



تحقیقات آب و خاک ایران | دوره ۵۳ | شماره ۳ | خرداد ۱۴۰۱ (ص ۶۰۲-۵۸۵)

[DOI:https://dx.doi.org/10.22059/ijswr.2022.338007.669195](https://dx.doi.org/10.22059/ijswr.2022.338007.669195)

(مقاله علمی - پژوهشی)

Digital Mapping of Top-soil Thickness and Associated Uncertainty Using Machine Learning Approach in Some Part of Arid and Semi-arid Lands of Qazvin Plain

ASGHAR RAHMANI¹, FERAYDOUN SARMADIAN^{*1}, HOSSIEN AREFI²

1. Department of Soil Science, College of Agriculture and Natural Resource, University of Tehran, Karaj, Iran.
2. Department of Remote Sensing and Photogrammetry, Geospatial and Surveying Faculty, College of Engineering, University of Tehran, Tehran, Iran.

(Received: Jan. 25, 2022- Revised: March. 16, 2022- Accepted: Apr. 9, 2022)

ABSTRACT

The present study was carried out to model topsoil thickness using machine learning models (MLM) including random forest (RF) and artificial neural network (ANN) in around 60,000 hectares of Qazvin plain lands (intermediate of Abyek and Nazarabad) with an observational density of 278 profiles during 2016 until 2020, and 17 environmental covariates extracted from Landsat 8 satellite images, primary and secondary derivatives from Digital elevation model, climate data, land use and geology maps. Boruta supervised algorithm and expert knowledge were used to select the best relevant environmental covariates. Two functions include "nnet" and "random forest" (RF) by "caret" package in the R software were used. Modeling of topsoil thickness carried out based on 80% of the data in the calibration subset and 20% of the data was used for model validation. The uncertainty of the output maps was quantified using two methods of "bootstrapping and k-fold". A number of 10 environmental covariates selected among 17 variables, and the relative importance introduced the greenness index, wind effect, diffused radiation, and Mrvbf as the most important covariates, respectively. The validation results indicate that the RF model with R^2 of 0.8 and RMSE less than 3 cm and the bias is 0.63 cm in compare to the ANN, With R^2 , RMSE, and Bias 0.43, 0.05, and 0.004, respectively was outperform. Also, the CCC for the RF model increased by 50% compared to the ANN. The uncertainty estimated by the bootstrapping method was 7 cm lower compared to k-fold in the regions with 10-15 cm thickness and both of two methods show the same spatial pattern in other parts. The RF model along with selected covariates environmental variables and quantified uncertainties of output maps can be used to model the topsoil thickness and management decision making in areas similar to this study in future studies.

Keywords: Random Forest, Artificial Neural Network, Environmental variables, Top-Soil Thickness, Uncertainty.

نقشه‌برداری رقومی ضخامت خاک سطحی و عدم قطعیت وابسته به آن با استفاده از رویکرد یادگیری ماشین در بخشی از اراضی خشک و نیمه‌خشک دشت قزوین

اصغر رحمانی^۱، فریدون سرمیدیان^{۱*}، حسین عارفی^۲

۱. گروه علوم و مهندسی خاک، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.

۲. گروه سنجش از دور و فتوگرامتری، دانشکده نقشه‌برداری و اطلاعات مکانی، دانشکده‌گان فنی، دانشگاه تهران، تهران، ایران.

(تاریخ دریافت: ۱۴۰۰/۱۱/۵ - تاریخ بازنگری: ۱۴۰۰/۱۲/۲۵ - تاریخ تصویب: ۱۴۰۱/۱/۲۰)

چکیده

پژوهش حاضر باهدف مدل‌سازی رقومی ضخامت خاک سطحی با استفاده از مدل‌های یادگیری ماشین جنگل تصادفی (RF) و شبکه عصبی مصنوعی (ANN) در حدود ۶۰۰۰۰ هکتار از اراضی دشت قزوین (حد واسط آبیک و نظرآباد) با تراکم مشاهداتی ۲۷۸ پروفیل در بازه زمانی ۱۳۹۵-۱۳۹۹ و تعداد ۱۷ متغیر محیطی مستخرج از تصاویر ماهواره لندست-۸، مشتقات اولیه و ثانویه مدل رقومی ارتفاع، داده‌های اقلیمی، نقشه کاربری اراضی و زمین‌شناسی اجرا گردید. برای انتخاب متغیرهای کمکی از الگوریتم نظارت‌شده باروتا (Boruta) به همراه نظر کارشناس استفاده شد. از دو تابع "nnet" و "random forest" و بسته "caret" در محیط نرم‌افزار R برای مدل‌سازی بر اساس ۸۰ درصد داده‌ها در مرحله واسنجی و ۲۰ درصد برای اعتبارسنجی استفاده شد و عدم قطعیت نقشه‌های نهایی با دو روش بوتس‌سراپ (bootstrapping) و کا-مرتبه (k-fold) کمی‌سازی گردید. نتایج بیانگر انتخاب ۱۰ متغیر کمکی از میان ۱۷ متغیر بود و متغیرهای شاخص سبزی‌نگی، تأثیر باد، تابش پخشیده و شاخص همواری دره با قدرت تفکیک بالا (Mrvbf) به‌عنوان مهم‌ترین متغیرهای کمکی مشخص گردیدند. نتایج اعتبارسنجی مدل RF بیانگر ضریب تبیین (R^2) ۰/۸ و ریشه میانگین مربعات خطای (RMSE) کمتر از ۳ سانتی‌متر و اریبی (Bias) ۰/۶۳ سانتی‌متر است. در مدل شبکه عصبی مقادیر R^2 ، RMSE و Bias به ترتیب برابر ۰/۴۳، ۰/۰۵ سانتی‌متر و ۰/۰۴ سانتی‌متر حاصل گردید، همچنین ضریب همبستگی توافق (CCC) برای مدل RF در مقایسه با ANN به میزان ۵۰ درصد افزایش نشان می‌دهد، عدم قطعیت برآورد شده توسط روش bootstrapping در مقایسه با k-fold به در مناطق با ضخامت ۱۰ تا ۱۵ سانتی‌متر بیشتر است و در بخش زیادی از منطقه میزان پایین و دارای الگوی مکانی یکسانی می‌باشند. مدل جنگل تصادفی به همراه متغیرهای محیطی انتخاب‌شده و عدم قطعیت‌های کمی شده نقشه‌های خروجی می‌توانند برای مدل‌سازی ضخامت خاک سطحی در نواحی مشابه با این پژوهش در مطالعات آتی استفاده گردد.

واژه‌های کلیدی: جنگل تصادفی، شبکه عصبی مصنوعی، متغیرهای محیطی، ضخامت خاک سطحی، کمی‌سازی عدم قطعیت

مقدمه

(Rahmani et al., 2020; Zhu et al., 2001). به‌منظور برطرف نمودن این محدودیت‌ها، پژوهشگران بر استفاده از مدل‌های کمی-تجربی جهت بررسی تغییرات مکانی عمق خاک تمرکز نموده‌اند. برخی پژوهشگران، از روش‌های رگرسیون خطی چند متغیره (MLR) (Moor et al., 1993; Gessler et al., 1995; Mehnatkesh et al., 2013) رگرسیون خطی تعمیم‌یافته (GLM) و جنگل تصادفی (RF) (Tesfa et al., 2009)، مدل‌های خاک-زمین نمای فازی (Wang et al., 2018) برای مدل‌سازی عمق خاک بر مبنای ارتباط آن با عامل‌های زمین‌نما و پوشش اراضی استفاده نموده‌اند. در ایران، برخی از محققین از جمله (Sepahvand et al., 2017) مهم‌ترین پارامترهای کنترل‌کننده

عمق خاک نقش کلیدی در تنظیم فرآیندهای آب شناختی سطحی و زیرسطحی، مدل‌سازی رشد محصول، عمق ریشه دهی گیاهان، تناسب اراضی و شاخص سلامت خاک ایفا می‌نماید (Leenars et al., 2018; Chen et al., 2021; Zhao & Wu, 2021). این خصوصیت به‌عنوان یکی از ۱۲ ویژگی کلیدی پروژه جهانی تهیه نقشه‌های خاک در قدرت تفکیک مکانی ۹۰ متر، هدف‌گذاری شده است (Arrouays et al., 2014; Hengle et al., 2017). روش‌های مرسوم شناسایی خاک به دلیل نیاز به نمونه-برداری متراکم، وابستگی به دانش کارشناس و به‌روزرسانی دشوار، فرآیند نقشه‌برداری خاک‌ها را محدود و هزینه‌بردار می‌نماید

میانگین عملکرد مدل را نسبت به مدل‌های ناحیه‌ای و ملی مورد بررسی قابل توجه گزارش نمودند و عدم قطعیت نمونه‌های اعتبارسنجی را به دلیل قرار گرفتن ۷۰ درصد آن‌ها در فاصله پیش‌بینی ۹۰٪ منطقی عنوان نمودند.

مطالعات گذشته مدل‌سازی مکانی عمق خاک، به‌طور معمول در اراضی دارای پستی‌وبلندی شدید (تپه‌ها و کوهستان) و مساحت‌های کوچک اجرا گردیده و بیشتر آن‌ها بر روی مدل‌سازی عمق کل خاک تحقیق نموده‌اند، با این وجود ضخامت خاک سطحی به‌عنوان آسیب‌پذیرترین بخش آن نیازمند بررسی و مستندسازی است، از طرفی به دلیل فعالیت فشرده کشاورزی در منطقه مطالعاتی و دارا بودن سطوح ژئومورفیک متنوع با مساحت قابل توجه و نبود نقشه قبلی از ضخامت خاک سطحی، پژوهش حاضر باهدف مدل‌سازی و نقشه‌برداری رقومی ضخامت خاک سطحی افق‌های (A و Ap) با استفاده از دو مدل یادگیری ماشین جنگل تصادفی و شبکه عصبی مصنوعی به‌عنوان پرکاربردترین مدل‌ها در اکثر تحقیقات انجام‌شده در ایران (Zeraatpisheh et al., 2020) اجرا گردید.

