشیاء فیزیکی و مجازی، دارای هویت، ویژگی‌های فیزیکی و شخصیت‌های مجازی هستند و به صورت یکپارچه درون زیرساخت اطلاعاتی تجمیع شده‌اند. نیروگاه‌های خورشیدی فتوولتائیک نیز از این موضوع مستثنی نیستند و سیستم‌های اینترنت اشیاء با هدف پایش لحظه‌ای و جلوگیری از بروز خطاهای احتمالی در این نیروگاه‌ها مورد توجه قرار گرفته‌اند.

به عنوان نمونه، یک سیستم کنترل نیروگاه فتوولتائیک در سال ۲۰۱۸ در[۱] پیشنهاد شده است که دارای سه لایه مختلف است. لایه‌ی اول شامل تجهیزات محیطی در مکان نیروگاه اعم از پنل‌های فتوولتائیک، سنسورها، سوئیچ‌ها و غیره است. لایه‌ی دوم شامل سرور مرکزی و روتر برای اتصال امن به شبکه است و آخرین لایه شامل اینترنت و پایگاه‌های داده است که کنترل و پایش این سیستم را به عهده دارد. معماری کلی این سیستم در شکل ۱ نمایش داده شده است. در این معماری، لایه اول شامل محیط طراحی سیستم فتوولتائیک است که در آن تمام اجزا مطابق با تنظیمات مورد نیاز برای تکمیل کامل نیازهای کاربر به هم متصل شدند. لایه دوم که به عنوان اتصال بین طراحی‌های سخت‌افزاری سیستم فتوولتائیک به وب سرور از طریق یک روتر با گزینه فایروال اینترنتی عمل می‌کند. اطلاعات از سرور به لایه سوم که لایه نظارت و کنترل از راه دور است داده می‌شود. در اینجا سرور اطلاعات جمع‌آوری شده در مورد سیستم فتوولتائیک را به دستگاه‌های ذخیره‌سازی که می‌توان از آن‌ها گزارش‌های دوره‌ای تولید کرد، تحویل می‌دهد.

از آنجایی که با توجه به این معماری و روش‌های مشابه، کنترل سیستم‌های فتوولتائیک به کمک اینترنت اشیا مورد توجه قرار گرفته است، نیاز به توسعه مدل‌های متناسب با اینترنت اشیا که پارامترهای مهم را به منظور کنترل دقیق‌تر این نیروگاه‌ها پیش‌بینی می‌کند از اهمیت ویژه‌ای برخوردار است.



شکل ۱: معماری اینترنت اشیا برای کنترل نیروگاه فتوولتائیک[۱]

Figure 1. Internet of Things architecture for photovoltaic power plant control[1]

[1] Photovoltaic

[2] Internet of Things

پارامتر مقدار تابش خورشیدی یکی از مهم‌ترین پارامترها در تعیین مقدار توان خروجی پنل‌های فتوولتائیک است. این پارامتر که به شدت یا توان تابشی خورشید نیز شناخته می‌شود، با میزان انرژی تولیدی در نیروگاه‌های فتوولتائیک رابطه مستقیم دارد و آن را با W/m^2 نشان می‌دهند. اگر شدت تابش را در پارامتر زمان ضرب کنیم به انرژی تابشی خورشید می‌رسیم و واحد آن را با Wh/m^2 یا Kwh/m^2 نشان می‌دهند. معادله ۱، رابطه بین تابش و جریان و توان ماژول‌ها را نمایش می‌دهد که در آن، G1 و G2 میزان تابش‌ها (بر حسب W/m2)، I1 و I2 جریان متناظر ماژول‌ها (در A) هستند و P1 و P2 توان حاصل در هنگام تغییر تابش (بر حسب W) هستند. لازم به ذکر است که در این پژوهش تابش خورشیدی به عنوان پارامتر وابسته در نظر گرفته شده که مقادیر آن توسط سنسورهای مختلف اینترنت اشیا به همراه پارامترهای مستقلی که می‌توانند در آن مؤثر باشند اندازه‌گیری شده و مجموعه داده‌های مورد استفاده را تشکیل می‌دهد.

$$\frac{G1}{G2}=\frac{I2}{I1}=\frac{P2}{P1} \quad (۱)$$

پیش‌بینی دقیق این پارامتر برای برنامه‌ریزی در واحدهای دیسپچینگ[1] و مدیریت بار از اهمیت ویژه‌ای برخوردار است. این پارامتر همچنین در حفظ پایداری ولتاژ و فرکانس شبکه و آماده‌سازی نیروگاه‌های پشتیبان نیز حائز اهمیت است. عدم قطعیت در میزان تابش خورشیدی و سختی پیش‌بینی کوتاه مدت آن، مدیران و طراحان سیستم‌هایی که از انرژی خورشیدی بهره می‌گیرند را با چالش‌های اقتصادی و مدیریتی مواجه می‌کند. از این رو یک چارچوب دقیق و دارای عمومیت بالا که بتواند در سیستم‌های اینترنت اشیایی که محیط‌های این نیروگاه‌ها را کنترل می‌کنند به خوبی اجرا شود، می‌تواند کمک شایانی در این زمینه داشته باشد [۲]. در این پژوهش، یک چارچوب پیش‌بینی تابش خورشیدی با ویژگی‌های زیر ارائه می‌شود:

- قابلیت پیاده‌سازی در سیستم‌های اینترنت اشیا با مصرف بهینه منابع
- دقت و عمومیت بالا
- عدم وجود مشکل بیش برازش[2]

