

مقایسه مدل‌های یادگیری ماشینی جهت اولویت‌بندی مناطق مستعد تولید گرد و غبار

- ❖ **سروه داروند؛** دانشجوی کارشناسی ارشد، رشته مدیریت و کنترل بیابان، گروه احیای مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران.
- ❖ **حسن خسروی*؛** دانشیار، گروه احیای مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران.
- ❖ **حمیدرضا کشتکار؛** استادیار، گروه احیای مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران.
- ❖ **غلامرضا زهتابیان؛** استاد، گروه احیای مناطق خشک و کوهستانی، دانشکده منابع طبیعی، دانشگاه تهران، کرج، ایران.
- ❖ **امید رحمتی؛** استادیار، بخش تحقیقات حفاظت خاک و آبخیزداری، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان کردستان، سنندج، ایران.

چکیده

مدل‌سازی یکی از ابزارهای مناسب برای تصمیم‌گیری پدیده‌های محیط‌زیستی می‌باشد که به صورت مدل‌های مفهومی یا روابط ریاضی بیان می‌شوند. هدف از این تحقیق مقایسه مدل‌های یادگیری ماشینی شامل ماشین بردار پشتیبان، درخت طبقه‌بندی و رگرسیون، جنگل تصادفی و مدل آنالیز تشخیص ترکیبی جهت اولویت‌بندی مناطق مستعد گرد و غبار است. جهت تعیین روزهای گرد و غبار از داده‌های ساعتی هواشناسی استان‌های البرز و قزوین و تصاویر ماهواره‌ای مربوط به همان روزها برای دوره ۲۰۰۰ تا ۲۰۱۹ استفاده شد. ۴۲۰ نقطه برداشت گرد و غبار در منطقه شناسایی و نقشه پراکنش آن‌ها تهیه گردید. سپس نقشه‌های عوامل تأثیرگذار بر وقوع گرد و غبار شامل نقشه‌های کاربری اراضی، خاک‌شناسی، شیب، جهت، ارتفاع، پوشش گیاهی، رطوبت سطح توپوگرافیکی، نسبت سطح توپوگرافیکی و زمین‌شناسی تهیه گردید. با استفاده از مدل‌های ذکر شده تأثیر هر یک از عوامل مؤثر گرد و غبار مشخص و نقشه‌های اولویت‌بندی مناطق برداشت گرد و غبار تهیه شد. ارزیابی مدل‌ها با استفاده از منحنی راک صورت گرفت. طبق نتایج حاصل شده عامل ارتفاع در تمامی مدل‌ها نسبت به سایر پارامترهای مورد استفاده در مدل از اهمیت بیشتری برخوردار است. نتایج مدل‌سازی نیز نشان داد مدل‌های جنگل تصادفی (RF) و مدل آنالیز تشخیص ترکیبی (MDA) دارای بیشترین مقادیر صحت (۰/۹۶)، دقت (۰/۹۴)، احتمال آشکارسازی (۰/۹۸) و کمترین نرخ هشدار اشتباه (۰/۰۵۱) نسبت به بقیه مدل‌ها است. عملکرد مدل‌های RF و MDA نسبت به سایر مدل‌ها بهتر بوده و پس از آن‌ها به ترتیب مدل‌های ماشین بردار پشتیبان (SVM) و درخت طبقه‌بندی و رگرسیون (CART) قرار دارند. همچنین در ارزیابی مدل‌ها با استفاده از منحنی مشخصه عملکرد (ROC)، مدل RF به عنوان بهترین مدل انتخاب شد.

کلید واژگان: البرز و قزوین، تصاویر ماهواره‌ای، گرد و غبار، مدل‌های یادگیری ماشینی، منحنی راک.

