تأثیر تغییر اقلیم بر سلامت خاک: تحلیل واکنش میکروبیوم به گرمایش با بهره گیری از مدل های یادگیری ماشین

ویراستاری نشده

چکیده

تغییر اقلیم اثرات گسترده‌ای بر عملکرد خاک دارد؛ به‌ویژه از طریق حساسیت تنفس میکروبی به دما که با ضریب حساسیت دمایی (Q10) سنجیده می‌شود. این پژوهش با هدف شناسایی عوامل مؤثر بر این ضریب و ارزیابی کارایی الگوریتم‌های یادگیری ماشین انجام شد.در این مطالعه از داده‌های مربوط به ۳۳۲ نمونه خاک از ۲۹ کشور جهان استفاده شد و ضریب حساسیت دمایی در سه سطح کم، متوسط و زیاد طبقه‌بندی گردید. شش الگوریتم شامل جنگل تصادفی، ماشین بردار پشتیبان، نزدیک‌ترین همسایگی، رگرسیون لجستیک، تقویت گرادیانی و شبکه عصبی چندلایه به‌کار گرفته شدند. عملکرد مدل‌ها با شاخص‌های دقت، سطح زیر منحنی مشخصه عملکرد گیرنده (ROC) و سطح زیر منحنی دقت–بازخوانی (PRC) مقایسه شد. نتایج نشان داد جنگل تصادفی بالاترین کارایی را داشته است؛ به‌گونه‌ای که با دقت ۰/۱۱۷۰ ± ۰، مقدار ROC برابر با ۰/۸۱۵۰ ± ۰/۰۸۰ و مقدار PRC برابر با ۰/۷۵۳۰ ± ۰/۰۹۶۰ بر سایر مدل‌ها برتری داشت. برای تفسیر نتایج از روش SHAP به‌عنوان رویکرد هوش مصنوعی توضیح‌پذیر استفاده شد. این تحلیل نشان داد تنفس ناشی از گلوکز، غنای قارچی و شوری خاک مهم‌ترین عوامل مؤثر بر ضریب حساسیت دمایی هستند. این یافته‌ها با ارائه تصویری روشن از نقش ترکیبی ویژگی‌های خاکی و میکروبی، توان الگوریتم‌های یادگیری ماشین و روش‌های توضیح‌پذیر را در تحلیل پاسخ خاک به گرمایش جهانی نشان می‌دهند.

کلیدواژه‌ها: *یادگیری ماشین، جامعه میکروبی خاک، تغییر اقلیم، هوش مصنوعی توضیح‌پذیر*

The Impact of Climate Change on Soil Health: Analyzing Microbiome Response to Warming Using Machine Learning Models

ABSTRACT

Climate change has extensive effects on soil functioning, particularly through the temperature sensitivity of microbial respiration, which is measured by the temperature sensitivity coefficient (Q10). This study aimed to identify the factors influencing this coefficient and to evaluate the performance of machine learning algorithms. Data from 332 soil samples collected across 29 countries were used, and the temperature sensitivity coefficient was classified into three levels: low, medium, and high. Six algorithms—including Random Forest, Support Vector Machine, K-Nearest Neighbors, Logistic Regression, Gradient Boosting, and Multilayer Perceptron—were applied. Model performance was compared using accuracy, the area under the Receiver Operating Characteristic curve (ROC), and the area under the Precision–Recall curve (PRC). The results indicated that Random Forest had the highest performance, achieving an accuracy of 0.1170 ± 0, ROC of 0.8150 ± 0.080, and PRC of 0.7530 ± 0.096, outperforming the other models. The SHAP method was employed for interpreting the results as an explainable AI approach. This analysis revealed that glucose-induced respiration, fungal richness, and soil salinity were the most influential factors affecting the temperature sensitivity coefficient. These findings provide a clear picture of the combined role of soil and microbial features and demonstrate the potential of machine learning algorithms and explainable methods in analyzing soil responses to global warming.

Keywords: *Climate change, Explainable Artificial Intelligence (XAI), Machine learning, Soil microbial community*

EXTENDED ABSTRACT

Introduction:

Climate change is having profound effects on soil health and its functioning, particularly by influencing microbial respiration and enhancing carbon release through rising temperatures. The Q10 index, which indicates the temperature sensitivity of respiration, is a crucial metric for understanding soil carbon dynamics in the context of global warming. This study leverages machine learning models and explainable AI (using SHAP) to predict Q10 values based on a combination of biological, chemical, and physical soil properties. By employing a three-level classification system for Q10, the study offers a more nuanced perspective on microbial response and delivers interpretable insights into soil-climate interactions.

Objective(s):

The objective of this study was to predict microbial Q10 temperature sensitivity using various machine learning algorithms, incorporating biological, chemical, and physical soil data. Additionally, the research aims to pinpoint the most influential factors impacting Q10 through interpretable AI methods such as SHAP, thereby improving our understanding of soil carbon behavior in response to climate change.

Methods:

This research utilized a dataset containing soil biological, chemical, and physical property data to classify microbial Q10 values into three sensitivity levels. Multiple machine learning models, including Random Forest, XGBoost, SVM, and MLP, were applied, with class imbalance addressed using resampling techniques like SMOTE, SMOTEENN, and ADASYN. Model performance was assessed using precision, recall, and F1-score, leveraging nested cross-validation. SHAP analysis was conducted to interpret feature importance and model decisions.

Results:

Ensemble models, particularly Extra Trees, Random Forest, and XGBoost, outperformed other models in accuracy and AUC. Extra Trees achieved the best results, with an accuracy of 0.653 ± 0.117, AUROC of 0.815 ± 0.080, and AUPRC of 0.753 ± 0.096. Decision Tree and KNN performed the weakest. Feature importance analysis, both embedded and SHAP, highlighted microbial traits such as Bacteria_Positive and Mean_Glucose as key predictors. SHAP plots also revealed threshold effects for chemical features like SOC and Alkane. Statistical tests confirmed the superiority of ensemble models over simpler approaches ($p < 0.05$).

Conclusions:

This research employs machine learning and explainable AI to identify critical factors affecting Q10 in soils. Key microbial traits, including Bacteria_Positive and Mean_Glucose, and chemical features like SOC and Alkane, were found to be influential. SHAP analysis provided transparency in model predictions, uncovering complex relationships between soil properties and microbial respiration. These insights can inform soil management practices and climate adaptation strategies. Future research could focus on incorporating physical and hydraulic soil properties to refine Q10 predictions for specific ecological conditions.

Author Contributions

KeshikNevisRazavi, S.R. contributed to the conceptualization, methodology, software development, investigation, data curation, writing of the original draft, and manuscript review and editing. Farahani, E. was involved in validation, writing of the original draft, review and editing of the manuscript, supervision of the research process, and overall project administration. Abedinzadeh, N. participated in software development, data curation, and writing of the original draft. Abdolahi, M. contributed to

ویراستاری نشده

the methodology, software implementation, validation, formal analysis, investigation, data curation, writing of the original draft, review and editing, visualization of results, and research supervision. All authors have read and approved the final version of the manuscript.

Data Availability Statement

The dataset used in this study is publicly available on: https://figshare.com/articles/dataset/The_soil_microbiome_governs_the_response_of_microbial_respiration_to_warming_across_the_globe/20776243?file=42356520

Acknowledgements

The authors would like to thank the original authors of the publicly available dataset used in this study. We also appreciate the support provided by our affiliated institutions and colleagues who contributed insightful discussions during the research process.

Ethical Considerations

This study is based on secondary data analysis of a publicly available dataset. No human or animal subjects were involved, and therefore no ethical approval was required.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

مقدمه

تغییراقلیم یکی از چالش‌های اساسی قرن بیست‌ویکم به شمار می‌رود که پیامدهای عمیقی بر اکوسیستم‌های طبیعی و انسانی برجای می‌گذارد. در این میان، خاک به‌عنوان یکی از مهم‌ترین اجزای محیط‌زیست، مستقیما تحت‌تأثیر گرمایش جهانی قرار می‌گیرد. افزایش دما و نوسانات بارشی ناشی از تغییر اقلیم می‌تواند ویژگی‌های فیزیکی، شیمیایی و زیستی خاک را دگرگون ساخته و عملکردهای کلیدی آن مانند چرخه کربن را تغییر دهد (Allen et al., 2011; Lal, 2011).