روش بررسی

منطقه مورد مطالعه

پژوهش حاضر در اراضی دشت قزوین واقع در محدوده $35^{\circ}06'$ تا $50^{\circ}15'$ تا $50^{\circ}29'25/53''$ طول شرقی و $35^{\circ}54'30/38''$ تا $38^{\circ}18/83''$ عرض شمالی با مساحت تقریبی 60000 هکتار اجرا گردید. در مجموع از داده‌های ۲۷۸ پروفیل مشاهداتی که در طی پنج سال (۱۳۹۵-۱۳۹۹) بر اساس روش طبقه‌بندی تصادفی با متوسط فاصله ۱۲۰۰ متر که براساس میزان مساحت کاربری‌های موجود اراضی و نظر کارشناس تعیین گردیده بودند، استفاده گردید. از نظر وضعیت کاربری اراضی، کشاورزی آبی در مرکز دشت، دیم در شمال منطقه و مراتع شور و غیر شور در جنوب و شمال منطقه قرار دارند. متوسط ارتفاع منطقه ۱۴۴۰ متر از سطح دریای آزاد بوده و دارای بارندگی سالیانه ۲۸۵ میلی-متر و متوسط دمای سالانه $14/35$ درجه سانتی‌گراد است. رژیم رطوبتی خاک بر اساس داده‌های هواشناسی بلندمدت (۴۹ ساله- ایستگاه هواشناسی قزوین ۱۳۹۸-۱۳۴۹) و مدل نیوهال در نرم‌افزار jNSM نسخه ۱/۶/۱ (Van Wambeke, 2000) به ترتیب زیریک خشک، اریدیک ضعیف، آکوییک و رژیم حرارتی خاک‌ها ترمیک می‌باشد. سطوح ژئومورفیک دشت دامنه‌ای و بعد از آن دشت، تپه و کوهستان مهم‌ترین واحدهای زمین‌نمای قابل تفکیک در اراضی مورد مطالعه می‌باشند (Zinck et al, 2016; Mousavi

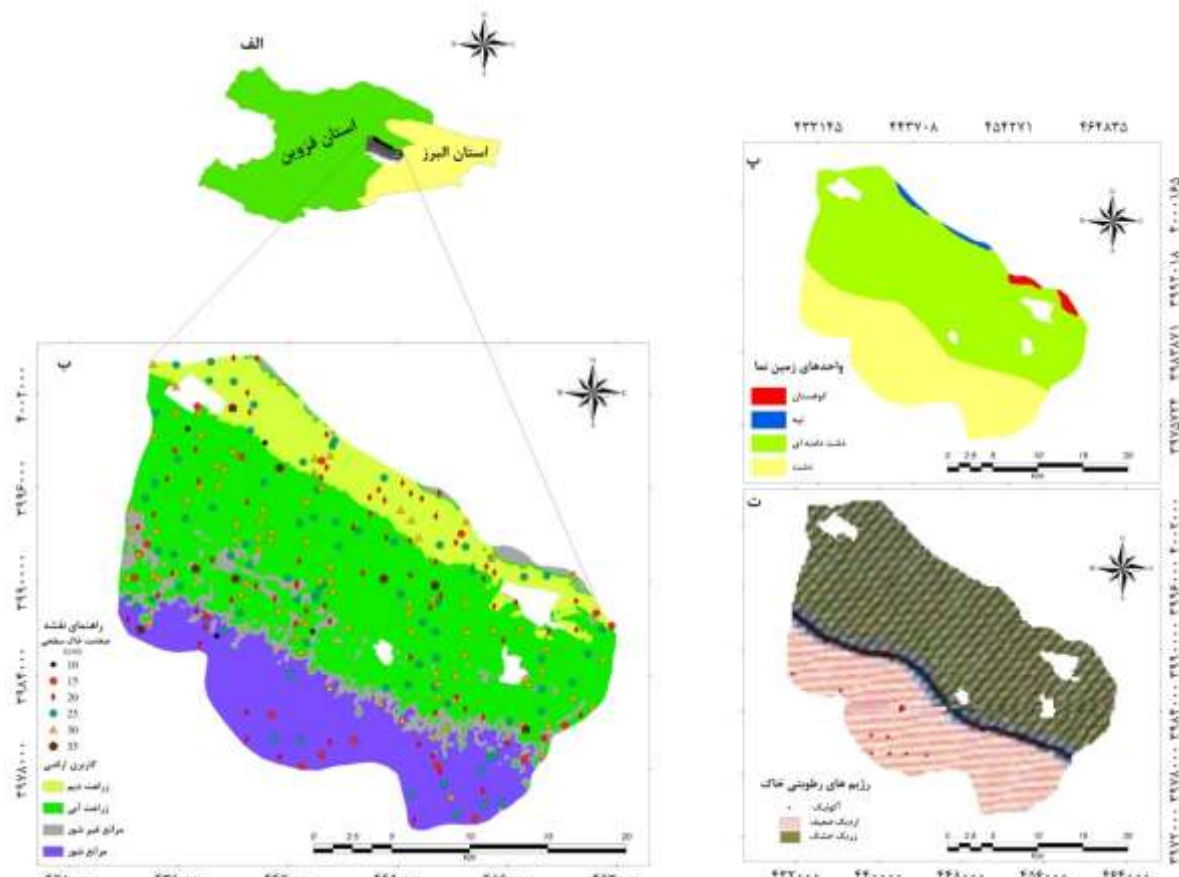
تغییرات مکانی عمق کل خاک را مشتقات مدل رقومی ارتفاع با استفاده از روش MLR عنوان نمودند. در تحقیقی دیگر نتایج مدل RF بیانگر ضریب تبیین مناسب بوده و شاخص‌های NDVI، سطح ویژه حوضه آبخیز، جهت شیب، گرادیان شیب و کاربری اراضی به‌عنوان مهم‌ترین متغیرهای کمکی برآورد کننده عمق خاک گزارش نمودند (Zahedi et al., 2017).

در مطالعات خارج از ایران، از پارامترهای درصد شیب، انحنای طولی و عرضی، شاخص خیزی توپوگرافی برای مدل‌سازی عمق خاک استفاده شده و نتایج بیانگر ضریب تبیین $0/64$ تا $0/77$ برای مدل ANN و $0/63$ تا $0/66$ برای مدل MLR بود (Han et al, 2018). مطالعه‌ای دیگر در حوضه کوچک جنوب غرب چین برای مدل‌سازی ضخامت لایه فعال عمق خاک نتایج بیانگر دقت بالاتر و خطای پیش‌بینی کمتر مدل جنگل تصادفی نسبت به سایر مدل‌ها بود (Li et al., 2017). همچنین برتری مدل RF در مقایسه با مدل‌های دیگر به‌منظور پیش‌بینی ضخامت کل خاک و خصوصیات هیدرولیکی در کشور هند توسط محققین گزارش گردید (Dharumarajan et al., 2021).

ارزیابی مدل‌های پیش‌بینی با استفاده از داده‌های مستقل نیازمند تفسیر با دقت و برآورد عدم قطعیت در نتایج اعتبارسنجی می‌باشد (Vanwinckelen and Blockeel, 2012). در فرایند مدل‌سازی رقومی نقشه‌های خاک، عدم قطعیت خروجی مدل، برآیندی از سه منبع اصلی شامل ساختار مدل، پارامتر مدل و ورودی مدل است. رویکرد عمومی این است که سهم هر کدام از منابع در عدم قطعیت نهایی به‌طور جداگانه تعیین گردد که به دلیل زمان‌بر بودن و هزینه به‌سختی انجام می‌گردد (McBratney et al., 2018). از روش‌های مختلفی مانند واریانس کریجینگ (Vaysse and Lagacherie, 2017) یا شبیه‌سازی (Heuvelink et al., 2015; Sztatmari et al., 2016) برای برآورد عدم قطعیت روش‌های زمین‌آماري تاکنون استفاده شده است، با این وجود برآورد عدم قطعیت مدل‌های یادگیری ماشین به‌عنوان یکی از هدف‌های اصلی نقشه‌برداری رقومی کمتر پرداخته شده و مطالعات بیشتر در این زمینه نیاز است (Sztatmari and Pasztor, 2019). در پژوهشی توسط (Wang et al., 2021) چهار روش نمونه‌گیری مجدد (۵ برابر و ۱۰ برابر) و راه‌اندازی (بوت استرپ و بوت استرپ optimism) به کار گرفته شد و نتایج بیانگر دقت بالاتر روش بوت استرپ نسبت به سایر روش‌ها در پیش‌بینی متغیر هدف بود. برای کمی‌سازی عدم قطعیت نقشه‌های ضخامت خاک پژوهشگرانی از جمله (Chen et al., 2021) ضخامت خاک تشکیل شده از رسوبات لسی از جنگل‌های رگرسیون چندک (QRF) به دلیل توانایی ارائه فواصل پیش‌بینی ۹۰٪ (PIs) برای مدل‌سازی مکانی و برآورد عدم قطعیت نقشه استفاده نمودند.

سولز، اینسپیتی سولز و انتی سولز) قرار دارند.

(et al, 2017). براساس سامانه آمریکایی رده‌بندی خاک‌ها (Soil Survey Staff, 2014) خاک‌های منطقه در سه رده‌ی (اریدی



شکل ۱- الف) - موقعیت نسبت به استان‌های البرز، قزوین ب) - نقاط نمونه‌برداری در منطقه مورد مطالعه با کاربری‌های مختلف اراضی پ) - نقشه واحدهای زمین نما ت) - نقشه رژیم‌های رطوبتی.

مطالعات میدانی

بر اساس اهداف این تحقیق همزمان با مطالعات میدانی، ضخامت خاک سطحی در محل هر یک از پروفیل‌های مطالعاتی مورد بررسی بر اساس تشریح صحرایی ۲۷۸ پروفیل اندازه‌گیری با استفاده از متر فلزی گردید.

متغیرهای کمکی و انتخاب آنها

برای مدل‌سازی از تعداد ۱۷ متغیر محیطی از منابع مختلف شامل شاخص‌های طیفی حاصل از داده‌های سنجش‌ازدور ماهواره لندست ۸ (OLI/TIRS) با قدرت تفکیک ۳۰ متر منطبق با بازه مطالعات میدانی (دوره ۵ ساله ۲۰۱۶-۲۰۲۰) شامل شاخص گیاهی تفاضلی نرمال شده^۱، شاخص نرمال شده رس^۲، شاخص سبزیگی حاصل از تبدیل تسلدکپ^۳، برگرفته از سایت سازمان زمین‌شناسی آمریکا (U.S Geology Survey 2014) در نرم‌افزار ERDAS IMAGINE 2014 پس انجام اصلاحات لازم

روند نمای پژوهش

به‌طور کلی روند انجام پژوهش در مراحل: ۱) - اندازه‌گیری ضخامت خاک سطحی در محل ۲۷۸ پروفیل خاک حفر شده ۲) - تهیه متغیرهای محیطی کمکی با استفاده از تصاویر ماهواره لندست ۸، مدل رقومی ارتفاع، نقشه رقومی کاربری اراضی به همراه زمین شناسی ۳) - انتخاب مهم‌ترین متغیرهای محیطی بر اساس نظر کارشناس و الگوریتم نظارت شده (Boruta ۴) - مدل‌سازی رابطه بین داده‌های ضخامت خاک سطحی و متغیرهای محیطی منتخب با استفاده از مدل‌های RF، ANN و بهینه‌سازی پارامترهای مدل‌ها بر مبنای روش اعتبارسنجی متقابل ۱۰ گام در ۱۰ تکرار با استفاده از ۸۰ درصد داده‌ها ۵) - اعتبارسنجی مدل‌ها با استفاده از ۲۰ درصد باقیمانده داده‌ها بر مبنای شاخص‌های آماری $(R^2, RMSE, CCC, Bias)$ ۶) - نقشه رقومی ضخامت خاک سطحی و کمی‌سازی عدم قطعیت نقشه پیش‌بینی (شکل ۲).