اهمیت قابل‌توجه تحلیل تابش خورشیدی حتی پیش از اینکه مدل‌های یادگیری ماشین در حوزه‌های مختلف علمی و مهندسی به طور گسترده به‌کار گرفته شوند، دانشمندان را بر آن داشت که مدل‌هایی برای تحلیل این پدیده ارائه کنند. مدل انگستروم [۳] و حالت ارتقا یافته‌ی آن یعنی مدل انگستروم-پرسکات [۴] را می‌توان اولین تلاش‌ها در این زمینه دانست. در طی سالیان متمادی، روش‌های متعدد دیگری بر مبنای همین پژوهش‌ها ارائه شد که برخی از آن‌ها در منابع [۵،۷] قابل مشاهده هستند. اما آنچه برای ما به عنوان پیشینه این پژوهش از اهمیت بیشتری برخوردار است استفاده از روش‌های یادگیری ماشین در این زمینه است که در ادامه به برخی از آن‌ها می‌پردازیم. شبکه‌های عصبی مصنوعی و سایر روش‌های مبتنی بر آن، پایه بسیاری از پژوهش‌ها در این زمینه قرار گرفته‌اند. از جمله ملیت و همکارانش در[۲] یک مدل شبکه عصبی برای مدل‌سازی داده‌های تابش خورشیدی با کاربرد سایزینگ[3] سیستم فتوولتائیک ارائه کرده‌اند. همچنین، کائو و همکارانش در[۸]، یک روش ترکیبی با استفاده از شبکه‌های عصبی مصنوعی و تحلیل موجک برای پیش‌بینی تابش خورشیدی در یک شبانه‌روز را توسعه داده‌اند. داده‌های استفاده شده در این تحقیق دیتای تاریخی روزانه مقدار تابش است. در[۹] نیز مدل شبکه‌های عصبی نوع LSTM[4] برای پیش‌بینی این پارامتر استفاده شده است با این تفاوت که داده‌های مورد استفاده در آن داده‌های هواشناسی مانند دما، رطوبت و غیره است. روش متفاوتی در[۱۰] برای پیش‌بینی نوسانات تابش ارائه شده است که بر تصویربرداری از آسمان مبتنی است. بازه زمانی پیش‌بینی در این تحقیق ۵ تا ۱۰ دقیقه است و بر همین اساس، پیش‌بینی کوتاه مدت به شمار می‌آید. تکنیک یادگیری ماشینی که در این تحقیق به کار رفته است، شبکه‌های عصبی کانولوشنی عمیق[5]

---

[1] Dispatching
[2] Overfitting
[3] Sizing
[4] Long-Short Term Memory
[5] Deap convolutional neural network

است. همچنین در[۱۱] مدلی را برای پیش‌بینی تابش خورشیدی در بازه‌های کوتاه مدت معرفی شده است که می‌تواند برای هر موقعیت جغرافیایی مورد استفاده قرار بگیرد، بدون اینکه به اندازه‌گیری‌های میدانی نیاز داشته باشد. این روش به جای اینکه استفاده از اطلاعات ماهواره‌ای کند پیش‌بینی‌های هواشناسی را در قالب شبکه‌های عصبی عمیق استفاده کرده است. همچنین در رویکردی که در[۱۲] پیشنهاد شده است، مدلی برای پیش‌بینی تابش با استفاده از شبکه‌های عصبی نوع کانولوشنی ارائه شده است. داده‌های ورودی به صورت ۱۵ حالت هواشناسی شامل ابری، نیمه‌ابری، صاف و غیره است که از داده‌های ایستگاه هواشناسی نمونه‌برداری می‌شوند. به منظور بهینه‌سازی‌های ابرپارامترهای شبکه‌ی عصبی پیشنهادی، از الگوریتم ترکیبی PSO/GA[1] استفاده شده است که از این جنبه شباهت بیشتری به پژوهش موجود دارد. ماشین بردار پشتیبان نیز در[۱۳] به منظور مدل‌سازی و پیش‌بینی تابش در بازه‌های زمانی ۳ ساعته به کار گرفته شده‌است. این پژوهش به طور خاص برای خانه‌های هوشمند که از انرژی خورشیدی استفاده می‌کنند طراحی شده است و پارامترهایی همچون دما، روز، سال، سرعت باد، پوشش ابر و غیره برای انجام پیش‌بینی مورد استفاده قرار گرفته‌اند.

در پژوهش‌های مشابه انجام شده که به مهم‌ترین آن‌ها اشاره شد، عموماً به دست آوردن یک مدل پیش‌بینی تابش خورشیدی با دقت بالا هدف قرار گرفته است. به همین دلیل، هرچند مدل‌های دقیقی برای پیش‌بینی تابش خورشیدی به دست آمده است در بسیاری از موارد فاقد تحلیل‌های مربوط به عمومیت مدل‌ها (بیش-برازش) هستند. از طرفی امکان عملی استفاده از مدل‌ها در سیستم‌های اینترنت اشیای در محیط نیروگاه‌های فتوولتائیک در مدل‌سازی در نظر گرفته نشده است. لذا در این پژوهش سعی شده است که با انتخاب مدل مناسب و روش مناسب برای بهینه‌سازی مدل به دست آمده دارای سرعت و سهولت پیاده‌سازی کافی برای استفاده در این نوع محیط‌ها مدنظر قرار داده شود. به طور خلاصه در این پژوهش، استفاده از روش‌های فرا ابتکاری به عنوان جنبه اصلی نوآوری، نه تنها برای بدست آوردن تنظیمات بهینه مدل‌های یادگیری ماشین بلکه به منظور کاهش اثر نویزها، داده‌های پرت و ورودی‌های کم اثر نیز به بهبود کیفیت خروجی نهایی مورد استفاده قرار گرفته است. علاوه بر این، مناسب‌سازی نتایج پیش‌بینی برای استفاده عملی در محیط‌های نیروگاه‌های فتوولتائیک، که در این پژوهش مورد توجه قرار گرفته است، جنبه‌ی نوآوری پراهمیتی است. با استفاده از تابع برازش نوآورانه مورد استفاده در این پژوهش برای بهینه‌سازی مدل‌ها، خروجی نهایی علاوه بر داشتن دقت بالا، از نظر سهولت پیاده‌سازی در محیط‌های واقعی نیروگاه‌های فتوولتائیک نیز بهبود یافته است.

۲-روش پیشنهادی

در این پژوهش، هدف اصلی توسعه‌ی یک مدل دقیق برای پیش‌بینی تابش خورشیدی است که قابلیت استفاده در اینترنت اشیا را دارد. لذا مدل‌های گروهی(تجمیعی) مبتنی بر درخت تصمیم به عنوان پایه‌ی اصلی مدل‌سازی انتخاب شدند. دلیل این انتخاب تفسیرپذیر بودن ذاتی روش‌های مبتنی بر درخت است. این روش‌ها در واقع از تعدادی درخت تصمیم مختلف که در کنار یکدیگر، یک مدل قوی‌تر را می‌سازند، تشکیل می‌شوند. در این بخش ابتدا به معرفی بلوک‌های سازنده چارچوب این پژوهش پرداخته می‌شود و در پایان روش پیشنهادی ارائه می‌شود.