میکروبیوم خاک، شامل باکتری‌ها، قارچ‌ها و سایر میکروارگانیسم‌ها، در چرخه‌های بیوژئوشیمیایی مانند چرخه کربن نقشی بنیادین ایفا می‌کنند. یکی از فرآیندهای اصلی این میکروارگانیسم‌ها، تنفس میکروبی است که طی آن مواد آلی اکسید شده و گازهای گلخانه‌ای همچون CO_2 آزاد می‌شود. این فرآیند تحت تأثیر دما قرار دارد و ضریب حساسیت دمایی (Q10) به‌عنوان معیاری برای سنجش واکنش‌پذیری آن نسبت به گرمایش به‌کار می‌رود (Patil and Lamnganbi, 2018; Haaf et al., 2021). در سال‌های اخیر، روش‌های یادگیری ماشین (Machine Learning) برای تحلیل داده‌های پیچیده و چند بعدی به‌طور فزاینده‌ای مورد توجه قرار گرفته‌اند. به‌کارگیری این روش‌ها در ترکیب با ابزارهای هوش مصنوعی قابل توضیح (Explainable Artificial Intelligence; XAI)، به درک عمیق‌تر الگوهای واکنش خاک در برابر گرمایش کمک می‌کند. با وجود این پیشرفت‌ها، بسیاری از پژوهش‌های پیشین فقط بر مقادیر افراطی ضریب حاسیت دمایی تمرکز کرده و داده‌های میانی را نادیده گرفته‌اند.

ویراستاری نشده

ضریب حساسیت دمایی در مدل‌سازی چرخه کربن خاک کاربرد گسترده‌ای دارد و شاخصی کلیدی برای ارزیابی واکنش‌پذیری تنفس میکروبی نسبت به دما محسوب می‌شود. این ضریب نشان می‌دهد که چگونه افزایش دما نرخ تنفس میکروبی را تغییر می‌دهد. مطالعات نظری بیان کرده‌اند که ضریب حساسیت دمایی تابعی از عوامل پیچیده‌ای همچون دما، رطوبت، ترکیب جامعه میکروبی، کیفیت مواد آلی و ویژگی‌های فیزیکوشیمیایی خاک نظیر pH، کربن آلی خاک و نسبت C:N است (Winkler et al., 1996; Mahecha et al., 2010) .

در مطالعه جدیدی (Saez-Sandino et al., 2023) با استفاده از یادگیری ماشین و روش‌های هوش مصنوعی توضیح پذیر، رابطه میان ویژگی‌های میکروبی و ضریب حساسیت دمایی بررسی شد. در این پژوهش داده‌های مربوط به دهک‌های بالا و پایین ضریب حساسیت دمایی حفظ و داده‌های میانی حذف گردید. گرچه این رویکرد دقت مدل را افزایش داد، اما توان تعمیم‌پذیری آن را کاهش داد. این مطالعه نشان داد که ساختار میکروبی یکی از عوامل اصلی در واکنش خاک به گرمایش است.

پژوهش‌های تجربی متعددی نیز ضریب حساسیت دمای را در شرایط اقلیمی و زیستی گوناگون بررسی کرده‌اند. برای مثال،(Haaf et al.,2021) گزارش کردند که ضریب حساسیت دمایی در مناطق مرطوب به‌دلیل فعالیت بالاتر میکروبی بیشتر است. همچنین (Meyer et al., 2018) با مطالعه بر ۹۰ سایت اروپایی نشان دادند که عواملی مانند pH، SOC و دمای سالانه از مهم‌ترین متغیرهای تعیین‌کننده این ضریب هستند. (Nottingham et al., 2015) در جنگل‌های آند پرو افزایش ضریب حساسیت دمایی را در ارتفاعات بالاتر به‌عنوان سازوکار تطبیقی جوامع میکروبی تفسیر کردند. افزون بر این، (Mahecha et al., 2010) در مطالعه‌ای در مقیاس جهانی، نوعی همگرایی در ضریب حساسیت دمایی را گزارش کردند.

در پژوهشی دیگر (Tong et al., 2021) اثر کاربری زمین را بر ضریب حساسیت دمایی بررسی کرده و نشان دادند که مدیریت‌های مختلف می‌تواند ترکیب میکروبی و حساسیت آن را به گرما تغییر دهد. همچنین (Yang et al., 2022) تأکید کردند که فصل نمونه‌برداری و استراتژی‌های زیستی میکروب‌ها، از عوامل کلیدی در تبیین تغییرپذیری این ضریب به شمار می‌روند.

مطالعات جدیدتر نیز این موضوع را گسترش داده‌اند. به‌عنوان نمونه، (Crowther et al. 2019) نشان دادند که جامعه میکروبی جهانی یکی از تعیین‌کننده‌های اصلی در چرخه کربن است. (Ahrens et al. 2020) نیز گزارش کردند که حساسیت دمایی در جنگل‌ها به‌شدت تحت تأثیر تنوع میکروبی و ویژگی‌های شیمیایی خاک قرار دارد. افزون بر این، (Xu et al. 2021) تأکید کردند که برهم‌کنش همزمان عوامل شیمیایی، فیزیکی و زیستی، نقش اساسی در تغییرپذیری ضریب حساسیت دمایی ایفا می‌کند.

با وجود این شواهد، همچنان چندین خلأ علمی مهم برجای مانده است. نخست، بسیاری از پژوهش‌های پیشین تنها بر مقادیر افراطی ضریب حساسیت دمایی تمرکز کرده و داده‌های میانی نادیده گرفته‌اند، در حالی که این داده‌ها برای ترسیم تصویری واقع‌بینانه‌تر از واکنش خاک به گرمایش ضروری‌اند. دوم، بیشتر مطالعات صرفاً به بررسی یک دسته از ویژگی‌های زیستی یا شیمیایی ـپرداخته‌اند، در حالی که حساسیت دمایی تنفس میکروبی حاصل تعامل پیچیده میان عوامل زیستی، شیمیایی و فیزیکی است. سوم، اغلب پژوهش‌ها در مقیاس‌های منطقه‌ای یا اکوسیستم‌های خاص انجام شده‌اند و نتایج آنها به‌سادگی به سطح جهانی تعمیم‌پذیر نیست. این خلأها موجب شده شناخت ما از دینامیک ضریب حساسیت دمایی همچنان ناقص باقی بماند؛ و در شرایطی که گرمایش جهانی به‌سرعت در حال تشدید است، این کاستی می‌تواند پیامدهای جدی برای پیش‌بینی چرخه کربن جهانی و تدوین سیاست‌های اقلیمی به همراه داشته باشد.

ویراستاری نشده

بر همین اساس، هدف اصلی پژوهش حاضر پیش‌بینی ضریب حساسیت دمایی در سه سطح کم، متوسط و زیاد است. برای تحقق این هدف، الگوریتم‌های یادگیری ماشین در ترکیب با روش‌های هوش مصنوعی توضیح‌پذیر (XAI) به‌کار گرفته شده‌اند تا علاوه بر افزایش دقت پیش‌بینی، شفافیت در نقش عوامل مؤثر نیز فراهم شود. در این چارچوب، داده‌های زیستی (مانند تنوع میکروبی و تنفس القاشده با گلوکز)، داده‌های شیمیایی (مانند pH وSOC) و فیزیکی (مانند رطوبت و بافت خاک) به‌طور همزمان وارد مدل شدند تا اثرات ترکیبی آنها آشکار شود. همچنین، عملکرد چندین الگوریتم مختلف با استفاده از اعتبارسنجی متقابل تودرتو ارزیابی و مقایسه گردید. افزون بر این، تحلیل SHAP برای شفاف‌سازی سهم هر متغیر و کشف روابط خطی، غیرخطی و آستانه‌ای به‌کار گرفته شد. این رویکرد جامع، شکاف علمی مطالعات پیشین را پوشش داده و می‌تواند تصویری تعمیم‌پذیرتر و کاربردی‌تر از دینامیک ضریب حساسیت دمایی در مقیاس جهانی ارائه دهد.

روش‌شناسی پژوهش

گردآوری داده‌ها

ویژگی‌های مورد مطالعه در این پژوهش (جدول ۱) ، از پایگاه داده عمومی Figshare دانلود شده است. این پایگاه شامل نمونه‌های خاک از ۲۹ کشور است (Saez-Sandino et al., 2023) و نمای کلی از شرایط متنوع آب‌و هوایی و ویژگی‌های خاک در سراسر جهان ارائه می‌دهند.