۱. مقدمه

توفان‌های گرد و غباری که تحت تأثیر مستقیم و غیر مستقیم تغییرات آب و هوایی در حال گسترش می‌باشند، اخیراً جوامع انسانی و سلامت آن‌ها را در معرض تهدید قرار داده‌اند. از نظر آماری، فراوانی و تعداد وقوع پدیده گرد و غبار در مناطق خشک و نیمه‌خشک بیشتر از نقاط دیگر دنیا بوده است [۸]. دلیل اصلی نگرش‌های اخیر در مورد توفان‌های گرد و غبار مربوط به آثار متفاوتی است که این پدیده بر سیستم‌ها و زیر سیستم‌های کره زمین می‌گذارد [۲۳، ۱۴]. از جمله این اثرات می‌توان به جذب و پراکنده نمودن انرژی خورشید بر دمای هوا، تشکیل ابر، فعالیت‌های هم‌رفتی، غلظت دی‌اکسید کربن و گوگرد در اتمسفر، چرخه‌های زمین-شیمی و تکامل خاک و بیابان‌زایی اشاره کرد [۶، ۱۱]. در این میان کشور ایران بیش از پیش طی سال‌های اخیر با پدیده گرد و غبار دست و پنجه نرم کرده است.

اولین و یکی از مهم‌ترین گام‌های مقابله با پدیده گرد و غبار و آثار مخرب آن، شناسایی کانون‌های گرد و غبار، اولویت‌بندی و تشخیص بحرانی‌ترین کانون‌های شکل‌گیری آن است. در واقع بدون اطلاع از محل دقیق کانون‌های شکل‌گیری توفان‌های گرد و غبار، نمی‌توان به مقابله سریع با آن‌ها برخاست [۴]. به دلیل گستردگی محدوده تحت تأثیر این پدیده، شناسایی کانون‌های شکل‌گیری گرد و غبار و همچنین اولویت‌بندی آن‌ها با استفاده از امکانات زمینی و یا مطالعات میدانی، بسیار پرهزینه و زمان‌بر است. از طرف دیگر، شناسایی و به ویژه اولویت‌بندی کانون‌های گرد و غبار در یک مقطع کوتاه از زمان امکان‌پذیر نیست؛ زیرا برای دستیابی به این امر نیاز است تا منطقه در بلند مدت بررسی شود تا بتوان کانون‌ها را از نظر تعداد دفعات وقوع توفان و همچنین شدت هر کدام از رخداد‌های ثبت شده در آن‌ها طبقه‌بندی و اولویت‌بندی نمود.

تصاویر ماهواره‌ای و تکنیک‌های سنجش از دور، بخاطر فراهم آوردن داده‌های به‌هنگام و قابلیت بالای پردازش

تصاویر، کاربرد وسیعی در تمام بخش‌ها از جمله منابع طبیعی دارند [۲۵، ۳]. وسعت منطقه تحت باندهای متفاوت برای پوشش، سنجش، قدرت تفکیک مکانی خوب، پوشش و تکرار زیاد در زمان‌های مختلف از نکاتی است که می‌تواند اطلاعاتی جامع در شناسایی محل تشکیل، چگونگی برداشت ذرات و مسیر حرکت گرد و غبار در اختیار محققین قرار دهد و آنان را در تحت کنترل قرار دادن این پدیده یاری رساند [۲۶، ۱۶]. نکته بسیار مهم این است که تعداد روزهای گرد و غبار رخ داده در سال‌های اخیر در ایران مرتباً در حال افزایش می‌باشد؛ که دو استان البرز و قزوین به دلیل وجود دشت‌های بایر در اطراف آن‌ها که یکی از کانون‌های گرد و غبار داخلی هستند از این قضیه مستثنی نیستند. لذا شناخت منابع تولید گرد و غبار و اولویت‌بندی کانون‌های بحرانی آن یک ضرورت اجتناب‌ناپذیر است.

مدلسازی یکی از ابزارهای مناسب برای تصمیم‌گیری و پیش‌بینی پدیده‌های محیط‌زیستی و طبیعی می‌باشد که اغلب به صورت مدل‌های مفهومی یا روابط ریاضی بیان می‌شوند [۳۰]. برای تعیین منابع اصلی گرد و غبار، ارزیابی و اولویت‌بندی مناطق مستعد پدیده گرد و غبار از مدل‌های یادگیری ماشینی استفاده می‌شود [۱، ۱۳]. یادگیری ماشین، زیرمجموعه‌ای از هوش مصنوعی است. در این تکنیک داده‌های آموزشی به الگوریتم‌های یادگیری ماشین تزریق می‌شوند و این الگوریتم‌ها، وظیفه یادگیری و واکنشی الگوهای مختلف، در داده‌ها را دارند. بعد از به دست آوردن الگوها توسط الگوریتم‌ها (معمولاً یکی از الگوریتم‌ها مورد استفاده قرار می‌گیرد)، یک مدل ساخته می‌شود. در فرآیندهای یادگیری ماشین دو نوع یادگیری مهم وجود دارد: ۱: یادگیری نظارت‌شده ۲: یادگیری نظارت‌نشده [۲۷، ۱۹، ۱۰]. در فرآیندهای یادگیری ماشین، داده‌ها، بسیار اهمیت دارند. در واقع این داده‌ها هستند که به الگوریتم تزریق می‌شوند و الگوریتم از روی آن‌ها یادگیری را انجام می‌دهد [۲۶]. مدل ماشین بردار پشتیبان یکی از مدل‌های ماشینی یادگیری نظارت شده