۲-۱-پیش پردازش

پیش‌پردازش داده‌ها یک مرحله مهم در فرآیند یادگیری ماشین است که شامل تبدیل و تهیه داده‌ها قبل از ورود به مدل یادگیری می‌شود. دو فعالیت مهم این مراحل مهم شامل حذف داده‌های پرت و نرمال‌سازی هستند. حذف داده‌های پرت به منظور حذف داده‌های نامناسب، نادرست یا غیرعادی استفاده می‌شود. این داده‌ها ممکن است ناقص یا با مقادیر غیرمنطقی باشند و ممکن است تحت تأثیر خطاهای اندازه‌گیری، نویزهای تصادفی یا خطاهای دیگر قرار گرفته باشند. حذف این داده‌ها می‌تواند باعث بهبود کیفیت داده‌ها و دقت مدل یادگیری شود، زیرا داده‌های پرت ممکن است در تصمیم‌گیری مدل تأثیر منفی داشته باشند [۸].

[1] Particle Swarm Optimization/Genetic Algorithm

نرمال‌سازی داده‌ها به منظور قرار دادن داده‌ها در یک مقیاس یکسان و قابل مقایسه استفاده می‌شود. این فرآیند معمولاً با استفاده از روش‌های مختلف مانند مقیاس‌بندی، زیرگروه‌بندی یا تبدیلات ریاضی انجام می‌شود. با نرمال‌سازی داده‌ها، واحدهای مختلف و مقیاس‌های متفاوت داده‌ها به یک مقیاس مشترک تبدیل می‌شوند. این موضوع باعث می‌شود که مدل‌های یادگیری ماشین بتوانند به خوبی الگوها و روابط بین داده‌ها را درک کنند و عملکرد بهتری در پیش‌بینی و تصمیم‌گیری داشته باشند [۱۴].

۲-۲-درخت تصمیم

مدل درخت تصمیم یک الگوریتم یادگیری ماشین است که برای مسائل طبقه‌بندی و پیش‌بینی استفاده می‌شود. این مدل بر اساس یک ساختار درختی تصمیم‌گیری ساخته می‌شود و به صورت سلسله مراتبی قواعد و شرایط را بررسی می‌کند تا به تصمیمات نهایی برسد. در مدل درخت تصمیم، هر گره از درخت نمایانگر یک ویژگی است که به عنوان یک شرط بر روی داده‌ها بررسی می‌شود. با توجه به مقدار ویژگی در داده‌ها، درخت به شاخه‌های مختلف تقسیم می‌شود [۱۵] هر شاخه از درخت نمایانگر یک مقدار ممکن برای ویژگی و یا یک مجموعه از مقادیر است. در نهایت، برگ‌ها (گره‌های پایانی درخت که فاقد گره‌های فرزند هستند) نمایانگر کلاس‌های طبقه‌بندی شده یا مقدار پیوسته پیش‌بینی شده برای داده‌ها هستند. مزیت اصلی مدل درخت تصمیم این است که آن راحت قابل فهم و تفسیر است و همچنین می‌تواند به صورت مجموعه‌ای از قوانین بیان شود. به علاوه، درخت تصمیم قابلیت کاربرد در داده‌های عددی و دسته‌ای را دارد و به خوبی با داده‌های کم‌کاربرد یا ناقص سازگاری دارد. در کل، مدل درخت تصمیم به عنوان یک ابزار قوی در یادگیری ماشین مورد استفاده قرار می‌گیرد که قابلیت طبقه‌بندی و پیش‌بینی دقیق را در مسائل مختلف ارائه می‌دهد [۱۶].

۲-۳-جنگل تصادفی

مدل جنگل تصادفی یک الگوریتم یادگیری ماشین است که بر پایه ترکیب چندین درخت تصمیم ساخته می‌شود. این مدل از ترکیب تصمیم‌های گوناگون چندین درخت تصمیم بهره می‌برد تا تصمیم‌های بهتر و قابل اطمینان‌تری را در مورد داده‌ها ارائه دهد. در مدل جنگل تصادفی، هر درخت تصمیم به صورت مستقل ساخته می‌شود. این درخت‌ها بر اساس تصادفی‌سازی در فرآیند ساخت شاخه‌بندی مورد استفاده قرار می‌گیرند. به عبارت دیگر، در هر مرحله از ساخت درخت، فقط یک زیرمجموعه تصادفی از ویژگی‌ها برای تقسیم داده‌ها استفاده می‌شود. این عمل تصادفی باعث ایجاد تنوع در درخت‌ها می‌شود و از بیش‌برازش جلوگیری می‌کند. در نهایت، تصمیم نهایی مدل جنگل تصادفی، بر اساس آرایه‌ای از تصمیم‌های هر درخت گرفته می‌شود. با ترکیب نتایج چندین درخت، مدل جنگل تصادفی قادر است به صورت هم‌زمان بر روی داده‌های آموزش و داده‌های تست پیش‌بینی دقیقی ارائه دهد [۱۷].

مزیت اصلی مدل جنگل تصادفی این است که از کاهش واریانس[1] و بیش‌برازش در مدل‌های تصمیم‌گیری تک درخت استفاده می‌کند. همچنین، این مدل قابلیت مقابله با داده‌های پرت، توانایی برخورد با داده‌های بزرگ، و قابلیت استخراج اهمیت ویژگی‌ها را داراست [۱۸].

۲-۴-ادابوست

مدل ادابوست[2] یک الگوریتم یادگیری ماشین مبتنی بر مجموعه‌های ضعیف است که با استفاده از ترکیب آن‌ها یک مدل قوی‌تر و دقیق‌تر را ایجاد می‌کند. ایده اصلی این الگوریتم این است که ترکیب چندین تصمیم‌گیری ضعیف به وسیله وزن‌دهی به هر یک از آن‌ها، منجر به ساخت یک تصمیم‌گیری قوی می‌شود [۱۹]. در مدل ادابوست، درخت‌های تصمیم به عنوان تصمیم‌گیری‌های ضعیف مورد استفاده قرار می‌گیرند. ابتدا هر درخت تصمیم با استفاده از مجموعه داده‌های آموزش ساخته