جدول ۱. متغیرهای ورودی مورد استفاده در تحلیل

نام ویژگی	معادل فارسی	واحد
Longitude_c	طول جغرافیایی	-
Forest	وجود پوشش جنگلی	-
MAP_wc2	میانگین بارندگی سالانه	میلی‌متر
MAT_wc2	میانگین دمای سالانه	سانتی‌گراد
Plant_richness	تعداد گونه‌های گیاهی	-
Plant_cov_er_v3	نسبت پوشش گیاهانی که در خاک نفوذ کرده‌اند	–
Soil_pH	pHخاک	–
Soil_salinity	هدایت الکتریکی خاک	دسی‌زیمنس بر متر
Fine_texture	رس و سیلت در خاک	گرم در صد گرم خاک
Soil_P	فسفر کل خاک	میلی‌گرم فسفر در کیلوگرم خاک خشک
Soil_CN	نسبت کربن آلی کل به نیتروژن کل	–
SOC	کربن آلی خاک	گرم کربن در کیلوگرم خاک
MAOC/POC_Ratio	نسبت کربن آلی به کربنمعدنی	–
Aromatic	ترکیبات آروماتیک	درصد
Alkanes	آلکان‌ها	درصد
Polysaccharide	پلی‌ساکاریدها	درصد
Amide	آمیدها	درصد

ویراستاری نشده

Mean_Glucose	تنفس خاک القاشده توسط گلوکز	میکروگرم کربن‌دی‌اکسید-کربن در هر گرم خاک در ساعت
Richness_bacteria	غنای باکتری‌ها	تعداد OTU نرمال‌شده
Richness_fungi	غنای قارچ‌ها	تعداد OTU نرمال‌شده
Richness_protist	غنای پروتیست‌ها	تعداد OTU نرمال‌شده
Bacteria_Negative	استاندارد گونه‌های باکتری دارای ارتباط منفی با Q10	درصد
Fungi_Negative	استاندارد گونه‌های قارچی دارای ارتباط منفی با Q10	درصد
Protists_Negative	استاندارد گونه‌های پروتیستی دارای ارتباط منفی با Q10	درصد
Bacteria_Positive	استاندارد گونه‌های باکتریایی دارای ارتباط منفی با Q10	درصد
Fungi_Positive	استاندارد گونه‌های قارچی دارای ارتباط منفی با Q10	درصد
Protists_Positive	استاندارد گونه‌های پروتیستی دارای ارتباط منفی با Q10	درصد

مجموعه داده شامل طیفی گسترده از عوامل محیطی؛ از جمله میانگین دمای سالانه در محدوده ۷- تا ۳۰ سانتی‌گراد، انواع مختلف پوشش گیاهی و خصوصیات شیمیایی و فیزیکی خاک است. برای هر نمونه خاک، ۲۷ ویژگی با ماهیت‌های متفاوت شناسایی شد که عوامل محیطی، میکروبیوتای خاک، مقاومت بیوشیمیایی، کمیت بستر را پوشش می‌دهد. هر نمونه خاک با مقدار ضریب حساسیت دمایی همراه است و تصویری از حساسیت تنفسی خاک نسبت به گرمایش ارائه می دهد. این اطلاعات در قالب یک جدول سازماندهی شده‌اند؛ به‌گونه‌ای که هر ردیف بیانگر یک نمونه خاص و هر ستون نشان دهنده یکی از ویژگی‌های خاک است (داده‌های این پژوهش به‌صورت آزاد از طریق پایگاه Figshare در دسترس هستند). عوامل محیطی شامل متغیرهای خاک (خصوصیات فیزیکی و شیمیایی)، کاربری اراضی، متغیرهای هواشناسی و موقعیت جغرافیایی محل نمونه برداری‌اند. به طور مشخص، طول جغرافیایی، وجود جنگل، میانگین بارندگی سالانه، میانگین دمای سالانه ، غنای گیاهی، نوع پوشش گیاهی، pH، هدایت الکتریکی، درصد رس و سیلت خاک، فسفر کل خاک و نسبت C:N (بر اساس کربن آلی و نیتروژن کل) در این دسته قرار می‌گیرند (Nottingham et al., 2015). شاخص‌های مقاومت بیوشیمیایی شامل درصد آروماتیک، آلکان، پلی ساکارید و آمید است. مواد معدنی شامل نسبت کربن آلی خاک و نسبت کربن آلی به کربن معدنی می‌شود. (Winkler et al., 1996) همچنین داده‌های میکروبیوم خاک نیز شامل میانگین تنفس خاک ناشی از گلوکز (به عنوان شاخص کل زیست توده میکروبی)، غنای باکتری ها، قارچ ها و پروتیست ها است. در نهایت در این مطالعه نسبت‌های استاندارد شده از گونه‌های باکتریایی، قارچی و پروتیست که با ضریب حساسیت دمایی دارای ارتباط مثبت و منفی هستند نیز اندازه گیری شده‌است. ترکیب این مجموعه متنوع از شاخص‌ها امکان بررسی جامع اثر متغیرهای گوناگون بر ضریب حساسیت دمایی و شناسایی عوامل کلیدی مؤثر بر آن را فراهم می‌سازد.

پیش‌پردازش داده‌ها

در گام نخست، داده‌های خام شامل متغیرهای زیستی، فیزیکی، شیمیایی و اقلیمی نمونه‌های خاک بازبینی قرار شدند تا از یکپارچگی و نبود داده‌های پرت یا گمشده اطمینان حاصل شود. برای پرهیز از حذف داده‌های میانی و دستیابی به مدلی واقع‌بینانه‌تر، مقادیر اولیه ضریب حساسیت دمایی (دریافت‌شده از پایگاه داده عمومی؛ Saez-Sandino et al., 2023) توسط نویسندگان این پژوهش

ویراستاری نشده

بر اساس صدک‌های ۳۳ و ۶۶ به سه دسته‌ی مجزا تقسیم شدند. این دسته‌بندی سه‌کلاسه امکان تحلیل پیوسته‌تری از پاسخ میکروبی خاک به گرمایش فراهم ساخت (Steyerberg, 2019; Ibrahimi et al., 2023) .

الگوریتم‌های یادگیری ماشین

برای پیش‌بینی مقادیر ضریب حساسیت دمایی و شناسایی عوامل مؤثر، از مجموعه‌ای از الگوریتم‌های یادگیری ماشین استفاده شد. ماشین‌های بردار پشتیبان (SVM) طبقه‌بندهای قدرتمندی هستند که به دلیل اثربخشی در فضاهایی با ابعاد بالا شناخته شده‌اند. .ماشین‌های بردار پشتیبان یک ابرصفحه یا مجموعه‌ای از ابرصفحه‌ها را در یک فضا با ابعاد بالا برای جداسازی کلاس‌های مختلف ایجاد می‌کند (Cortes and Vapnik 1995) . این مدل با استفاده از هسته‌های مختلف (خطی، چندجمله‌ای، تابع پایه شعاعی و سیگموید) و پارامترهای منظم‌سازی مختلف برای بهینه‌سازی عملکرد مدل بررسی شد:

- C ∈ {1.0, 0.1, 0.01, 0.5}
- kernel ∈ {‘linear’, ‘poly’, ‘rbf’, ‘sigmoid’}

الگوریتم نزدیک‌ترین همسایگی (K-Nearest Neighbors) یک روش غیرپارامتریک برای طبقه‌بندی است. این الگوریتم یک نقطه داده را بر اساس اکثریت کلاس همسایگان آن طبقه‌بندی می کند (Guo et al., 2003). عملکرد مدل را با استفاده از تعداد مختلفی از همسایگان، معیارهای فاصله و توابع وزن‌گذاری ارزیابی شد:

- n_neighbors ∈ {3, 5, 7, 9, 11}
- weights ∈ {‘uniform’, ‘distance’}
- metric ∈ {‘euclidean’, ‘manhattan’}

رگرسیون لجستیک یک مدل خطی است که برای طبقه‌بندی استفاده می‌شود و احتمال تعلق یک نمونه به هر یک از کلاس‌ها را پیش‌بینی می‌کند. در این تحقیق، انواع مختلف جریمه‌ها از جمله L1، L2 و ElasticNet مورد بررسی قرار گرفتند. علاوه بر این، پارامتر C و انواع solvers نیز به دقت تنظیم شدند (LaValley 2008).

- penalty ∈ {‘l1’, ‘l2’, ‘elasticnet’}
- C ∈ {1.0, 0.1, 0.01, 0.5}
- solver ∈ {‘lbfgs’, ‘liblinear’, ‘saga’}

مدل درخت تصمیم یک تکنیک یادگیری نظارت‌شده غیرپارامتریک است که برای انجام وظایف طبقه‌بندی و رگرسیون به‌کار می‌رود (Swain and Hauska 1977). هدف این مدل، پیش‌بینی مقدار یک متغیر هدف با استخراج قوانین تصمیم‌گیری ساده از

ویراستاری نشده

ویژگی‌های داده است. ساختار درخت تصمیم به صورت سلسله‌مراتبی و درخت‌مانند است که شامل گره ریشه، شاخه‌ها، گره‌های داخلی و برگ‌ها می‌شود. هایپرپارامترهای درخت تصمیم با استفاده از گرید جستجو تنظیم شدند:

• max_depth ∈ {None, 3, 5, 7, 9},
• criterion ∈ {‘gini’, ‘entropy’, ‘log_loss’}.