ایجاد کانون‌های برداشت گرد و غبار در دو مدل نسبت فراوانی و وزن واقعه داشته‌اند و همچنین مدل نسبت فراوانی و وزن واقعه به ترتیب با سطح زیر منحنی ۰/۸۱۸ و ۰/۸۲۵ دقت قابل قبولی دارند.

به دلیل تغییرات کاربری اراضی که در سال‌های اخیر در دو استان البرز و قزوین رخ داده است و همچنین وجود دشت‌های وسیع عاری از پوشش گیاهی این مناطق مستعد توفان‌های گرد و غبار هستند. ضروری است جهت مدیریت بهتر، مناطق مستعد گرد و غبار شناسایی و جهت مقابله با آن راهکارهای مدیریتی در سستی ارائه شود. بنابراین در این تحقیق سعی شده است با مقایسه مدل‌های یادگیری ماشینی نسبت به اولویت‌بندی مناطق مستعد تولید گرد و غبار در این دو استان اقدام شود و پس از تولید نقشه‌های پهنه‌بندی و ارزیابی مدل‌های یادگیری ماشینی بهترین مدل جهت اولویت‌بندی تعیین شود.

۲. روش شناسی

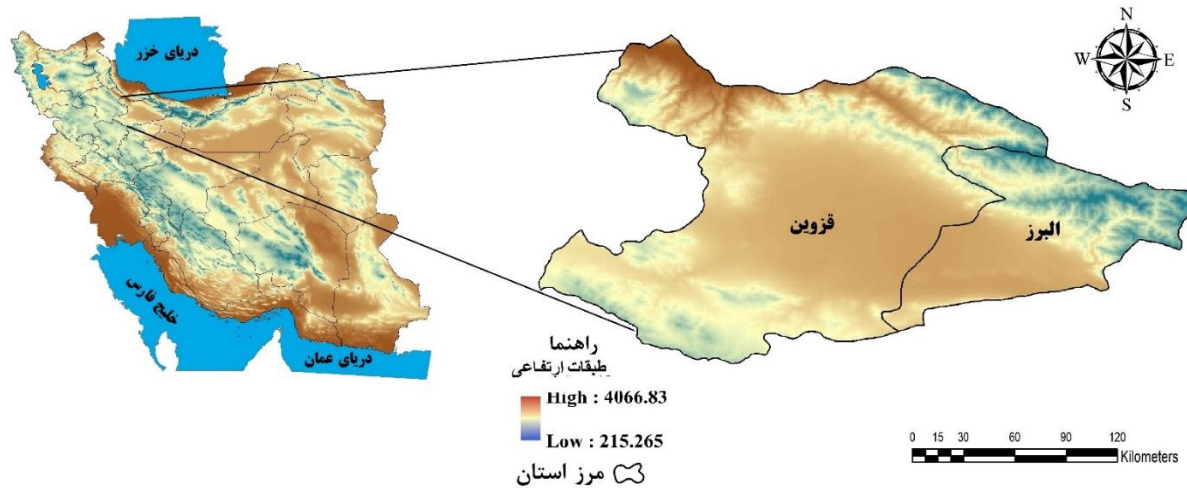
۲.۱. منطقه مورد مطالعه

منطقه مورد مطالعه در این تحقیق استان‌های البرز و قزوین می‌باشد. استان البرز در شمال ایران با وسعت ۵۸۳۳ کیلومتر مربع ۰/۳۱ درصد از مساحت کشور را به خود اختصاص داده است. بین ۳۵ درجه و ۲۸ دقیقه تا ۳۶ درجه و ۳۰ دقیقه عرض شمالی، ۵۰ درجه و ۱۰ دقیقه تا ۵۱ درجه و ۳۰ دقیقه طول شرقی در دامنه رشته کوه‌های البرز مرکزی واقع شده است. این استان از شمال با استان مازندران، از غرب با استان قزوین، از شرق و جنوب شرقی با استان تهران و از جنوب غربی با استان مرکزی هم‌سایه است [۲۱]. استان قزوین که در حوزه مرکزی ایران با مساحتی معادل ۱۵۸۲۱ کیلومتر مربع بین ۴۸ درجه و ۴۵ دقیقه تا ۵۰ درجه و ۵۰ دقیقه طول شرقی و ۳۵ درجه و ۳۷ دقیقه تا ۳۶ درجه و ۴۵ دقیقه عرض شمالی قرار دارد. از شمال به استان‌های مازندران و گیلان، از غرب به

است که جهت طبقه‌بندی تفکیک گروه‌ها به کار می‌رود [۳۲، ۱۵]. به عبارت دیگر، پس از مشخص شدن عوامل محیطی اثرگذار (متغیرهای مستقل) و (متغیر وابسته)، مدل ماشین‌بردار پشتیبان به تجزیه و تحلیل روابط میان آن‌ها پرداخته و داده‌ها را به گروه‌های متمایزی تقسیم می‌کند. یکی از محبوب‌ترین و در عین حال از ساده‌ترین الگوریتم‌های درخت‌های تصمیم، درخت تصمیم CART است که کاربردهای زیادی در طبقه‌بندی و رگرسیون دارد. الگوریتم جنگل تصادفی