---

[1] Variance

[2] AdaBoost

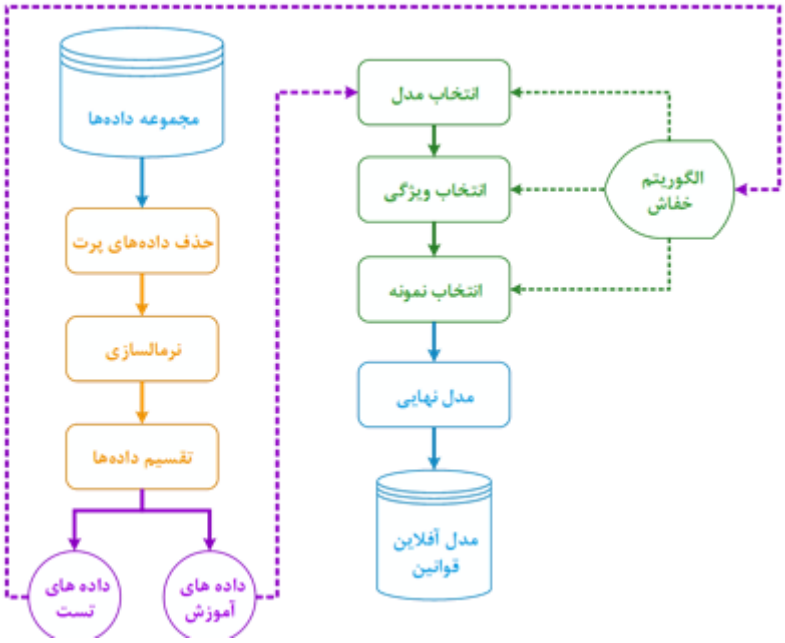
می‌شود. سپس به هر داده‌ی آموزشی وزنی اختصاص داده می‌شود و درخت با استفاده از این وزن‌ها آموزش داده می‌شود. در این مرحله، تمرکز بر داده‌هایی که به درستی تشخیص داده نشده‌اند، بیشتر می‌شود. سپس درخت تصمیم نهایی با توجه به وزن‌های داده‌ها ساخته می‌شود. به درخت‌هایی که توانستند داده‌های نادرست را به خوبی تشخیص دهند، در ادامه فرآیند وزن بیشتری اختصاص می‌دهند. این روند تکرار می‌شود و در هر مرحله، درختی جدید با توجه به وزن‌های جدید ساخته می‌شود. در نهایت، تمام درخت‌های ساخته شده با استفاده از روش اجماع ترکیب می‌شوند تا مدل نهایی ادابوست به دست آید. این مدل، با توجه به وزن‌های اختصاص داده شده به هر درخت، تصمیماتی را اتخاذ می‌کند که به دقت بیشتری در پیش‌بینی و تصمیم‌گیری می‌انجامد. در واقع، این وزن‌ها به مدل اجازه می‌دهند تا در تصمیم‌گیری بیشترین تأثیر را داشته باشند. به عبارت دیگر، درخت‌هایی که در تشخیص داده‌های پیچیده و دشوار موفقیت بیشتری دارند، نقش مهم‌تری در تصمیم‌گیری نهایی دارند [۲۰].

مزیت اصلی مدل ادابوست این است که با ترکیب چندین درخت تصمیم ضعیف، یک مدل قوی‌تر و مقاوم‌تر در برابر بیش‌برازش ایجاد می‌کند. همچنین، با تمرکز بر داده‌هایی که توسط درخت‌های قبلی اشتباه تشخیص داده شده‌اند، مدل توانایی بهبود دقت و عملکرد خود را افزایش می‌دهد. این الگوریتم قابلیت استفاده در مسائل دسته‌بندی و رگرسیون[1] را دارد و در بسیاری از موارد با عملکرد برتر نسبت به الگوریتم‌های تک درخت مانند درخت تصمیم استفاده می‌شود [۲۱].

۲-۵-روش پیشنهادی

در اینجا از دو روش جنگل تصادفی(که در آن تعدادی درخت تصمیم به صورت مستقل از هم پیش‌بینی را انجام می‌دهند و در مدل نهایی نتایج آن‌ها با هم ترکیب می‌شود) و ادابوست(که در آن هر مدل درخت تصمیم بر اساس داده‌هایی که در مراحل قبل با خطای بالا پیش بینی شده است توسعه می‌یابد) استفاده می‌کنیم. اما از آنجایی که این مدل‌ها خود دارای ابر-پارامترهایی هستند که مقدار آن‌ها تأثیر مستقیمی در نتیجه نهایی دارد، انتخاب این پارامترهای حیاتی و همچنین انتخاب ویژگی‌ها(ورودی‌ها) و نمونه‌های مورد استفاده در مرحله‌ی آموزش از اهمیت خاصی برخوردار است. به عنوان جنبه‌ی اصلی نوآوری، در این پژوهش بر مبنای ایده‌ی مطرح شده در [۱۹] برای بهینه‌سازی این موارد از الگوریتم فرا ابتکاری خفاش [۲۲] استفاده شده است که همانند الگوریتم ژنتیک یک الگوریتم الهام گرفته شده از طبیعت است و برای مسائل بهینه‌سازی بسیار پرکاربرد است. در ادامه به توضیح جزئیات روش پیشنهادی که معماری کلی آن در شکل ۲ نمایش داده شده است، می‌پردازیم. یکی از اولین اقدامات در هر مسئله یادگیری نظارت‌شده، پیش‌پردازش است. در اینجا در مرحله پیش‌پردازش ابتدا داده‌های پرت شناسایی و حذف می‌شوند. به این منظور از روش فاصله‌ی کوک[۲۴،۲۳] استفاده می‌شود. به علاوه در این مرحله عمل نرمال‌سازی داده‌ها نیز انجام می‌شود، عدم نرمال‌سازی ورودی‌ها می‌تولند باعث شود که ویژگی‌هایی که بازه‌ی تغییر گسترده‌تری دارند تأثیر زیادی در مدل نهایی به صورت غیر عادی ایفا کرده و از دقت مدل بکاهد. لذا مقادیر همه ورودی‌ها در این مرحله نرمال‌سازی می‌شوند. در آخرین فعالیت از مراحل پیش‌پردازش، داده‌ها به دو زیر مجموعه داده‌های آموزش و داده‌های تست تقسیم می‌شوند. از آنجایی که داده‌های تست برای آزمایش و محاسبه‌ی معیارهای مختلف در مدل‌های تولید شده مورد استفاده قرار می‌گیرند و در مرحله آموزش نقشی ندارد، تعداد این داده‌ها باید محدود و بسیار کوچک‌تر از داده‌های آموزش باشد. لذا این تقسیم با نسبت ۹ به ۱ انجام می‌پذیرد و صرفاً ۱۰ درصد از داده‌ها به صورت تصادفی به عنوان داده‌های تست انتخاب می‌شوند.

