



## Digital modeling and prediction of soil subgroup classes using deep learning approach in a part of arid and semi-arid lands of Qazvin Plain

ASGHAR RAHMANI <sup>1</sup> | FERAYDOON SARMADIAN <sup>2</sup> | HOSSIEN AREFI <sup>3</sup>

1. Department of Soil Science, College of Agriculture and Natural Resource, University of Tehran, Karaj, Iran. E-mail:

[a.rahmani@ut.ac.ir](mailto:a.rahmani@ut.ac.ir)

2. Department of Soil Science, College of Agriculture and Natural Resource, University of Tehran, Karaj, Iran. E-mail:

[fsarmad@ut.ac.ir](mailto:fsarmad@ut.ac.ir)

3. Department of Remote Sensing and Photogrammetry, Geospatial and Surveying Faculty, College of Engineering, University of Tehran, Tehran, Iran. E-mail: [hossein.arefi@ut.ac.ir](mailto:hossein.arefi@ut.ac.ir)

### Article Info

**Article type:** Research Article

#### Article history:

**Received:** Jan. 3, 2023

**Revised:** Feb. 15, 2023

**Accepted:** March. 11, 2023

**Published online:** Jan. 22, 2022

#### Keywords:

Convolutional Neural Network,  
Machine learning,  
Scale effect,  
Soil class

### ABSTRACT

#### Introduction

Soil class maps contain useful information that helps stakeholders to understand soil behavior in response to different management programs. As well as, their numerical prediction is dependent on the appropriate scale of environmental variables. Therefore, the current research intends to use the deep learning approach (CNN) and the spatial information of geomorphometric attributes and the sentinel 1/2 satellite images along with band ratios to predict the soil subgroup classes with its uncertainty map. Also, comparing the results of CNN and the random forest (RF) model in prediction of soil classes and different environmental variables was not well documented.

#### Material and Methods

CNN model was run in the Google Collaboratory online environment and the RF model was performed by the "rf" function in the "caret" package in the RStudio environment. The models were calibrated with 80% of the data set along with six different window sizes and validated according to 20% of rest data based on two indices of overall accuracy (OA) and F1-Score.

#### Results and Discussion

Six covariates i.e., DEM, SWI, WE, SH, MRVBF, DIFF were selected as the most effective variables among 33 geomorphometric attributes, with 12 individual bands and the indices of sentinel 1/2. Totally, 13 soil subgroups including nine from Aridisols, three Inceptisols subgroups and, one Entisols subgroup are recognized in the study area. The overall accuracy for two models with a slightly difference of 7% in the window size (15\*15) was observed with 43% and 50% for CNN and RF models, respectively. The CNN model has three patterns (increasing-decreasing), small and large optimal window size, and the same pattern observed in the scaled RF model, too. The OA was zero in all window sizes for the *Sodic Xeric Calcigypsids* subgroup in the CNN model and the *Xeric Calcigypsids*, and *Typic Xerorthents* subgroups in the RF model. In addition, the *Xeric Haplocalcids* and *Xeric Haplogypsids* only predicted by the RF model in 3\*3 and 5\*5 window size, respectively. By increasing the window size from three to nine, and 15, the *Typic Calcixerepts* shows a mild increasing trend in the F1-Score and also a mild decreasing trend after reaching the peak. The amount of F1-score for *Typic Calcixerepts* in CNN and RF models was 69% and 77%, respectively. The F1-Score values of *Gypsic Aquisalids* and *Xeric Haplogypsids* increase by 30% and 17%, by increasing the window size from three to five, and immediately a sharp downward trend, which indicates the appropriateness of the small window size in order to predict.

#### Conclusion

In general, despite the limited number of observation profiles (n=278), the CNN model provides an acceptable prediction in mapping the soil subgroup classes, and although a slight difference in the overall accuracy with the RF model, while, the CNN presents a lower uncertainty map in comparison to RF. In future studies, this model and its procedure can be used to predict soil class maps in other arid and semi-arid regions.

Cite this article: Rahmani, A., Sarmadian, F., Arefi, H. (2023). Digital modeling and prediction of soil subgroup classes using deep learning approach in a part of arid and semi-arid lands of Qazvin Plain, *Iranian Journal of Soil and Water Research*, 53 (11), 2477-2499. <https://doi.org/10.22059/ijswr.2023.353339.669426>



## مدل سازی و پیش بینی رقومی کلاس های زیر گروه خاک با استفاده از رویکرد یادگیری عمیق در بخشی از اراضی خشک و نیمه خشک دشت قزوین

اصغر رحمانی<sup>۱</sup> | فریدون سرمدیان<sup>۲</sup> | حسین عارفی<sup>۳</sup><sup>۱</sup>. گروه علوم و مهندسی خاک، دانشکده گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: [a.rahmani@ut.ac.ir](mailto:a.rahmani@ut.ac.ir)<sup>۲</sup>. گروه علوم و مهندسی خاک، دانشکده گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران. رایانامه: [fsarmad@ut.ac.ir](mailto:fsarmad@ut.ac.ir)<sup>۳</sup>. گروه سنجش از دور و فتوگرامتری، دانشکده نقشه برداری و اطلاعات مکانی، دانشکده گان فنی، دانشگاه تهران، تهران، ایران. رایانامه: [hossein.arefi@ut.ac.ir](mailto:hosseini.arefi@ut.ac.ir)

### اطلاعات مقاله

### چکیده

نوع مقاله: مقاله پژوهشی

تاریخ دریافت: ۱۴۰۱/۱۰/۱۳

تاریخ بازنگری: ۱۴۰۱/۱۱/۲۶

تاریخ پذیرش: ۱۴۰۱/۱۲/۲۰

تاریخ انتشار: ۱۴۰۱/۱۱/۱

### واژه های کلیدی:

اثر مقیاس، شبکه عصبی کانولوشن، کلاس خاک، یادگیری ماشین.

در پژوهش حاضر به مقایسه کارایی رویکردهای یادگیری عمیق و روش های یادگیری ماشین به منظور تهیه نقشه کلاس های خاک پرداخته شد. جهت تحقق این هدف از اطلاعات حاصل از ۲۷۸ خاکرخ مشاهداتی، و متغیرهای ژئومورفومتری حاصل از مدل رقومی ارتفاع، باندها و شاخص های مستخرج از ماهواره های سنتیل ۱ و ۲ در فرآیند مدل سازی استفاده گردید. مدل یادگیری عمیق در محیط آنلاین Google Collaboratory و مدل جنگل تصادفی (نماینده یادگیری ماشین) با استفاده از تابع "rf" در بسته "caret" در محیط RStudio بر مبنای ۸۰ درصد داده ها و اندازه پنجره های ۳، ۵، ۷، ۹، ۱۵، ۲۱ اجرا شد. مدل ها با ۲۰ درصد باقی مانده داده ها بر اساس دو شاخص صحت عمومی و F1-Score اعتبارسنجی گردیدند. عدم قطعیت پیش بینی نیز با استفاده از نقشه های احتمال هر زیر گروه و شاخص آنتروپی محاسبه گردید. صحت عمومی پیش بینی دو مدل یادگیری عمیق و جنگل تصادفی در اندازه پنجره بهینه ۱۵×۱۵ به ترتیب ۴۳ و ۵۰ درصد برای به دست آمد. نتایج نشان داد که زیر گروه Typic Calcixerepts با افزایش اندازه پنجره محاسباتی از ۳ تا ۹ و ۱۵ روند افزایشی در شاخص F1-Score و پس از رسیدن به قله یک روند کاهشی مشاهده گردید. میزان شاخص F1-score این زیر گروه در دو مدل به ترتیب مقادیر ۶۹ و ۷۷ درصد به دست آمد. به طور کلی مدل یادگیری عمیق با وجود تعداد محدود خاکرخ های مشاهداتی توانسته در پیش بینی کلاس های پیش بینی قابل قبولی را ارائه نماید و با وجود اختلاف اندک در شاخص صحت عمومی با مدل جنگل تصادفی، نقشه های نهایی کلاس های زیر گروه خاک با عدم قطعیت کمتری پیش بینی نماید.

استناد: رحمانی؛ اصغر، سرمدیان؛ فریدون، عارفی؛ حسین، (۱۴۰۱). مدل سازی و پیش بینی رقومی کلاس های زیر گروه خاک با استفاده از رویکرد یادگیری عمیق در بخشی از اراضی خشک و نیمه خشک دشت قزوین، مجله تحقیقات آب و خاک ایران، ۵۳ (۱۰)، ۲۴۹۹-۲۴۷۷.

<https://doi.org/10.22059/ijswr.2023.353339.669426>



© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijswr.2023.353339.669426>

## مقدمه

کلاس‌های خاک حاوی اطلاعاتی هستند که به مدیران در شناخت رفتار خاک، در پاسخ به اقدامات مدیریتی کمک می‌کنند؛ بنابراین شناسایی و نقشه‌برداری دقیق آن‌ها در مکان‌های متفاوت زمین‌نما برای اجرای مدیریت پایدار خاک حیاتی است (Maynard et al., 2020). فن‌های نقشه‌برداری رقومی خاک تغییرپذیری خاک را با استفاده از الگوریتم‌های مختلف مانند الگوریتم‌های یادگیری ماشین<sup>۱</sup> بر اساس رابطه بین خصوصیات<sup>۲</sup> کلاس‌های<sup>۳</sup> خاک و اطلاعات کمکی، پیش‌بینی می‌کنند (Esfandiarpour-Boroujeni et al., 2020). مطالعات بی‌شماری برای نقشه‌برداری رقومی کلاس‌های خاک در سطوح مختلف تاکسونومیک در مناطق خشک و نیمه‌خشک ایران و جهان با استفاده از الگوریتم‌های منفرد و تلفیقی یادگیری ماشین انجام شده‌است. در این مطالعات، همه مراحل نقشه‌برداری رقومی شامل روش‌های نوین نمونه‌برداری، استفاده از الگوریتم‌های انتخاب متغیر و متغیرهای محیطی متنوع همراه با دامنه گسترده‌ای از مدل‌های داده‌کاوی استفاده شده و صحت مناسبی را نیز گزارش نموده‌اند. اما تأثیر استفاده از پیکسل‌های مجاور به جای استفاده از الگوریتم‌های انتخاب متغیر و متغیرهای محیطی متنوع همراه با دامنه گسترده‌ای از مدل‌های داده‌کاوی (Jafari et al., 2014; Taghizadeh-Mehrjardi et al., 2015; Brungard et al., 2015; Heung et al., 2016; Zeraatpisheh et al., 2017; Mirakzehi et al., 2018; Taghizadeh-Mehrjardi et al., 2019; Esfandiarpour-Boroujeni et al., 2020). پیکسل محل رخداد خاک‌رخ‌ها در مطالعات اشاره شده چندان مورد بررسی قرار نگرفته است.

در مشاهدات نقطه‌ای متغیرهای محیطی به دلیل اینکه بخش بزرگی از اطلاعات زمینه مکانی از دست می‌رود به‌طور تقریبی می‌توانند ویژگی خاک را توصیف نماید. چندین مطالعه نشان داده‌اند که مشارکت اطلاعات زمینه‌ای متغیرهای محیطی سبب بهبود پیش‌بینی کلاس و یا ویژگی‌های خاک می‌گردد (Behrens et al., 2014; Gallant and Dowling, 2003).

Smith et al. (2006) اندازه همسایگی‌های مختلف را در محاسبه عوارض سرزمین به‌منظور شناسایی خاک به کار گرفتند و بیان داشتند که میزان اطلاعات زمینه‌ای<sup>۴</sup> به کار رفته برای مدل‌سازی، تأثیر معنی‌داری بر خروجی‌های شناسایی خاک دارد (Behrens et al., 2010). در زمینه مشارکت اطلاعات مکانی متغیرهای محیطی، مدل نقشه‌ضمنی<sup>۵</sup> ConMAP را که اختلاف ارتفاع از پیکسل مرکزی برای هر پیکسل در یک همسایگی و مدل آمار ضمنی<sup>۶</sup> ConStat را که اندازه‌های آماری ارتفاع در داخل یک شعاع در حال رشد از پیکسل مرکزی را استخراج می‌کند ابداع شد (Behrens et al., 2014). این روش‌ها تعداد بی‌شماری از فرامتغیرهای کمکی برای نمایش اطلاعات زمینه‌ای را فراهم می‌کنند که به‌عنوان پیش‌بینی کننده در مدل‌های یادگیری ماشین استفاده می‌شوند. این روش‌ها دارای معایب پیش‌پردازش‌های زمان‌بر متغیرهای محیطی، تصمیم‌های ذهنی کارشناس در انتخاب اندازه پیکسل ورودی و پنجره مناسب جهت محاسبه متغیرهای محیطی می‌باشند (Miller et al., 2015; Behrens et al., 2014). در جهت برطرف نمودن این معایب استفاده از روش‌های یادگیری عمیق خصوصاً شبکه‌های عصبی کانولوشن (CNN)<sup>۷</sup> به‌عنوان یک جایگزین به دلیل به‌کارگیری صریح اطلاعات زمینه‌ای محلی متغیرهای محیطی برای نقشه‌برداری متداول گردیده است (Padarian et al., 2019). از مدل‌های یادگیری عمیق در سالین اخیر در علوم محیطی و کشاورزی در زمینه‌های طبقه‌بندی تالاب (Jamali et al., 2020) پیش‌بینی عملکرد محصولات زراعی (Zhang et al., 2022) نقشه‌برداری و پیش‌بینی انواع محصولات زراعی (Kwak et al., 2021) استفاده شده است.

در اکثر مطالعات پیشین تمرکز اصلی پژوهشگران روی مدل‌سازی و پیش‌بینی ویژگی‌های خاک به‌صورت منفرد و یا چندگانه در اعماق منفرد و یا چندین عمق برای برآورد آن‌ها بوده است و در زمینه طبقه‌بندی کلاس‌های خاک و استفاده از مدل‌های یادگیری عمیق خصوصاً مدل شبکه‌های عصبی کانولوشن در پیش‌بینی و برآورد کلاس‌های خاک تنها یک مطالعه اخیراً توسط (Beucher et al., 2022) در کشور دانمارک با حجم بالای داده بر روی طبقه‌بندی خاک‌های اسید سولفات و غیر اسید سولفات با استفاده از ۵۹۰۰ داده و ۱۴ متغیر محیطی اجرا گردیده است. نتایج پژوهش آن‌ها بیانگر برتری ۷ درصدی مدل شبکه‌های عصبی کانولوشن در مقایسه با مدل‌های رایج جنگل تصادفی (RF)<sup>۸</sup> و مقیاس‌گذاری شده جنگل تصادفی برای پیش‌بینی کلاس‌ها می‌باشد. همچنین از الگوریتم شاپ (SHAP) برای واضح‌سازی پیش‌بینی مدل شبکه‌های عصبی کانولوشن با استفاده از تفسیر مکانی و اهمیت متغیرهای محیطی برای پیش‌بینی در نواحی

1. Machine Learning
2. Properties
3. Classes
4. Contextual information
۵. Contextual map
۶. Contextual satatistic
7. Convolutional neural network
8. Random Forest



مختلف نقشه‌های تولیدشده کلاس‌های خاک به‌خوبی استفاده نموده و قادر به تفسیر مدل شبکه‌های عصبی کانولوشن گردیده‌اند. با این اوصاف هنوز به دلیل تعداد اندک مطالعات پیش‌بینی کلاس‌های خاک با استفاده از مدل شبکه‌های عصبی کانولوشن جهت رفع خلا موجود در مطالعات نقشه‌برداری تحقیق حاضر برای اولین بار در کشور ایران با هدف پیش‌بینی رقوم کلاس‌های زیرگروه خاک در بخشی از اراضی خشک و نیمه‌خشک دشت قزوین با تعداد داده محدود و مقایسه آن با مدل جنگل تصادفی مقیاس‌گذاری شده بر روی متغیرهای محیطی منتخب اجرا شد.