الگوریتم طبقه‌بندی جنگل تصادفی یک الگوریتم یادگیری ماشین است که مدل‌های فردی مختلف را ترکیب می‌کند تا یک مدل قوی‌تر و دقیق‌تر ایجاد کند. جنگل تصادفی از مجموعه‌ای از درخت‌های تصمیم ساخته شده است. هر درخت تصمیم با استفاده از یک زیرمجموعه از داده‌های آموزشی و انتخاب تصادفی ویژگی‌ها ساخته می‌شود. این الگوریتم از تکنیکی به نام "بوت‌استرپینگ" استفاده می‌کند که با نمونه‌برداری با جایگزینی، چندین زیرمجموعه تصادفی از داده‌های آموزشی ایجاد می‌کند. هر زیرمجموعه برای آموزش یک درخت تصمیم جداگانه استفاده می‌شود. (Breiman 2001) در هر گره از درخت تصمیم، به جای در نظر گرفتن همه ویژگی‌ها برای انجام تقسیم، جنگل تصادفی به‌طور تصادفی یک زیرمجموعه از ویژگی‌ها را انتخاب می‌کند. این کار به کاهش همبستگی بین درخت‌ها و کاهش حساسیت آنها به نویز در داده‌ها کمک می‌کند. جنگل تصادفی پیش‌بینی‌های چندین درخت را ترکیب می‌کند که باعث کاهش بیش‌برازش و بهبود عملکرد تعمیم‌پذیری می‌شود:

• n_estimators ∈ {100, 200}
• max_depth ∈ {None, 5, 10}

مدل درخت‌های فوق تصادفی یک الگوریتم یادگیری ماشین ترکیبی است که برای وظایف طبقه‌بندی به‌کار می‌رود. این الگوریتم از روش‌های مبتنی بر درخت تصمیم استفاده می‌کند و برای بهبود عملکرد و کاهش واریانس، از تصادفی‌سازی استفاده می‌کند. مانند دیگر الگوریتم‌های ترکیبی، این مدل مجموعه‌ای از درختان تصمیم می‌سازد. هر درخت بر روی یک زیرمجموعه تصادفی از داده‌های آموزشی آموزش داده می‌شود و برای تقسیم گره‌ها از انتخاب ویژگی‌های تصادفی استفاده می‌کند. با استفاده از تصادفی‌سازی گره‌های تقسیم و استفاده از زیرمجموعه‌های تصادفی از داده‌ها، درخت‌های فوق تصادفی در برابر بیش‌برازش حساسیت کمتری نسبت به درختان تصمیم سنتی دارند. این ویژگی آن را در شرایطی که داده‌ها دارای نویز یا واریانس بالایی هستند، بسیار مفید می‌کند . این مدل قادر است داده‌هایی با تعداد زیادی ویژگی را به‌طور مؤثر پردازش کند و به‌خوبی برای مسائل پیچیده مناسب است (Geurts et al., 2006). در نهایت، فرآیند انتخاب ویژگی در حین ساخت درخت‌ها به ارزیابی اهمیت هر ویژگی در مدل کمک می‌کند:

• n_estimators ∈ {100, 200}
• max_depth ∈ {None, 10, 20}
• min_samples_split ∈ {2, 5, 10}
• criterion ∈ {'gini', 'entropy'}

مدل گرادیان تقویت شده (Extreme Gradient Boosting) به عنوان یک الگوریتم یادگیری ماشین پرکاربرد شناخته می‌شود که برای حل مسائل رگرسیون و طبقه‌بندی به کار می‌رود. این الگوریتم به دلیل توانمندی‌اش در انجام رقابت‌های داده‌کاوی و یادگیری ماشین شهرت دارد و عملکرد برجسته‌ای را در زمینه‌های مختلف مسائل و مجموعه‌های داده از خود نشان می‌دهد. ویژگی‌های

ویراستاری شده

انعطاف‌پذیری و مقیاس‌پذیری گرادیان تقویت شده آن را برای پردازش حجم زیادی از داده‌ها ضروری می‌کند (Chen and Guestrin,2016). به عنوان یک روش یادگیری تجمیعی، گرادیان تقویت شده پیشرفت‌هایی نسبت به الگوریتم‌های افزایشی (Gradient Boosting)دارد. با ساختن مجموعه‌ای از درختان تصمیم، که هر درخت در پیش‌بینی نهایی سهیم است، مدل گرادیان تقویت شده توان پیش‌بینی مدل را تقویت می‌کند. قابلیت ارزیابی متقابل داخلی (cross-validation) از ویژگی‌های داخلیاین مدل است که به ارزیابی مدل و تنظیم پارامترها کمک می‌کند تا عملکرد بهینه حاصل شود. علاوه بر این، روش تجمیعی گرادیان تقویت شده که مدل‌ها اشتباهات یکدیگر را اصلاح می‌کنند، دقت پیش‌بینی بالاتری نسبت به بسیاری از الگوریتم‌های دیگر به ارمغان می‌آورد (Aydın et al., 2023). همچنین، استفاده از تکنیک‌های منظم‌سازی (regularization) به مقابله با چالش بیش‌برازش کمک می‌کند، به طوری که مدل می‌تواند به داده‌های جدید سازگار شود:

• max_depth ∈ {None, 3, 5, 7, 9}
• n_estimators ∈ {50, 100, 150, 200, 250}

شبکه‌های عصبی مصنوعی (ANN) با ساختار Multi-layer Perceptron (MLP) که در این مطالعه استفاده شد، مدل‌های یادگیری عمیقی هستند که توانایی مدل‌سازی روابط پیچیده غیرخطی بین متغیرها را دارند. این شبکه‌ها از لایه‌های متوالی نورون‌ها شامل لایه ورودی، لایه‌های پنهان و لایه خروجی تشکیل شده‌اند. هر نورون از طریق وزن‌ها و تابع فعال‌سازی با نورون‌های لایه بعدی متصل است. مدل شبکه عصبی چندلایه به‌عنوان طبقه‌بندی‌کننده غیرخطی برای پیش‌بینی مقادیر ضریب حساسیت دمایی خاک به کار رفت. از آنجا که شبکه‌های عصبی به مقیاس ویژگی‌ها حساس هستند، ویژگی‌های ورودی با استفاده از استانداردسازی و نرمال‌سازی مقیاس‌بندی شدند. همچنین، برای جلوگیری از بیش‌برازش و بهبود عملکرد، تنظیم دقیق پارامترهایی همچون تعداد نورون‌ها، تعداد لایه‌ها، تابع فعال‌سازی و نرخ یادگیری انجام شد (Novielli et al., 2024). مطابق با مطالعات پیشین، شبکه‌های عصبی پس از تنظیم بهینه، می‌توانند در مسائل طبقه‌بندی چندکلاسه عملکرد رقابتی و قابل توجهی داشته باشند:

• hidden_layer_sizes ∈ {(50,), (100,), (50, 50), (100, 50)}
• activation ∈ {'relu', 'tanh', 'logistic'} • solver ∈ {'adam', 'sgd', 'lbfgs'} • alpha ∈ {0.0001, 0.001, 0.01}
• learning_rate ∈ {'constant', 'adaptive'}

برای اطمینان از پایداری و استحکام نتایج، یک روش اعتبارسنجی متقابل تودرتو (nested cross-validation) با ۱۰ بخش (K=10) و با ۲۰ تکرار پیاده‌سازی شد. در این فرآیند، تنظیم پارامترهای مدل‌های یادگیری ماشین در حلقه داخلی اعتبارسنجی انجام گرفت. پس از انتخاب مدل بهینه، مقادیر SHAP محاسبه گردید تا میزان تأثیر هر متغیر در پیش‌بینی کلاس‌ها شفاف‌سازی شده و جهت تأثیر آن مشخص شود.

تمامی تحلیل‌ها در محیط پایتون (نسخه ۱۲٫۳) و با بهره‌گیری از کتابخانه‌های scikit-learn برای مدل‌سازی، xgboost برای پیاده‌سازی الگوریتم تقویت گرادیانی و shap برای تحلیل هوش مصنوعی قابل توضیح انجام گرفت. برای هر یک از این مدل‌ها، تنظیم دقیق پارامترها (hyperparameter tuning) با استفاده از شبکه جست‌وجوی مشبک (GridSearchCV) در درون حلقه اعتبارسنجی انجام شد.همچنین برای بهبود عملکرد مدل، داده‌ها با مقیاس‌بندی‌های مختلف از جمله StandardScaler ، MinMaxScaler،

ویراستاری نشده

RobustScaler، MaxAbsScaler، Normalizer، PowerTransformer و QuantileTransformer (با توزیع نرمال) نرمال‌سازی گردیدند. این مقیاس‌بندی برای همه مدل‌ها انجام شد تا بهترین مقیاس‌بندی انتخاب شود.

معیارهای ارزیابی عملکرد مدل‌ها

به‌منظور ارزیابی دقت و قابلیت اطمینان مدل‌ها، از معیارهای استاندارد طبقه‌بندی استفاده گردید. معیارهای ارزیابی ابزارهای مهمی برای ارزیابی عملکرد و اثر بخشی مدل‌های یادگیری ماشین هستند (Ferrer, 2022). این معیارها مقادیر کمی هستند که به ارزیابی عملکرد مدل در یک وظیفه خاص کمک می‌کنند. رایج‌ترین معیارهای ارزیابی در تحلیل طبقه‌بندی شامل دقت (رابطه ۱)، مساحت زیر منحنی مشخصه عملیاتی گیرنده (Receiver Operating Characteristic) و مساحت زیر منحنی دقت–یادآوری (Precision-Recall) هستند.