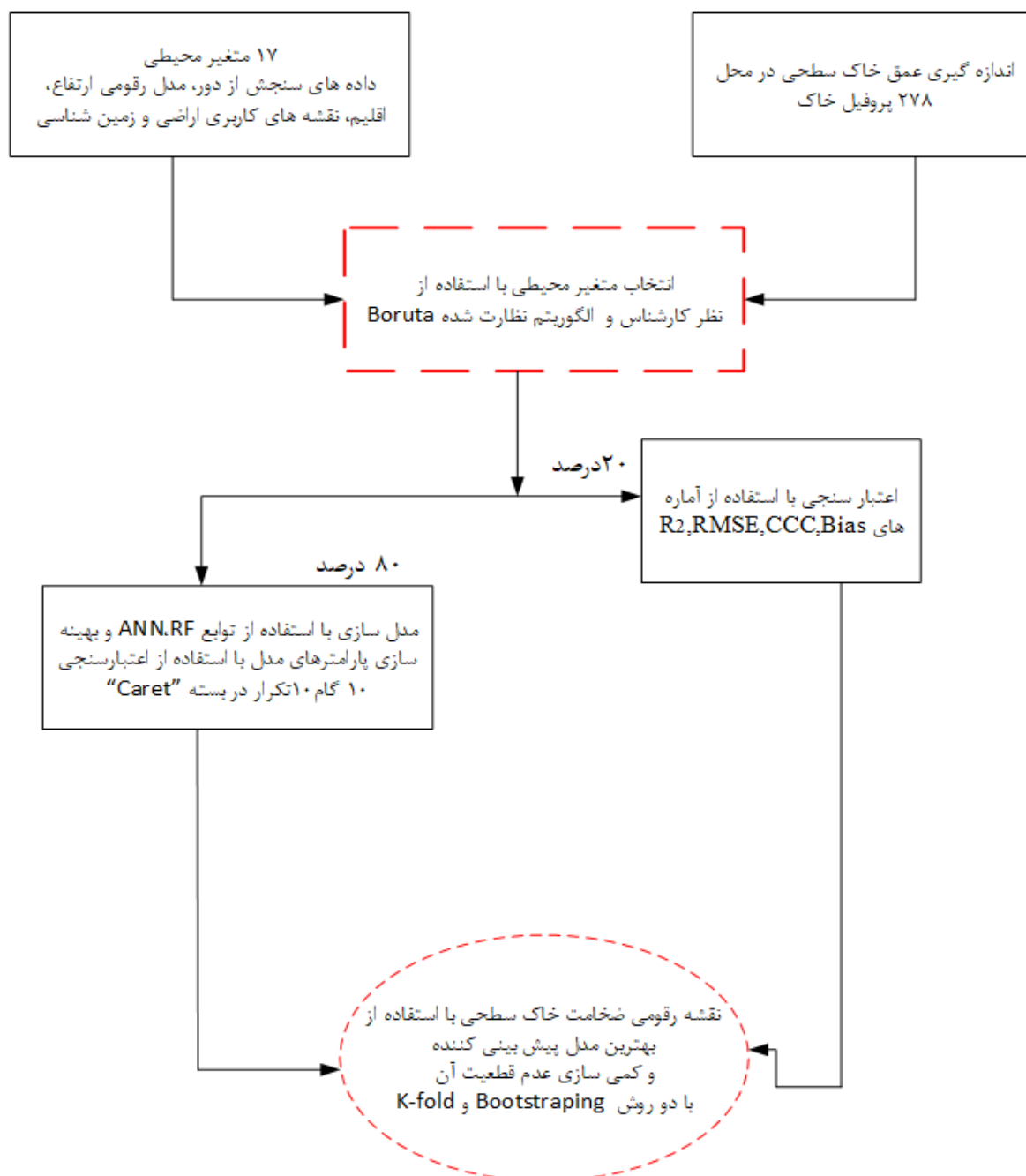
1. Normalized difference vegetation index

2. Clay normalized index

3. Tasseled Cap Transformation

داده‌های مستخرج از ۲۴ ایستگاه هواشناسی موجود در منطقه طی یک دوره آماری ۴۹ ساله (۱۳۴۹-۱۳۹۸) (Iran Meteorological organization, 2018) با استفاده از روش کریگینگ بی‌زین^۲ تجربی پهنه‌بندی گردیدند. همچنین تمامی لایه‌های محیطی در محیط نرم‌افزار ArcGIS با استفاده از دستورالعمل نزدیکترین همسایگی^۳ به قدرت تفکیک مکانی ۳۰ متر با نمونه‌گیری گردیدند.

(رادپومتري) تهیه گردیدند. مشتقات اولیه و ثانویه مشتقات پستی‌وبلندی بر پایه روش‌های مبتنی بر تجزیه پستی‌وبلندی (Wilson et al., 2018) از مدل رقومی ارتفاع ماهواره آلوس پالسار^۱ با قدرت تفکیک ۱۲/۵ متر در محیط نرم‌افزار SAGA GIS نسخه ۷/۹/۱ (Olaya and Conrad, 2015)، زمین‌شناسی با مقیاس ۱:۱۰۰۰۰۰ به همراه نقشه کاربری اراضی تهیه‌شده از تصاویر ماهواره لندست ۸ استفاده شد (جدول ۱)، دو عامل میانگین بارندگی سالیانه و میانگین درجه حرارت سالیانه بر اساس



شکل ۲- روندنمای اجرای پژوهش

جدول ۱- متغیرهای محیطی مورد استفاده برای مدل سازی مکانی ضخامت خاک سطحی

نام متغیر و نماد	متغیر کمکی محیطی / شاخص	عامل خاک سازی در مدل scorpan
Wind effect (WE)	شاخص اثر باد	
Multi-Resolution Index of Valley Bottom Flatness (Mrvbf)	شاخص همواری دره با وضوح مکانی بالا	
Diffuse insolation (Diffuse)	شدت تابش پخشیده	توپوگرافی (r)
Standard Height	ارتفاعات استاندارد	
Mid-slope position	موقعیت میانی شیب	
Valley_depth	عمق دره	
SAGA wetness index	شاخص خیسگی ساگا	
Topographic wetness index (wetness)	شاخص خیسگی توپوگرافی	
Catchment (CAT)	حوزه های آبخیز	
Mean annual temperature (MAT)	میانگین دمای سالانه	اقلیم
Mean annual precipitation (MAP)	میانگین بارندگی سالانه	(c)
$NDVI = NIR - RED / NIR + RED$	شاخص تفاضلی پوشش گیاهی نرمال شده	
$Greenness = -0.2848 B2 - 0.2435 B3 - 0.5436 B4 + 0.7243 B5 + 0.0840 B6 - 0.1800 B7$	شاخص سبزیگی	موجودات زنده (o)
$Brightness = 0.3029Blue + 0.2786Green + 0.4733Red + 0.5599NIR + 0.508SWIR1 + 0.1872SWIR2$	شاخص روشنایی	
Land use1	کاربری اراضی	
$Clay_index = B6 - B7 / B6 + B7$	شاخص رس	مواد مادری
Geology	نقشه زمین شناسی	(p)

Bands numbers, B2: Blue, B3: Green, B4: RED, B5: NIR, B6: SWIR1, B7: SWIR2. S: برای تعدیل اثر خاک پس زمینه

مدلسازی مکانی

تمامی فرآیند مدل سازی در نرم افزار آماری R، با استفاده از بسته های تخصصی "caret" و تابع "random forest" با تعداد درخت بهینه ۱۵۰۰ و mtry برابر با ۲ با حداقل RMSE (شکل ۳ ب) انجام گردید (Horst-Heinen et al., 2021). در فرآیند مدل سازی از ۸۰ درصد داده ها (۲۲۲ نمونه) برای واسنجی و همچنین بهینه سازی پارامترهای مدل RF و ANN با استفاده از روش اعتبارسنجی متقابل ۱۰ گام با ۱۰ تکرار در بسته "caret" و از ۲۰ درصد باقیمانده داده ها (۵۶ نمونه) برای اعتبارسنجی نقشه نهایی پیش بینی شده عمق سطحی خاکها استفاده شد. از تابع "nnet" بسته "caret" تعداد نورون های لایه مخفی با حداقل RMSE برای تعداد ۵ نورون در ضریب واپاشی وزنی ۰/۰۰۱ بهینه گردید (شکل ۳ الف). هنگامی که یک شبکه عصبی (nnet) با استفاده از Caret

انتخاب متغیر محیطی

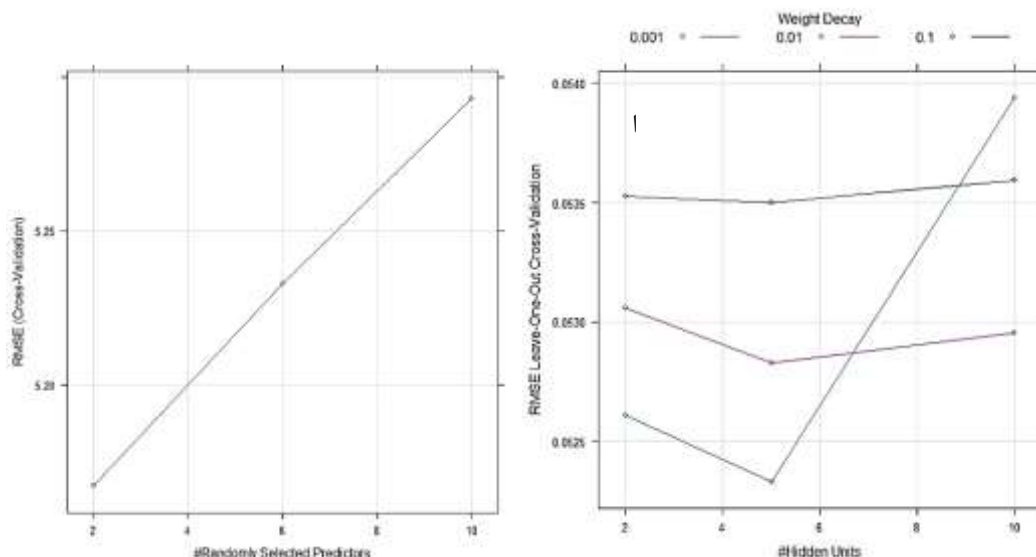
برای جلوگیری از طولانی شدن زمان مدل سازی و بیش برآزش در این پژوهش از الگوریتم "Boruta" برای انتخاب متغیرهای محیطی بهینه، استفاده شد (Kursa and Rudnicki, 2010). این روش جزء روش های نظارت شده - نیمه خودکار انتخاب متغیرهای کمکی می باشد. مبنای اجرای این روش بر اساس مدل RF می باشد که در آن مهمترین متغیرهای کمکی محیطی بر اساس یک سیستم رفت و برگشتی تکرار پذیر در حضور متغیر هدف مورد بررسی قرار می گیرند. در نهایت خروجی فرآیند انتخاب متغیر بر اساس مقدار عامل Z که در چهار کلاس کلی تعیین می گردد. در صورتی که برای یک متغیر کمکی مقدار عامل Z به ترتیب کمتر از ۵، حدواسط ۵ تا ۱۰، ۱۰ تا ۱۵ و بیش از ۱۵ باشد، غیر مرتبط، کمی مرتبط، نسبتاً مرتبط و کاملاً مرتبط محسوب می شود (Keskin et al., 2019). فرآیند انتخاب متغیر بر اساس بسته تخصصی "Boruta" در نرم افزار R صورت پذیرفت.

که از مدل درخت طبقه بندی و رگرسیون (CART) مشتق شده است. درختان منفرد با استفاده از یک نمونه بوت استرپ از داده های آموزشی آموزش داده می شوند به طوری که هر بار زیر مجموعه ای از داده های اصلی (در کیسه) برای آموزش استفاده می شود. باقیمانده برای اعتبارسنجی متقاطع داخلی (خارج از کیسه) استفاده می شود این مدل روابط غیر خطی و پیچیده بین خصوصیات خاک و عامل های محیطی ثبت می کند و برای تنظیم به پارامترهای کمی نیاز دارد و پیاده سازی آن آسان است (شفیع زاده مقدم، ۲۰۲۱). پارامتر تنظیمی بحرانی در RF به عنوان mtry شناخته می شود که تعداد متغیرهای کمکی به طور تصادفی انتخاب شده را برای تقسیم در هر گره تعیین می کند. در مدل سازی رگرسیونی این مدل از میانگین پاسخ یادگیرنده پایه که بارها اجرا می شود ارائه می نماید (Breiman 2001).

آموزش داده می شود، باید دو پارامتر اندازه^۱ و ضریب واپاشی وزنی^۲ بهینه گردد. اندازه تعداد نورون های لایه پنهان است و ضریب واپاشی وزنی، پارامتر تنظیم برای جلوگیری از بیش برآزش بوده به صورتی که در هر به روزرسانی، وزن ها در یک ضریب کمتر از ۱ ضرب می شوند. این عمل از بزرگ تر شدن بیش از حد وزن ها جلوگیری می کند که معمولاً به صورت لگاریتمی تغییر می نماید در این تحقیق ۰/۱، ۰/۰۱ و ۰/۰۰۱ توسط بسته Caret به طور خودکار تست و بهینه ترین جواب در حالت ۵ نورون و ضریب ۰،۰۰۱ است.