### پیشینه پژوهش

در زمینه علوم خاک تحقیقات کاربرد مدل‌های یادگیری عمیق برای پیش‌بینی و برآورد ویژگی‌های خاک از سال ۲۰۱۸ شروع گردیده‌است. برخی از پژوهشگران از مدل‌های یادگیری عمیق برای پیش‌بینی تنها یک ویژگی در یک عمق مشخص از خاک استفاده نموده‌اند. (et al. 2020) Tziolas در پیش‌بینی میزان رس خاک سطحی از داده‌های سری زمانی ماهواره‌های سنتیل ۱ و ۲ مستخرج شده از محیط گوگل ارث انجین و داده‌های بانک جهانی LUCAS و مدل شبکه‌های عصبی کانولوشن را در مقایسه با مدل‌های ANN و RF در محدوده کشور یونان مورد مقایسه قرار دادند. نتایج بیانگر برتری مدل شبکه‌های عصبی کانولوشن نسبت به دو مدل دیگر بر اساس آماره‌های دقت و خطا ( $R^2$ , RMSE, RPIQ) است. از ترکیب داده‌های بلندمدت فنولوژی پوشش گیاهی حاصل از ماهواره MODIS در مدت ۱۰ سال به همراه سایر متغیرهای محیطی برای پیش‌بینی کربن آلی خاک از مدل شبکه‌های عصبی کانولوشن در استان آن‌هوی چین استفاده نمودند و این مدل را با مدل RF در یک سناریو متغیر محیطی مورد مقایسه قرار دادند. نتایج بیانگر بهبود ۵/۵۷ درصد در RMSE و ۳۱/۲۹ درصد در ضریب  $R^2$  با مشارکت داده‌های فنولوژی در مدل شبکه‌های عصبی کانولوشن می‌باشد. مدل شبکه‌های عصبی کانولوشن بدون در نظر گرفتن سناریوهای متغیر محیطی صحت پیش‌بینی بالاتری را نشان داد.

طی پژوهشی که در بخشی از اراضی منطقه مورد مطالعه توسط (موسوی و همکاران، ۱۳۹۸) با هدف مقایسه کارایی مدل‌های یادگیری ماشین جنگل تصادفی (RF) و رگرسیون درختی توسعه یافته (BRT) در پیش‌بینی مکانی کلاس‌های خاک در دو سطح تاکسونومیک زیرگروه و فامیل پرداختند؛ و نتایج ایشان بیانگر کارایی بالاتر مدل RF در هر دو سطح تاکسونومیک نسبت به مدل BRT در پیش‌بینی کلاس‌های خاک بود.

در تحقیقی دیگر سطوح شوری و سدیمی خاک با استفاده از مدل‌های جنگل تصادفی و شبکه‌های عصبی کانولوشن بر اساس تعداد داده محدود میدانی (۸۸ نمونه) و ۱۶ متغیر محیطی در شهر دان چین پیش‌بینی و نقشه‌برداری گردید. نتایج نشان‌دهنده برتری مدل جنگل تصادفی و استفاده از الگوریتم متعادل‌سازی داده نسب به مدل شبکه‌های عصبی کانولوشن می‌باشد (Yan et al., 2022). در پژوهشی به‌منظور شناسایی ویژگی‌های دینامیک خاک در ۱۶۰ خاکرخ با رده‌های خاک مختلف در چندین استان چین و بر اساس نرم‌افزاری هوشمند کاربردی بر پایه تلفن همراه مدل شبکه‌های عصبی کانولوشن قابلیت ویژه‌ای را در تشخیص و مرزبندی افق‌های خاک به‌صورت درجا ارائه نمود (Jiang et al., 2021).

در چندین مطالعه از توانایی پیش‌بینی هم‌زمان مدل شبکه‌های عصبی کانولوشن برای چندین ویژگی در چندین عمق و یا یک ویژگی خاک در چندین عمق استفاده گردیده است. برای مثال (Padarian et al., 2018) از داده‌های طیفی بانک جهانی LUCAS به‌صورت اسپکتروگرام دو بعدی و شبکه‌های عصبی کانولوشن برای پیش‌بینی هم‌زمان شش ویژگی (OC, CEC, Clay, Sand, pH, total N) استفاده نمودند. در مقایسه با مدل‌های حداقل مربعات جزئی<sup>۱</sup> (PLSR) و کوپبیست (Cubist) عملکرد بهتری را گزارش نمودند. در استفاده از همین داده‌های بانک جهانی پروژه مطالعه کاربردی و پوشش اراضی<sup>۲</sup> (LUCAS) در کشور فرانسه (Wadoux et al., 2019) علاوه بر استفاده از یک مدل شبکه‌های عصبی کانولوشن برای پیش‌بینی هم‌زمان شش ویژگی خاک اشاره شده، از مجموعه متنوعی از متغیر محیطی به‌عنوان نمایندگان فاکتورهای خاک‌سازی استفاده نموده و علاوه بر تهیه نقشه‌های پیش‌بینی مکانی آنها، عدم قطعیت این نقشه‌ها را با استفاده از یک روش دو مرحله‌ای کمی‌سازی نمودند. نتایج بیانگر نقشه‌های با صحت بیشتر بر اساس آماره‌های ضریب تبیین و ضریب همبستگی تطابق در مقایسه با روش جنگل تصادفی بوده و نقشه تولیدی روش شبکه‌های عصبی کانولوشن دارای الگوی جزئی با تغییرپذیری مکانی معنی‌دار می‌باشند. (Padarian et al. 2019) در کشور شیلی مدل شبکه‌های عصبی کانولوشن را برای پیش‌بینی هم‌زمان کربن آلی خاک در چندین عمق به کار گرفتند. نتایج نشان داد که در مقایسه با روش جنگل تصادفی، ۳۰ درصد کاهش خطا مشاهده شده و اندازه همسایگی

1 . Partial least squares regression  
2 . Land use and land cover survey

با پیکسل‌ها ۳ تا ۹ مؤثرترین اندازه همسایگی نسبت به مکان‌های نقطه‌ای و اندازه‌های بزرگ‌تر همسایگی می‌باشند. همچنین این مدل دارای عدم قطعیت کمتر و توانایی بالاتر در پیش‌بینی کربن آلی در اعماق پایین‌تر است. (Taghizadeh-Mehrjardi et al., 2020) از توانایی پیش‌بینی هم‌زمان مدل شبکه‌های عصبی کانولوشن برای پیش‌بینی اندازه ذرات رس، سیلت و شن بر اساس اعماق استاندارد جهانی نقشه‌برداری رقومی خاک و متغیرهای محیطی متنوع در ایران مرکزی استفاده نمود و کارایی این مدل را با مدل جنگل تصادفی که در آن متغیرهای محیطی از قبل در اندازه همسایگی‌های مختلف تهیه‌شده بودند، مقایسه نمود. بر اساس شاخص‌های آماری  $R^2$  و RMSE، نتایج بیانگر برتری مدل شبکه‌های عصبی کانولوشن نسبت به مدل جنگل تصادفی در حالت‌های با و بدون متغیرهای محیطی مقیاس‌گذاری شده، می‌باشد.

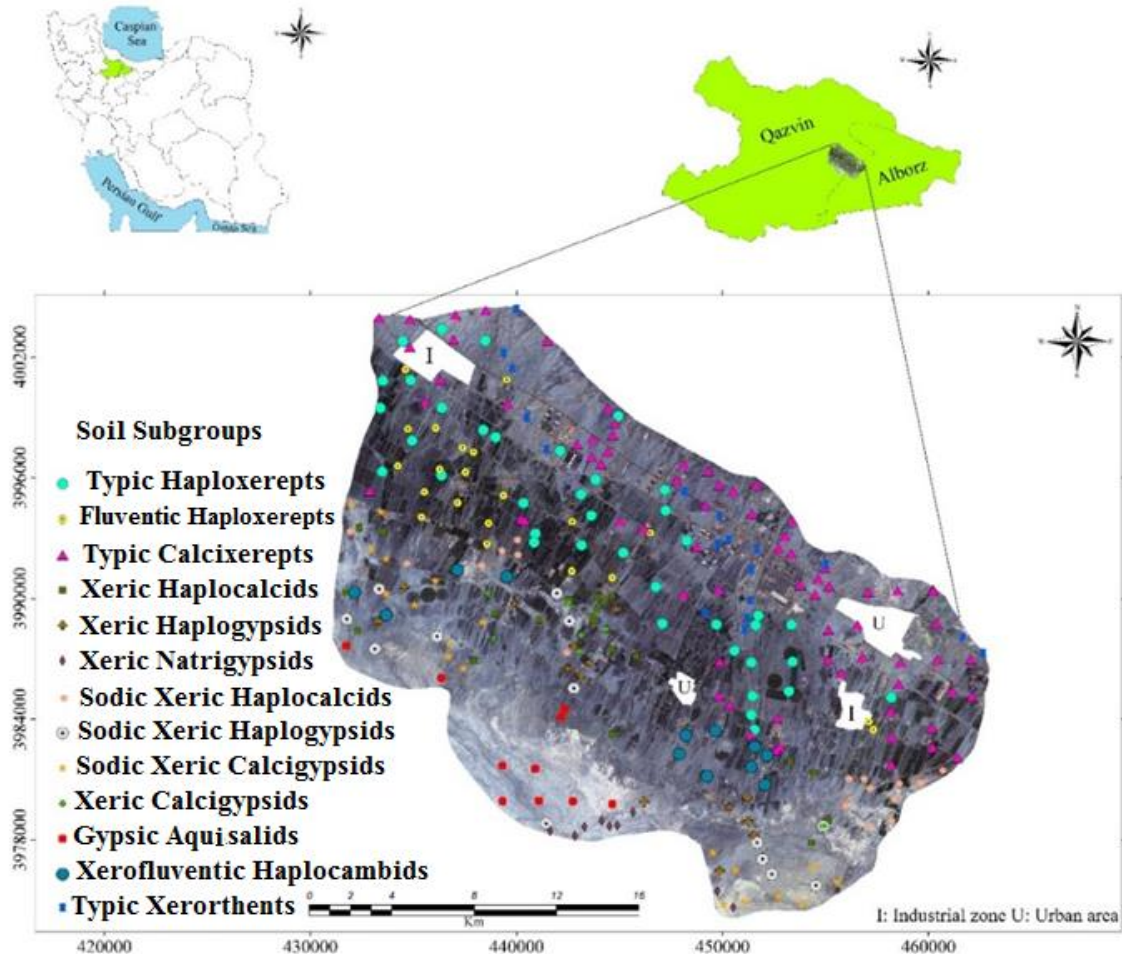
## روش‌شناسی پژوهش

### منطقه مورد مطالعه

پژوهش حاضر در محدوده‌ای با مساحت حدود ۶۰۰۰۰ هکتار از اراضی دشت قزوین در ساختار زمین‌شناسی ایران مرکزی واقع در پهنه ۱۹۸۴ WGS (UTM) N39 با ۴۳۰۰۰۰ تا ۴۷۰۰۰۰ متر طول شرقی و ۳۹۷۸۰۰۰ تا ۴۰۰۲۰۰۰ متر عرض جغرافیایی شمالی انجام شد (شکل ۱). موقعیت مکانی خاک‌رخ‌های خاک به همراه طبقات زیرگروه خاک آن‌ها در شکل ۱ ارائه شده است. حداقل مقدار بارندگی سالانه در نواحی مرکزی به سمت جنوب ۲۵۰ میلی‌متر و از بخش‌های مرکزی به سمت شمال به مقادیر حداکثر تا ۳۱۰ میلی‌متر متغیر است. بر اساس اطلاعات هواشناسی منطقه دو ماه اسفند و شهریور بیشترین و کمترین میزان بارندگی را دارند. میانگین دمای سالانه ۱۳ تا ۱۵ درجه سانتی‌گراد است (ایستگاه سینوپتیک قزوین، ۱۹۷۰-۲۰۱۹). بر اساس محاسبات صورت پذیرفته بر اساس مدل نیوهال<sup>۱</sup> در نرم‌افزار jNSM (1.6.1v)، خاک‌ها در منطقه مورد مطالعه به ترتیب دارای رژیم‌های رطوبتی زیریک خشک-اریدیک ضعیف و اکوتیک و رژیم حرارتی ترمیک است (Van Wambeke., 2000). واحدهای غالب زمین‌نمای منطقه در بخش‌های شمالی به تپه‌ها و فلات محدود می‌شوند و در بخش‌های میانی و جنوبی دشت دامنه‌ای و دشت هستند (Zinck et al., 2015). اختلاف ارتفاع از ۱۱۳۹ متر در جنوب تا ۱۷۴۸ متر در نواحی شمالی و قسمت میانی به‌طور متوسط ۱۳۰۰ متر است. شیب اراضی ۰ تا ۸ درصد است. کاربری اراضی اصلی در قسمت میانی، محصولات آبی گندم، جو، یونجه و ذرت برای کشت زمستانه و تابستانه و در قسمت‌های شمالی با مرتع غیر شور و عمدتاً دیم، در نواحی جنوبی منطقه شوری مشاهده می‌شود. مراتع با پوشش گیاهی غالب شامل علف شور، خارستر و سالیکورنیا می‌باشد. مواد مادری سازنده خاک‌های منطقه در واحدهای ژئومورفیک تپه و فلات سنگ‌های آذرین بازالتی آهکی و در دشت دامنه‌ای و دشت به‌صورت رسوبات کواترنر از سنگ‌های بازالتی آهکی و سنگ‌های گچی نمکی و کفه رسی هستند (سازمان زمین‌شناسی ایران، ۱۹۹۵).

### داده‌های خاک

تعداد ۲۷۸ خاک‌رخ در ابتدا به‌صورت شبکه منظم با میانگین فاصله ۱۲۰۰ متر طراحی و با سیستم موقعیت‌یاب جهانی به مدت ۵ سال (۱۳۹۵-۱۳۹۹) از شهریور تا دی ماه تعیین موقعیت گردید. فاصله و محل خاک‌رخ‌ها در برخی موارد به‌ویژه در زمین‌های نظامی و مالکیت شخصی و در اراضی شور به دلیل رطوبت زیاد خاک و سطح آب زیرزمینی و عدم امکان تردد وسایل نقلیه تغییر کرد. خاک‌رخ‌ها تا عمق ۲ متر یا بالای لایه محدودکننده و بر اساس دستورالعمل توصیف و نمونه‌برداری از خاک‌ها، نسخه ۳ (Schoeneberger et al., 2012) و با توجه به ویژگی‌های مورفولوژیکی و وجود یا عدم وجود افق‌های ژنتیکی تشریح شدند. از خاک هر افق ژنتیکی برای تجزیه و تحلیل ویژگی‌های فیزیکوشیمیایی نمونه‌برداری شد. خصوصیات فیزیکوشیمیایی خاک‌ها شامل (اسیدیته خاک، هدایت الکتریکی، نسبت سدیم جذب سطحی شده، آهک معادل، کربن آلی، وزن مخصوص ظاهری، گچ و بافت خاک) طبق روش استاندارد در آزمایشگاه خاک تعیین شد (Dane and Toppp., 2021; Spark et al., 2020). در نهایت، خاک‌رخ‌ها با توجه به افق‌های مشخصه سطحی و زیرسطحی، رژیم رطوبتی و حرارتی، آنالیزهای آزمایشگاهی و مشخصات مورفولوژیکی شناسایی شده در برگ تشریح مطالعه صحرائی بر اساس کلید طبقه‌بندی خاک آمریکایی ۲۰۱۴ تا سطح زیرگروه طبقه‌بندی شدند (Soil Survey Staff, 2014).