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$ رابطه ۱)

در محاسبه دقت، TP (True Positives) تعداد نمونه‌های مثبت به‌درستی طبقه‌بندی‌شده، TN (True Negatives) تعداد نمونه‌های منفی به‌درستی طبقه‌بندی‌شده، FP (False Positives) تعداد نمونه‌های منفی که به اشتباه مثبت طبقه‌بندی‌شده‌اند و FN (False Negatives) تعداد نمونه‌های مثبت هستند که به اشتباه منفی طبقه‌بندی‌شده‌اند.

مساحت زیر منحنی مشخصه عملیاتی گیرنده (ROC) از ابزارهای ارزیابی است که برای سنجش اثربخشی مدل‌های طبقه‌بندی دودویی به‌کار می‌روند. منحنی ROC یک نمایش گرافیکی از چگونگی تغییر حساسیت (مثبت‌های واقعی) و ویژگی (منفی‌های واقعی) در آستانه‌های مختلف است. در واقع، این منحنی تعادلی بین شناسایی دقیق نمونه‌های مثبت و منفی توسط مدل را نشان می‌دهد. این پارامترعملکرد کلی مدل را با اندازه‌گیری مساحت زیر منحنی ROC کمی می‌کند. مقداری نزدیک به ۱ عملکرد عالی مدل را نشان می‌دهد، در حالی که مقدار حدود ۰,۵ نشان‌دهنده طبقه‌بندی تصادفی است. به طور خلاصه، این معیارها برای ارزیابی و مقایسه توانایی طبقه‌بندی مدل‌های دودویی ضروری هستند (Ozenne et al., 2015).

مساحت زیر منحنی دقت–یادآوری (PRC) از ابزارهای ارزیابی عملکرد مدل‌های طبقه‌بندی دودویی است. منحنی PRC رابطه بین دقت (نرخ مثبت‌های واقعی) و بازخوانی (حساسیت) مدل در آستانه‌های مختلف تصمیم‌گیری را نشان می‌دهد. دقت معیاری است که نشان می‌دهد از تعداد نمونه‌هایی که به‌عنوان مثبت طبقه‌بندی شده‌اند، چه میزان واقعاً مثبت هستند، در حالی که بازخوانی معیاری است که نشان می‌دهد از تعداد نمونه‌های مثبت واقعی موجود در مجموعه داده، چه میزان به‌درستی توسط مدل شناسایی شده‌اند. مساحت زیر منحنی PRC یک اندازه‌گیری تجمیع‌شده از عملکرد مدل در رابطه با دقت و بازخوانی است. مساحت بزرگ‌تر نشان‌دهنده عملکرد بهتر است، به‌طوری‌که حداکثر مساحت برابر با ۱ به مدل کاملاً درست اشاره دارد (Wen et al., 2024).

شناسایی عوامل کلیدی با رویکرد هوش مصنوعی قابل توضیح

ویراستاری نشده

برای شناسایی عوامل کلیدی مؤثر بر ضریب حساسیت دمایی، از رویکرد هوش مصنوعی قابل توضیح بهره گرفته شد.هوش مصنوعی قابل توضیح (XAI) به‌عنوان یک پیشرفت اساسی در سیستم‌های هوش مصنوعی مطرح است که اطمینان حاصل می‌کند مدل‌ها برای انسان‌ها قابل درک و تفسیر باقی بمانند. در هوش مصنوعی قابل توضیح معمولاً SHAP و اهمیت ویژگی استفاده می‌شود. مقادیر SHAP بینش‌های ارزشمندی در مورد دلایل نتایج مدل ارائه می‌دهند (Lundberg and Novielli et al., 2024). مقدار SHAP برای ویژگی j از نمونه x با تجمیع آن در تمام مجموعه‌های ممکن بر اساس فرمول (رابطه ۲) زیر محاسبه می‌شود:

$$\Phi_j(x) = \sum_{F\subseteq S-\{j\}} \frac{|F|!\,(|S| - |F| - 1)!}{|S|!} [\, fx(F \cup j) - fx(F)] \qquad \text{رابطه ۲}$$

در اینجا، !|F| تعداد ترتیب‌های ویژگی‌ها در زیرمجموعه F، !(|S| − |F| − 1) ,F تعداد ترتیب‌های ویژگی‌ها در زیرمجموعه (S − (F ∪ {j} و !|S| تعداد کل ترتیب‌های ویژگی‌ها را نشان می‌دهد. در انتها برای مقایسه دقت مدل‌ها و بررسی تفاوت‌های معنی‌دار میان آن‌ها، از آزمون تحلیل واریانس یک‌طرفه (ANOVA) استفاده شد. در صورتی که p-value به‌دست‌آمده از آزمون ANOVA کمتر از ۰٫۰۵ باشد، تفاوت معنی‌دار بین مدل‌ها تایید می‌شود.

نتایج

نتایج نشان داد که مدل‌های ترکیبی، به‌ویژه جنگل تصادفی، درخت‌های فوق تصادفی، و تقویت گرادیان بهینه‌شده، به‌طور کلی عملکرد بهتری از نظر دقت، ROC و PRC نسبت به سایر مدل‌ها دارند (جدول ۲). بالاترین دقت مربوط به مدل جنگل تصادفی (۰/۱۱۷ ± ۰/۶۵۳) بود، پس از آن مدل درخت‌های فوق تصادفی و تقویت گرادیان بهینه‌شده به ترتیب دقت‌های۰/۱۱۲ ± ۰/۶۴۲ و ۰/۱۰۵ ± ۰/۶۴۱را داشتند. در مقابل، مدل‌های غیرترکیبی همچون درخت تصمیم (۰/۰۸۸ ± ۰/۶۰۶)، SVM (۰/۱۴۸ ± ۰/۶۴۱)، KNN (۰/۱۳۸ ± ۰/۵۸۹) و رگرسیون لجستیک (۰/۰۹۳ ± ۰/۶۴۷) عملکرد ضعیف‌تری داشتند. همچنین کمترین دقت در شبکه‌های عصبی (۰/۱۰۰ ± ۰/۷۰۰) مشاهده شد.

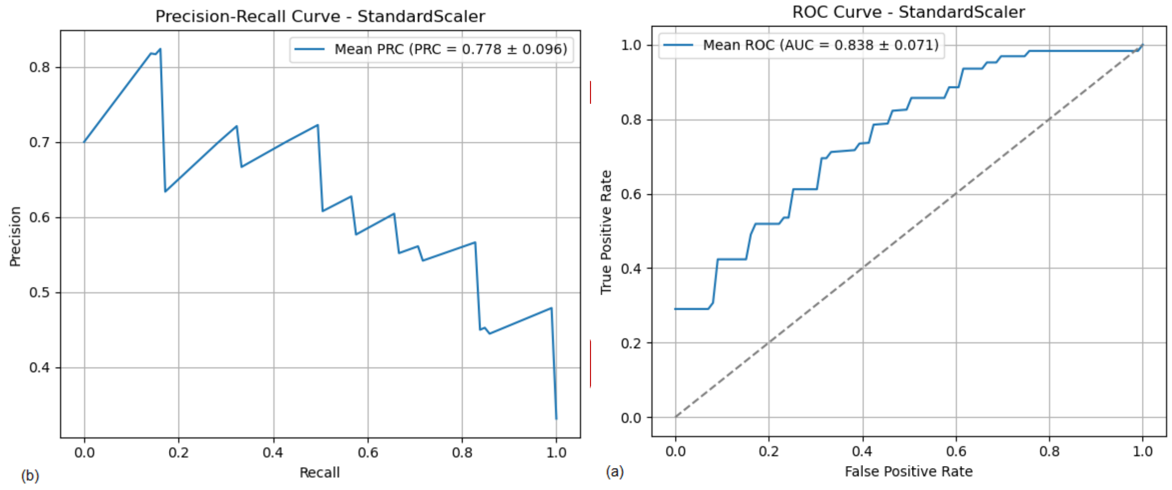
ویراستاری نشده

جدول ۲. میانگین عملکرد مدل های یادگیری ماشین به روش اعتبارسنجی متقاطع ۱۰ مرحله‌ای با ۲۰ تکرار