مدل جنگل تصادفی

یادگیرنده جنگل تصادفی (RF) از رویکردهای غیر پارامتریک است که از مجموعه ای از مدل های مبتنی بر درخت استفاده می کند



شکل ۳- نمودار مقادیر پارامترهای بهینه شده مدل های الف) - شبکه عصبی ب) - جنگل تصادفی

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (Y_i - P_i) \quad (\text{رابطه ۳})$$

$$\text{CCC} = \frac{2r \partial_{Y_i} \partial_{P_i}}{\partial_{Y_i}^2 + \partial_{P_i}^2 + (Y + P)^2} \quad (\text{رابطه ۴})$$

در معادلات فوق Y_i : مقادیر اندازه گیری شده، P_i : مقادیر پیش بینی شده ضخامت خاک سطحی \bar{Y} میانگین مقادیر اندازه گیری شده، \bar{P} میانگین مقادیر پیش بینی شده و n : تعداد مشاهدات است.

کمی سازی عدم قطعیت

رویکردهای متنوعی برای کمی سازی عدم قطعیت خروجی مدل

اعتبارسنجی

ارزیابی کارایی مدل RF بر اساس شاخص های آماری ضریب تبیین (R^2) و ریشه میانگین مربعات خطا (RMSE)، آماره اریبی (Bias) و ضریب تطابق همبستگی (CCC) صورت پذیرفت (به ترتیب روابط ۱ تا ۴).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - P_i)^2} \quad (\text{رابطه ۱})$$

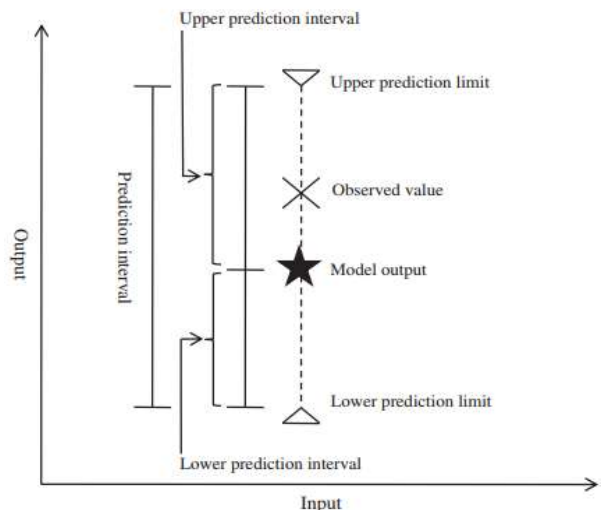
$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - P_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (\text{رابطه ۲})$$

برای هر یک از پیکسل‌ها فراهم می‌گردد. توضیحات بیشتر در مطالعات (Viscarrá Rossel et al., 2015) و (Liddicoat et al., 2015) برای نقشه‌برداری رقومی ویژگی‌های خاک ارائه شده است. اعتبار سنجی متقاطع k-fold به شرح زیر عمل می‌کند. کل داده‌های موجود به k زیر بخش با اندازه مساوی تقسیم می‌شوند که به آن folds یا چینه نیز می‌گویند. در هر مرحله یک چینه کنار گذاشته شده و مدل با توجه به سایر چینه‌ها آموزش می‌بیند که در نهایت منجر به ساخت یک مدل یاد گیرنده می‌شود. سپس دقت این مدل بر روی هر دسته از داده‌های کنار گذاشته شده بررسی شده و در نهایت میانگین تمام این دقت‌ها به عنوان تخمین مقدار متغیر هدف در هر پیکسل محاسبه می‌شود (Vanwinckelen and Blockeel, 2012). در این پژوهش از ۱۰ فولد برای تخمین عدم قطعیت استفاده گردید. جهت اجرای این روش‌ها از مدل RF و کد نویسی در نرم افزار متن‌باز آماری R استفاده گردیده است.

نتایج و بحث خلاصه آماری

خلاصه آماری داده‌های واسنجی و اعتبار سنجی ضخامت سطحی خاک‌ها در دشت قزوین در جدول ۱ ارائه شده است. همان‌طور که مشاهده می‌گردد داده‌های واسنجی و اعتبارسنجی دارای مقادیر حداقل، حداکثر مشابه بوده و میانگین داده نیز اندکی باهم اختلاف دارد. از طرفی میزان انحراف معیار داده‌های اعتبارسنجی به میزان ۱ سانتی‌متر در عمق سطحی از داده‌های واسنجی بیشتر و ضریب تغییرات آن به میزان ۷ درصد از داده‌های واسنجی بیشتر است، باین‌وجود دودسته داده از نظر میزان ضریب تغییراتی پذیری در کلاس طبقه‌بندی متوسط (۱۵ تا ۳۵ درصد) بر اساس روش (Wilding, 1945) قرار می‌گیرند. (Zahedi et al., 2017^a) ضریب تغییرات بیش از ۶۰ درصد را برخلاف نتایج این مطالعه برای عمق کل خاک گزارش نموده و میزان بالای تغییرپذیری را ناشی از شدت توپوگرافی در تیپ اراضی تپه‌ای بیان داشتند. میزان میانگین مشاهده‌شده در دودسته داده با مقدار ۲۰ سانتی متر نشان‌دهنده وضعیت مناسب سلامت خاک‌های این منطقه بر اساس استانداردهای موجود است (Zhao and Wu, 2021). توزیع فراوانی داده‌های مرحله واسنجی و اعتبارسنجی نشان‌دهنده انتخاب بهینه دسته‌های داده جداگانه از هر کدام از ضخامت‌های خاک سطحی در این پژوهش می‌باشد (شکل ۵).

وجود دارد، یکی از این روش‌ها، رویکرد تجربی است که باقیمانده‌های بین خروجی‌های مدل‌سازی شده و داده‌های مشاهداتی را برای کمی‌سازی فاصله پیش‌بینی^۱ (PI) مورد استفاده قرار می‌دهد. چنین رویکردی توسط Solomatine, Shrestha and (2006) پیشنهاد شد و بیان نمودند که فاصله پیش‌بینی به‌صراحت تمام منابع عدم قطعیت را در نظر می‌گیرد. ویژگی‌های یک فاصله پیش‌بینی شامل هر دو حد پیش‌بینی بالا و پایین است (شکل ۴). در DSM یک راه متداول برای نمایش واضح عدم قطعیت یک پیش‌بینی مکانی، ترسیم حد بالایی و پایینی فاصله پیش‌بینی ۹۰ درصدی (PI) است. PI محدوده مقادیری را گزارش می‌کند که انتظار می‌رود مقدار واقعی در آن ۹ بار از ۱۰ مرتبه رخ دهد. حد پایین و بالای ۹۰٪ PI را می‌توان به راحتی با تفریق و اضافه ۱,۶۴ برابر انحراف استاندارد پیش‌بینی به میانگین پیش‌بینی هر پیکسل محاسبه نمود.

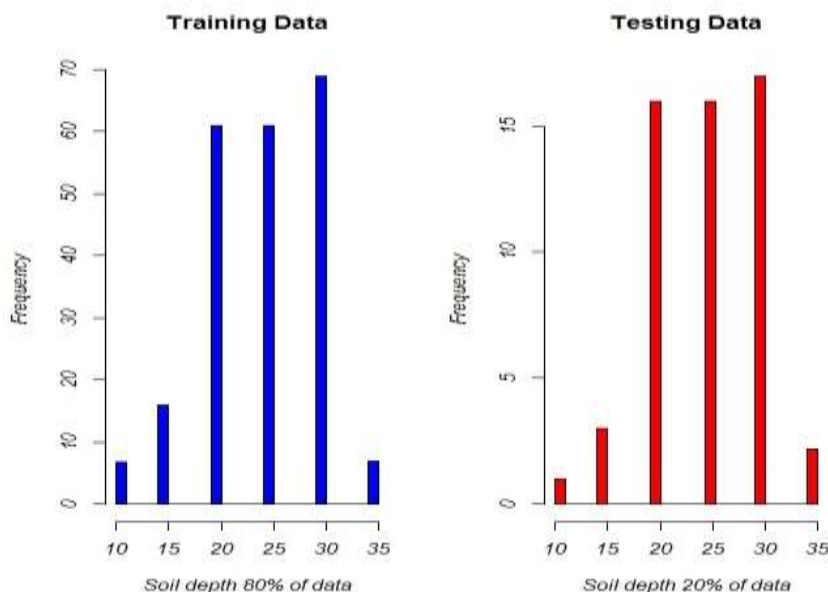


شکل ۴: شماتیک فاصله پیش‌بینی، ویژگی‌های مشخصه و روش‌شناسی تشریحی عمومی اقتباس شده از Shrestha & Solomatine (2006)

برای کمی‌سازی عدم قطعیت در این پژوهش از روش‌های Bootstrapping و k-fold استفاده گردید. بوت‌سرایت یکی از روش‌های غیرپارامتریک و شناخته‌شده برای محاسبه عدم قطعیت مدل‌سازی است (Tibshirani and Efron, 1993). کارکرد این روش شامل یک نمونه‌برداری تصادفی همراه با قابلیت جایگزینی از داده‌های موجود می‌باشد. بر اساس هر بار انتخاب تصادفی از داده‌های موجود در بخش واسنجی یک مدل برازش می‌گردد که منجر به پیش‌بینی متغیر هدف می‌گردد؛ بنابراین با تکرار هر یک از مدل‌ها در هر مرحله در نهایت امکان دست‌یابی به نقشه احتمال

جدول ۱- خلاصه آماری داده‌های واسنجی و اعتبارسنجی ضخامت خاک سطحی

اعتبارسنجی (۲۰ درصد)	واسنجی (۸۰ درصد)	پارامتر
۱۰	۱۰	کمینه
۲۳/۳۶	۲۴/۷۲	میانگین
۳۵	۳۵	بیشینه
۶/۳۸	۵/۰۷	انحراف معیار
۲۷/۳	۲۰/۵	ضریب تغییرات (درصد)



شکل ۵- توزیع فراوانی داده‌های واسنجی و اعتبارسنجی ضخامت خاک سطحی

ارزیابی عملکرد مدل‌های RF و ANN

نتایج اعتبارسنجی داخلی^۱ و خارجی^۲ ارائه شده در جدول ۲، برای مدل RF بیانگر دقت بالا و میزان خطای پایین این مدل برای پیش‌بینی ضخامت خاک سطحی است. میزان خطای پیش‌بینی در هر دو بخش به میزان کمتر از ۳ سانتی‌متر و ضریب تبیین نیز توجیه‌کننده بیش از ۸۰ درصد تغییرات داده‌های واقعی ضخامت خاک سطحی است که بر اساس طبقه‌بندی ارائه شده توسط (Vissara Rossel and MacBratney, 2008) در کلاس پیش‌بینی دقت عالی قرار می‌گیرند. نمودار پراکنش داده‌های پیش‌بینی شده در مقابل داده‌های واقعی بر اساس شکل ۵ الف برای مدل RF در دو دامنه مقادیر بالا و پایین ضخامت خاک سطحی نسبت به مدل ANN به خط ۱:۱ نزدیک‌تر و دامنه خطای کمتری برخوردار است.