شکل ۱. نقشه موقعیت منطقه مورد مطالعه در کشور و استان های مورد مطالعه به همراه پراکنش مکانی طبقه بندی زیر گروه های خاک بر روی تصویر ماهواره سنتینل ۲

### متغیرهای محیطی کمکی

در این تحقیق، از منابع داده های تصویری سنجنش از دور (RS) شامل سنتینل ۱ و ۲، همچنین از مدل رقومی ارتفاع (DEM) به دست آمده از ماهواره به ترتیب ALOS PALSAR به عنوان نماینده ارگانیک (O) و توپوگرافی (R) بر اساس معادله اسکورپن (McBratney et al., 2003) استفاده شد. DEM با وضوح مکانی ۱۲/۵ متر از سایت (<https://asf.alaska.edu>) دانلود شد. برای سنتینل ۱ و ۲ از میانگین سری زمانی از سپتامبر تا ژانویه با توجه به فاصله نمونه برداری در ۵ سال (۲۰۱۶-۲۰۲۰) استفاده گردید. در این پژوهش ۱۵۸ و ۱۴۸ تصویر در موتور (GEE) Google Earth Engine در این بازه زمانی به دست آمد (Gorelick et al., 2017). برای ماهواره سنتینل ۱، محصول ("COPERNICUS/S1\_GRD") در حالت (IW) با قطبش دوگانه (VV + VH) در وضوح فضایی ۱۰ متر و تصاویر سنتینل ۲ با استفاده از محصول ("COPERNICUS/S2") ee.ImageCollection فیلتتر پوشش ابر کمتر از ۱۰ درصد در باندهای ۲، ۳، ۴، ۸ با ۱۰ متر و ۱۱ و ۱۲ با وضوح ۲۰ متر استفاده گردید. همچنین برای یکسان سازی وضوح مکانی داده های تصویر ماهواره ای و DEM بر اساس ۱۵ متر از تابع نمونه برداری مجدد و روش bilinear از بسته "raster" (Robert and Jacob., 2012) در RStudio نسخه 1.3.959 (RStudio Team., 2020) استفاده شد. تعداد ۳۳ متغیر کمکی اولیه و ثانویه از DEM با استفاده از تجزیه و تحلیل زمین (ویژگی های پایه، مورفومتریک با مقیاس های محلی و زمین نما، توابع هیدرولوژی - اقلیم و تشعشع - دید) در نرم افزار SAGA GIS نسخه ۷/۳ استخراج شد (Conrad et al., 2015). علاوه بر باندهای VV و VH ماهواره سنتینل ۱، ترکیب VV/VH، VV-VH و (Nguyen et al., 2022) VV+VH/2 با استفاده از تابع محاسبه بستری در نرم افزار ArcGIS نسخه ۱۰/۷ محاسبه شد. شاخص سبزینگی نیز به عنوان یکی از سه مؤلفه Tasseled cap (Shi and Xu., 2019) در SAGA GIS از ترکیب باندهای ۲، ۳، ۴، ۸، ۱۱ و ۱۲ تهیه شد.

### انتخاب متغیر محیطی

انتخاب مناسب‌ترین متغیرهای محیطی در دو گام اساسی زیر در نرم‌افزار R صورت پذیرفت.

### بررسی اثر هم خطی

روش شاخص تورم واریانس (VIF) بر اساس یک رویکرد گام‌به‌گام اقدام به حذف آن دسته از متغیرهایی می‌نماید که در مجموعه متغیرها دارای بیشترین همبستگی با یکدیگر هستند (Dormann et al., 2013). در واقع این شاخص بیان می‌دارد چه مقدار از تغییرات مربوط به ضرایب برآورد شده بابت هم خطی افزایش یافته است. VIF مطابق با رابطه ۱ محاسبه می‌شود.

$$\text{VIF} = \frac{1}{1-R_j^2} \quad (\text{رابطه ۱})$$

در این معادله  $R_j^2$  برازش مدل خطی با یک متغیر بر اساس سایر متغیرها می‌باشد.

شدت هم خطی چندگانه در تحلیل‌های محاسباتی رگرسیونی مقدار عددی فاکتور تورم واریانس ( $\text{VIF} < 10$ ) را مدنظر قرار می‌دهد و برای مواردی که مقادیر تورم واریانس ( $\text{VIFmax} \geq 10$ ) باشد فاقد اعتبار و قطعیت محاسباتی بوده و از روند تحلیلی حذف می‌شود (بی‌همتا و زارع چاهوکی: ۱۳۸۹). بر اساس میزان شاخص VIF، ۱۰ متغیری که دارای بالاترین میزان اثر هم خطی و مقادیر بالاتر از ۱۰ بودند در گام نخست حذف شدند.

### روش حذف ویژگی برگشتی<sup>۱</sup> (RFE)

روش RFE یک روش انتخاب متغیرهای کمکی بر اساس مدل‌های یادگیری ماشین است که وزن هر متغیر با توجه به مدل یادگیری ماشین محاسبه می‌شود. سپس متغیرها بر اساس وزن رتبه‌بندی شده و متغیرهایی با وزن کمتر حذف می‌شوند. این روند تا حذف تمام متغیرها ادامه پیدا می‌کند. متغیرهایی که دیرتر حذف شوند، اهمیت بیشتری دارند. در نهایت زیرمجموعه‌ای به‌عنوان مجموعه نهایی انتخاب می‌شود که مربوط به حداکثر دقت مدل یادگیری با بالاترین وزن باشند (Lin et al., 2017).

### روش یادگیری عمیق شبکه عصبی کانولوشن (CNN)

در این تحقیق از مراحل رویکرد یادگیری عمیق با استفاده از مدل CNN و بر اساس (Taghizadeh-Mehrjardi et al., 2020) و برای جزئیات مدل‌سازی ریاضیات این شبکه‌ها از منبع (Wadoux et al., 2019) استفاده شد. در شکل ۲ معماری کامل یک شبکه عصبی کانولوشن ارائه شده است. CNN از دو بخش تشکیل شده است بخش اول: شامل فیلتر کانولوشن برای استخراج ویژگی‌ها به‌صورت سلسله مراتبی است و لایه دوم یک‌لایه کاملاً متصل برای محاسبه مقدار خروجی از مقادیر ورودی چندگانه متشکل از لایه‌های کاملاً متصل از نورون‌ها است. شبکه‌های عصبی مصنوعی معمولی (ANN) فقط از بخش دوم تشکیل شده‌اند. بنابراین مرحله استخراج ویژگی، تمایز اصلی بین CNNها و ANNهای معمولی می‌باشد. استخراج ویژگی با فیلتر کانولوشن شامل مراحل مختلفی است که می‌تواند چندین بار تکرار شود و شامل لایه‌های کانولوشن و ادغام<sup>۲</sup> می‌شود. هدف استخراج مرتبط‌ترین ویژگی‌ها و نمایش‌های چندگانه مقیاس‌های مکانی از کل مجموعه داده‌های کمکی است.

### لایه کانولوشن<sup>۳</sup>

مرحله کانولوشن شامل یک مرحله فیلتر و اعمال یک تابع فعال‌سازی می‌باشد. یک پنجره متحرک، فیلتر با اندازه‌های معمولی  $3 \times 3$  یا  $5 \times 5$  پیکسل در کل مجموعه داده‌های کمکی حرکت می‌کند (معمولاً از چپ به راست و از بالا به پایین). مقادیر ارزش داخل فیلتر که در ابتدا به‌طور تصادفی تخصیص داده می‌شود و وزن هر پیکسل را نشان می‌دهد، در مقادیر داده‌های کمکی مربوطه ضرب می‌شود و به یک مقدار جدید خلاصه می‌شود که به مرکز فعلی فیلتر متحرک اختصاص داده می‌شود و به‌طور متوالی یک نقشه ویژگی جدید را تشکیل می‌دهد. از آنجایی که این "نقشه" جدید از ترکیبی از تمام داده‌های کمکی ورودی تشکیل شده است، نتیجه را نقشه ویژگی می‌نامند و نه داده‌های کمکی فیلتر شده<sup>۴</sup>. تعداد فیلترهای کانولوشن توسط کاربر تعریف می‌گردد و به‌عنوان "عمق" CNN نامیده می‌شود. وزن فیلتر معمولاً به‌طور تصادفی مقداردهی اولیه شده و با انتشار پس‌انداز تنظیم می‌شود. مرحله کانولوشن را می‌توان چندین بار متوالی اعمال کرد. با این حال، آن را تنها به داده‌های کمکی ورودی در مرحله اول اعمال می‌شود. مراحل کانولوشن بعدی برای نقشه‌های ویژگی که قبلاً

7. Recursive Feature Elimination

2. Pooling

3. Convolutional

8. filtered auxiliary data



ایجاد شده‌اند اعمال می‌شوند و بنابراین، سطوح عمومی‌تر و درشت‌تری از اطلاعات مکانی ایجاد می‌کنند. یک مرحله کانولوشن بلافاصله با یک مرحله تبدیل دنبال می‌شود که خروجی نهایی نورون‌ها (پیکسل‌ها) نقشه‌های ویژگی را محاسبه می‌کند. مانند تمام شبکه‌های عصبی مصنوعی، این کار با استفاده از یک تابع فعال‌سازی انجام می‌شود. در مورد CNN، معمولاً از یک واحد خطی اصلاح‌شده (ReLU) برای ارائه غیرخطی بودن مدل استفاده می‌شود.

#### لایه ادغام<sup>۱</sup>

ادغام مرحله اصلی نمونه‌گیری مجدد است که تعداد پیکسل‌ها را در نقشه‌های ویژگی کاهش می‌دهد تا بعد مکانی و در نتیجه هزینه محاسبات را کاهش می‌دهد، برای جلوگیری از برازش بیش از حد و تعمیم بیشتر اطلاعات (رضایی و همکاران، ۲۰۱۸) و برای هر یک از نقشه ویژگی‌ها به‌طور جداگانه انجام می‌شود. از نظر فنی، همچنین نوعی فیلتر است که معمولاً از اندازه‌های هسته ۲×۲ پیکسل و گام ۲ استفاده می‌شود. مقادیر کرنل یا میانگین گرفته می‌شوند یا از حداکثر مقدار استفاده می‌شود (max-pooling).

#### لایه کاملاً به‌هم‌پیوسته<sup>۲</sup>

پس از چند مرحله پیچیدگی و ادغام بعدی، بخش دوم CNN (طبقه‌بندی یا رگرسیون) از لایه‌های شبکه کاملاً متصل معمولی استفاده می‌کند. در اینجا، ویژگی‌های مکانی-محلی استخراج‌شده از بخش کانولوشن با هم ترکیب می‌شوند. تفاوت اصلی بین لایه‌های کانولوشن و لایه‌های کاملاً به‌هم‌پیوسته این است که نورون‌های لایه کانولوشن فقط به نورون‌های میدان گیرنده متصل هستند و نه به همه نورون‌های لایه قبلی. تبدیل یک لایه کانولوشن به یک لایه کاملاً متصل را صاف کردن<sup>۳</sup> می‌گویند.

#### حذف تصادفی<sup>۴</sup>

یک رویکرد برای جلوگیری از کاهش بیش از حد برازش، استفاده از فنی به نام «حذف تصادفی» است. حذف تصادفی به‌طور تصادفی نورون‌ها را در لایه‌های کانولوشن یا کاملاً متصل قطع می‌کند. این انصراف برای هر یک از  $n$  مرحله آموزش اعمال می‌شود و در نتیجه  $n$  شبکه مختلف ایجاد می‌شود که در نهایت برای پیش‌بینی میانگین می‌شوند. برای پیش‌بینی، مجموعه شبکه‌های پراکنده حاصل از فرآیند حذف تصادفی با استفاده از میانگین هندسی وزن ورودی نورون‌ها میانگین می‌شود.

#### افزایش داده‌ها<sup>۵</sup>

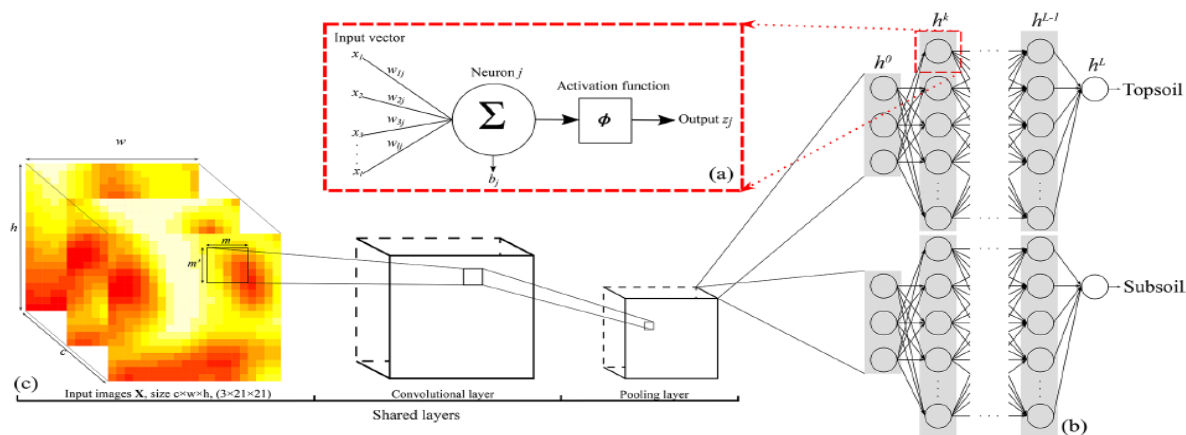
علاوه بر لایه‌های حذف تصادفی، از فن‌های تقویت داده‌ها برای جلوگیری از مشکلات بیش از حد برازش و افزایش قابلیت تعمیم مدل استفاده می‌شود. افزایش داده‌ها فرآیند تولید نمونه‌های آموزشی بیشتر از داده‌های آموزشی موجود با استفاده از تعدادی تبدیل تصادفی است. هدف اصلی این است که به مدل اجازه داده شود تا جنبه‌های بیشتری از نمونه‌های آموزشی را بررسی کند و در نتیجه قابلیت تعمیم شبکه را افزایش دهد. این از اهمیت ویژه‌ای برخوردار است زیرا موفقیت CNN تا حد زیادی به تعداد نمونه‌های آموزشی موجود بستگی دارد.

به‌منظور مدل‌سازی و پیش‌بینی کلاس‌های زیرگروه خاک با استفاده از روش یادگیری عمیق و خصوصاً شبکه‌های عصبی کانولوشن از محیط آنلاین سامانه (https://colab.research.google.com/notebooks/intro.ipynb) Google Collaboratory استفاده گردید. در این سامانه از کتابخانه‌های "numpy" برای محاسبات ریاضیاتی، "pandas" برای چارچوب داده‌سازی<sup>۶</sup>، کتابخانه "geopandas" برای پردازش داده‌های برداری، "rasterio" برای پردازش داده‌های رستری و از بسته‌های "TensorFlow" و "Tensorflow.keras" برای مدل‌سازی شبکه CNN استفاده گردید. در جهت افزایش تعداد داده‌های آموزشی از ۷ مرحله شامل روش‌های تبدیل flip چپ به راست و بالا به پایین و قطری از بالا راست به سمت پایین چپ و قطری از بالا چپ به سمت پایین راست و چرخش<sup>۷</sup> ۹۰ درجه در جهت عقربه‌های ساعت، ۹۰ درجه در جهت خلاف عقربه‌های ساعت و چرخش ۱۸۰ درجه نمونه‌های آموزشی قطعات<sup>۸</sup> اطراف پروفیل‌های زیرگروه‌های خاک استفاده گردید. قطعات آموزشی با اندازه‌های پنجره‌های ۳،۵،۷،۹،۱۵،۲۱ پیکسل در نظر گرفته شد و از تعداد ۱۰۰ اپوک<sup>۹</sup> در مرحله آموزش با در

9. Pooling  
2. Fully-connected layer  
11. flattening  
12. Dropout.  
13. Data augmentation  
14. Data frames  
15. rotate  
16. patch  
17. epoch



نظر گرفتن اپوک ۱۰ به‌عنوان مرحله محاسبه بیش‌برازش استفاده شد. از تابع "adam" برای بهینه‌سازی مدل و تابع "Softmax" به‌عنوان تابع فعال‌سازی برای خروجی کلاس‌های زیرگروه و از تابع "relu" به‌عنوان فعال‌ساز بعد از لایه‌های کانولوشن استفاده گردید. معماری شبکه ارائه شده در جدول ۱ برای مدل‌سازی شبکه CNN در این تحقیق به کار گرفته شد.