[1] Regression



شکل ۲: معماری کلی روش پیشنهادی

Figure 2. The general architecture of the proposed method

پس از پیش‌پردازش، نوبت به فعالیت‌های سه‌گانه‌ای می‌رسد که تحت عنوان کلی انتخاب مدل شناخته می‌شوند. اما این فعالیت‌ها صرفاً شامل انتخاب یک الگوریتم برای انجام مدل‌سازی نیستند. در این پژوهش انتخاب الگوریتم مورد استفاده(یکی از دو الگوریتم جنگل تصادفی و ادابوست که پیش‌تر ذکر شد)، انتخاب مقادیر ابر پارامترها و انتخاب نمونه‌ها و ویژگی‌های مورد استفاده در این مرحله انجام می‌پذیرد. در واقع اگر هر ترکیب از مواردی که ذکر شد را به عنوان یک راه حل در نظر بگیریم این راه حل یک آرایه عددی است که اولین عنصر آن الگوریتم مورد استفاده را به صورت ۰ و ۱ مشخص می‌کند. سپس ۵ عنصر بعدی مقادیر ابر پارامترها را مشخص می‌کند. این ابر پارامترها عبارتند از،

- تعداد درخت تصمیم(مدل‌های پایه)
- حداکثر عمق درخت تصمیم
- حداکثر تعداد ویژگی مورد بررسی در هر گره از درخت تصمیم
- تعداد نمونه لازم برای تشکیل یک انشعاب در درخت تصمیم
- نرخ یادگیری(برای ادابوست)

سایر عناصر یک راه‌حل نیز حضور-عدم حضور ویژگی‌ها و نمونه‌های موجود در مجموعه داده‌های آموزش را برای الگوریتم مشخص می‌کند. بنابراین با هر یک راه حل که در بالا توضیح داده شد می‌توانیم یک مدل کامل بسازیم که بر اساس مقادیر موجود در آرایه ذکر شده پیکره‌بندی شده است. با این توضیحات در واقع ما یک مساله بهینه‌سازی داریم که با روش‌های تکاملی و فرا ابتکاری قابل حل است. در اینجا ما الگوریتم خفاش را انتخاب کرده‌ایم. انگیزه‌ی این انتخاب همگرایی سریع این الگوریتم است.

از آنجایی که در مدل نهایی به دست آمده از این پژوهش باید علاوه بر دقت پیش‌بینی بالا، تا جای ممکن از پیچیدگی کمی برخوردار باشد. لذا در مقایسه ترکیب‌های مختلف با یکدیگر علاوه بر پارامتر دقت مدل باید تعداد درخت تصمیم و حداکثر عمق درخت را نیز مورد بررسی قرار دهیم. به دست آوردن نوعی تعادل بین دقت مدل و پیچیدگی آن باعث می‌شود مدل نهایی از نظر مصرف منابع محاسباتی در زمان پیاده‌سازی در این اینترنت اشیاء بهینگی بیشتری داشته باشد. به بیان دیگر، داشتن مدلی که علاوه بر دقت بالا دارای پیچیدگی بهینه نیز باشد هدف قابلیت پیاده‌سازی در سیستم‌های کنترلی مبتنی بر

اینترنت اشـیا را تأمین می‌کند. بر همین اسـاس، تابع برازش را با اسـتفاده از امتیاز R^2 که نمایانگر دقت مدل اسـت و تعداد درخت تصمیم و حداکثر عمق که پیچیدگی مدل را نشان می‌دهند، این‌گونه تعریف می‌کنیم،

$$Fitness = \frac{100R^2}{MaxDept + MaxTree} \quad (۲)$$

$$R^2 = \frac{(n\sum t_i o_i - \sum t_i \sum o_i)^2}{(n\sum {t_i}^2 - (\sum t_i)^2) - (n\sum {o_i}^2 - (\sum o_i)^2)} \quad (۳)$$

در معادلات فوق n تعداد کل نقاط داده را مشـخص می‌کند، t مقدار اندازه‌گیری شـده(مقدار مورد انتظار) را نشـان می‌دهد و o مقدار پیش‌بینی شده توسط مدل را مشخص می‌کند.

پس از انجام بهینه‌سـازی، یک مدل نهایی به‌دسـت می‌آید که بهترین و دقیق‌ترین پیش‌بینی را از میان راه حل‌های مختلف تولید شـده انجام می‌دهد. در این مرحله با توجه به اینکه هر درخت تصـمیم(مدل پایه‌ی روش‌های مورد اسـتفاده) می‌تواند به صـورت مجموعه‌ای از قوانین بیان شـود، می‌توانیم قوانین همه‌ی درخت‌های تولید شـده توسـط مدل نهایی را اسـتخراج کرده و مجموعه قوانین از پیش تعیین شده‌ای را برای استفاده در محیط اینترنت اشیاء داشته باشیم. این قوانین با توجه به اینکه ثابت بوده و نیازی به تغییر منظم و مدل‌سـازی در سـمت گره‌های اینترنت اشـیا نیسـت. بنابراین سـربار اضـافی نیز برای پیش بینی تابش خورشید اعمال نمی‌شود.

با توجه به امکان استخراج قوانین در روش پیشـنهادی، این مدل به صورت خودکار برای تشخیص آفلاین تابش به کار می‌رود و می‌تواند در برابر شـکسـت بسـتر مخابراتی و قطعی ارتباطات کمک کننده باشـد. در واقع با این روش، قوانین پیش‌بینی‌کننده می‌تواند جایگزین توسـعه تکراری مدل‌ها شـده و این می‌تواند به بهبود قابل توجه ارتباطات و کاهش ریسـک ناشـی از قطعی ارتباطات در برابر شکست بستر مخابراتی منجر شود.