میزان صحت مدل بر اساس شاخص آماری CCC به میزان بالای ۰/۸۴ به دست آمد. میزان ضریب تبیین به دست آمده برای مدل ANN به میزان ۵۰ درصد از مدل RF پایین‌تر و خطای

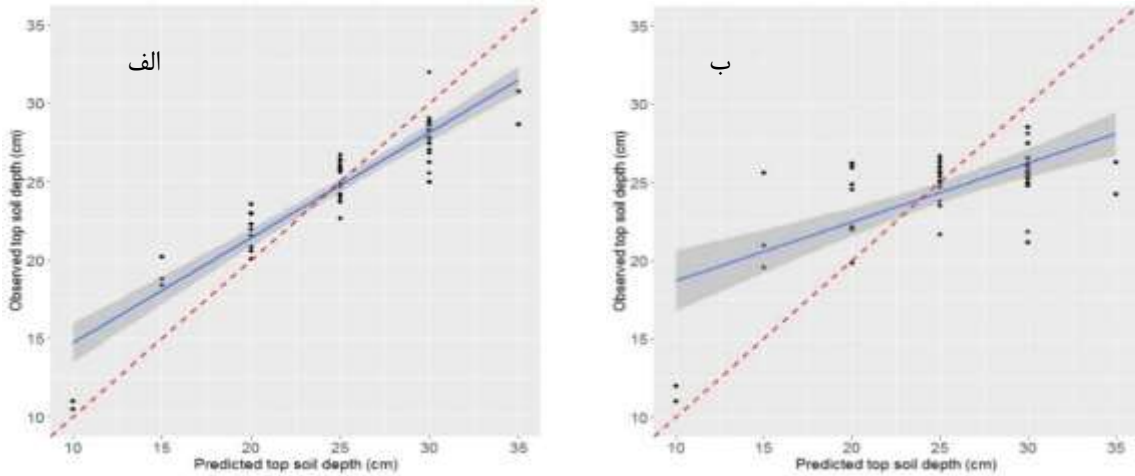
مدل‌سازی کمتر حاصل گردید. برخی محققین به ترتیب ضریب تبیین ۰/۷۵ و ۰/۶۲ با استفاده از تعداد داده بیشتر در مرحله اعتبارسنجی برای مدل RF گزارش نمودند (Zahedi et al., 2017). (Zahedi et al., 2017; Li et al., 2017) که در مقایسه با پژوهش حاضر (با تعداد ۵۵ داده اعتبارسنجی) با مساحت کوچک‌تر و تعداد کل داده بیشتر برای مدل‌سازی ضخامت خاک دارای دقت پایین‌تری می‌باشند. همچنین (Sepahvand et al., 2013; Mehnatkesh et al., 2017) به ترتیب میزان ضریب تبیین ۰/۶۵ و ۰/۷۶ را با تعداد داده و مساحت کمتر نسبت به پژوهش حاضر و شرایط توپوگرافی متفاوت با پژوهش حاضر برای مدل MLR متغیره گزارش نمودند. از طرفی (Zahedi et al., 2017^b) برخلاف پژوهش حاضر، مدل ANN برای کاربری کشاورزی دارای خاک عمیق و مدل RF برای پیش‌بینی نواحی کم‌عمق با پستی‌وبلندی شدید به‌عنوان مدل‌های برتر گزارش نمودند. همچنین مشابه با نتایج به‌دست‌آمده برای مدل RF در این پژوهش، (Horst-Heinen et al., 2021) میزان ضریب تبیین بالا و حتی بیشتر $R^2=0/91$ و

2 External validation

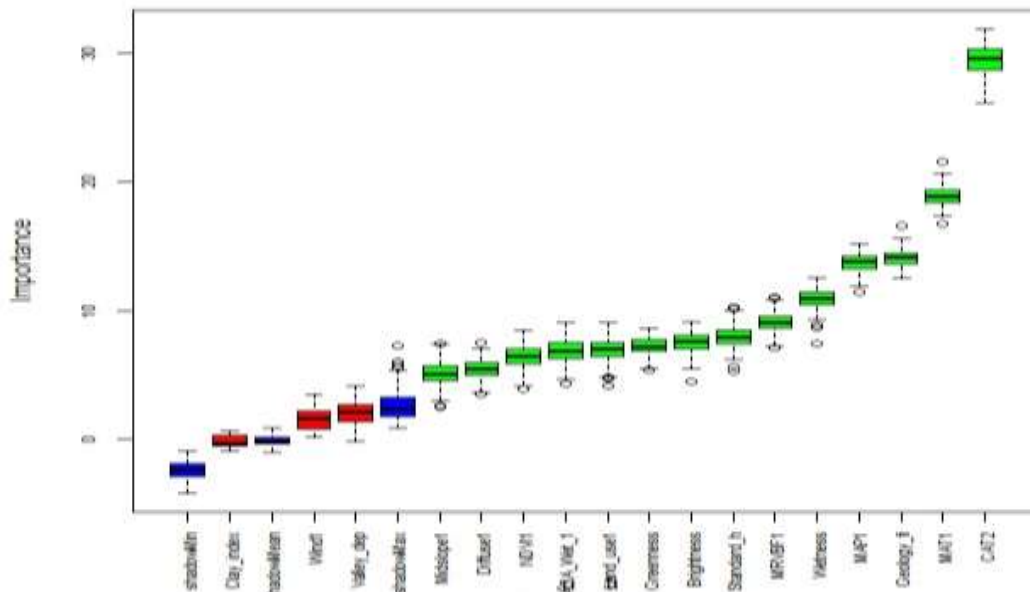
1 Internal validation

شرایط تعداد داده کمتر و مساحت کوچکتر گزارش نمودند.

خطای کمتر $RMSE = 0.17$ سانتی‌متر در پیش‌بینی عمق خاک تا سنگ‌بستر در اراضی با پستی‌وبلندی شدید با کاربری جنگل در



شکل ۶- نمودار پراکنش داده‌های پیش‌بینی شده ضخامت خاک سطحی در مقابل داده‌های اندازه‌گیری شده الف)- مدل RF ب)- مدل ANN



شکل ۷- نتایج الگوریتم Boruta برای شناسایی مهم‌ترین متغیرهای محیطی به منظور پیش‌بینی ضخامت خاک سطحی

CAT: حوضه‌های آبخیز؛ MAT: میانگین دمای سالیانه؛ Geology: زمین‌شناسی؛ MAP: میانگین بارش ماهانه Wetness شاخص خیسگی توپوگرافی؛ Mrvbf: شاخص همواری دره با قدرت تفکیک بالا؛ Standard_h: ارتفاعات استاندارد شده؛ Brightness: شاخص روشنایی؛ Greenness: شاخص سبزیگی؛ Land_use: کاربری اراضی؛ SAGA_wet: شاخص خیسگی ساگا؛ NDVI: شاخص اختلافات نرمال شده پوشش گیاهی؛ Diffuse1: میزان تابش پخشیده؛ Midslope1: موقعیت‌های میانی شیب؛ Valley_dep: عمق دره؛ Wind: شاخص تأثیر باد؛ Clay_index: شاخص تأثیر رس

*جعبه‌های سبز و قرمز نشان دهنده به ترتیب نشان دهنده ویژگی‌های تایید و رد شده می‌باشد.

جدول ۲- اعتبارسنجی داخلی و خارجی پیش‌بینی ضخامت خاک سطحی با استفاده از مدل‌های RF و ANN

مدل	R^2		CCC		RMSE (cm)		Bias (cm)	
	داخلی	خارجی	داخلی	خارجی	داخلی	خارجی	داخلی	خارجی
RF	کل داده‌های ضخامت							
	میانگین	۰/۸۴	۰/۸۰	۰/۸۱	۰/۸۰	۰/۸۱	۰/۸۴	۰/۸۰
	انحراف معیار	۰/۰۳	۰/۰۵	۰/۰۴	۰/۰۵	۰/۲۰	۰/۷۲	۰/۳۲
	حداقل	۰/۸۰	۰/۷۱	۰/۷۴	۰/۷۱	۲/۲۸	۱/۴۰	۰/۸۸
حداکثر	۰/۸۸	۰/۸۵	۰/۸۶	۰/۸۵	۲/۷۹	۳/۴۲	۰/۰۲	

۱۲۰/۶۹	۷۲/۹۲	۲۷/۸۹	۷/۸۸	۶/۳۲	۴/۶۸	۶/۳۲	۳/۹۰	ضریب تغییرات(%)
۱۰ سانتی متر								
-۲/۰۳	-۲/۲۸	۳/۴۲	۲/۳۹	-	-	-	-	میانگین
۰/۴۸	۰/۲۷	۰/۱۷	۰/۲۵	-	-	-	-	انحراف معیار
-۳/۰۰	-۲/۷۹	۳/۱۸	۲/۱۵	-	-	-	-	حداقل
-۱/۰۰	-۲/۰۰	۳/۷۵	۲/۸۷	-	-	-	-	حداکثر
۲۳/۷۶	۱۱/۸۷	۵/۰۷	۱۰/۵۱	-	-	-	-	ضریب تغییرات(%)
۱۵ سانتی متر								
-۳/۰۷	-۳/۱۷	۳/۴۲	۳/۳۲	-	-	-	-	میانگین
۰/۱۴	۰/۰۹	۰/۱۷	۰/۰۹	-	-	-	-	انحراف معیار
-۳/۳۳	-۳/۳۵	۳/۱۸	۳/۱۸	-	-	-	-	حداقل
-۲/۸۹	-۳/۰۱	۳/۷۵	۳/۴۹	-	-	-	-	حداکثر
۴/۶۴	۲/۹۳	۵/۰۷	۲/۸۴	-	-	-	-	ضریب تغییرات(%)
۲۰ سانتی متر								
-۲/۰۳	-۱/۷۹	۲/۳۸	۲/۲۳	-	-	-	-	میانگین
۰/۴۶	۰/۵۷	۰/۶۲	۰/۷۹	-	-	-	-	انحراف معیار
-۳/۰۰	-۲/۷۳	۱/۸۶	۱/۵۴	-	-	-	-	حداقل
-۱/۵۸	-۱/۲۳	۳/۷۱	۳/۵۳	-	-	-	-	حداکثر
۲۲/۵۴	۳۱/۷۴	۲۵/۹۸	۳۵/۲۸	-	-	-	-	ضریب تغییرات(%)
۲۵ سانتی متر								
-۰/۳۶	۰/۱۷	۱/۴۰	۱/۵۴	-	-	-	-	میانگین
۰/۲۶	۰/۱۹	۰/۲۱	۰/۲۷	-	-	-	-	انحراف معیار
-۰/۶۸	-۰/۰۹	۱/۰۵	۱/۲۵	-	-	-	-	حداقل
۰/۱۴	۰/۴۴	۱/۷۱	۱/۹۷	-	-	-	-	حداکثر
۷۳/۲۴	۱۰۹/۸۶	۱۵/۲۵	۱۷/۷۲	-	-	-	-	ضریب تغییرات(%)
۳۰ سانتی متر								
۲/۷۹	۲/۴۶	۳/۰۵	۲/۷۴	-	-	-	-	میانگین
۰/۵۷	۰/۵۴	۰/۶۷	۰/۶۷	-	-	-	-	انحراف معیار
۲/۴۳	۲/۰۱	۲/۶۳	۲/۲۰	-	-	-	-	حداقل
۳/۹۵	۳/۴۵	۴/۴۸	۴/۰۴	-	-	-	-	حداکثر
۲۰/۴۵	۲۱/۹۲	۲۲/۰۶	۲۴/۶۳	-	-	-	-	ضریب تغییرات(%)
۳۵ سانتی متر								
۲/۹۹	۱/۹۷	۳/۰۱	۲/۷۶	-	-	-	-	میانگین
۰/۱۹	۰/۲۵	۰/۱۸	۰/۳۹	-	-	-	-	انحراف معیار
۲/۶۸	۱/۶۲	۲/۶۹	۱/۹۷	-	-	-	-	حداقل
۳/۳۱	۲/۴۵	۳/۳۴	۳/۳۲	-	-	-	-	حداکثر
۶/۳۵	۱۲/۵۳	۶/۰۶	۱۴/۰۶	-	-	-	-	ضریب تغییرات(%)
۰/۰۰۴	-۰/۰۰۱	۰/۰۵	۰/۰۵	۰/۵۹	۰/۸۲	۰/۴۳	۰/۸۱	ANN