نوعی الگوریتم است که از درخت‌های تصمیم، برای الگوریتم‌های ساده و ضعیف خود استفاده می‌کند. در الگوریتم جنگل تصادفی از چندین درخت تصمیم استفاده می‌شود. در واقع مجموعه‌ای از درخت‌های تصمیم، با هم یک جنگل را تولید می‌کنند و این جنگل می‌تواند طبقه‌بندی بهتری را (نسبت به یک درخت) اتخاذ نماید. MDA یک جزء توسعه یافته مدل آنالیز تشخیص خطی (LDA) است [۲۰، ۵]. در ایران و جهان مطالعات گوناگونی در این زمینه صورت گرفته است، که می‌توان به موارد زیر اشاره کرد:

محققانی [۲۸] در پژوهشی به شناسایی کانون‌های گرد و غبار به کمک داده‌های دورسنجی پرداختند و به این نتیجه رسیدند که نتایج اعتبار سنجی نشان دهنده دقت بالای مناطق شناسی شده و کاهش معنی‌دار پوشش گیاهی، رطوبت خاک و دمای سطح زمین در محل کانون‌های شناسایی شده در طی دوره زمانی مورد مطالعه است. در حوضه آبریز جازموریان [۲۴] به بررسی آشکار سازی گرد و غبار در حوضه آبریز جازموریان با استفاده تکنیک‌های چند طیفی در تصاویر سنجنده مودیس پرداخته شد. نتایج حاصل از پژوهش نشان داد الگوریتم‌های مورد مطالعه عملکرد مناسبی در شناسایی پیکسل‌های آلوده به گرد و غبار دارند. اخیراً [۴] به بررسی اولویت‌بندی مناطق برداشت گرد و غبار با استفاده از مدل‌های آماری پرداختند نتایج آن‌ها نشان داد که ژئومرفولوژی، شیب و کاربری اراضی بیش‌ترین نقش را در

استان‌های همدان و زنجان، از جنوب به استان مرکزی و از شرق به استان تهران محدود می‌شود [۹].



شکل ۱. موقعیت منطقه مورد مطالعه

برداشت گرد و غبار شامل (BTD2931, BTD3132), NDDI و D_2 می‌باشند. اختلاف دمای درخشندگی پدیده گرد و غبار در طول موج‌های ۱۱ و ۱۲ میکرومتری منفی است و دلیل آن آلبدو بیشتر گرد و غبار در طول موج ۱۲ میکرومتر نسبت به ۱۱ میکرومتر بیان شده است. تفاوت درخشانی ذرات جامد معلق در طول ۸/۵، ۱۱ و ۱۲ میکرومتر، شناسایی پدیده مزبور امکان می‌شود. بدین صورت که با افزایش مقدار گرد و غبار اختلاف $(BT_{8.5}-BT_{11})$ و $(BT_{11}-BT_{12})$ افزایش و مقدار $(BT_{11}-BT_{12})$ کاهش می‌یابد [۳۱]. با استفاده از این روش و اعمال آن بر روی تصاویر موجود، توده گرد و غباری بر روی تصاویر آشکار و در نهایت نقشه پراکنش کانون‌های تولید گرد و غبار تهیه گردید.