مدل		PRC	ROC	دقت
ماشین بردار پشتیبان	۰/۶۴۱ ± ۰/۱۳۸	۰/۷۲۹ ± ۰/۰۹۹	۰/۸۰۷ ± ۰/۰۸۰	۰/۶۴۱ ± ۰/۱۳۸
نزدیک‌ترین همسایگی	۰/۵۸۹ ± ۰/۱۳۸	/۶۷۲ ± ۰/۱۲۲	۰/۷۶۲ ± ۰/۱۰۲	۰/۵۸۹ ± ۰/۱۳۸
رگرسیون لجستیک	۰/۶۴۷ ± ۰/۰۹۳	۰/۷۱۰ ± ۰/۰۸۲	۰/۸۰۲ ± ۰/۰۷۲	۰/۶۴۷ ± ۰/۰۹۳
جنگل تصادفی	۰/۶۵۳ ± ۰/۱۱۷	۰/۷۵۳ ± ۰/۰۹۶	۰/۸۱۵ ± ۰/۰۸۰	۰/۶۵۳ ± ۰/۱۱۷
تقویت گرادیان بهینه‌شده	۰/۶۴۱ ± ۰/۱۰۵	۰/۷۲۹ ± ۰/۰۹۱	۰/۸۰۴ ± ۰/۰۶۹	۰/۶۴۱ ± ۰/۱۰۵

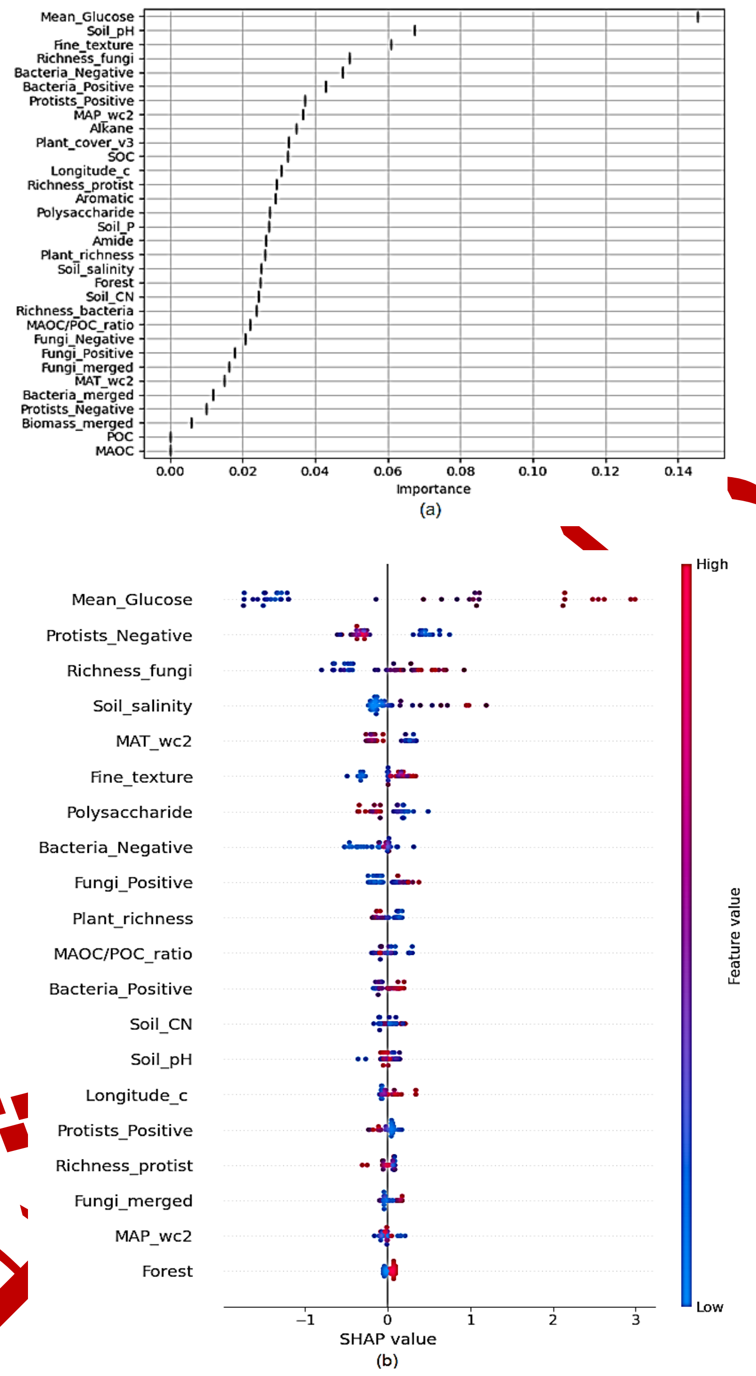
مدل	PRC	ROC	دقت
درخت تصمیم ۰/۶۰۶ ± ۰/۰۸۸	۰/۶۱۸ ± ۰/۰۷۳	۰/۷۶۲ ± ۰/۰۵۸	۰/۶۰۶ ± ۰/۰۸۸
درخت‌های فوق تصادفی ۰/۶۴۲ ± ۰/۱۱۲	۰/۷۷۸ ± ۰/۰۹۶	۰/۸۳۸ ± ۰/۰۷۱	۰/۶۴۲ ± ۰/۱۱۲
شبکه عصبی چندلایه ۰/۵۹۴ ± ۰/۱۱۵	۰/۷۰۰ ± ۰/۱۰۰	۰/۷۸۱ ± ۰/۰۷۳	۰/۵۹۴ ± ۰/۱۱۵

نمودار منحنی ROC و منحنی PRC (شکل ۱) برای مدل درخت‌های فوق تصادفی با استفاده از ۱۰ بار تقسیم تصادفی داده‌ها با ۲۰ تکرار جهت ارزیابی عملکرد مدل رسم گردید. مقدار ROC و PRC برای مدل درخت‌های فوق تصادفی با مقیاس‌بندی MinMaxScaler به ترتیب برابر با ۰/۰۷۱ ± ۰/۸۳۸ و ۰/۰۹۶ ± ۰/۷۷۸ بود که نشان‌دهنده توانایی بالای مدل در پیش‌بینی صحیح طبقات بر اساس مقادیر مختلف Q10 می‌باشد. پس از مدل درخت‌های فوق تصادفی، مدل جنگل تصادفی بالاترین ROC (۰/۰۹۶ ± ۰/۷۵۳) و PRC (۰/۰۸۰ ± ۰/۸۱۵) داشت. همچنین کمترین ROC (۰/۰۵۸ ± ۰/۷۶۲) در و کمترین PRC (۰/۰۷۳ ± ۰/۶۱۸) در درخت تصمیم مشاهده شد.



شکل ۱. منحنی مشخصه عملیاتی گیرنده ROC (a) و منحنی دقت-یادآوری PRC (b).

نمودار اهمیت ویژگی‌ها با استفاده از مدل تقویت گرادیان بهینه‌شده و روش SHAP رسم شد (شکل ۲). مقادیر SHAP با ساختن مدل‌های ساده اطراف هر نمونه، اهمیت هر ویژگی را به طور محلی تخمین می‌زنند. در این تحلیل مهم‌ترین متغیرها از نظر الگوریتم SHAP برای طبقه‌بندی نمونه‌ها مشخص می‌شود. در هر دو روش ویژگی تنفس میکروبی ناشی از گلوکز (Mean_Glucose) بیشترین نقش را در تعیین ضریب حساسیت دمایی داشت. چهار ویژگی اصلی در مدل به ترتیب شامل تنفس القاشده با گلوکز، pH خاک، بافت خاک و غنای قارچی بودند. این ویژگی‌ها نشان می‌دهند که ویژگی‌های زیستی و شیمیایی نقش غالبی در پیش‌بینی حساسیت دمایی تنفس میکروبی دارند