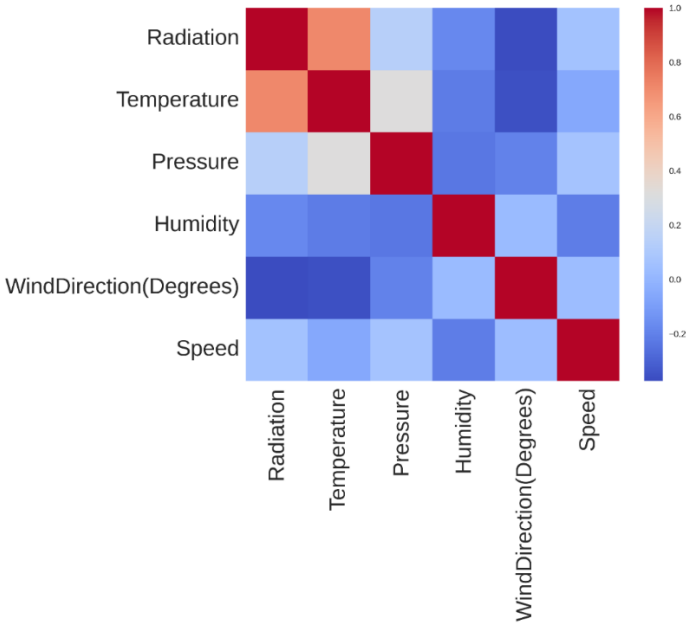
۲-۶-مجموعه داده‌ها و روش ارزیابی

به منظور آزمایش روش پیشـنهادی که در بخش قبل بیان شـد، مجموعه داده‌های ارائه شـده در یکی از چالش‌های ناسـا در پیش‌بینی شدت تابش در نظر گرفته شده است که در تعدادی از کارهای مشابه مانند [۲۵] استفاده شده است. این مجموعه داده‌ها مشـتمل بر بیش از ۳۰ هزار بردار داده اسـت که دارای ۱۱ سـتون(ویژگی) مختلف اسـت و در اکثر پژوهش‌های اخیر به عنوان یک مجموعه داده‌ای استاندارد به منظور مقایسه روش‌ها استفاده شده است. در این پژوهش ویژگی‌های مربوط به تاریخ و زمان ثبت رکورد مربوط حذف شـده‌اند تا مدل‌سـازی صـرفاً بر اسـاس پارامترهای آب و هوایی دیگر انجام شـود. در نهایت خروجی هدف مقدار تابش و ورودی‌های مدل‌سازی عبارتند از،

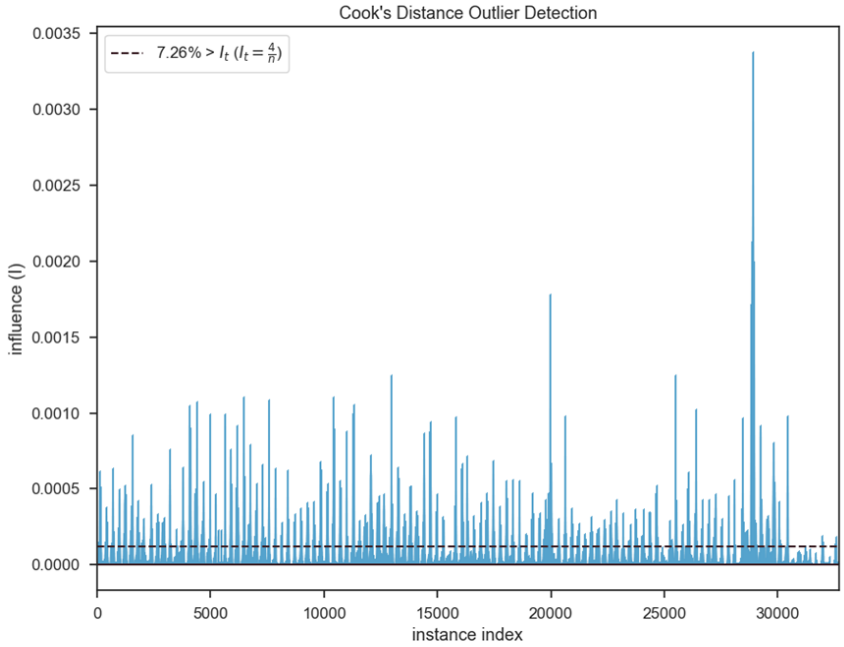
- دما
- فشار
- رطوبت هوا
- جهت باد
- سرعت باد

تحلیل ضرایب همبستگی پیرسون[1] پارامترهای ورودی و خروجی در شکل ۳ نمایش داده شده است. از این نمودار می‌توانیم نتیجه بگیریم که هیچ رابطه کاملاً خطی بین مقادیر پارامترها وجود ندارد. لذا با توجه به این‌که تعداد پارامترها به نسـبت کم اسـت، حذف هیچ‌کدام از ورودی‌ها نیاز نیسـت. هر چند در ادامه در فعالیت انتخاب ویژگی ممکن اسـت بعضـی از این ویژگی‌ها حذف شوند اما در اینجا از هیچ یک از ورودی‌ها صرف نظر نمی‌کنیم.

[1] Pearson



شکل ۳: تحلیل همبستگی پیرسون
Figure 3. Pearson correlation analysis



شکل ۴: تحلیل داده‌های پرت با روش فاصله کوک
Figure 4. Analyzing outlier data with Cook's distance method

۳-نتایج شبیه‌سازی

در این بخش نتایج روش پیشنهادی را که به کمک زبان برنامه‌نویسی پایتون پیاده‌سازی شده است بررسی می‌کنیم. در مرحله پیش‌پردازش تحلیل کوک همان طور که در بخش روش پیشنهادی گفته شد انجام شده که نتایج آن در شکل ۴ نمایش داده شده است. بر این اساس ۷/۲ درصد از داده‌ها به عنوان داده پرت تشخیص داده شد و از مرحله آموزش مدل‌ها کنار گذاشته شد. بعد از حذف داده‌های پرت، بر مبنای آنچه در روش پیشنهادی بیان شد با اجرای تکرارهای الگوریتم خفاش، بهترین پیکره‌بندی برای پیش‌بینی شدت تابش به دست آمد. خروجی نهایی این بهینه‌سازی مدل جنگل تصادفی با ابرپارامترهای نمایش داده شده در جدول ۱ است.