در مرحله بعد استفاده از مرور منابع و مطالعات قبلی و شرایط کنونی منطقه مورد مطالعه، مهم‌ترین متغیرهای تأثیر گذار بر وقوع گرد و غبار شناسایی و تهیه گردید. لایه‌های مورد استفاده در این مطالعه شامل: کاربری اراضی،

۲.۲. روش تحقیق

در این پژوهش با استفاده از داده‌های مورد نیاز و تلفیق داده‌ها برای تولید نقشه پراکنش کانون‌های برداشت گرد و غبار نسبت به مقایسه مدل‌های یادگیری ماشینی اقدام شد. بدین جهت از داده‌های روزانه ایستگاه‌های هواشناسی و تصاویر ماهواره‌ای سنجنده مودیس سطح یک ماهواره ترا استفاده شد.

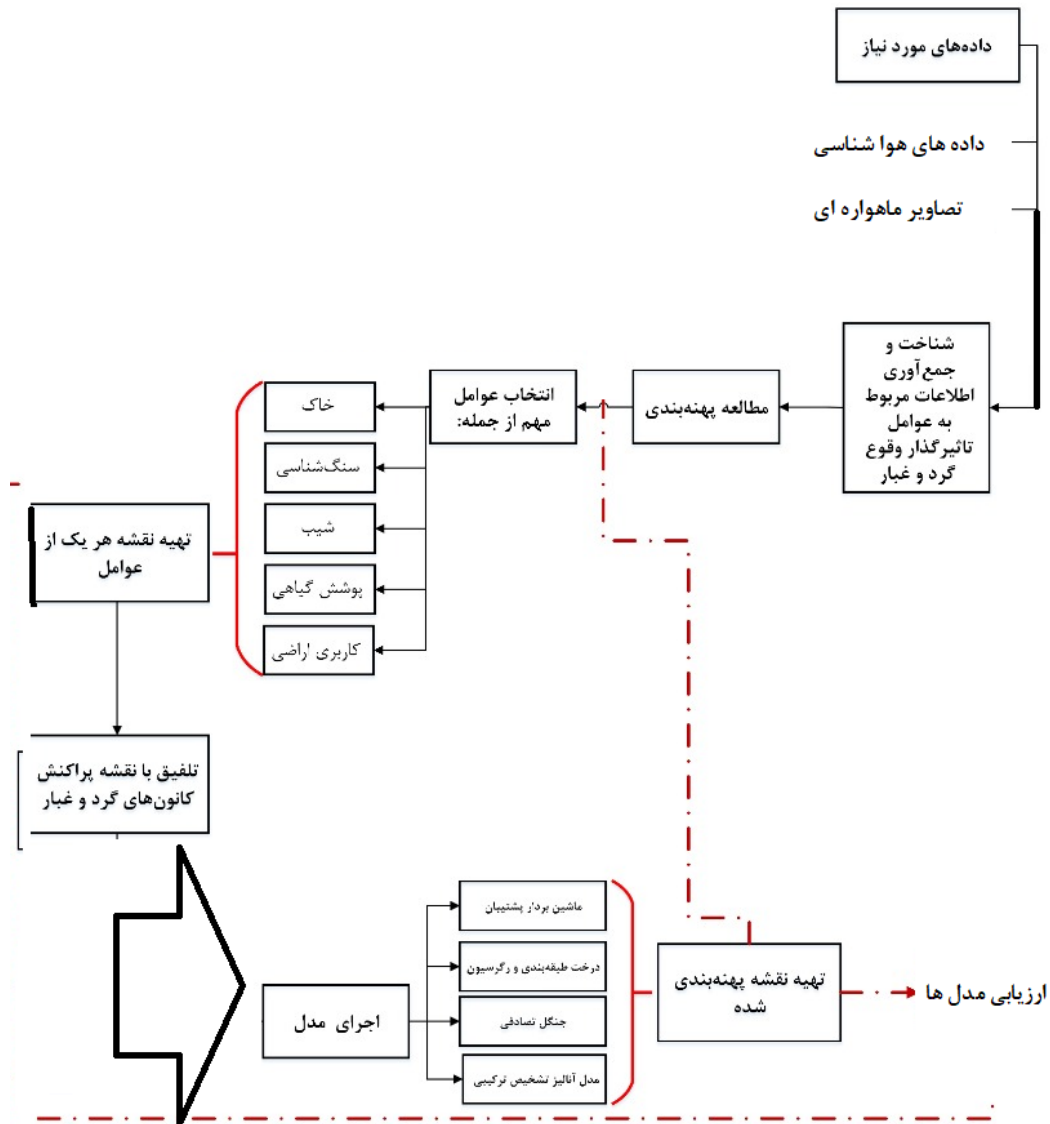
فلوچارت روش تحقیق به صورت شکل (۲) می‌باشد.