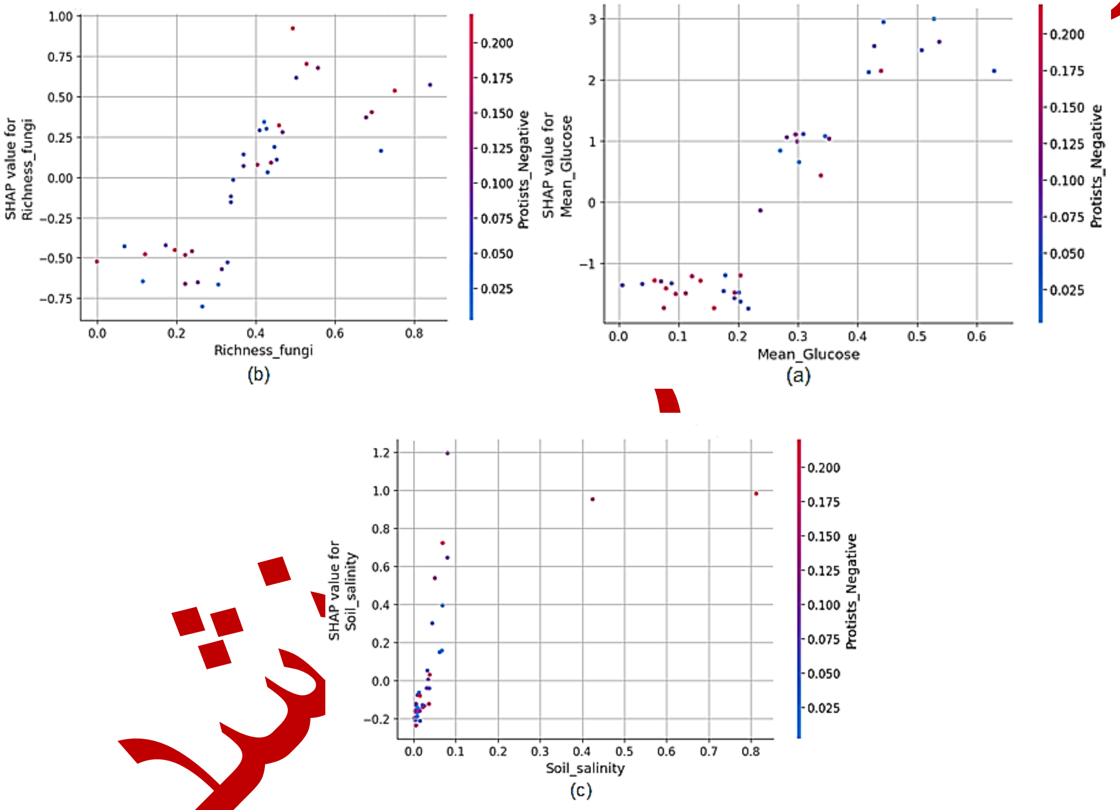


شکل ۲. نمودار اهمیت ویژگی‌ها با مدل تقویت گرادیان بهینه‌شده (a) و نمودار اهمیت ویژگی‌ها با SHAP (b)

در مقابل تحلیل SHAP علاوه بر تأیید نقش کلیدی تنفس میکروبی ناشی از گلوکز و غنای قارجی، نشان داد که شوری خاک نیز اهمیت بالایی در تعیین حساسیت میکروبی خاک به دما دارد. حضور شوری به‌عنوان یک عامل مهم در SHAP نشان‌دهنده تأثیر بالقوه

تنش‌های یونی بر عملکرد میکروبی است. تفاوت بین دو تحلیل بیانگر این است که برخی ویژگی‌ها ممکن است اثرات غیرخطی یا وابسته به تعامل با دیگر ویژگی‌ها داشته باشند.

نمودارهای وابستگی حاصل از تحلیل SHAP (شکل ۳) نشان می‌دهند که متغیرهای میکروبی دارای ارتباط مستقیم با مقادیر SHAPهستند. به‌عبارت دیگر، با افزایش مقادیر این متغیرها، مقادیر SHAP نیز افزایش یافته و در نتیجه، احتمال طبقه‌بندی نمونه‌ها در گروه دارای ضریب حساسیت دمایی بالا افزایش می‌یابد. این روند بر نقش کلیدی متغیرهای مربوط به میکروبیوم خاک در افزایش حساسیت تنفس میکروبی به گرمایش تأکید دارد. افزون بر این، تحلیل اهمیت ویژگی‌ها نشان داد که متغیرهای شوری نیز در میان مهم‌ترین عوامل شناسایی‌شده قرار دارند. این یافته‌ها، ضمن هم‌راستایی با مطالعات پیشین، بر اهمیت ویژگی‌های فیزیکی و شیمیایی خاک در تنظیم پاسخ تنفسی میکروبی به تغییرات دمایی تأکید دارند.



شکل ۳. نمودارهای وابستگی که سهم ویژگی های مهم را بر اساس تحلیل SHAP نشان می‌دهد. تنفس حاصل از گلکوز (a)، غنای قارچی (b)، شوری خاک (c)

بحث

نتایج به‌دست‌آمده در این پژوهش نشان داد که مدل‌های یادگیری ماشین، به‌ویژه الگوریتم‌های ترکیبی مانند درخت‌های فوق تصادفی و تقویت گرادیان بهینه‌شده توانایی بالایی در پیش‌بینی ضریب حساسیت دمایی دارند. این یافته‌ها با مطالعه (Saez-Sandino et al., 2023) هم‌راستا است. آن‌ها با استفاده ازهوش مصنوعی قابل توضیح و حذف داده‌های میانی ضریب حساسیت دمایی نشان

دادند که ویژگی‌های میکروبی مانند تنوع باکتریایی و قارچی نقش مهمی در پیش‌بینی ضریب حساسیت دمایی دارند. در پژوهش حاضر، با حفظ کل طیف ضریب حساسیت دمایی (شامل مقادیر میانی)، نیز مشخص شد که متغیرهای میکروبی در میان مهم‌ترین عوامل پیش‌بینی قرار دارند. در واقع، رویکرد سه‌کلاسه این پژوهش، با در نظر گرفتن کل طیف داده‌ها، توانست تصویری جامع‌تر و تعمیم‌پذیرتر از دینامیک ضریب حساسیت دمایی ارائه دهد. همچنین، بهره‌گیری از روش SHAP نه‌تنها موجب شفافیت مدل شد، بلکه امکان تحلیل رفتار آستانه‌ای متغیرها را نیز فراهم کرد. افزون بر این، عواملی مانند pH نیز سهم بالایی در پیش‌بینی داشتند که نشان‌دهنده تعامل چندعاملی میان ویژگی‌های فیزیکوشیمیایی و زیستی است.

برتری مدل‌های ترکیبی نسبت به مدل‌های ساده‌تر مانند نزدیک‌ترین همسایگی و درخت تصمیم را می‌توان به توانایی این الگوریتم‌ها در ترکیب چندین درخت و کاهش واریانس نسبت داد (Breiman, 2001; Geurts et al., 2006). این موضوع توضیح می‌دهد چرا مدل‌های ساده‌تر در پژوهش حاضر دقت پایین‌تری نشان دادند. همچنین مدل تقویت گرادیان بهینه شده به دلیل مکانیزم‌های منظم‌سازی و بهینه‌سازی تدریجی توانست ساختارهای پیچیده داده را بهتر از سایر مدل‌ها تبیین کند (Chen & Guestrin, 2016).

تحلیل مبتنی بر SHAP نشان داد که سه ویژگی کلیدی شامل تنفس میکروبی القاشده با گلوکز، غنای قارچی و شوری خاک بیشترین نقش را در طبقه‌بندی سطوح ضریب حساسیت دمایی ایفا می‌کنند. افزایش تنفس میکروبی به‌صورت یکنواخت با افزایش مقدار SHAP و در نتیجه با احتمال بالاتر قرار گرفتن نمونه‌ها در طبقه ضریب حساسیت دمایی بالا همراه بود. (Saez-Sandino et al., 2023)نیز نشان دادند افزایش فعالیت میکروبی، به‌ویژه در مناطق گرم‌تر، منجر به افزایش حساسیت تنفسی خاک می‌شود. همچنین (Meyer et al., 2018)نیز تأکید کردند که شاخص‌های تنفسی نقش کلیدی در تبیین واریانس ضریب حساسیت دمایی ایفا می‌کنند.

غنای قارچی (Richness_fungi) نیز به‌عنوان یکی از مهم‌ترین متغیرها توسط مدل شناسایی شد. افزایش تعداد zOTUs قارچی با افزایش مقدار SHAP همراه بود؛ این موضوع نشان‌دهنده نقش مهم تنوع قارچی در تنظیم نرخ تنفس خاک است. نتیجه‌ای که با یافته‌های (Crowther et al., 2019) مطابقت دارد و آن‌ها نیز ارتباط مستقیم ترکیب جامعه میکروبی به‌ویژه قارچ‌ها با ضریب حساسیت دمایی را گزارش کردند. این همگرایی یافته‌ها نشان می‌دهد که قارچ‌ها، بیش از بسیاری از گروه‌های میکروبی دیگر، در پایداری کربن خاک تحت شرایط تغییر اقلیم نقش دارند.

یکی از یافته‌های قابل‌توجه پژوهش حاضر، رفتار آستانه‌ای شوری خاک بود. طبق تحلیل SHAP، افزایش شوری تا محدوده‌ای خاص (تقریباً بین ۰٫۵ تا ۲ دسی‌زیمنس بر متر) با افزایش ضریب حساسیت دمایی همراه بود، اما در مقادیر بالاتر، اثر شوری معکوس شد و ضریب حساسیت دمایی کاهش یافت. این رفتار احتمالاً ناشی از محدودیت‌های فیزیولوژیکی ناشی از شوری بالا بر فعالیت میکروبی است. این یافته با مطالعه (Rolnick et al., 2022) که نشان دادند تنفس میکروبی در شرایط شوری بالا مهار می‌شود، هماهنگ است.

همچنین نتایج مدل نشان‌دهنده اثر pH روی پارامتر حساسیت میکروبی به دما بود (Cascio et al., 2016; Ahsan et al., 2021). نیز نشان دادند pH خنثی (بازه ۶/۵ تا ۷/۵) شرایط بهینه‌ای برای رشد و فعالیت میکروبی فراهم می‌کند و در نتیجه ضریب حساسیت دمایی را افزایش می‌دهد.