شکل ۲. معماری یک شبکه عصبی کانولوشن (منبع: Wadoux et al., 2019)

جدول ۱. معماری شبکه CNN استفاده شده برای پیش‌بینی زیرگروه‌های خاک در منطقه مورد مطالعه

پارامتر	شکل خروجی	نوع لایه
۸۸۰	(۱۶, ۳, ۳)	Conv2D
.	(۱۶, ۲, ۲)	MaxPooling2D
.	(۱۶, ۲, ۲)	Dropout
۴۶۴۰	(۳۲, ۲, ۲)	Conv2D
.	(۱۲۸)	Flatten
۱۲۹۰	(۱۰)	Dense
۱۴۳	(۱۳)	Dense

### روش جنگل تصادفی (RF)

بر اساس نتایج مطالعات قبلی (Brungard et al., 2015; Heuang et al., 2016)، الگوریتم جنگل تصادفی (RF) با توجه به جلوگیری از برازش بیش از حد یک درخت تصمیم واحد و ارائه اهمیت نسبی متغیرهای محیطی در پیش‌بینی نهایی زیرگروه‌های خاک مورد استفاده در این مطالعه بر اساس ۸۰ درصد (n=) داده‌های آموزشی با استفاده از بسته "caret" و روش اعتبارسنجی متقاطع ۱۰-فولد با ۱۰ تکرار برای به دست آوردن یک مدل پایدار و قابل اعتماد استفاده شد. پارامترهای بهینه مدل با استفاده از روش جستجوی شبکه‌ای (Grid search) به صورت  $n_{tree} = 250$ ،  $m_{try} = 3$  بهینه شد و در نهایت مدل با مجموعه ۲۰٪ داده‌های مستقل (n=?) با استفاده از آماره‌های ماتریس خطا شامل شاخص صحت عمومی (Overall accuracy) و شاخص F1-Score اعتبارسنجی شد. همچنین جهت مقایسه کارایی مدل RF با مدل CNN از متغیرهای محیطی مقیاس‌گذاری شده در اندازه پنجره‌های محاسباتی مختلف با استفاده از تابع Filter و روش Smooth در اندازه‌های ۳، ۵، ۷، ۹، ۱۵، ۲۱ از نوع مربع در محیط نرم‌افزار SAGAGIS قبل از مدل‌سازی تهیه و به‌عنوان ورودی مدل RF استفاده گردیدند.

### اعتبار سنجی و عدم قطعیت

دقت پیش‌بینی‌ها با استفاده از یک ماتریس خطا ارزیابی شد (Jensen., 2015). دقت کلی نقشه طبقه‌بندی با تقسیم تعداد کل پیکسل‌های به‌درستی طبقه‌بندی شده (یعنی مجموع قطر در ماتریس خطا) بر تعداد کل پیکسل‌های ماتریس خطا (N) تعیین می‌شود (رابطه ۲).

$$OA = \sum_{i=1}^k \frac{X_{ii}}{N} \quad \text{رابطه ۲}$$

که در آن  $X_{ii}$  تعداد مشاهدات در ردیف  $i$  و ستون  $i$  است،  $k$  تعداد سطرها (به عنوان مثال، طبقات خاک) در ماتریس خطا است. همچنین برای ارزیابی هر طبقه خاک، دو شاخص شامل دقت تولیدکننده (PA) و دقت کاربر (UA) محاسبه شد. تعداد کل پیکسل‌های صحیح در یک کلاس تقسیم بر تعداد کل پیکسل‌های آن کلاس که از داده‌های مرجع زمین (کل ستون) تعیین می‌شود، دقت تولیدکننده نامیده می‌شود (معادله ۳).

$$\text{Producer's accuracy (PA) for class } i = \frac{X_{ii}}{X_{+i}} \quad \text{رابطه ۳}$$

اگر تعداد کل پیکسل‌های صحیح در یک کلاس بر تعداد کل پیکسل‌هایی که واقعاً در آن دسته طبقه‌بندی شده‌اند (کل ردیف) تقسیم شود (رابطه ۴).

$$\text{User's accuracy (UA) for class } j = \frac{X_{jj}}{X_{+j}} \quad \text{رابطه ۴}$$

F1-Score به عنوان یک میانگین هارمونیک از شاخص‌های تولیدکننده و کاربر حاصل مستخرج از ماتریس خطا بر اساس رابطه ۵ محاسبه می‌گردد.

$$F1\text{-Score} = 2 \times \frac{(PA \times UA)}{(PA + UA)} \quad \text{رابطه ۵}$$

میزان F1-Score در دامنه ۰ تا ۱۰۰ نوسان دارد و هر چه میزان آن به ۱۰۰ نزدیک‌تر باشد صحت مدل و نقشه تولیدی بیشتر است. یکی از مزایای روش رقومی در مقایسه با روش معمول نقشه‌برداری خاک این است که می‌تواند عدم قطعیت مرتبط با پیش‌بینی کلاس‌های خاک را ارائه دهد. در روش‌های CNN و RF کلاس نهایی هر پیکسل بر اساس بیشترین تعداد رأی تخمین زده می‌شود. با این حال، با کمی کردن توزیع آرا برای کلاس‌های مختلف، می‌توان میزان قابلیت اطمینان و عدم قطعیت پیش‌بینی هر پیکسل در نقشه نهایی را تعیین کرد (Chaney et al., 2016; Stumpf et al., 2017; Heung et al., 2016). عدم قطعیت پیکسل‌های (x) نقشه‌های تهیه‌شده در این تحقیق با توزیع آرای مربوط به طبقات مختلف با استفاده از شاخص آنتروپی برابر با عدم قطعیت نادیده گرفتن (IU<sup>1</sup>) بر اساس رابطه ۶ کمی‌سازی شد.

$$IU = -\frac{1}{\ln(n)} \sum_{i=1}^n P_i(x) \times \ln(P_i(x)) \quad \text{رابطه ۶}$$

جایی که  $n$  تعداد کلاس‌های پیش‌بینی شده در هر پیکسل (x) است و  $P(x)$  نشان‌دهنده نسبت آرای است که کلاس  $i$  از مدل جنگل تصادفی دریافت می‌کند. وقتی همه آرا به یک طبقه اختصاص داده می‌شود، میزان آن برابر با ۰ است و در صورت توزیع یکنواخت آرا به طبقات پیش‌بینی شده، نسبت برابر با ۱ خواهد بود. در این تحقیق از نقشه‌های سطح احتمال زیرگروه‌های مختلف خاک به عنوان یکی از خروجی‌های مدل‌های CNN و RF برای تعیین کمی عدم قطعیت پیکسل‌ها در نقشه نهایی طبقات خاک استفاده شده است.

## یافته‌های پژوهش و بحث

### خصوصیات مورفولوژیکی و فیزیکوشیمیایی و طبقه‌بندی کلاس زیرگروه‌های خاک

نتایج مهم‌ترین ویژگی‌های فیزیکوشیمیایی خاک‌های شاهد هر یک از زیرگروه‌های خاک به همراه طبقه‌بندی آن‌ها بر اساس سامانه آمریکایی طبقه‌بندی خاک‌ها (۲۰۱۴) در جدول ۲ ارائه شده است. به طور کلی، ۱۳ زیرگروه خاک با توجه به تنوع افق‌های مشخصه سطحی و زیرسطحی و رژیم‌های رطوبتی (زیریک خشک، اریدیک ضعیف و آکوئیک) مشاهده شد. افق سطحی همه زیرگروه‌ها به دلیل قرار گرفتن در اقلیم‌های خشک و نیمه‌خشک با محتوای کم کربن آلی خاک به عنوان اپی‌پدون اُکریک شناسایی شد. خاک‌ها در سه زیرگروه در رده Inceptisols، یک زیرگروه در رده Entisols و نه زیرگروه در رده Aridisols قرار گرفتند. از نظر تکامل، Inceptisols (Typic)

بیش از ۱ متر می‌باشد. pH افق سطحی و زیرسطحی به ترتیب با دامنه ۷/۴ تا ۷/۹۱ و ۸/۰۲ تا ۸/۰۲ مشاهده شد. همه افق‌ها دارای  $EC < 4ds/m$  و  $SAR < 13$  بدون مشکل شوری و سدیمی هستند. مقدار کربن آلی با عمق روند کاهشی دارد و گاهی خاک‌های دارای شرایط فلوونتییک به‌طور نامنظم تغییر می‌کنند. بافت خاک سطحی متوسط تا سبک (لوم، لوم شنی) است و بافت افق‌های زیرسطحی به دلیل تنوع شرایط رسوب‌گذاری، اغلب مشابه با افق سطحی و در مواردی لوم رسی و رسی مشاهده شد. به‌استثنای زیرگروه Typic Calcixerepts، درصد آهک معادل بقیه خاک‌ها کمتر از ۱۰ درصد می‌باشد. اکثر زیرگروه‌های Typic Calcixerepts در اکثر موارد دارای بیش از ۱۵ درصد سنگریزه در سطح و افق زیرسطحی هستند. زیرگروه‌های Inceptisols با فراوانی ۱۳۳ خاکرخ دارای فراوانی تخمینی نزدیک به ۵۰ درصد زیرگروه‌های خاک بوده و اغلب به دلیل وجود رژیم رطوبتی Xeric در منطقه نیمه‌خشک منطقه مورد مطالعه مشاهده شد. Aridisols دارای افق‌های زیرسطحی کلسیک، جیپسیک، سالیک و گاهی حضور هم‌زمان افق‌های جیپسیک-سالیک و جیپسیک-کلسیک می‌باشد. دو مورد از Aridisols با توالی از افق‌های کلسیکی بیش از ۱۵ درصد آهک معادل داشتند و با توجه به اینکه یکی از خاک‌ها دارای  $SAR > 13$  و  $pH > 8.5$  است، در زیر گروهی با شرایط سدیمی Sodic Xeric Haplocalcids طبقه‌بندی شد؛ اما شوری دو زیرگروه کمتر از ۴ ds/m بوده و در طبقه خاک غیر شور قرار گرفت. خاک با افق سطحی (Sodic Xeric Haplocalcids) دارای کاربری مرتعی و Xeric Haplocalcids به دلیل قلیائیت کم برای کشاورزی آبی استفاده می‌شود. این زیرگروه‌ها اغلب دارای بافت سطحی و زیرسطحی نسبتاً سنگین لوم رسی بوده و با ۳۹ مشاهده بیش از ۱۴ درصد فراوانی خاک‌های مشاهده شده را دارند. خاک‌های با تجمع افق‌های گچی در دو حالت با و بدون افق‌های کلسیک در منطقه مشاهده می‌شوند که با تعداد ۵۳ رخداد نزدیک به ۲۰ درصد از خاک‌های مشاهده‌شده را به خود اختصاص داده‌اند. این خاک‌ها در هنگامی که همراه افق کلسیک در منطقه مشاهده می‌شوند بیش از ۱۵ درصد گچ و در شرایط فاقد افق کلسیک کمتر از ۱۵ درصد گچ دارند. در تعدادی از خاکرخ‌های مشاهده‌شده میزان SAR بیش از ۱۳ و pH کمتر از ۸ و در رده خاک‌های سدیمی با پسوند Sodic قرار گرفتند. خاک‌ها غالباً عمیق و بیش از ۱۰۰ سانتی‌متر عمق داشته و برخی دارای کاربری مرتع با افق‌های سطحی A و با تجمع گچ Ay می‌باشند. بافت سطحی خاک‌های دارای گچ، نسبتاً سنگین Clay Loam و بافت زیرسطحی Clay است. یک زیرگروه دارای افق‌های متوالی سالیک همراه با تجمع گچ در افق‌های سطحی و زیرسطحی با میزان شوری بیشتر از ۳۰ دسی‌زیمنس بر متر و SAR بیش از ۱۳، در رده خاک‌های شور و سدیمی با ضخامت خاک سطحی کم و بافت سطحی و زیرسطحی خیلی سنگین Clay بیش از ۵۰ درصد قرار دارد.

یک زیرگروه هم فاقد سایر افق‌های کلسیک، جیپسیک و سالیک بوده و دارای توالی از افق‌های کلسیک و تحت کشت زراعت آبی دارای بافت سطحی و زیرسطحی سنگین نزدیک به ۵۰ درصد رس با pH بیش از ۸ خاکی است عمیق که میزان کربن آلی آن تحت تأثیر رسوب‌گذاری اراضی بالادست در گذشته دارای تغییر نامنظمی با عمق است. زیرگروه خاک Entisols شامل یک افق با ضخامت کم A با کاربری مرتع بر روی افق BC که میزان سنگ و سنگریزه آن بیش از ۵۰ درصد بافت متوسط لوم در سطح و لوم شنی در افق زیر سطح و pH آن زیر ۸ و میزان کربن آلی خاک با عمق کاهش می‌یابد و شدت در معرض فرسایش است و از نشانه‌های آن می‌توان به میزان ضخامت خاک سطحی کم و کربن آلی پایین و درصد شن بیشتر اشاره نمود.