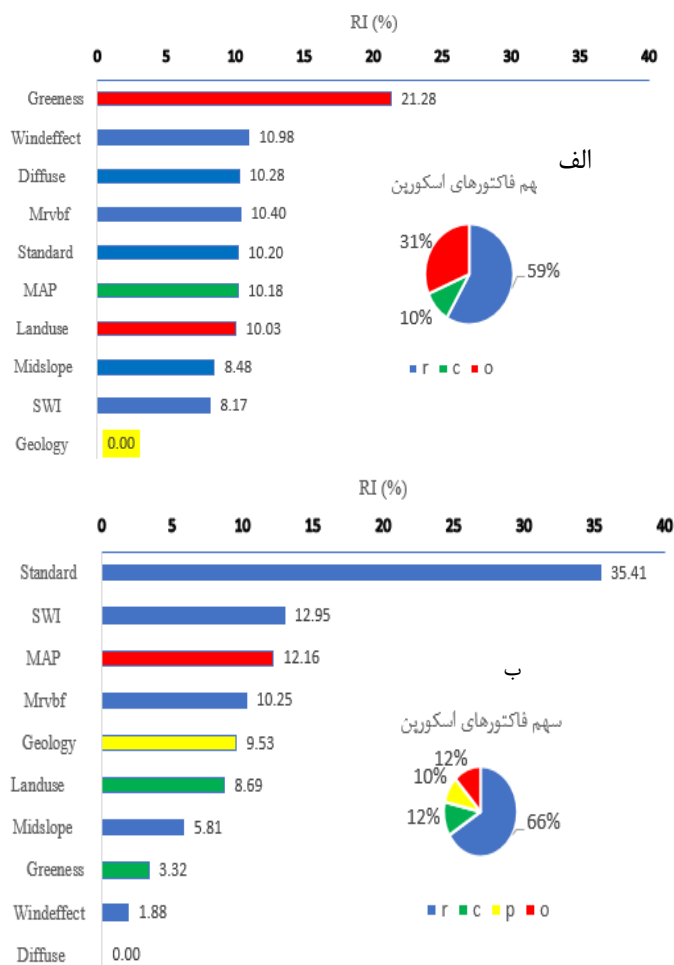
متغیرهای محیطی منتخب و اهمیت نسبی

نتایج انتخاب متغیرهای محیطی بر اساس الگوریتم Boruta نشان می‌دهد که از ۱۷ متغیر محیطی (شکل ۷)، ۱۰ متغیر شامل شاخص سبزی‌نگی (Greenness) و کاربری اراضی، زمین‌شناسی، میانگین بارش سالیانه (MAP)، تاثیر باد "Wind effect"، شاخص همواری کف دره چندمقیاسه "Mrvbf"، تابش پخشیده "Diffuse"، ارتفاعات استاندارد شده "Standard"، شاخص خیسی ساگا "SWI" و موقعیت‌های میانی شیب "Midslop" به ترتیب به عنوان نمایندگان عامل‌های موجودات زنده (o)، مواد مادری (p)،

اقلیم (c) و توپوگرافی (r) با ۱۰، ۳۱ و ۵۹ درصد دارای بیشترین اهمیت نسبی در پیش‌بینی ضخامت خاک سطحی می‌باشند. این نتایج نشان‌دهنده حداکثر استفاده از نمایندگان عامل‌های خاک-ساز بر اساس معادله اسکورپن $S_{(a,c)} = f(s, c, o, p, a, n) + e$ برای مدل‌سازی ضخامت خاک سطحی در دشت قزوین نسبت به مطالعات قبلی (Sepahvad et al., 2017; Zahedi et al., 2017^{a,b}) در سایر مناطق ایران می‌باشد.

همچنین بر اساس نتایج آنالیز اهمیت نسبی حاصل از مدل RF، چهار پارامتر Greenness، Wind effect، Diffuse و Mrvbf

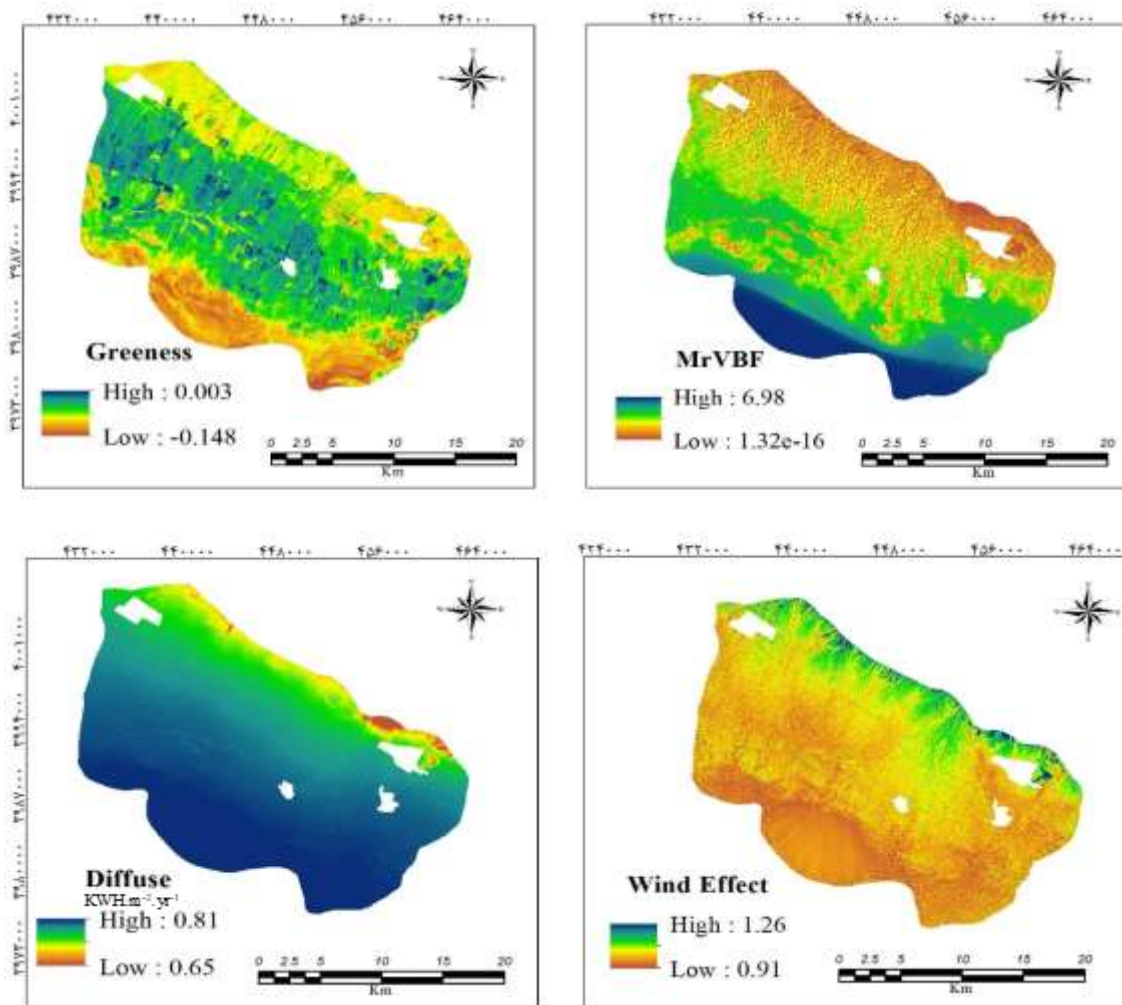
بیشترین تأثیر را بر روی تغییرات مکانی ضخامت خاک سطحی داشته و نقشه تغییرات مکانی آن‌ها در شکل ۴ ب نشان داده شده است.



شکل ۸- نمودار اهمیت نسبی متغیرهای محیطی منتخب توسط الف-مدل RF ب-مدل ANN

جنگل تصادفی، پارامترهای دارای درجه اهمیت بالا در بخش پایین اهمیت قرار گرفته و تنها متغیر Mrvbf در دو مدل در موقعیت یکسانی از نظر درجه اهمیت قرار دارد که به نظر می‌رسد این متغیر محیطی با توجه مطالعه قبلی (Rahmani et al., 2019) در این منطقه کارایی بالایی در پیش‌بینی کلاس‌های خاک دارد. بیشترین میزان اهمیت نسبی با بیش از ۳۵ درصد در مدل شبکه عصبی به متغیر محیطی ارتفاعات استاندارد "Standard" تعلق دارد. پارامتر Standard در مطالعه (Mousavi et al., 2019) به‌عنوان پارامتر تأثیرگذار بر تغییرات مکانی کلاس‌های خاک در دشت قزوین گزارش گردیده است. همچنین (Mousavi et al., 2021) پارامترهای توپوگرافی را از مهمترین پیش‌ران‌های محیطی برای مدلسازی ویژگی‌های خاک در منطقه مورد مطالعه گزارش نمودند.

در مطالعه‌ی (Rahmani et al., 2019) دو پارامتر Diffuse و Mrvbf را به‌عنوان پارامترهای بهینه پیش‌بینی کلاس‌های خاک در دشت قزوین گزارش نمودند. با توجه به غالبیت پراکنش داده‌های ضخامت خاک سطحی ۲۵ تا ۳۰ سانتی‌متر در ناحیه مرکزی دشت که دارای کاربری کشاورزی آبی متمرکز می‌باشد، شاخص Greenness به دلیل بارز نمودن نقش پوشش گیاهی و دارای رشد سبزی‌نگی مناسب نسبت به سه شاخص دیگر اختلاف قابل توجهی در پیش‌بینی مکانی ضخامت خاک سطحی دارد و بیش از ۲۰ درصد از اهمیت نسبی متغیرها را در توزیع و پیش‌بینی مکانی به خود اختصاص داده است و مشابه با این پژوهش (Zahedi et al., 2017^b) بر اهمیت بالای کاربری اراضی به همراه شاخص پوشش گیاهی NDVI در مدل‌سازی عمق خاک تأکید نمودند. سایر متغیرها به‌استثنای زمین‌شناسی تقریباً از یک اهمیت یکسان برخوردارند، در مدل شبکه عصبی برعکس مدل