جدول ۱: پارامترهای مدل نهایی
Table 1. parameters of the final model

ابر پارامتر	تعداد درخت تصمیم	حداکثر عمق	حداکثر تعداد ویژگی	نمونه لازم برای انشعاب
مقدار	۸۹	۲۵	۴	**۳۲**

مدل نهایی به دست آمده دارای امتیاز R^2 معادل ۰/۹۵۶ در مرحله آموزش و ۰/۹۴۸ برای داده‌های تست است که نشان‌دهنده‌ی یک دقت قابل قبول برای پیش‌بینی شدت تابش است. لازم به ذکر است که در تکرارهای الگوریتم بهینه‌سازی تعدادی از مدل‌های کاندید دقت بالاتری نیز از خود نشان دادند اما با توجه به اینکه فاصله دقت آموزش و تست در آن‌ها قابل توجه بود(بیش برازش) و یا مدل پیچیدگی بالایی داشت(تعداد بالای درخت تصمیم)، این مدل‌ها با توجه به تابع برازشی که در این پژوهش تعریف کرده‌ایم به عنوان خروجی نهایی بهینه‌سازی انتخاب نشدند.

مدل نهایی جنگل تصادفی به دست آمده در این پژوهش دارای حداکثر عمق ۲۵ و تعداد درخت ۸۹ است، بنابراین استخراج همه‌ی قوانین برای پیش‌بینی شدت تابش به تعداد محدودی قانون منجر شده و می‌تواند به صورت آفلاین و مستقر در حافظه در دستگاه‌های اینترنت اشیاء مورد استفاده قرار گیرد.

همچنین به منظور اعتبارسنجی مدل نهایی از روش k-fold استفاده شد. این روش به کمک تقسیم داده‌ها به k بخش، هر بار یکی از بخش‌ها را به عنوان داده‌های اعتبارسنجی و بقیه بخش‌ها را به عنوان داده‌های آموزش استفاده کرده و مدل را با استفاده از داده‌های آموزش، آموزش داده و عملکرد آن را با داده‌های اعتبارسنجی ارزیابی می‌کند. این فرایند را k بار(در اینجا k=10) تکرار می‌کند و در هر مرحله، بخش‌های دیگر را به عنوان داده‌های آموزش و اعتبارسنجی استفاده می‌کند. سپس عملکرد مدل در هر بار تکرار، میانگین شده و به عنوان عملکرد نهایی مدل گزارش می‌شود. این روش اعتبارسنجی مدل را بهبود می‌بخشد و به دقت و استنباط قوی‌تری از عملکرد مدل در مواجهه با داده‌های تست می‌انجامد. نتیجه این الگوریتم اعتبارسنجی در مدل نهایی نشان دهنده انحراف معیار ۰/۰۰۱۴۸ در شاخص R^2 است. این عدد نشان می‌دهد که مدل به دست آمده از عمومیت بالایی برخوردار است.

جدول ۲: مقایسه با روش‌های موجود
Table 2. Comparison with existing methods

روش	میانگین امتیاز R^2 در ۱۰ تکرار	انحراف معیار R^2 در ۱۰ تکرار
روش پیشنهادی	۰/۹۵۶	**۰/۰۰۱۴۸**
شبکه‌های عصبی مصنوعی[۲]	۰/۹۴۸	**۰/۰۰۹۴۲**
شبکه‌های عصبی مصنوعی و تحلیل موجک[۳]	۰/۸۹۳	**۰/۰۱۲۴۰**
LSTM [۹]	۰/۹۴۱	**۰/۰۱۱۰۷**
شبکه‌های عصبی کانولوشنی[۱۱]	۰/۹۷۱	**۰/۰۹۸۴۹**
شبکه‌ی عصبی و الگوریتم ترکیبی PSO/GA [۱۲]	۰/۹۳۹	**۰/۰۰۱۴۴**
ماشین بردار پشتیبان[۱۳]	۰/۹۲۱	**۰/۰۰۱۶۶**

همچنین برخی دیگر از روش‌های ارائه شده بر روی داده‌های مشابه مورد آزمایش قرار گرفتند که مقایسه آن‌ها با روش پیشنهادی در جدول ۲ نمایش داده شده و نشان‌دهنده این واقعیت است که روش پیشنهادی علی‌رغم بهینه‌سازی برای حفظ سادگی مدل نهایی در بدترین حالت مشابه روش‌های موجود عمل کرده و در بسیاری از موارد از این روش‌ها دقیق‌تر است.

در این مقایسه، تنها در یک مورد امتیاز روش پیشنهادی از سایر روش‌ها بر روی مجموعه‌داده مشابه کمتر است که با توجه به انحراف معیار کمتر و همچنین سادگی بیشتر مدل بهینگی روش پیشنهادی قابل توجیه است. لازم به ذکر است که با توجه به یکسان نبودن مجموعه داده‌ها و همچنین معیارهای ارزیابی گزارش شده در پژوهش‌های موجود، این روش‌ها مجدداً بر اساس پارامترهای موجود در مقالات مربوط پیاده‌سازی شده و نتایج بر مجموعه داده‌های مشابه در این جدول گزارش شده است.

۴-نتیجه‌گیری

در این پژوهش، یک روش مبتنی بر درخت تصمیم به منظور پیش‌بینی شدت تابش ارائه شده‌است. روش پیش‌بینی با استفاده از الگوریتم فرا ابتکاری خفاش به منظور حفظ تعادل میان دقت مدل و سادگی و قابلیت پیاده‌سازی آن در دستگاه‌های اینترنت اشیاء بهینه‌سازی شده است. در نهایت به یک مدل قوی که با معیار مربع-R دارای امتیاز ۰/۹۵ است و همچنین از نظر سادگی مدل گروهی تا حدود زیادی بهینه است، به دست آمد. لذا از روش پیشنهادی می‌توان به منظور ساختن مدل‌های آفلاین در محیط‌های نیروگاه‌های فتوولتائیک که با اینترنت اشیاء مدیریت می‌شوند استفاده کرد. آنچه به‌عنوان آینده‌ی این روش پیشنهادی به منظور بهبود و گسترش نتایج آن می‌توان بیان نمود عبارتند از ارائه یک معماری جامع اینترنت اشیاء برای نیروگاه‌های فتوولتائیک بر مبنای مدل ارائه شده در این پژوهش، ارائه یک مکانیزم هرس و بهینه‌سازی برای قوانین استخراج شده از مدل‌های مبتنی بر درخت و توسعه یک مجموعه داده‌های بومی از پارامترهای هواشناسی داخلی به منظور استفاده به عنوان ورودی روش پیشنهادی که می‌تواند کارایی مدل را برای نیروگاه‌های داخلی افزایش دهد.