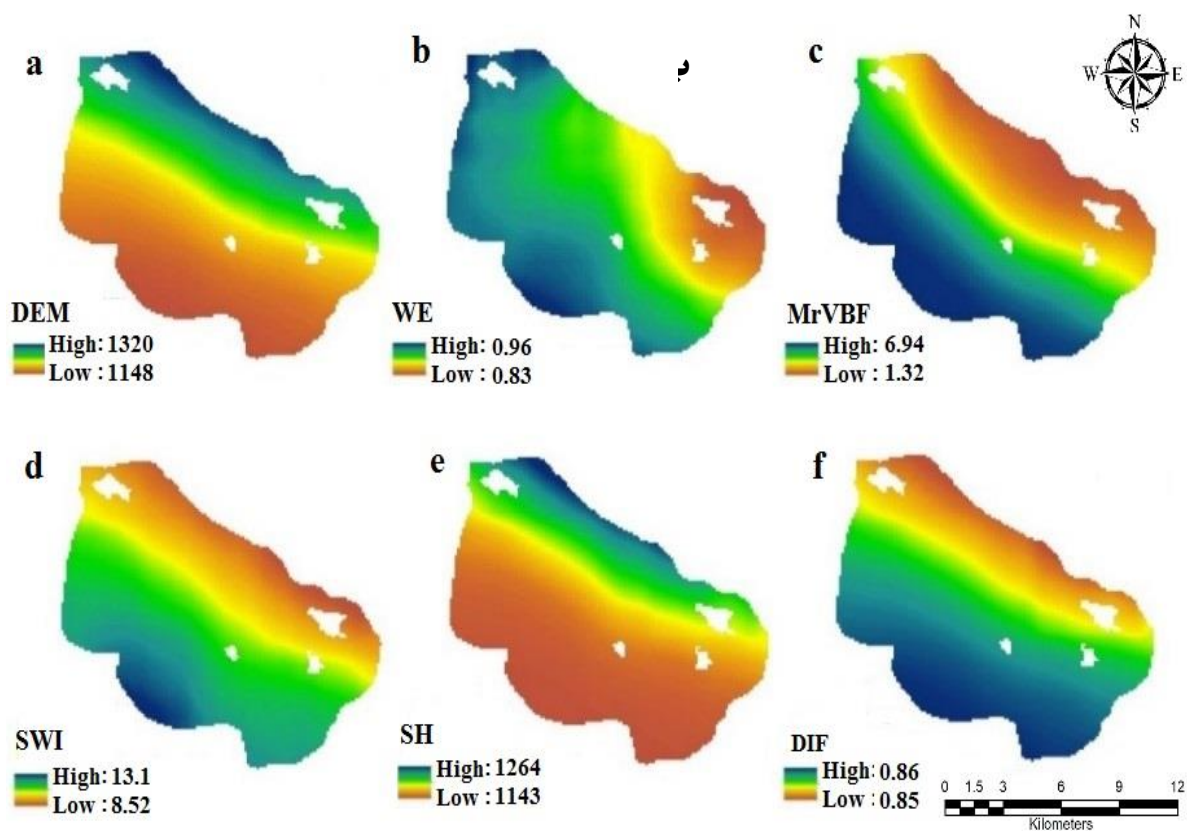
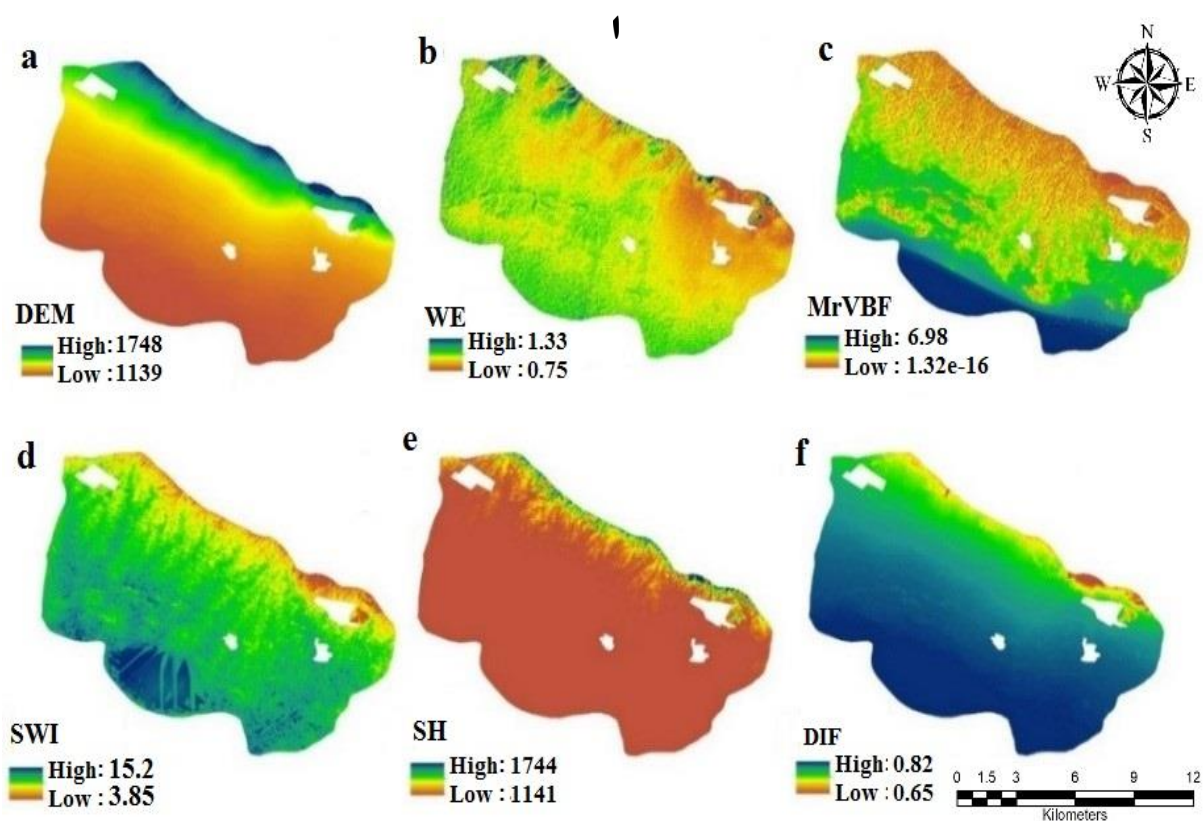
### انتخاب متغیر محیطی

همان‌طور که در شکل ۴ ارائه شده است در مرحله اول انتخاب متغیرهای محیطی با استفاده از شاخص هم‌راستایی خطی توسط شاخص VIF، ۱۰ متغیر از میان ۳۳ متغیر ژئومورفومتری دارای میزان شاخص بالاتر از ۱۰ بوده و حذف گردید، در مرحله دوم ۲۳ متغیر باقیمانده توسط مدل RFE مورد بررسی قرار گرفت که در نهایت شش متغیر (شکل ۳) شامل شاخص همواری دره با درجه وضوح بالا (MrVBF)، شاخص خیزی ساگا (SWI)، ارتفاع استاندارد (SH)، اثر باد (WE)، شدت تابش پخشیده (DIF) و ارتفاع (DEM) دارای بالاترین ضریب اهمیت نسبی و در مدل نهایی نگه داشته شد. در مرحله آخر این شش متغیر همراه با باندها و شاخص‌های تصاویر سنتیل ۱ و ۲ دوباره توسط مدل RFE مورد بررسی قرار گرفتند که در نهایت همان متغیرهای مرحله قبل بالاترین ضریب اهمیت نسبی را نشان دادند و برای مدل‌سازی و پیش‌بینی کلاس‌های زیرگروه خاک توسط مدل‌های CNN و RF استفاده شدند. در تحقیقی مشابه توسط (Rahmani et al., 2019) بخش غالبی از همین متغیرهای محیطی برای پیش‌بینی فازی کلاس‌های زیرگروه خاک در ناحیه‌ای با مساحت کمتر گزارش شد.

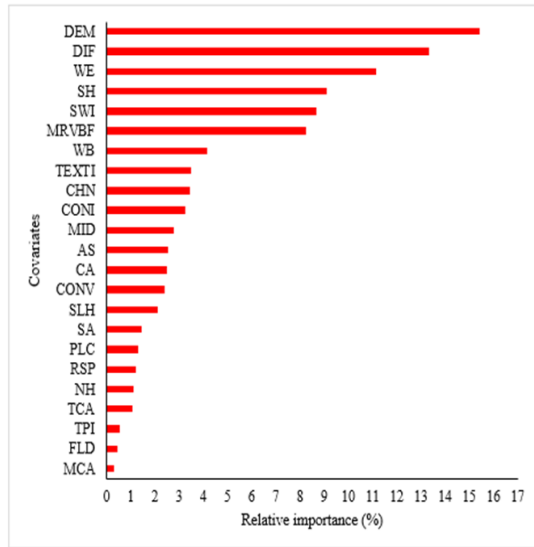


جدول ۲. نتایج خصوصیات فیزیکوشیمیایی و طبقه‌بندی کلاس‌های زیرگروه خاک‌رخ‌های شاهد منطقه مورد مطالعه

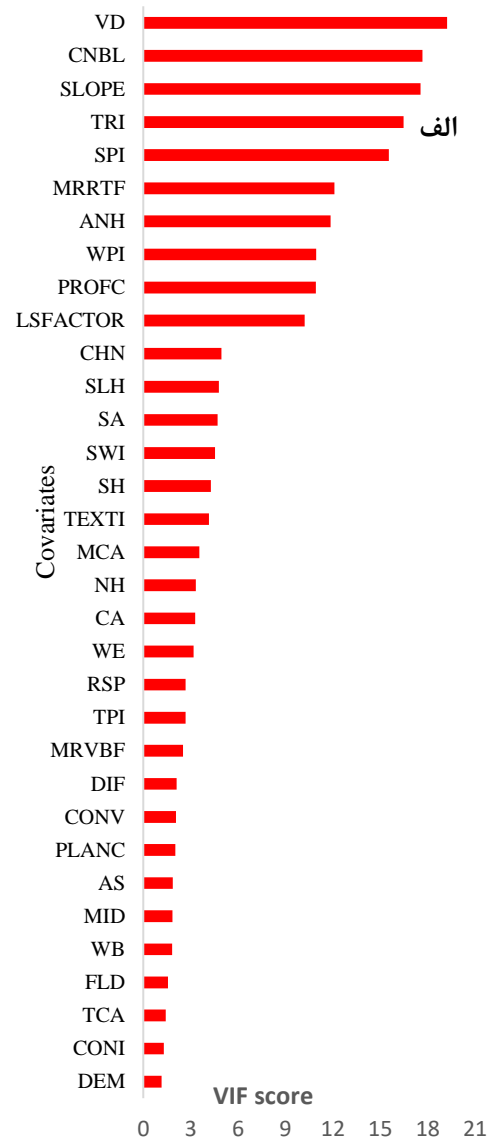
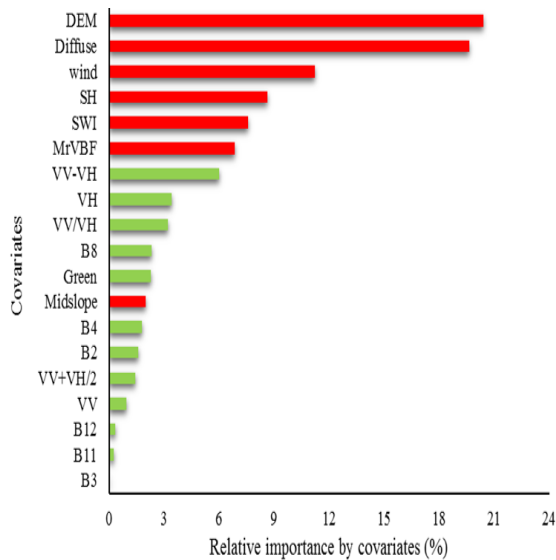
Horizon	Depth (cm)	pH	EC (ds/m)	SAR	OC	CCE	Gypsum %	Sand %	Clay	Silt	C.F	Texture
Typic Haploxerepts (n=۳۹، ۱۴/۲۴%)												
Ap	۰-۳۰	۷/۶۴	۲/۳۰	۷/۹۴	۰/۳۵	۹/۳۵	-	۳۳	۲۰	۴۷	-	Loam
Bw1	۳۰-۷۰	۷/۸۲	۳/۴۴	۵/۶۳	۰/۲۶	۷/۳۲	-	۴۴	۲۰	۳۶	-	Loam
Bw2	۷۰-۱۵۰	۷/۸۵	۲/۱۲	۵/۴۸	۰/۲۳	۷/۹۷	-	۵۸	۱۵	۲۷	-	Sandy Loam
Fluventic Haploxerepts (n=۲۱، ۷/۶۷%)												
Ap	۰-۳۰	۷/۹۴	۰/۵۸	۴/۲۵	۰/۷۱	۵/۷۶	-	۳۳	۲۹	۳۸	-	Clay Loam
Bw1	۳۰-۶۰	۷/۵۶	۰/۰۹	۵/۶۸	۰/۵۵	۸/۰۷	-	۲۵	۴۵	۳۰	-	Clay
Bw2	۶۰-۹۰	۷/۵۵	۱/۳۰	۶/۳۵	۰/۶۷	۷/۳۰	-	۳۱	۳۵	۳۴	-	Clay Loam
Bw3	۹۰-۱۵۰	۷/۸۹	۱/۳۲	۷/۵۴	۰/۴۶	۸/۰۷	-	۲۹	۲۳	۴۸	-	Loam
Typic Calcixerepts (n=۷۳، ۲۶/۶۵%)												
Ap	۰-۲۰	۷/۷۱	۰/۴۱	۳/۲۶	۰/۷۰	۶/۸۲	-	۵۲	۲۰	۲۸	-	Loam
Bw	۲۰-۶۰	۷/۹۸	۰/۳۰	۲/۳۵	۰/۵۳	۱۶/۲۹	-	۳۶	۲۹	۳۵	-	Clay Loam
Bk	۶۰-۱۵۰	۸/۰۲	۰/۳۰	۶/۴۱	۰/۳۵	۲۴/۲۱	-	۳۳	۲۹	۳۸	-	Clay Loam Sandy Loam Sandy Loam
Ap	۰-۲۰	۷/۶۹	۰/۵۰	۱/۳۵	۰/۶۴	۱۳/۶۵	-	۶۷	۱۳	۱۹	۱۵	Sandy Loam
Bw	۲۰-۲۵	۷/۶۸	۰/۴۰	۱/۵۴	۰/۵۳	۱۴/۵۸	-	۶۹	۱۵	۱۶	۲۵	Sandy Loam
Bk	۲۵-۵۵	۷/۶۸	۰/۷۱	۲/۲۱	۰/۴۳	۱۹/۴۹	-	۶۹	۱۲	۱۹	۲۰	Sandy Loam
BC	۵۵-۱۱۰	۷/۶۷	۰/۳۳	۱/۲۰	۰/۰۹	۱۲/۷۱	-	۸۰	۱۲	۹	۵۰	Sandy Loam
Xeric Haplocalcids (n=۲۱، ۷/۶۶%)												
Ap	۰-۲۰	۷/۸	۱/۷۴	۲/۸۰	۱/۰۱	۱۷/۳۱	-	۲۹	۳۵	۳۶	-	Clay Loam
Bk1	۲۰-۷۰	۷/۷	۰/۹۸	۲/۵۷	۰/۹۲	۲۱/۲۵	-	۲۳	۳۹	۳۸	-	Clay Loam
Bk2	۷۰-۱۴۰	۷/۸۳	۱/۰۵	۴/۰۳	۰/۱۷	۲۳/۶۱	-	۲۹	۳۷	۳۴	-	Clay Loam
Sodic Xeric Haplocalcids (n=۱۸، ۶/۵۷%)												
A	۰-۳۰	۸/۰۹	۲/۴۵	۸/۱۴	۰/۶۵	۲۰/۴۰	-	۳۳	۳۷	۳۰	-	Clay Loam
Bk1	۳۰-۷۵	۸/۵۲	۳/۰۵	۱۶/۸۵	۰/۵۰	۳۰/۷۰	-	۴۱	۲۹	۳۰	-	Clay Loam
Bk2	۷۵-۱۴۰	۸/۹۲	۱/۸۶	۲۲/۴۷	۰/۱۸	۴۱/۲۰	-	۲۵	۵۱	۲۴	-	Clay
Sodic Xeric Haplogypsids (n=۱۳، ۴/۷۳%)												
A	۰-۱۵	۷/۷۶	۲۴	۲۸/۷۰	۰/۶۷	۱۰/۴۹	-	۲۶	۴۰	۳۵	-	Clay Loam
Bk	۱۵-۵۵	۷/۶۶	۱۱/۰۹	۱۷/۳۷	۰/۵۹	۱۱/۲۸	-	۱۱	۶۴	۲۵	-	Clay
By1	۵۵-۹۰	۷/۹۸	۱۹/۱۵	۲۹/۰۰	۰/۵۰	۹/۹۷	۴/۴۰	۲۹	۴۵	۲۶	-	Clay
By2	۹۰-۱۳۰	۸/۲۱	۱۹/۹۷	۳۵/۹۷	۰/۴۲	۹/۹۷	۷/۲۳	۳۹	۲۱	۴۱	-	Loam
Xeric Haplogypsids (n=۱۴، ۵/۱۱%)												
Ap	۰-۲۰	۷/۲۰	۳/۶۴	۵/۹۴	۰/۷۹	۱۱/۱۳	-	۳۲	۳۵	۳۳	-	Clay Loam
Bw	۲۰-۳۵	۷/۵۰	۱/۵۶	۴/۲۵	۰/۵۳	۱۵/۶۲	-	۳۸	۳۸	۲۴	-	Clay Loam
By	۳۵-۸۵	۷/۱۰	۷/۴۹	۷/۱۴	۰/۴۴	۱۵/۳۴	۸/۸۰	۴۵	۳۰	۲۵	-	Clay Loam
BCy	۸۵-۱۵۰	۷/۷۰	۵/۵۴	۸/۱۰	۰/۲۳	۲۹/۱۲	۲/۳۰	۴۷	۲۷	۲۶	-	Sandy Clay Loam
Sodic Xeric Calcigypsids (n=۱۹، ۶/۹۳%)												
Ap	۰-۲۰	۷/۸۰	۰/۸۵	۴/۵۰	۰/۴۴	۱۴/۹۱	-	۲۴	۴۰	۳۶	-	Clay
Bky1	۲۰-۶۵	۷/۸۳	۰/۷۱	۳۱/۶۰	۰/۳۶	۱۶/۴۲	۲۰/۱	۱۴	۵۸	۲۸	-	Clay
Bky2	۶۵-۱۵۰	۷/۵۶	۰/۴۳	۲۷/۶۷	۰/۲۴	۱۲/۲۰	۳۳/۱	۱۱	۶۴	۲۵	-	Clay
Xeric Calcigypsids (n=۷، ۲/۵۵%)												
Ay	۰-۱۵	۷/۶۷	۲/۵۳	۰/۵۲	۰/۳۳	۱۷/۲۸	۲۵/۲	۲۳	۳۸	۳۵	-	Clay Loam
Bky	۱۵-۷۰	۷/۷۹	۳/۳۷	۱/۳۸	۰/۲۸	۱۷/۶۲	۲۳/۷	۱۵	۵۵	۳۰	-	Clay
By	۷۰-۱۵۰	۷/۶۱	۴/۳۲	۲/۸۵	۰/۲۴	۱۳/۵۵	۲۵/۲	۱۶	۵۲	۳۲	-	Clay
Gypsic Aquisalids (n=۱۰، ۳/۶۵%)												
Ayz	۰-۱۵	۷/۴۵	۸۶/۴	۸۶/۱۸	۰/۶۱	۷/۷۵	۱۳/۹	۲۱	۴۹	۳۰	-	Clay
Byz1	۱۵-۴۵	۷/۴۸	۹۷/۵	۹۲/۵۹	۰/۴۴	۱۲/۰۳	۱۴/۸	۲۵	۶۱	۱۴	-	Clay
Byz2	۴۵-۱۵۰	۷/۴۴	۱۰۰/۷	۹۱/۷۰	۰/۳۵	۱۳/۲۶	۱۲/۹	۱۷	۶۳	۲۰	-	Clay
Xerofluventic Haplocambids (n=۱۲، ۴/۳۸%)												
Ap	۰-۱۸	۸/۴۷	۱/۸۴	۷/۲۷	۰/۹۱	۲۰/۰۰	-	۱۷	۴۵	۳۸	-	Clay
Bw1	۱۸-۷۰	۸/۱۸	۲/۳۴	۹/۳۵	۰/۴۳	۲۰/۰۰	-	۱۲	۴۹	۳۹	-	Clay
Bw2	۷۰-۱۵۰	۸/۶۳	۳/۴۲	۲۴/۳۲	۰/۶۴	۲۱/۶۹	-	۹	۵۲	۳۹	-	Clay
Typic Xerorthents (n=۱۸، ۶/۵۷%)												
A	۰-۱۵	۷/۹۷	۰/۴۸	-	۰/۵۴	۱۸/۸۰	-	۵۲	۱۴	۳۴	۳۰	Loam
BC	۱۵-۴۵	۷/۸۳	۰/۴۱	-	۰/۳۵	۱۹/۷۸	-	۵۳	۱۵	۳۲	۶۰	Sandy Loam