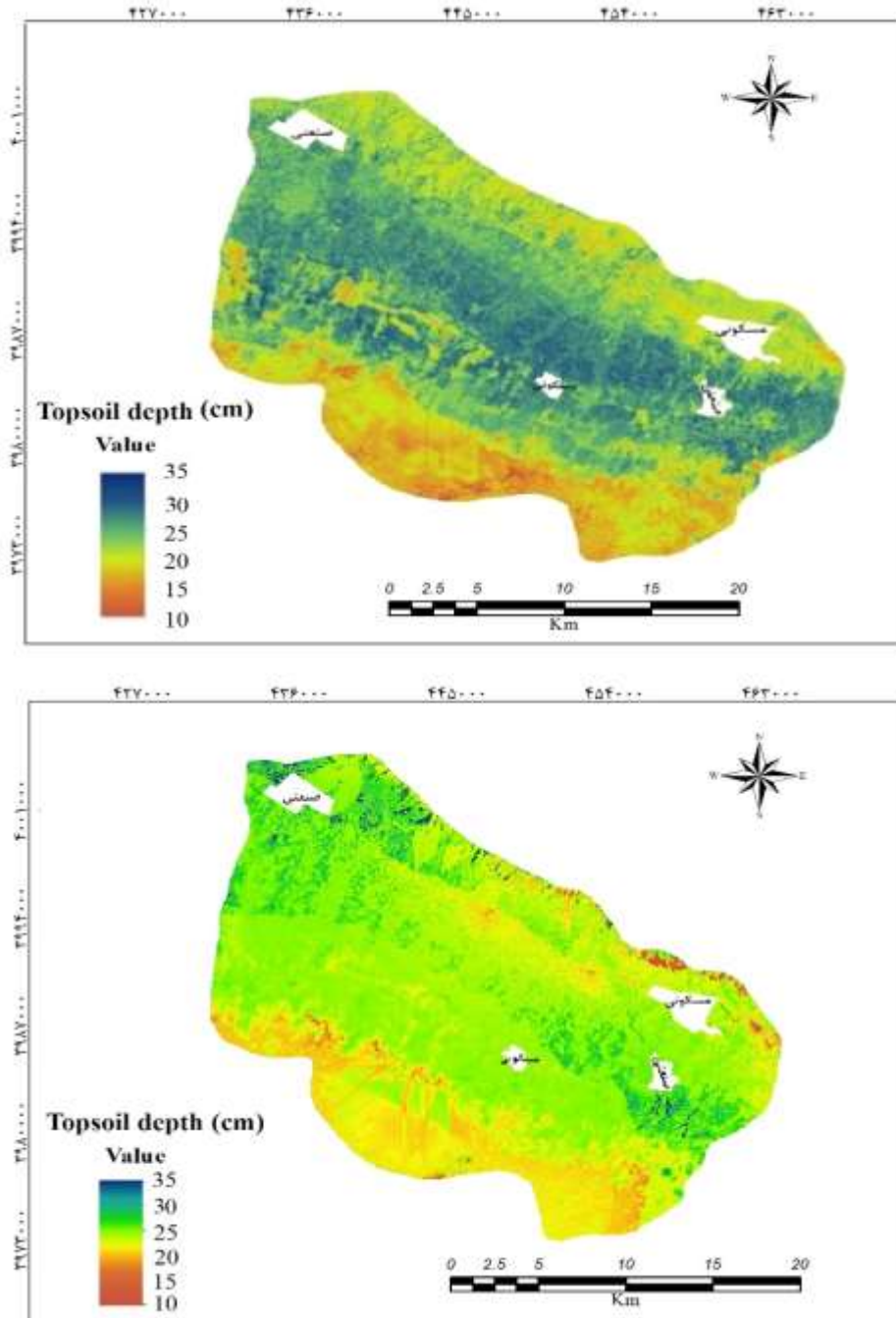


شکل ۹- نقشه چهار متغیر محیطی دارای بیشترین اهمیت براساس مدل RF

مکانی عمق خاک تأکید دارند. بخش جنوبی دشت دارای حداقل ضخامت خاک سطحی (۱۰ سانتی متر) است که مطابق با بالاترین مقدار در نقشه Mrvbf می باشد، این ناحیه به دلیل دارا بودن مشکلات زهکشی و شوری ناشی از خاک های سنگین بافت، فرصت تکامل و خاک سازی نداشته و پارامتر Mrvbf با توجه به قابلیت ارائه پهنه های رسوب گذاری متفاوت (Adhikari et al., 2018; Gallant and Dowling, 2003; Rasaie et al., 2020; Mousavi et al., 2021) به خوبی توانسته این پهنه ها را جداسازی نماید.

تغییرپذیری مکانی ضخامت خاک سطحی

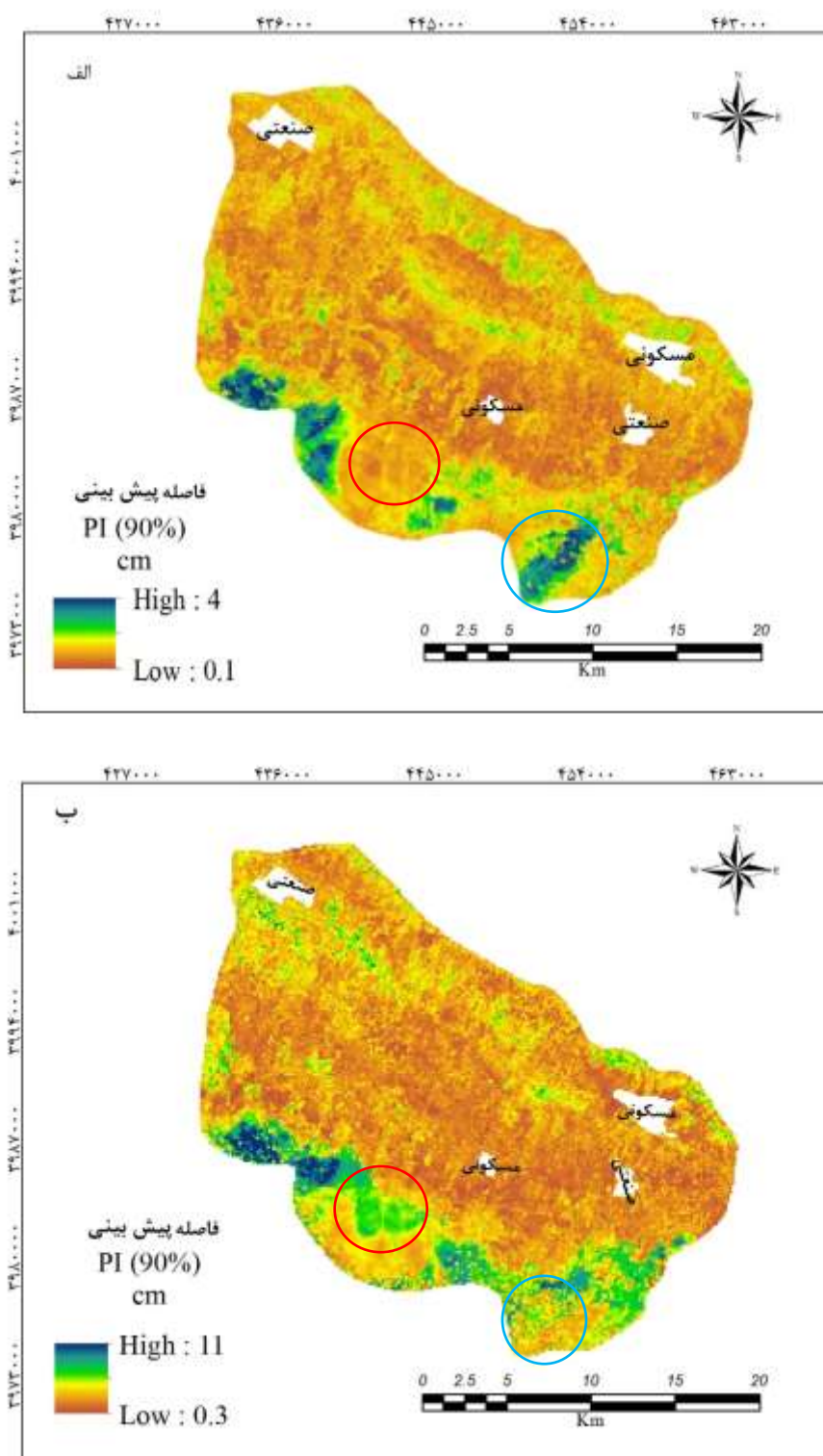
مناطق مرکزی دشت به دلیل تمرکز و سابقه بالای کشاورزی آبی و عملیات های مدیریتی خاک ورزی و آبیاری دارای بیشترین میزان ضخامت خاک سطحی پیش بینی شده (۲۵ تا ۳۰ سانتی متر) با استفاده از مدل جنگل تصادفی بوده و این نواحی همسو با میزان بالای شاخص Greenness به عنوان نماینده ای از عامل خاک سازی موجودات زنده مخصوصا انسان در این بخش از دشت می باشد، در همین راستا (Wang et al., 2018) بر اهمیت کاربری های اراضی مختلف و تأثیر عملیات کشت و کار، بر تغییرپذیری



شکل ۱۰- نقشه رقمی ضخامت خاک سطحی دشت قزوین مدل سازی شده الف (RF) - ANN

اثر باد به دلیل تأثیر بر بالا رفتن میزان تبخیر و تعرق و کاهش ظرفیت رطوبتی خاک سبب کاهش روند خاک سازی می‌گردد و از طرفی این بخش مطابق با کشاورزی دیم و میزان پایین شاخص سبزیگی است. نقشه رقمی حاصل از مدل شبکه عصبی مصنوعی (شکل ۶ ب) به جزء در نواحی جنوبی و بخشی از نواحی جنوب شرقی دارای برآورد مناسب و در نواحی مرکزی دشت، مدل کم برآورد و در نواحی شمال غربی در پیش‌بینی ضخامت خاک سطحی بیش برآورد نموده است.

همچنین در همین ناحیه، شدت تابش پخشیده بالاترین میزان را نشان می‌دهد که می‌توان بیان داشت که افزایش میزان تابش بر بالا رفتن تبخیر و تعرق سالیانه تأثیر داشته و در نهایت روند خاک سازی به آهستگی پیش می‌رود و این نواحی مطابق با رژیم رطوبتی اریدیک ضعیف و اکوئیک می‌باشد. بخش شمالی دشت دارای میزان متوسط ۱۵ تا ۲۰ سانتی‌متر ضخامت خاک سطحی است که هم‌راستا با پهنه‌های دارای حداقل Mrvbf و حداکثر شاخص Wind effect است. پایین بودن میزان Mrvbf که نشان از درشت بودن میزان رسوبات و بالا بودن میزان شاخص



شکل ۱۱- فاصله پیش بینی ۹۰٪ ضخامت خاک سطحی مدل سازی شده توسط مدل RF با استفاده از الف) bootstrapping (ب) - k-fold

عدم قطعیت را در ناحیه مرکزی منطقه مورد مطالعه برآورد نموده اند که در این ناحیه مقدار آن ۰/۳ سانتی متر می باشد که مطابق با خاک های با ضخامت ۲۵ تا ۳۰ سانتی متر است و از دلایل پایین بودن میزان عدم قطعیت در این ناحیه می توان به تعداد بالای نقاط مشاهداتی و همبستگی بالای متغیرهای کمکی و ضخامت

مقایسه دو روش کمی سازی عدم قطعیت

میزان عدم قطعیت کمی سازی شده توسط روش k-fold در مقایسه با روش bootstrapping براساس فاصله پیش بینی ۹۰ درصد، بیشتر است. بطور کلی بیش از ۷ سانتی متر در دامنه بالای این شاخص اختلاف وجود دارد، هر دو روش حداقل میزان

دشت قزوین با سطح دقت عالی و متوسط تهیه شد. کمی‌سازی عدم قطعیت نقشه‌های پیش‌بینی‌شده توسط روش‌های bootstrapping و k-fold بیانگر دقت مناسب و قابل اطمینان مدل جنگل تصادفی به عنوان مدل برتر برای برآورد ضخامت خاک سطحی بخش زیادی از منطقه مورد مطالعه می‌باشد و در بخش‌های از منطقه نیاز به داده بیشتر و استفاده از متغیرهای محیطی مرتبط برای نقشه‌برداری دقیق‌تر می‌باشد. نقشه پیش-بینی مکانی ضخامت خاک سطحی تهیه‌شده توسط مدل RF بیانگر وجود مناطق با ضخامت خاک سطحی بیشتر در نواحی مرکزی مطابق با کاربری کشاورزی آبی است، نواحی جنوبی با زهکشی نامناسب و شوری بالا دارای حداقل ضخامت (۱۰ تا ۱۵ سانتی‌متر) و بخش شمالی منطقه با کاربری اراضی دیم دارای ضخامت متوسط (۱۵ تا ۲۰) است. چهار متغیر کمی شاخص سبزی‌نگی، اثر باد، شدت تابش پخشیده و شاخص همواری دره با قدرت تفکیک بالا به ترتیب اهمیت دارای بیشترین تأثیر بر پراکنش مکانی ضخامت خاک سطحی می‌باشند، بنابراین توصیه می‌گردد در مطالعات آتی از رویکرد استفاده‌شده در این پژوهش و پارامترهای بهینه‌شده در مدل RF به همراه متغیرهای محیطی منتخب برای مدل‌سازی ضخامت خاک سطحی در مناطق با شرایط محیطی مشابه استفاده گردد و با توجه به تحقیقاتی بودن مطالعه حاضر، مقیاس بهینه برای استفاده کاربران در حدود ۷۰ متر می‌باشد. همچنین به منظور اعتبارسنجی نقشه‌های عدم قطعیت بهتر است از روش‌های کمی در مطالعات آتی استفاده گردد.