مراجع

[1] N. M. Kumar, K. Atluri and S. Palaparthi, “Internet of Things (IoT) in Photovoltaic Systems,” *National Power Engineering Conference (NPEC)*, 2018, pp. 1-4, doi: 10.1109/NPEC.2018.8476807.

[2] A. Mellit, M. Menghanem and M. Bendekhis, “Artificial neural network model for prediction solar radiation data: application for sizing stand-alone photovoltaic power system,” *IEEE Power Engineering Society General Meeting*, 2005, San Francisco, CA, USA, 2005, pp. 40-44 Vol. 1, doi: 10.1109/PES.2005.1489526.

[3] A. Angstrom, “Solar and terrestrial radiation: Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation,” *Quarterly Journal of the Royal Meteorological Society*, 1924, pp. 121-126,1924, doi:10.1002/qj.49705021008.

[4] J. Prescott, “Evaporation from a water surface in relation to solar radiation,” *Transactions of the Royal Society of South Australia*, vol. 64, pp 114-118, 1940, doi: 10.4236/epe.2017.912045.

[5] A. A. Trabea and M.M. Shaltout, “Correlation of global solar radiation with meteorological parameters over Egypt,” *Renewable Energy*, vol. 21, pp. 297-308, Oct. 2000, doi:10.1016/S0960-1481(99)00127-5.

[6] D.Yeboah-Amankwah and K.Agyeman, “Differential Ångstrom model for predicting insolation from hours of sunshine,” *Solar Energy*, vol. 45, pp. 371-377, 1990, doi:10.1016/0038-092X(90)90158-9.

[7] H. Ninomiya, “Study on application of amedas meteorological data to the simulation of building heat environment,” University of Tokyo, 1994.

[8] J. C . Cao and S. H. Cao, “Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis,” *Energy*, vol. 31, pp. 3435-3445, Dec. 2006, doi: 10.1016/j.energy.2006.04.001.

[9] X. Qing and Y. Niu, “Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM,” *Energy*, vol. 148, pp. 461-468, Apr. 2018, doi: 10.1016/j.energy.2018.01.177.

[10] D. Pothineni, M. R. Oswald, J. Poland and M. Pollefeys,“KloudNet: Deep Learning for Sky Image Analysis and Irradiance Forecasting,” *German Conference on Pattern Recognition*, 2018, pp. 535-551, doi:10.1007/978-3-030-12939-2_37.

[11] J. Lago, K. D. Brabandere, F. D. Ridder and B. D. Schuttera, “Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data*,” SolarEnergy*, vol. 173, pp. 566-577, Oct. 2018, doi: 10.1016/j.solener.2018.07.050.

[12] N. Dong, J. F. Chang, A. G. Wu and Z. K. Gao, “A novel convolutional neural network framework based

solar irradiance prediction method," *International Journal of Electrical Power & Energy Systems*, vol. 114, pp. 1-19, Jan. 2020, doi: 10.1016/j.ijepes.2019.105411.

[13] N. Sharma, P. Sharma, D. Irwin and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," *IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 528-533, doi: 10.1109/SmartGridComm.2011.6102379.

[14] P. J. M. Ali and R. H. Faraj, "Data normalization and standardization: a technical report," *Mach. Learn. Tech. Rep.*, vol. 1, no. 1, pp. 1-6, 2014, doi: 10.13140/RG.2.2.28948.04489.

[15] Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry*, vol. 27, no. 2, pp. 130-135, Apr. 2015, doi: 10.11919/j.issn.1002-0829.215044.

[16] N. Bhargava, G. Sharma, R. Bhargava and M. Mathuria, "Decision tree analysis on j48 algorithm for data mining," *Intrenational Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 6, 2013.

[17] D. Borup, B.J. Christensen, N.S. Mühlbach and M.S. Nielsen, "Targeting predictors in random forest regression," *International Journal of Forecasting*, vol. 39, pp. 841–868, 2023, doi:10.1016/j.ijforecast.2022.02.010.

[18] Y. Li *et al.*, "Random forest regression for online capacity estimation of lithium-ion batteries," *Applied energy*, vol. 232, pp. 197-210, Dec. 2018, doi: 10.1016/j.apenergy.2018.09.182.

[19] R. Seyghaly, J. Garcia, X. Masip-Bruin and M. M. Varnamkhasti, "Interference Recognition for Fog Enabled IoT Architecture using a Novel Tree-based Method," *IEEE International Conference on Omni-layer Intelligent Systems (COINS)*, 2022, pp. 1-6, doi: 10.1109/COINS54846.2022.9854944.

[20] G. Sembina, "Building a scoring model using the ADABOOST ensemble model," *International Conference on Smart Information Systems and Technologies (SIST)*, 2022, doi:10.1109/sist54437.2022.9945713.

[21] D. P. Solomatine and D. L. Shrestha, "AdaBoost.RT: a boosting algorithm for regression problems," *IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, Budapest, Hungary, 2004, pp. 1163-1168 vol.2, doi: 10.1109/IJCNN.2004.1380102.

[22] X. S. Yang, "A New Metaheuristic Bat-Inspired Algorithm," *Nature Inspired Cooperative Strategies for Optimization (NICSO 2010)*, vol. 284, pp. 65–74, 2010, doi: 10.1007/978-3-642-12538-6_6.

[23] M.Banerjee, "Cook's distance in linear longitudinal models," *Communications in Statistics-Theory and Methods*, vol. 27, pp. 2973-2983, 1998, doi: 10.1080/03610929808832267.

[24] J. Carmona, I. Lopez, J. Mateo, L. Jimenez and E. Aldana, "A distance-based method for outlier detection on high dimensional datasets," *IEEE Latin America Transactions*, vol. 18, pp. 589–597, 2020, doi:10.1109/tla.2020.9082731.

[25] E. -S. M. El-Kenawy *et al.*, "Advanced Ensemble Model for Solar Radiation Forecasting Using Sine Cosine Algorithm and Newton's Laws," in *IEEE Access*, vol. 9, pp. 115750-115765, 2021, doi: 10.1109/ACCESS.2021.3106233.

COPYRIGHTS
©2024 by the authors. Published by the Islamic Azad University Bushehr Branch. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) https://creativecommons.org/licenses/by/4.0