شکل ۳. شش متغیر محیطی مهم و انتخاب شده در مقیاس اندازه پیکسل الف) - ۱۵ متر و ب) - پنجره محاسباتی بهینه ۱۵\*۱۵،  
a: DEM, b: WE (Wind effect), c: MRVBF (Multiresolution valley bottom flatness), d: SWI (Saga wetness index), e: SH (Standardized height),  
f: DIF (Diffuse insolation),



The relative importance of remained covariates after collinearity test by RF model



شکل ۴. نمودارهای بیانگر انتخاب متغیرهای محیطی منتخب بر اساس الف) - شاخص VIF (ب) - اهمیت نسبی RFE1 (ج) - متغیرهای محیطی

منتخب ژئومورفومتری همراه با مقادیر منفرد طیفی باندها و شاخص‌های مستخرج از ماهواره های سنتینل ۱ و ۲ (RFE2)

Stream power index, MRRTF:Multiresolution :VD: Valley Depth, CNBL:Chanel network base level,TRI: Terrain ruggedness index, SPI Hieghts, ruggedeness terrain flatness, ANH: Analitical hillshading, PROFC: Profile curvature, CHN: Channel network distance, SLH: Slope NH: ,SA:Surface Area, SWI: Saga wetness index, SH: Standardized Hieght, TEXTI: Texture index, MCA: Modified catchment area Normalized Hieght, CA: Catchment area,WE: Wind effect, RSP: Relative slope position, TPI: Topographic position index, MrVBF: Multiresolution valley bottom flatness, DIF: Diffuse insolation, CONV: Convexity, PLANC: Plan curvature, AS:Aspect, green: Greennes index

### عملکرد مدل‌های شبکه CNN و RF در پیش‌بینی زیرگروه‌های خاک متأثر از مقیاس‌های مختلف

نتایج عملکرد مدل‌های RF و CNN در پیش‌بینی کلاس‌های زیرگروه خاک در جدول ۳ ارائه شده است. نتایج صحت عمومی مدل CNN بیانگر افزایش صحت از اندازه پنجره ۳ تا ۵ و از اندازه ۵ تا ۹ میزان صحت عمومی ثابت (۳۸ درصد) می‌باشد و در اندازه پنجره ۱۵ میزان آن به حداکثر ۴۳ و سپس میزان آن کاهش می‌یابد. در مدل RF نیز شرایط مشابه با مدل CNN در روند صحت عمومی مشاهده گردید با این تفاوت که میزان صحت عمومی در کل بیشتر بوده و در اندازه پنجره ۳ نیز در ابتدا میزان صحت عمومی نسبت به سایر اندازه پنجره‌ها بیشتر می‌باشد. همچنین بیشترین میزان صحت عمومی با ۵۰ درصد در اندازه پنجره ۱۵ مشاهده شد. میزان اختلاف صحت دو مدل در حالت بهینه به میزان ۷ درصد بوده که با توجه به نیاز بالای مدل CNN به تعداد داده و پایین بودن فراوانی داده‌های مشاهده‌ای (۲۷۸ خاکرخ) در تحقیق حاضر صحت قابل قبولی مشاهده شده‌است. در تحقیق (Yan et al., 2022) در استفاده از داده محدود و تکنیک

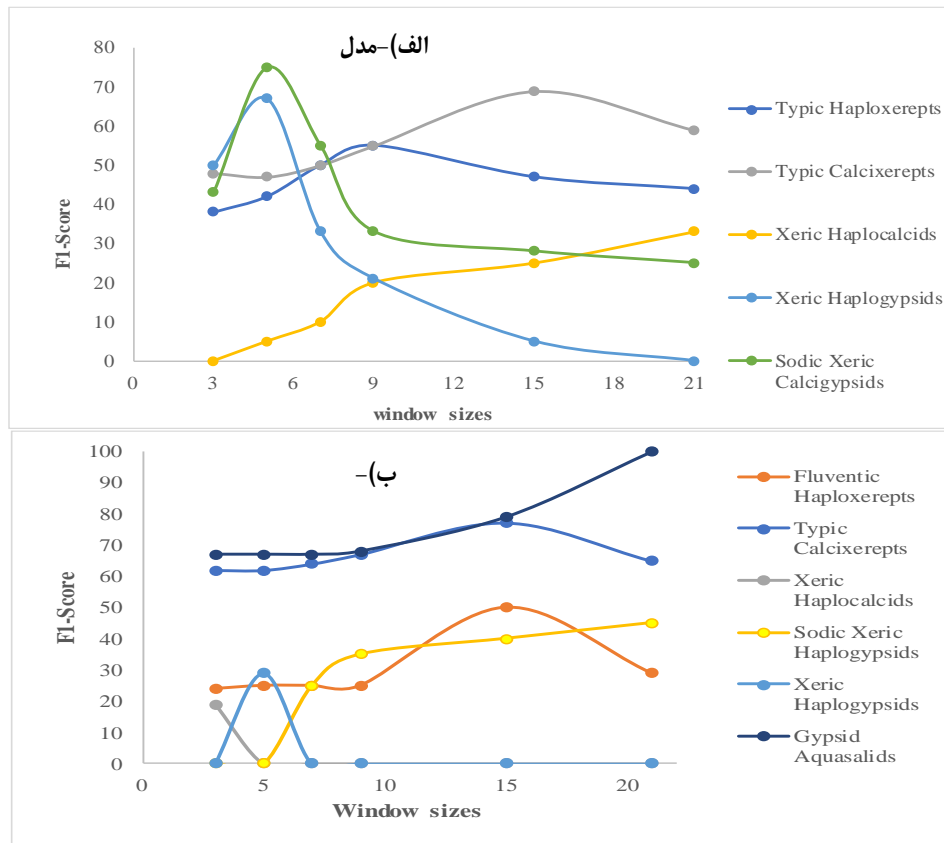
متعادل‌سازی کلاس‌های پیش‌بینی مدل RF نسبت به مدل CNN برتری نشان داد. در تحقیق (Beucher et al., 2022) برتری ۷ درصدی صحت عمومی مدل CNN در حضور تعداد ۵۰۰۰ داده نسبت به مدل RF مقیاس‌گذاری شده متغیرهای محیطی و رویکرد انتخاب متغیر مناسب RFE گزارش نمودند. با توجه به نتایج جدول ۳ برای زیرگروه Sodic Xeric Calcigypsids در مدل CNN و زیرگروه‌های Xeric Calcigypsids و Typic Xerorthents در مدل RF در تمامی اندازه پنجره‌های محاسباتی پیش‌بینی صورت نگرفته است و کلاس‌های Xeric Haplocalcids و Xeric Haplogypsids تنها در یک پنجره قابلیت پیش‌بینی توسط مدل RF را دارا می‌باشند. مطابق با این نتایج (Smith et al., 2006) در پیش‌بینی سری‌های خاک و (Yang et al., 2021) در پیش‌بینی کربن آلی خاک روند نوسانی در ضریب تبین پیش‌بینی مدل CNN را مشاهده نمودند.

جدول ۳. عملکرد مدل‌های CNN و RF در پیش‌بینی زیرگروه‌های خاک در اندازه پیکسل ۱۵ متر با در نظر گرفتن اندازه پنجره‌های محاسباتی مختلف

CNN	RF	CNN	RF	CNN	RF	CNN	RF	CNN	RF	CNN	RF	Window Sizes
۲۱×۲۱		۱۵×۱۵		۹×۹		۷×۷		۵×۵		۳×۳		
F1-score												Subgroup level
۴۴	۳۱	۴۷	۳۶	۵۵	۲۱	۵۰	۱۹	۴۲	۵۴	۳۸	۶۲	Typic Haploxerepts
.	۲۹	۲۰	۵۰	.	۲۵	۲۹	۲۵	۲۹	۲۵	.	۲۴	Fluventic Haploxerepts
۵۹	۶۵	۶۹	۷۷	۵۵	۶۷	۵۰	۶۴	۴۷	۶۲	۴۸	۶۲	Typic Calcixerepts
۳۳	.	۲۵	.	۲۰	.	۱۰	.	۵	.	.	۱۹	Xeric Haplocalcids
۳۳	۸۶	۲۵	۷۵	۵۷	۸۶	۲۹	۹	۴۰	۸۶	۲۵	۱۰۰	Sodic Xeric Haplocalcids
۵۰	۴۵	۵۰	۴۰	.	۳۵	۲۹	۲۵	.	.	۲۹	.	Sodic Xeric Haplogypsids
.	.	۵	.	۲۱	.	۳۳	.	۶۷	۲۹	۵۰	.	Xeric Haplogypsids
۲۵	.	۲۸	۵۰	۳۳	۲۲	۵۵	۲۹	۷۵	.	۴۳	.	Gypsic Aquisalids
۳۳	.	۴۰	.	۲۲	.	۲۲	.	۶۰	.	.	.	Xeric Calcigypsids
.	۱۰۰	.	۷۹	.	۶۸	.	۶۷	.	۶۷	.	۶۷	Sodic Xeric Calcigypsids
۳۳	۵۰	۵۰	۳۳	.	۱۷	.	۶۷	۵۰	۳۳	۵۰	۵۰	Xerofluventic Haplocambids
۳۳	.	۲۹	.	.	.	۴۰	.	۳۳	.	۲۹	.	Typic Xerorthents
۴۱	۴۵	۴۳	۵۰	۳۸	۴۳	۳۸	۴۲	۳۸	۴۳	۳۴	۴۹	Overall accuracy (OA%)

تعداد هفت زیرگروه در مدل CNN و شش زیرگروه در مدل RF دارای روند نوسانی افزایشی-کاهشی در شاخص F1-Score می‌باشد. شاید یکی از عوامل این روند را در تنوع بالای زیرگروه‌های خاک و توزیع در یک مساحت قابل توجه با تنوع رسوب‌گذاری و زمین‌نماهای متفاوت بیان داشت و بر اساس نتایج (Duan and Zhang., 2021) بهتر است از ادغام اندازه پنجره‌های مختلف بهینه‌شده برای هر زیرگروه خاک در مدل‌سازی و پیش‌بینی کلاس‌های خاک به‌طور هم‌زمان استفاده نمود. زیرگروه‌های دارای روند قابل‌ملاحظه در شاخص F1-Score برای دو مدل CNN و RF در پنج و شش زیرگروه در شکل ۵ ارائه گردیده است.

در مدل CNN سه شکل تغییرات مشاهده شد، در گروه اول: زیرگروه‌های Gypsic Aquisalids و Xeric Haplogypsids با افزایش پنجره محاسباتی از ۳ تا ۵، افزایش ۳۰ و ۱۷ درصدی در میزان F1-Score و بلافاصله یک‌روند نزولی شدید مشاهده گردید. این روند بیانگر مناسب بودن اندازه پنجره‌های کوچک در جهت پیش‌بینی این زیرگروه‌ها و به‌کارگیری اطلاعات با همسایگی نزدیک از متغیرهای محیطی منتخب در جهت جلوگیری از کاهش صحت پیش‌بینی این کلاس‌های خاک می‌باشد. (Behrens et al. (2010) در پیش‌بینی کلاس‌های خاک‌های با الگوی توزیع با شکل کشیده و یا پراکنده در منطقه مورد مطالعه خود استفاده از اندازه پنجره کوچک را در پیش‌بینی توصیه نمودند. هر چند که در زیرگروه Xeric Haplogypsids میزان شاخص به صفر می‌رسد ولی در زیرگروه Gypsic Aquisalids روند با افزایش اندازه پنجره بعد از قله به میزان تقریبی ۳۰ درصد ثابت می‌رسد. زیرگروه Xeric Haplogypsids دارای یک توزیع پراکنده در میان سایر کلاس‌های خاک بوده و این شکل توزیع خود گویای این است که استفاده از پنجره‌های کوچک بهترین گزینه بوده که همسو با تغییرات میزان شاخص F1-Score می‌باشد.



شکل ۵. نمودار تغییرات شاخص F1-Score برخی از زیرگروه‌های خاک دارای روند با استفاده از مدل‌های (الف-CNN) (ب-RF)

اگرچه زیرگروه Gypsic Aqualids در بخش جنوبی منطقه بر اساس شکل ۱ توزیع جغرافیایی دارد و فاقد تداخل با سایر زیرگروه‌ها بوده اما به دلیل تعداد فراوانی کم، مدل CNN ترجیح داده در جهت جلوگیری از کاهش دقت پیش‌بینی، اندازه پنجره کوچک را به‌عنوان پنجره بهینه برای پیش‌بینی نهایی انتخاب نماید. گروه دوم شامل دو زیرگروه Typic Calcixerepts و Typic Haploxerepts بوده که در آن‌ها با افزایش اندازه پنجره محاسباتی از ۳ تا ۹ و ۱۵ روند افزایشی ملایم در شاخص F1-Score و همچنین پس از رسیدن به قله یک‌روند کاهشی ملایم مشاهده گردید. این روند بیانگر این است که این دو زیرگروه با توجه به تعداد قابل توجه مشاهده در منطقه مورد مطالعه و پراکنش ناحیه‌ای در شمال منطقه با توجه به شکل ۱ نیازمند استفاده از اندازه پنجره‌های متوسط تا بزرگ و اطلاعات بیشتر از همسایگی خاک‌های مشاهده‌ای برای این کلاس‌های خاک می‌باشد. (Wadoux et al., 2019) در استفاده از مدل CNN برای پیش‌بینی اجزای بافت و درصد کربن آلی خاک اندازه پنجره ۹ پیکسل را به‌عنوان پنجره با حداقل خطای پیش‌بینی مشاهده نمودند. استفاده از اندازه پنجره‌های بزرگ در نواحی صاف و یا مناطق وسیع دارای تنوع کلاس کم در مطالعه (Cavazzi et al., 2013) نیز مشاهده گردید.