سپاس‌گذاری

از کارکنان فرمانداری شهرستان نظرآباد و کارشناسان محترم سازمان جهاد کشاورزی، دهیار روستای صالحیه و کشاورزان عزیز جهت همکاری در انجام مطالعات میدانی کمال تشکر و قدردانی را داریم.

"هیچ‌گونه تعارض منافع بین نویسندگان وجود ندارد"

REFERENCES

- Adhikari, K., Owens, P. R., Ashworth, A. J., Sauer, T. J., Libohova, Z., Richter, J. L., & Miller, D. M. (2018). Topographic controls on soil nutrient variations in a silvopasture system. *Agrosystems, Geosciences & Environment*, 1(1), 1-15.
- Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B., Hong, S. Y., ... & Zhang, G. L. (2014). GlobalSoilMap: Toward a fine-resolution global grid of soil properties. *Advances in agronomy*, 125, 93-134.
- Baltensweiler, A., Walthert, L., Hanewinkel, M., Zimmermann, S., & Nussbaum, M. (2021). Machine learning based soil maps for a wide range of soil properties for the forested area of Switzerland. *Geoderma Regional*, 27, e00437.
- Chen, S., Richer-de-Forges, A. C., Mulder, V. L., Martelet, G., Loiseau, T., Lehmann, S., & Arrouays, D. (2021). Digital mapping of the soil thickness of loess deposits over a calcareous bedrock in central France. *Catena*, 198, 105062.
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., ... & Böhner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1. 4. *Geoscientific Model Development*, 8(7), 1991-

- 2007.
- Dharumarajan, S., Hegde, R., & Lalitha, M. (2021). Modelling of soil depth and hydraulic properties at regional level using environmental covariates-A case study in India. *Geoderma Regional*, 27, e00439.
- Gessler, P. E., Moore, I. D., McKenzie, N. J., & Ryan, P. J. (1995). Soil-landscape modelling and spatial prediction of soil attributes. *International journal of geographical information systems*, 9(4), 421-432.
- Han, X., Liu, J., Mitra, S., Li, X., Srivastava, P., Guzman, S. M., & Chen, X. (2018). Selection of optimal scales for soil depth prediction on headwater hillslopes: A modeling approach. *Catena*, 163, 257-275.
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., ... & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS one*, 12(2), e0169748.
- Horst-Heinen, T. Z., Dalmolin, R. S. D., ten Caten, A., Moura-Bueno, J. M., Grunwald, S., de Araújo Pedron, F., ... & da Silva-Sangoi, D. V. (2021). Soil depth prediction by digital soil mapping and its impact in pine forestry productivity in South Brazil. *Forest Ecology and Management*, 488, 118983.
- Kursa, M.B.; Rudnicki, W.R. Feature selection with the Boruta package. *J. Stat. Softw.* 2010, 36, 1–13.
- Lalitha, M., Dharumarajan, S., Suputhra, A., Kalaiselvi, B., Hegde, R., Reddy, R. S., ... & Dwivedi, B. S. (2021). Spatial prediction of soil depth using environmental covariates by quantile regression forest model. *Environmental Monitoring and Assessment*, 193(10), 1-10.
- Leenaars, J. G., Claessens, L., Heuvelink, G. B., Hengl, T., González, M. R., van Bussel, L. G., ... & Cassman, K. G. (2018). Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa. *Geoderma*, 324, 18-36.
- Li, A., Tan, X., Wu, W., Liu, H., & Zhu, J. (2017). Predicting active-layer soil thickness using topographic variables at a small watershed scale. *Plos one*, 12(9), e0183742.
- Li, X., Luo, J., Jin, X., He, Q., & Niu, Y. (2020). Improving Soil Thickness Estimations Based on Multiple Environmental Variables with Stacking Ensemble Methods. *Remote Sensing*, 12(21), 3609.
- Liddicoat, C., Maschmedt, D., Clifford, D., Searle, R., Herrmann, T., Macdonald, L. M., & Baldock, J. (2015). Predictive mapping of soil organic carbon stocks in South Australia's agricultural zone. *Soil Research*, 53(8), 956-973.
- McBratney, A. B., Minasny, B., & Stockmann, U. (Eds.). (2018). *Pedometrics*. Springer.
- Mehnatkesh, A., Ayoubi, S., Jalalian, A., & Sahrawat, K. L. (2013). Relationships between soil depth and terrain attributes in a semi arid hilly region in western Iran. *Journal of Mountain Science*, 10(1), 163-172.
- Moore, I. D., Gessler, P. E., Nielsen, G. A. E., & Peterson, G. A. (1993). Soil attribute prediction using terrain analysis. *Soil science society of America journal*, 57(2), 443-452.
- Mousavi, S. R., Sarmadian, F., & Rahmani, A. (2020). Modelling and Prediction of Soil Classes Using Boosting Regression Tree and Random Forests Machine Learning Algorithms in Some Part of Qazvin Plain. *Iranian Journal of Soil and Water Research*, 50(10), 2525-2538. (In Farsi).
- MOUSAVI, S. R., Sarmadian, F., Omid, M., & Bogaert, P. (2021). Digital Modeling of Three-Dimensional Soil Salinity Variation Using Machine Learning Algorithms in Arid and Semi-Arid lands of Qazvin Plain. *Iranian Journal of Soil and Water Research*, 52(7), 1915-1929. (In Farsi).
- Mousavi, S., Sarmadian, F., Alijani, Z., & Taati, A. (2017). Land suitability evaluation for irrigating wheat by geopedological approach and geographic information system: A case study of Qazvin plain, Iran. *Eurasian Journal of Soil Science*, 6(3), 275-284.
- Rahmani, A., Sarmadian, F., Mousavi, S.R. 2020. Application of geomorphometric features in digital soil mapping using fuzzy logic and machine learning. *Journal of Rangeland and Watershed Management*. Vol 73, (1):105-124. (In Farsi).
- Rahmani, A., Sarmadian, F., Mousavi, S. R., & Khamoshi, S. E. (2019). Digital soil mapping using geomorphometric analysis and case-based fuzzy logic approach. In: *Proceedings of 5th the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 863-866.
- Rasaei, Z., Rossiter, D. G., & Farshad, A. (2020). Rescue and renewal of legacy soil resource inventories in Iran as an input to digital soil mapping. *Geoderma Regional*, 21, e00262.
- Rossel, R. V., Chen, C., Grundy, M. J., Searle, R., Clifford, D., & Campbell, P. H. (2015). The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. *Soil Research*, 53(8), 845-864.
- Sepahvand, M., Khormali, F., Kiani, F., & Eftekhari, K. (2017). Modeling the Relationship between Soil Depth and Topography Characteristics in order to Predict Soil Depth in Rimleh Basin of Lorestan Province. *Journal of Soil Research* 31 (4), 601-611. (In Farsi).
- Shrestha, D. L., & Solomatine, D. P. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks*, 19(2), 225-235.
- Staff, S.S., 2014. *Keys to Soil Taxonomy*, 12th Edn Washington. DC: Natural Resources Conservation Service, United States Department of Agriculture.
- Szatmári, G., & Pásztor, L. (2019). Comparison of various uncertainty modelling approaches based on geostatistics and machine learning



- algorithms. *Geoderma*, 337, 1329-1340.
- Tesfa, T. K., Tarboton, D. G., Chandler, D. G., & McNamara, J. P. (2009). Modeling soil depth from topographic and land cover attributes. *Water Resources Research*, 45(10).
- Thompson, A. L., Gantzer, C. J., & Anderson, S. H. (1991). Topsoil depth, fertility, water management, and weather influences on yield. *Soil Science Society of America Journal*, 55(4), 1085-1091.
- Tibshirani, R. J., & Efron, B. (1993). An introduction to the bootstrap. *Monographs on statistics and applied probability*, 57, 1-436.
- Van Wambeke, A. R. (2000). The Newhall Simulation Model for estimating soil moisture and temperature regimes. Department of Crop and Soil Sciences. Cornell University, Ithaca, NY, USA.
- Vanwinckelen, G., & Blockeel, H. (2012). On estimating model accuracy with repeated cross-validation. In *BeneLearn 2012: Proceedings of the 21st Belgian-Dutch conference on machine learning* (pp. 39-44).
- Vickers, N. J. (2017). Animal communication: when i'm calling you, will you answer too?. *Current biology*, 27(14), R713-R715.
- Minasny, B., McBratney, A. B., & Lark, R. M. (2008). Digital soil mapping technologies for countries with sparse data infrastructures. *Digital soil mapping with limited data*, 15-30.
- Wang, F., Sahana, M., Pahlevanzadeh, B., Pal, S. C., Shit, P. K., Piran, M. J., ... & Mosavi, A. (2021). Applying different resampling strategies in machine learning models to predict head-cut gully erosion susceptibility. *Alexandria Engineering Journal*, 60(6), 5813-5829.
- Wang, Q., Wu, B., Stein, A., Zhu, L., & Zeng, Y. (2018). Soil depth spatial prediction by fuzzy soil-landscape model. *Journal of soils and sediments*, 18(3), 1041-1051.
- Wilson, J. (2018). Environmental applications of digital terrain modeling. *John Wiley & Sons*. 359pp.
- Zahedi, p. Shahedi, K. Nejad Roshan, H. & Soleimani, K. (2017a). Estimation of soil depth using environmental variables obtained from digital elevation model and remote sensing data. *Journal of Soil and Water Sciences*, 21 (4), 111-127. (In Farsi).
- Zahedi, S., Shahedi, K., Rawshan, M. H., Solimani, K., & Dadkhah, K. (2017b). Soil depth modelling using terrain analysis and satellite imagery: the case study of Qeshlaq mountainous watershed (Kurdistan, Iran). *Journal of Agricultural Engineering*, 48(3), 167-174.
- Zeraatpisheh, M., Jafari, A., Bodaghabadi, M. B., Ayoubi, S., Taghizadeh-Mehrjardi, R., Toomanian, N., and Xu, M. (2020). Conventional and digital soil mapping in Iran: Past, present, and future. *Catena*, 188, 104424.
- Zhao, R., & Wu, K. (2021). Soil Health Evaluation of Farmland Based on Functional Soil Management—A Case Study of Yixing City, Jiangsu Province, China. *Agriculture*, 11(7), 583.
- Zhu, A. X., Hudson, B., Burt, J., Lubich, K., & Simonson, D. (2001). Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal*, 65(5), 1463-1472.
- Zinck, J. A., Metternicht, G., Bocco, G., & Del Valle, H. F. (Eds.). (2015). *Geopedology: An integration of geomorphology and pedology for soil and landscape studies*. Springer.