این پژوهش با به‌کارگیری طیف کامل داده‌های ضریب حساسیت دمایی و استفاده همزمان از مدل‌های ترکیبی و روش‌های هوش مصنوعی قابل توضیح، نسبت به مطالعات پیشین جامع‌تر است. نوآوری اصلی در آن است که برخلاف پژوهش (Saez-Sandino et

al.,2023) داده‌های میانی حذف نشدند و بنابراین امکان تفسیر تعمیم‌پذیرتر فراهم شد. این رویکرد می‌تواند در توسعه مدل‌های پیش‌بینی چرخه کربن جهانی و سیاست‌گذاری‌های اقلیمی نقش مؤثری داشته باشد.

نتیجه‌گیری و پیشنهاد

این مطالعه با بهره‌گیری از الگوریتم‌های یادگیری ماشین و رویکردهای هوش مصنوعی توضیح‌پذیر توانست بینشی دقیق از عوامل مؤثر بر حساسیت تنفس میکروبی خاک به دما ارائه دهد. یافته‌ها نشان داد که ترکیبی از ویژگی‌های میکروبی، فیزیکی و شیمیایی از جمله تنفس القاشده با گلوکز، غنای قارچی و شوری، بیشترین نقش را در طبقه‌بندی ضریب حساسیت دمایی ایفا می‌کنند. تحلیل‌های مبتنی بر SHAP به شفاف‌سازی نحوه تأثیر متغیرها بر خروجی مدل کمک کرده و روابط غیرخطی یا آستانه‌ای مانند اثر شوری را آشکار ساختند. نکته قابل توجه در این پژوهش، استفاده از داده‌های کامل ضریب حساسیت دمایی به‌صورت سه‌کلاسه بود، در حالی که مطالعات پیشین فقط از مقادیر افراطی استفاده کرده‌اند. این رویکرد جامع موجب افزایش توان تعمیم مدل و ارائه تحلیلی واقع‌بینانه‌تر شد پیشنهاد می‌شود در مطالعات آتی، متغیرهای فیزیکی و هیدرولیکی نظیر پایداری خاکدانه‌ها، ظرفیت نگهداری آب، و هدایت هیدرولیکی نیز به مدل افزوده شود تا تصویر کامل‌تری از عملکرد خاک در مواجهه با گرمایش به‌دست آید. همچنین به‌کارگیری پایگاه‌های داده بومی و ناحیه‌ای برای بررسی دقیق‌تر تفاوت‌های منطقه‌ای در ضریب حساسیت دمایی می‌تواند به سیاست‌گذاری بهتر در حوزه کشاورزی و مدیریت خاک کمک کند. در انتها، ترکیب مدل‌های پیش‌بینی ضریب حساسیت دمایی با مدل‌های اقلیمی و چرخه کربن جهانی، می‌تواند در طراحی سیاست‌های اقلیمی مؤثر نقش ایفا کند.

"هیچگونه تعارض منافع بین نویسندگان وجود ندارد"

References

Ahrens, B., et al. (2020). Temperature responses of soil respiration are contingent on the microbial community composition. *Global Change Biology*, 26(7), 3732–3746.

Ahsan, M. M., Mahmud, M. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. *Technologies*, 9(3), 52.

Allen, D. E., Singh, B. P., & Dalal, R. C. (2011). Soil health indicators under climate change: A review of current knowledge. In B. P. Singh, A. L. Cowie, & K. Y. Chan (Eds.), *Soil health and climate change* (pp. 25–45). Springer.

Aydın, Y., Işıkdağ, U., Bekdaş, G., Nigdeli, S. M., & Geem, Z. W. (2023). Use of machine learning techniques in soil classification. *Sustainability*, 15(3), 2374.

Biecek, P., & Burzykowski, T. (2021). *Explanatory model analysis: Explore, explain, and examine predictive models*. CRC Press.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

ویراستاری نشده

Cascio, D., Taormina, V., Cipolla, M., Bruno, S., Fauci, F., & Raso, G. (2016). A multi-process system for HEp-2 cells classification based on SVM. *Pattern Recognition Letters*, 82, 56–63.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM.
Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.

Crowther, T. W., Van den Hoogen, J., Wan, J., Mayes, M. A., Keiser, A., Mo, L., Averill, C., & Maynard, D. S. (2019). The global soil community and its influence on biogeochemistry. *Science*, 365(6455), eaav0550.

Ferrer, L. (2022). Analysis and comparison of classification metrics. *arXiv preprint arXiv:2209.05355*.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42.

Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. In R. Meersman, Z. Tari, & D. C. Schmidt (Eds.), *On the move to meaningful internet systems 2003* (pp. 986–996). Springer.

Haaf, D., Six, J., & Doetterl, S. (2021). Global patterns of geo-ecological controls on the response of soil respiration to warming. *Nature Climate Change*, 11(7), 623–627.

Ibrahimi, E., Lopes, M. B., Dhamo, X., Simeon, A., Shigdel, R., Hron, K., et al. (2023). Overview of data preprocessing for machine learning applications in human microbiome research. *Frontiers in Microbiology*, 14, 1250909.

Lal, R. (2011). Soil health and climate change: An overview. In B. P. Singh, A. L. Cowie, & K. Y. Chan (Eds.), *Soil health and climate change* (pp. 3–24). Springer.

LaValley, M. P. (2008). Logistic regression. *Circulation*, 117(18), 2395–2399.

Mahecha, M. D., Reichstein, M., Carvalhais, N., Lasslop, G., Lange, H., Seneviratne, S. I., et al. (2010). Global convergence in the temperature sensitivity of respiration at ecosystem level. *Science*, 329(5993), 838–840.

Meyer, N., Welp, G., & Amelung, W. (2018). The temperature sensitivity (Q10) of soil respiration: Controlling factors and spatial prediction at regional scale based on environmental soil classes. *Global Biogeochemical Cycles*, 32(3), 306–323.

Nottingham, A. T., Whitaker, J., Turner, B. L., Salinas, N., Zimmermann, M., Malhi, Y., & Meir, P. (2015). Climate warming and soil carbon in tropical forests: Insights from an elevation gradient in the Peruvian Andes. *Bioscience*, 65(9), 906–921.

Novielli, P., Romano, D., Magarelli, M., Bitonto, P. D., Diacono, D., Chiatante, A., Lopalco, G., Sabella, D., Venerito, V., Filannino, P., et al. (2024). Explainable artificial intelligence for microbiome data analysis in colorectal cancer biomarker identification. *Frontiers in Microbiology*, 15, 1348974.

ویراستاری نشده

Ozenne, B., Subtil, F., & Maucort-Boulch, D. (2015). The precision–recall curve overcame the optimism of the receiver operating characteristic curve in rare diseases. *Journal of Clinical Epidemiology*, 68(8), 855–859.

Patil, A., & Lamnganbi, M. (2018). Impact of climate change on soil health: A review. *International Journal of Chemical Studies*, 6(1), 2399–2404.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., et al. (2022). Tackling climate change with machine learning. *ACM Computing Surveys*, 55(2), 1–96.

Saez-Sandino, T., Garcia-Palacios, P., Maestre, F. T., Plaza, C., Guirado, E., Singh, B. K., et al. (2023). The soil microbiome governs the response of microbial respiration to warming across the globe. *Nature Climate Change*, 13(11), 1382–1387.

Steyerberg, E. W. (2019). Coding of categorical and continuous predictors. In *Clinical prediction models* (pp. 175–190). Springer.

Swain, P. H., & Hauska, H. (1977). The decision tree classifier: Design and potential. *IEEE Transactions on Geoscience Electronics*, 15(3), 142–147.

Tong, D., Li, Z., Xiao, H., Nie, X., Liu, C., & Zhou, M. (2021). How do soil microbes exert impact on soil respiration and its temperature sensitivity? *Environmental Microbiology*, 23(6), 3048–3058.

Wen, P., Xu, Q., Yang, Z., He, Y., & Huang, Q. (2024). Algorithm-dependent generalization of AUPRC optimization: Theory and algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(9), 5062–5079.

Wilhelm, R. C., van Es, H. M., & Buckley, D. H. (2022). Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biology and Biochemistry*, 164, 108472.

Winkler, J. P., Cherry, R. S., & Schlesinger, W. H. (1996). The Q10 relationship of microbial respiration in a temperate forest soil. *Soil Biology and Biochemistry*, 28(8), 1067–1072.

Xu, M., et al. (2021). Interactive effects of soil properties on temperature sensitivity of microbial respiration. *Soil Biology and Biochemistry*, 156, 108220.

Yang, S., Wu, H., Wang, Z., Semenov, M. V., Ye, J., Yin, L., Wang, X., Kravchenko, I., Semenov, V., Kuzyakov, Y., et al. (2022). Linkages between the temperature sensitivity of soil respiration and microbial life strategy are dependent on sampling season. *Soil Biology and Biochemistry*, 172, 108758.

ویراستاری نشده