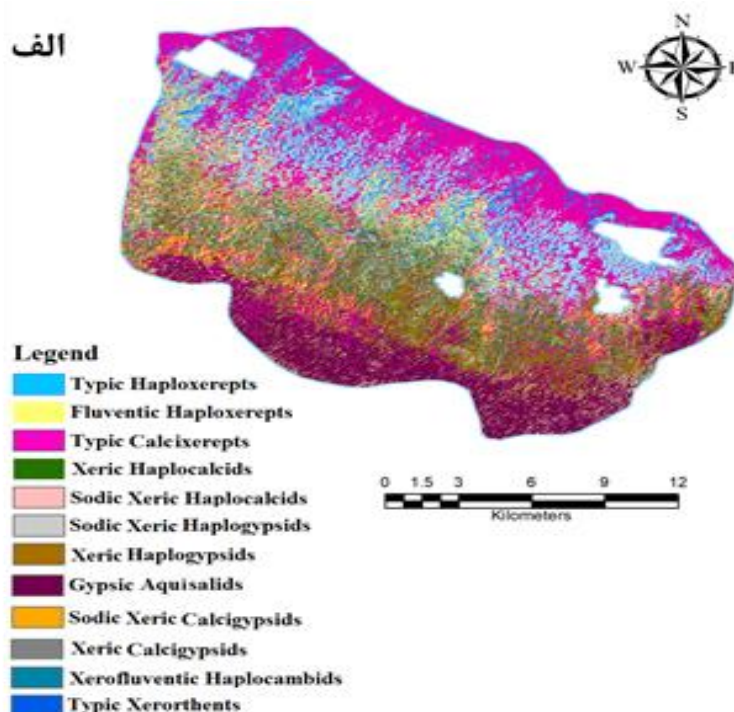
در گروه سوم که غالب خاک‌های آن در مرکز منطقه مورد مطالعه به‌صورت یک توده با توجه به شکل ۱ پراکنده شده‌اند نشان می‌دهد که نیازمند استفاده از پنجره‌های محاسباتی با اندازه متوسط و بزرگ بوده و با افزایش اندازه پنجره شاخص F1-Score در آن افزایش می‌دهند اما با توجه به تعداد اندک مشاهدات این زیرگروه میزان F1-score در حداکثر میزان خود کمتر از ۳۰ درصد است. در مدل RF دو زیرگروه Fluventic Haploxerepts و Typic Calcixerepts (به‌مانند مدل CNN) با افزایش اندازه پنجره محاسباتی روند افزایشی و در اندازه پنجره ۱۵ به حداکثر میزان F1-Score می‌رسد و سپس یک کاهش ملایم را دنبال می‌نمایند. این دو زیرگروه بخش زیادی از نواحی شمالی منطقه با تعداد فراوانی مناسب پروفیل خاک را بر اساس شکل ۱ به خود اختصاص داده و همچنین کلاس‌های خاک دیگر کمتر در داخل محدوده مشاهداتی این زیرگروه‌ها تداخل دارند. به این دلیل با به‌کارگیری بیشتر اطلاعات همسایگی و افزایش اندازه پنجره محاسباتی همچنان دقت پیش‌بینی این زیرگروه‌ها نیز افزایش می‌یابد. بنابراین استفاده از متغیرهای محیطی در مقیاس درشت‌تر می‌تواند به‌خوبی نمایانگر فرآیندهای خاک‌سازی این زیرگروه‌ها باشد. همچنین زیرگروه Xeric Haplogypsids به‌مانند مدل CNN یک الگوی مشابه را نشان می‌دهد به طوری که از اندازه پنجره ۳ تا ۵ روند افزایشی میزان شاخص F1-Score و برخلاف مدل CNN یک‌باره در سایر اندازه پنجره‌های میزان شاخص صفر می‌گردد.



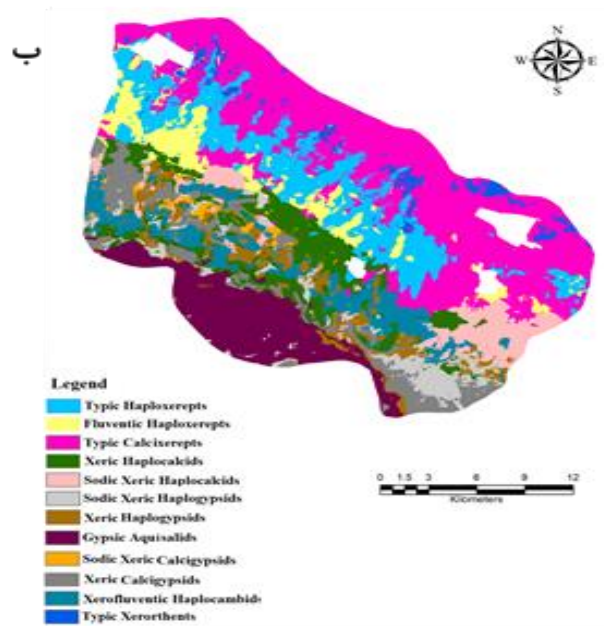
زیرگروه Sodic Xeric Haplogypsids در اندازه پنجره‌های کوچک ۳ و ۵ دارای شاخص F1-Score برابر با صفر و سپس دارای یک‌روند افزایشی با شیب نزولی است. این خاک در نواحی مختلفی از منطقه دارای توزیع پراکنده با فراوانی کمتر و در بیشتر موارد با خاک‌های دیگر دارای تداخل است و تنها در بخش جنوب شرقی منطقه از فراوانی مناسبی برخوردار است که به نظری رشد مدل این روند افزایشی دقت پیش‌بینی را همراه با افزایش اندازه پنجره در این ناحیه مدنظر داشته که زیرگروه از تجمع یکجا برخوردار است. زیرگروه Gypsic Aquisalids دارای یک‌روند افزایشی تند در میزان شاخص F1-Score همراه با افزایش اندازه پنجره برای محاسبه متغیرهای محیطی بوده و برخلاف مدل CNN ترجیح مدل RF بر انتخاب اندازه پنجره‌های بزرگ با توجه به عدم تداخل سایر زیرگروه‌ها برای این زیرگروه و واقع شدن در یک ناحیه جغرافیایی مشخص (اراضی پست) در جهت پیش‌بینی آن می‌باشد.

#### پیش‌بینی رقومی و عدم قطعیت کلاس‌های زیر گروه خاک

نقشه رقومی پیش‌بینی شده دو مدل CNN و RF زیرگروه‌های خاک در اندازه پنجره ۱۵\*۱۵ پیکسل در شکل ۶ ارائه شده است. همان‌طور که در راهنمای نقشه ملاحظه می‌گردد در هر دو مدل ۱۲ کلاس زیرگروه در منطقه مورد مطالعه پیش‌بینی شد. در نقشه حاصل از مدل CNN کلاس‌های Typic Calcixerepts، به دلیل تعداد رخداد مناسب و پیدا نمودن رابطه خاک با متغیرهای محیطی منتخب مخصوصاً میزان بالای شاخص MRVBF که بیانگر عمق کمتر رسوبات و درشتی بافت است که فرصت شستشوی کربنات‌های کلسیم را فراهم نموده و در اعماق تجمع پیدا نموده و منجر به تشکیل این زیرگروه شده است اشاره نمود (Mousavi et al., 2022). همچنین کلاس‌های Typic Haploxerepts در برخی نواحی و Typic xerorthents به درستی در محل رخداد خود پیش‌بینی گردیدند. هر چند کلاس زیرگروه Fluventic Haploxerepts هم در همین ناحیه پیش‌بینی گردیده با این وجود تحت پوشش زیرگروه Typic Haploxerepts قرار گرفته و در نقشه از پراکندگی لکه‌ای برخوردار می‌باشد. پیش‌بینی زیرگروه‌های خاک توسط مدل CNN در نواحی مرکزی دارای پراکنش پراکنده لکه‌ای و در هم است که از دلایل آن می‌توان به تنوع تعداد کلاس‌های زیرگروه و فراوانی کمتر این کلاس‌ها و نیاز بالای مدل‌های یادگیری عمیق به داده بیشتر جهت نقشه خروجی با صحت بالاتر اشاره نمود؛ بنابراین مرز بین کلاس‌های خاک به درستی متمایز نگردیده است.



شکل ۶- الف. نقشه رقومی کلاس‌های زیرگروه خاک در اندازه پیکسل ۱۵ متر برای اندازه پنجره بهینه (۱۵×۱۵) با استفاده از دو مدل (الف) - CNN



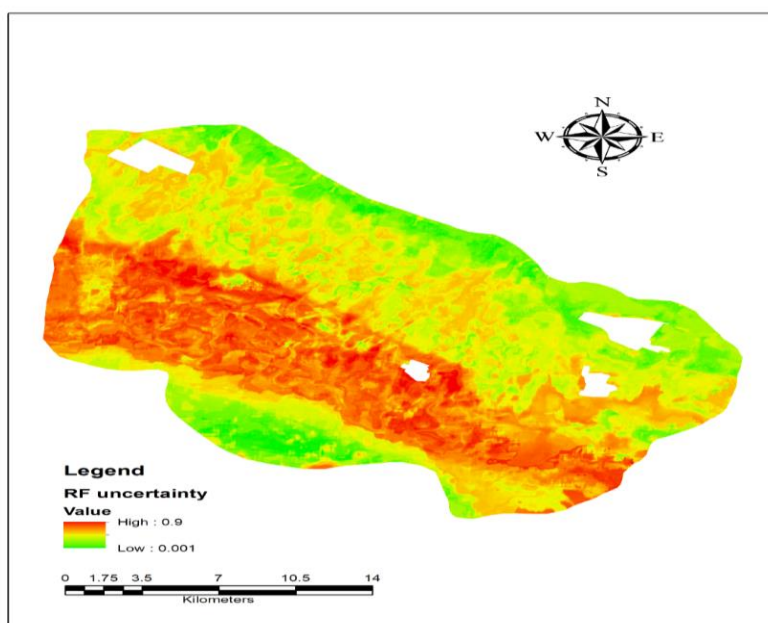
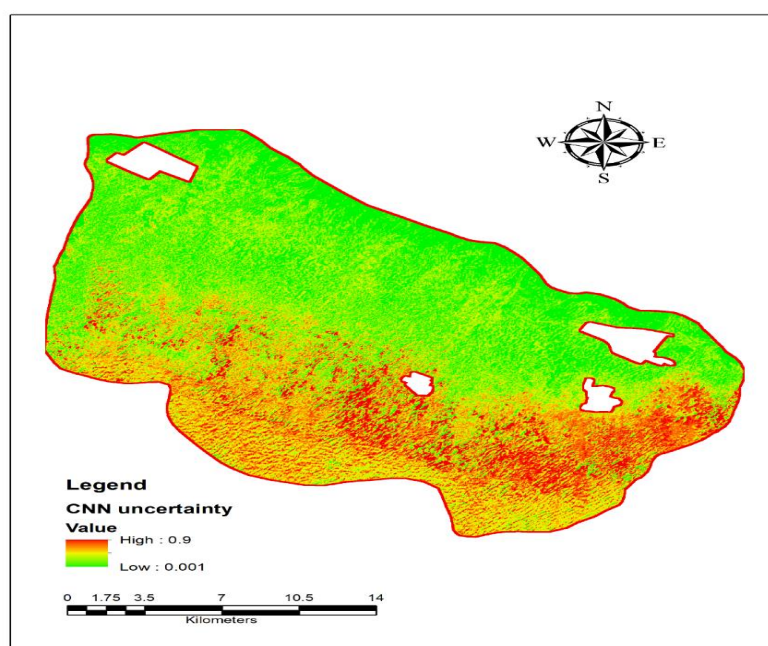
شکل ۶- ب. نقشه رقومی کلاس‌های زیرگروه خاک در اندازه پیکسل ۱۵ متر برای اندازه پنجره بهینه (۱۵×۱۵) با استفاده از دو مدل RF -

در ناحیه جنوب منطقه مدل CNN توانسته از رابطه بین زیرگروه Gypsic Aquisalids و متغیرهای محیطی شاخص SWI (بالاترین میزان خیس) حداقل MRVBF (که بیانگر نواحی با عمق تجمع بیشتر رسوبات ریزبافت است) و هم‌راستایی زیرگروه با نواحی با حداکثر میزان شدت تابش پخشیده بر اساس شاخص Diffuse Insolation استفاده نموده که همگی از پیش‌ران‌های فرآیند خاک‌ساز Salinization این زیرگروه می‌باشند (Nayestani et al., 2022; Mousavi et al., 2020). مرز پیش‌بینی این زیرگروه با وجود تعداد فراوانی کم به‌خوبی متمایز گردیده هرچند که در درون محدوده پیش‌بینی تداخل با تعدادی از زیرگروه‌های دیگر هنوز مشاهده می‌گردد و سبب بالا رفتن عدم قطعیت بر اساس شکل ۷ شده است. این زیرگروه در ناحیه شرقی منطقه مورد مطالعه بیش برآزش داشته است و به جای خاک Sodic Xeric Haplocalcids پیش‌بینی شده است و میزان عدم قطعیت به همین دلیل بالاتر رفته است. مدل RF در پیش‌بینی محل رخداد اکثر زیرگروه‌ها با بالاتر رفتن اندازه پنجره و افزایش اطلاعات همسایگی از متغیرهای محیطی منتخب به‌خوبی توانسته مرز متمایزی برای ۱۲ زیرگروه پیش‌بینی نماید و در استفاده از داده‌های محدود برای کلاس‌های با فراوانی کمتر و تنوع بیشتر در مرکز منطقه مورد مطالعه و ناحیه شرقی برای خاک Sodic Xeric Haplocalcids کارایی بالاتری نشان داده (Rad et al., 2016) و محدوده‌های مشخص‌تری را نسبت به مدل CNN جداسازی نموده است. با این وجود همان‌طور که در نقشه عدم قطعیت آن در شکل ۷ قابل مشاهده است این پیش‌بینی خصوصاً در ناحیه مرکزی با عدم قطعیت بالاتری نسبت به مدل CNN همراه می‌باشد. همچنین در شمال منطقه مورد مطالعه در شرایطی که هر دو مدل با تعداد فراوانی بیشتر از زیرگروه‌های مشخص مواجه هستند، پیش‌بینی مدل CNN نسبت به مدل RF از عدم قطعیت کمتری برخوردار می‌باشد.