مراحل انجام تحقیق به شرح زیر است:

جهت تعیین روزهای گرد و غبار از داده‌های ساعتی هواشناسی شامل دید افقی، سرعت و جهت باد و وضعیت ابرناکی برای ۱۴ ایستگاه هواشناسی در استان‌های البرز و قزوین برای دوره ۲۰۰۰ تا ۲۰۱۹ استفاده شد. سپس تصاویر مرتبط با روزهای تعیین شده گرد و غبار به وسیله داده‌های هواشناسی، از سایت (<http://ladsweb.nasacom.nasa.gov/data>) شناسایی و دانلود گردید. پارامترهای استفاده شده برای ردیابی کانون

تأثیر را بر وقوع گرد و غبار در منطقه دارند انتخاب شده‌اند. تمام لایه‌های اطلاعاتی به صورت نقشه‌های رستری در محیط GIS با پیکسل سائز ۳۰×۳۰ متر تهیه گردیدند.

خاک شناسی، شیب، جهت، ارتفاع، پوشش گیاهی، رطوبت توپوگرافیکی (TWI)، زبری توپوگرافیکی (VRM) و زمین شناسی می‌باشد. به دلیل اینکه این عوامل بیش‌ترین



شکل ۲. فلوجارت روش تحقیق

نقشه شیب، جهت، رطوبت توپوگرافیکی و زبری توپوگرافیکی: به منظور تهیه لایه شیب، جهت، رطوبت توپوگرافیکی و زبری توپوگرافیکی از مدل رقومی ارتفاع با تفکیک مکانی ۳۰ متر استفاده گردید که برای تهیه آن از ابزار Google Earth Engine استفاده شد. سپس بر اساس مدل رقومی ارتفاع در نرم‌افزار ArcGIS 10.7 نسبت به

نقشه لایه‌های اطلاعاتی به صورت زیر تهیه شدند: نقشه کاربری اراضی: از نقشه کاربری اراضی استان‌های البرز و قزوین در ده کلاس شامل اراضی بایر، کشاورزی دیم، مراتع مرغوب، جنگل نامتراکم، کشاورزی-باغ، مراتع متوسط مراتع فقیر، پیکره‌های آبی، اراضی سنگی و اراضی شهری با مقیاس ۱:۲۵۰۰۰ مورد استفاده قرار گرفت.

تصاویر نقاط منشاء تشخیص داده شد. در مدل‌ها نقاط کانون‌های تولید گرد و غبار را به عنوان نقاط یک و نقاطی را که منشاء تولید گرد و غبار محسوب نمی‌شوند به عنوان نقاط صفر در نظر گرفته شدند. نکته‌ای که در این مرحله حائز اهمیت است این است که تعداد نقاطی که به عنوان کانون‌های تولید گرد و غبار و نقاطی که به عنوان عدم تولید گرد و غبار شناسایی شده‌اند بایستی برابر باشد.

۲.۲.۲. استفاده از مدل‌های یادگیری به ماشینی به

منظور تهیه نقشه پهنه‌بندی کانون‌های گرد و غبار

در این مطالعه به منظور تهیه نقشه پهنه‌بندی کانون‌های گرد و غبار از چهار مدل ماشین بردار پشتیبان (SVM)، درخت طبقه‌بندی و رگرسیون (CART)، جنگل تصادفی (RF) و آنالیز تشخیص ترکیبی (MDA) استفاده شد. بر اساس روش طبقه‌بندی Natural Break [۷] در نقشه‌ها خطر تولید گرد و غبار به ۵ طبقه خیلی کم، کم، متوسط، زیاد و خیلی زیاد طبقه‌بندی شدند.

۳.۲.۲. اجرای مدل

قبل از اجرای مدل داده‌ها به صورت تصادفی به دو سری تقسیم‌بندی شد که ۷۰ درصد برای اجرای مدل و ۳۰ درصد برای ارزیابی کارایی مدل استفاده شدند. در مرحله اجرای تمام لایه‌هایی که به عنوان متغیر مستقل شناخته شدند به همراه ۷۰ درصد داده‌های وابسته برای مدل تعریف شد. در این مرحله بر اساس ۳۰ درصد از داده‌هایی که از قبل جدا شد و در مرحله آموزش مورد استفاده قرار نگرفتند، دقت پیش‌بینی مدل ارزیابی شد.

۴.۲.۲. ارزیابی و اعتبارسنجی مدل‌های یادگیری ماشینی

یکی از مراحل مهم بعد از مدل‌سازی، ارزیابی و اعتبارسنجی مدل‌سازی است. منحنی ROC^۱ و

تهیه لایه‌های مذکور اقدام گردید.

نقشه زمین شناسی: نقشه زمین شناسی منطقه از سازمان زمین‌شناسی کشور با مقیاس ۱:۲۵۰۰۰ تهیه گردید. سپس سازندهای تشکیل دهنده بر اساس دوران‌های زمین شناسی تفکیک شد.

نقشه خاک شناسی: به منظور شناختی از خاک منطقه مورد مطالعه و مناطقی که از نظر خاک‌شناسی مستعد فرسایش هستند از نقشه خاک شناسی سازمان جنگل‌ها، مراتع و آبخیزداری کشور با مقیاس ۱:۱۰۰۰۰۰ استفاده شد. NDVI: در این مطالعه از شاخص نرمال شده تفاوت پوشش گیاهی به عنوان معرف پوشش گیاهی منطقه استفاده گردید. برای استخراج شاخص NDVI از تصاویر لندست ۸ در نرم‌افزار ENVI 5.3 بر اساس معادله زیر استفاده شد:

$$NDVI = (B5 - B4) / (B5 + B4) \quad (۱) \text{ رابطه}$$

که در اینجا NDVI شاخص نرمال شده تفاوت پوشش گیاهی و B5، B4 به ترتیب باند طیفی قرمز باند ۴ و باند مادون قرمز نزدیک باند ۵ هستند [۲، ۱۵]. نقشه عوامل مذکور با نقشه پراکنش کانون‌های برداشت تلفیق شدند تا در مراحل بعد جهت انجام مدل‌سازی استفاده شوند.

۱.۲.۲. مدل‌سازی

جهت مدل‌سازی متغیرها به دو دسته مستقل و وابسته تقسیم شدند. نقاطی که کانون گرد و غبار در منطقه بودند تحت عنوان متغیر وابسته یا X در نظر گرفته شدند همچنین متغیرهایی که بر پدیده گرد و غبار اثرگذار هستند و خود از این پدیده تأثیر نمی‌پذیرد تحت عنوان متغیر مستقل یا Y در نظر گرفته شدند. پس از آشکار سازی توده گرد و غبار در تصاویر ماهواره‌ای با استفاده از گلباد روزانه مطابق با زمان تصویر برداری

۳. نتایج

۱,۳. نقاط منشاء گرد و غبار

در این پژوهش ۴۲۰ نقطه برداشت تشخیص داده شد که در شکل (۳) نشان داده شده است. به صورتی که نقاط سبز رنگ نشان دهنده عدم وقوع گرد و غبار در منطقه و نقاط قرمز رنگ روی نقشه نشان دهنده کانون‌های گرد و غبار شناسایی شده به وسیله داده‌های هوا شناسی و سنجش از دور هستند.