## نتیجه‌گیری و پیشنهادها

در این پژوهش تأثیرپذیری زیرگروه‌های مختلف خاک از اندازه پنجره‌های محاسباتی متغیرهای محیطی منتخب در اندازه پیکسل ۱۵ متر با استفاده از دو مدل RF و CNN مورد بررسی قرار گرفت. همچنین از دو رویکرد انتخاب متغیر نظارت نشده VIF و یادگیری ماشین RFE در دو مرحله برای انتخاب مؤثرترین متغیرها استفاده شده و در نهایت نقشه رقومی و عدم قطعیت حاصل از اندازه پنجره بهینه در دو مدل تهیه گردید. بر اساس رویکردهای انتخاب متغیر محیطی شش متغیر ژئومورفومتری DIFF, MRVBF, SH, WE, SWI, DEM به‌عنوان مؤثرترین متغیرها معرفی شد. اندازه پنجره محاسباتی ۱۵×۱۵ پیکسل به‌عنوان پنجره بهینه در دو مدل انتخاب شده که منجر به پیش‌بینی با صحت عمومی ۴۳ و ۵۰ درصد به ترتیب در دو مدل CNN و RF گردید. سه الگوی رفتاری مقیاس با افزایش اندازه پنجره محاسباتی متغیرهای محیطی در پیش‌بینی کلاس‌های زیرگروه به صورت فاقد روند (افزایشی-کاهشی)، اندازه پنجره بهینه کوچک و بزرگ در پنج و

شش زیرگروه به ترتیب در مدل‌های CNN و RF مشاهده و در چندین کلاس بر اساس شاخص F1-Score پیش‌بینی حاصل نشد. مدل CNN در پیش‌بینی زیرگروه‌های با تعداد داده مناسب دارای دقت کافی و در شرایطی که تعداد فراوانی کم و تنوع کلاس‌های خاک مانند مرکز منطقه مطالعاتی بالا بود پیش‌بینی لکه‌ای با مرز نامتمایز از خود نشان داد. هرچند که پیش‌بینی مدل RF در این مناطق دارای مرز مشخصی برای کلاس‌های زیرگروه بوده اما این پیش‌بینی همراه با عدم قطعیت بالاتری می‌باشد. بنابراین بهتر است برای افزایش قدرت پیش‌بینی مدل CNN در این نواحی از رویکردهای افزایش و متعادل‌سازی داده مانند الگوریتم‌های SMOTE<sup>۱</sup> و GAN<sup>۲</sup> استفاده نموده و در جهت جلوگیری از افزایش هزینه ناشی از نمونه‌برداری بیشتر در این نواحی و در مطالعات آتی نیز از معماری شبکه ساخته‌شده در این پژوهش برای پیش‌بینی کلاس فامیل‌های خاک استفاده شود.



شکل ۷. نقشه عدم قطعیت پیش‌بینی کلاس‌های زیرگروه خاک با استفاده از دو مدل (الف) - CNN (ب) - RF



## سیاس گزاری

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.

"هیچ گونه تعارض منافع بین نویسندگان وجود ندارد"

## فهرست منابع

- ایستگاه سینوپتیک قزوین، ۱۳۴۹-۱۳۹۹، سازمان هواشناسی ایران.  
بی همتا؛ محمدرضا، زارع چاهوکی؛ محمد علی. (۱۳۸۹). اصول آمار در علوم مرتع داری. ویرایش سوم. تهران. انتشارات دانشگاه تهران. ۳۰۰ صفحه.  
سازمان زمین شناسی ایران، ۱۳۷۴. نقشه چهارگوش زمین شناسی. شماره ۱۱۱. زمین شناسی.  
موسوی؛ سیدروح اله، سرمیدان؛ فریدون، رحمانی؛ اصغر. (۱۳۹۸). مدل سازی و پیش بینی مکانی کلاس خاک با استفاده از الگوریتم یادگیری رگرسیون درختی توسعه یافته و جنگل های تصادفی در بخشی از اراضی دشت قزوین. تحقیقات آب و خاک ایران، ۵۰ (۱۰)، ۲۵۲۵-۲۵۳۸. doi: 10.22059/ijswr.2019.280905.668198
- Behrens, T., Schmidt, K., Ramirez-Lopez, L., Gallant, J., Zhu, A. X., & Scholten, T. (2014). Hyper-scale digital soil mapping and soil formation analysis. *Geoderma*, 213(1), 578-588. <https://doi.org/10.1016/j.geoderma.2013.07.031>
- Bihamta, M., & Zare-chahooki, M. (2011). *Principles of Statistics in range sciences*. 3<sup>rd</sup> Ed. Tehran, University of Tehran Press: 300. (In Persian).
- Behrens, T., Zhu, A. X., Schmidt, K., & Scholten, T. (2010). Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma*, 155(3-4), 175-185. <https://doi.org/10.1016/j.geoderma.2009.07.010>
- Beucher, A., Rasmussen, C. B., Moeslund, T. B., & Greve, M. H. (2022). Interpretation of convolutional neural networks for acid sulfate soil classification. *Frontiers in Environmental Science*, 679(9), 1-14. <https://doi.org/10.3389/fenvs.2021.809995>
- Brungard, C. W., Boettinger, J. L., Duniway, M. C., Wills, S. A., & Edwards Jr, T. C. (2015). Machine learning for predicting soil classes in three semi-arid landscapes. *Geoderma*, 239(1), 68-83. <https://doi.org/10.1016/j.geoderma.2014.09.019>
- Cavazzi, S., Corstanje, R., Mayr, T., Hannam, J., & Fealy, R. (2013). Are fine resolution digital elevation models always the best choice in digital soil mapping?. *Geoderma*, 195(1), 111-121. <https://doi.org/10.1016/j.geoderma.2012.11.020>
- Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma*, 274(15), 54-67. <https://doi.org/10.1016/j.geoderma.2016.03.025>
- 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.
- Dane, J. H., & Topp, C. G. (Eds.). (2020). *Methods of soil analysis, Part 4: Physical methods* (Vol. 20). John Wiley & Sons.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., & Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27-46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Duan, M., & Zhang, X. (2021). Using remote sensing to identify soil types based on multiscale image texture features. *Computers and Electronics in Agriculture*, 187(1), 106272. <http://dx.doi.org/10.1016/j.compag.2021.106272>
- Esfandiarpour-Boroujeni, I., Shahini-Shamsabadi, M., Shirani, H., Mosleh, Z., Bagheri-Bodaghabadi, M., & Salehi, M. H. (2020). Assessment of different digital soil mapping methods for prediction of soil classes in the Shahrekord plain, Central Iran. *Catena*, 193(1), 104648. <https://doi.org/10.1016/j.catena.2020.104648>.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202(1), 18-27. <https://doi.org/10.1016/j.rse.2017.06.031>.
- Geological Survey of Iran, 1995. *Geological Quadrangle Map*. No111.Geology. (inPersian)
- Gallant, J. C., & Dowling, T. I. (2003). A multiresolution index of valley bottom flatness for mapping

- depositional areas. *Water resources research*, 39(12). <https://doi.org/10.1029/2002WR001426>.
- Heung, B., Hodúl, M., & Schmidt, M. G. (2017). Comparing the use of training data derived from legacy soil pits and soil survey polygons for mapping soil classes. *Geoderma*, 290, 51-68. <https://doi.org/10.1016/j.geoderma.2016.12.001>.
- Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, 265, 62-77. <https://doi.org/10.1016/j.geoderma.2015.11.014>.
- Jamali, A., Mahdianpari, M., Mohammadimanesh, F., & Homayouni, S. (2022). A deep learning framework based on generative adversarial networks and vision transformer for complex wetland classification using limited training samples. *International Journal of Applied Earth Observation and Geoinformation*, 115, 103095. <https://doi.org/10.1016/j.jag.2022.103095>.
- Jafari, A., Khademi, H., Finke, P. A., Van de Wauw, J., & Ayoubi, S. (2014). Spatial prediction of soil great groups by boosted regression trees using a limited point dataset in an arid region, southeastern Iran. *Geoderma*, 232, 148-163. <https://doi.org/10.1016/j.geoderma.2014.04.029>.
- Jiang, Z. D., Owens, P. R., Zhang, C. L., Brye, K. R., Weindorf, D. C., Adhikari, K., & Wang, Q. B. (2021). Towards a dynamic soil survey: Identifying and delineating soil horizons in-situ using deep learning. *Geoderma*, 401, 115341. <https://doi.org/10.1016/j.geoderma.2021.115341>.
- Jensen, J. R. 2015. Introductory digital image processing: a remote sensing perspective (No. Ed. 4). Prentice-Hall Inc.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, & C., Engelhardt, A., Team, R. C. (2020). Package 'caret'. *The R Journal*, 223, 7.
- Kwak, G. H., Park, C. W., Lee, K. D., Na, S. I., Ahn, H. Y., & Park, N. W. (2021). Potential of hybrid CNN-RF model for early crop mapping with limited input data. *Remote Sensing*, 13(9), 1629. <https://doi.org/10.3390/rs13091629>.
- Lin, X., Li, C., Zhang, Y., Su, B., Fan, M., & Wei, H. (2017). Selecting feature subsets based on SVM-RFE and the overlapping ratio with applications in bioinformatics. *Molecules*, 23(1), 52:1-10. <https://doi.org/10.3390/molecules23010052>.
- Maynard, J. J., Salley, S. W., Beaudette, D. E., & Herrick, J. E. (2020). Numerical soil classification supports soil identification by citizen scientists using limited, simple soil observations. *Soil Science Society of America Journal*, 84(5), 1675-1692. <https://doi.org/10.1002/saj2.20119>.
- McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2), 3-52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- Mirakzahi, K., Pahlavan-Rad, M. R., Shahriari, A., & Bameri, A. (2018). Digital soil mapping of deltaic soils: A case of study from Hirmand (Helmand) river delta. *Geoderma*, 313, 233-240. <https://doi.org/10.1016/j.geoderma.2017.10.048>.
- Miller, B. A., Koszinski, S., Wehrhan, M., & Sommer, M. (2015). Impact of multi-scale predictor selection for modeling soil properties. *Geoderma*, 239, 97-106. <https://doi.org/10.1016/j.geoderma.2014.09.018>.
- 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. <https://doi.org/10.22059/ijswr.2019.280905.668198>.
- Mousavi, S. R., Sarmadian, F., Omid, M., & Bogaert, P. (2022). Three-dimensional mapping of soil organic carbon using soil and environmental covariates in an arid and semi-arid region of Iran. *Measurement*, 201, 111706. <https://doi.org/10.1016/j.measurement.2022.111706>.
- Neyestani, M., Sarmadian, F., Jafari, A., Keshavarzi, A., & Sharififar, A. (2021). Digital mapping of soil classes using spatial extrapolation with imbalanced data. *Geoderma Regional*, 26, e00422. <https://doi.org/10.1016/j.geodrs.2021.e00422>.
- Nguyen, T. T., Pham, T. D., Nguyen, C. T., Delfos, J., Archibald, R., Dang, K. B., ... & Ngo, H. H. (2022). A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. *Science of the Total Environment*, 804, 150187. <https://doi.org/10.1016/j.scitotenv.2021.150187>.
- Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., & McBratney, A. B. (2019). Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma*, 352, 251-267. <https://doi.org/10.1016/j.geoderma.2019.06.016>.
- Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning to predict soil properties from



- regional spectral data. *Geoderma Regional*, 16, e00198. <https://doi.org/10.1016/j.geodrs.2018.e00198>.
- Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning for digital soil mapping. *Soil*, 5(1), 79-89. <https://doi.org/10.5194/soil-5-79-2019>.
- Rad, M. R. P., Toomanian, N., Khormali, F., Brungard, C. W., Komaki, C. B., & Bogaert, P. (2014). Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. *Geoderma*, 232, 97-106. <https://doi.org/10.1016/j.geoderma.2014.04.036>.
- Rahmani, A., Sarmadian, F., Mousavi, S. R., & Khamoshi, S. E. (2019). Digital soil mapping using geomorphometric analysis and case-based fuzzy logic approach. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 863-866.
- Qazvin synoptic station, metrological data from 1970-2019. Iranian metrological organization. (inPersian)
- Shi, T., & Xu, H. (2019). Derivation of tasseled cap transformation coefficients for Sentinel-2 MSI at-sensor reflectance data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(10), 4038-4048. <https://doi.org/10.1109/JSTARS.2019.2938388>.
- Shi, J., Yang, L., Zhu, A. X., Qin, C., Liang, P., Zeng, C., & Pei, T. (2018). Machine-Learning Variables at Different Scales vs. Knowledge-based Variables for Mapping Multiple Soil Properties. *Soil Science Society of America Journal*, 82(3), 645-656. <https://doi.org/10.2136/sssaj2017.11.0392>.
- Schoeneberger, P. J., Wysocki, D. A., & Benham, E. C. (Eds.). (2012). *Field book for describing and sampling soils*. Government Printing Office.
- Smith, M. P., Zhu, A. X., Burt, J. E., & Stiles, C. (2006). The effects of DEM resolution and neighborhood size on digital soil survey. *Geoderma*, 137(1-2), 58-69. <https://doi.org/10.1016/j.geoderma.2006.07.002>.
- Sparks, D. L., Page, A. L., Helmke, P. A., & Loeppert, R. H. (Eds.). (2020). *Methods of soil analysis, part 3: Chemical methods* (Vol. 14). John Wiley & Sons.
- Soil Survey Staff. (2014). *Keys to Soil Taxonomy*, 12th ed. USDA-Natural Resources Conservation Service.
- Stumpf, F., Schmidt, K., Goebes, P., Behrens, T., Schönbrodt-Stitt, S., Wadoux, A., & Scholten, T. (2017). Uncertainty-guided sampling to improve digital soil maps. *Catena*, 153, 30-38.
- Taghizadeh-Mehrjardi, R., Mahdianpari, M., Mohammadimanesh, F., Behrens, T., Toomanian, N., Scholten, T., & Schmidt, K. (2020). Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran. *Geoderma*, 376, 114552. <https://doi.org/10.1016/j.geoderma.2020.114552>.
- Taghizadeh-Mehrjardi, R., Minasny, B., Toomanian, N., Zeraatpisheh, M., Amirian-Chakan, A., & Triantafyllis, J. (2019). Digital mapping of soil classes using ensemble of models in Isfahan region, Iran. *Soil Systems*, 3(2), 37. <https://doi.org/10.3390/soilsystems3020037>.
- Taghizadeh-Mehrjardi, R., Nabiollahi, K., Minasny, B., & Triantafyllis, J. (2015). Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. *Geoderma*, 253, 67-77. <https://doi.org/10.1016/j.geoderma.2015.04.008>.
- Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., & Zalidis, G. (2020). Employing a multi-input deep convolutional neural network to derive soil clay content from a synergy of multi-temporal optical and radar imagery data. *Remote Sensing*, 12(9), 1389. <https://doi.org/10.3390/rs12091389>.
- Van Wambeke, A. R. (2000). The Newhall Simulation Model for estimating soil moisture & temperature regimes. Conservation Service: *Department of Crop and Soil Sciences Cornell University*, Ithaca, NY USA. <http://www.bfenvironmental.com/pdfs/nsmt.pdf>.
- Wadoux, A. M. C. 2019. Using deep learning for multivariate mapping of soil with quantified uncertainty. *Geoderma*, 351, 59-70. <https://doi.org/10.1016/j.geoderma.2019.05.012>.
- Yang, L., Cai, Y., Zhang, L., Guo, M., Li, A., & Zhou, C. (2021). A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102428. <https://doi.org/10.1016/j.jag.2021.102428>.
- Yang, J., Wang, X., Wang, R., & Wang, H. (2020). Combination of Convolutional Neural Networks and Recurrent Neural Networks for predicting soil properties using Vis-NIR spectroscopy. *Geoderma*, 380, 114616. <https://doi.org/10.1016/j.geoderma.2020.114616>.
- Yan, Y., Kayem, K., Hao, Y., Shi, Z., Zhang, C., Peng, J., & Li, B. (2022). Mapping the Levels of Soil Salination and Alkalinization by Integrating Machine Learning Methods and Soil-Forming Factors. *Remote Sensing*, 14(13), 3020. <https://doi.org/10.3390/rs14133020>.
- Zhang, J., Tian, H., Wang, P., Tansey, K., Zhang, S., & Li, H. (2022). Improving wheat yield estimates using data augmentation models and remotely sensed biophysical indices within deep neural networks in the Guanzhong Plain, PR China. *Computers and Electronics in Agriculture*, 192, 106616.

<https://doi.org/10.1016/j.compag.2021.106616>.

Zeraatpishah, M., Ayoubi, S., Jafari, A., & Finke, P. (2017). Comparing the efficiency of digital and conventional soil mapping to predict soil types in a semi-arid region in Iran. *Geomorphology*, 285, 186-204.

<https://doi.org/10.1016/j.geomorph.2017.02.015>.

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.