در شکل (۴) لایه‌های مختلف اطلاعاتی کاربری اراضی، خاک‌شناسی، شیب، جهت، ارتفاع، پوشش گیاهی، رطوبت توپوگرافیکی، زبری توپوگرافیکی و زمین‌شناسی به عنوان متغیر مستقل نشان داده شده است. پس از تهیه نقشه کاربری اراضی اقدام به محاسبه مساحت کاربری‌های موجود در منطقه شد. بیشترین مساحت مربوط به مراتع متوسط و کمترین مساحت نیز مربوط به اراضی سنگی می‌باشد که به ترتیب ۳۱/۳۷ و ۰/۰۱ درصد سطح منطقه را شامل می‌شوند. با تهیه نقشه خاک شناسی منطقه رده‌بندی آن رده‌های مختلف خاک در ۱۱ رده طبقه‌بندی شد. بیشترین مساحت مربوط به رده آریدیسول و اینسپتی سول به ترتیب با درصد ۳۷/۲۸ و ۲۶/۲۱ است. بررسی‌های زمین‌شناسی منطقه بر اساس دوران‌های زمین‌شناسی نشان داد که سازندهای مربوط به دوران سنوزوئیک دارای بیشترین مساحت در سطح منطقه هستند نشان دهنده این است که سازندهای مربوط به این دوره حساس به فرسایش بوده‌اند که باعث پدیده گرد و غبار در منطقه شده‌اند. همچنین نقشه شیب منطقه به پنج کلاس طبقه‌بندی شد. بیشترین مساحت منطقه مربوط به کلاس شیب ۵-۰ با مساحت ۱۰۱۱۸ کیلومتر مربع و کمترین مساحت مربوط به کلاس شیب بیشتر ۷۲ درصد با مساحت ۱۰۲۵ کیلومتر مربع است.

کاربردهای آن یادگیری ماشین به‌خصوص در شاخه نظارت شده آن، مفید بوده است. در این مطالعه نیز برای ارزیابی و اعتبارسنجی مدل‌سازی‌ها از پنج آماره خطا که شامل صحت، دقت، اریبی، احتمال آشکارسازی و نرخ هشدار اشتباه (Accuracy, Precision, Bias, Probability of Detection (POD) and False Alarm Ratio (FAR)) می‌باشد، استفاده شد:

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \quad \text{رابطه ۲}$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{رابطه ۳}$$

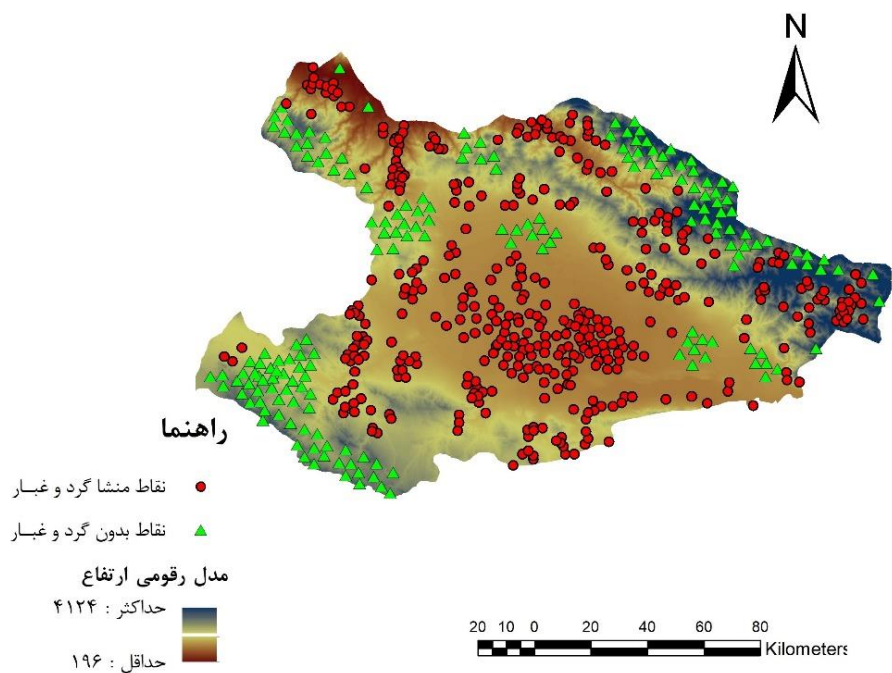
$$\text{Bias} = \frac{TP+FP}{TP+FN} \quad \text{رابطه ۴}$$

$$\text{POD} = \frac{TP}{TP+FN} \quad \text{رابطه ۵}$$

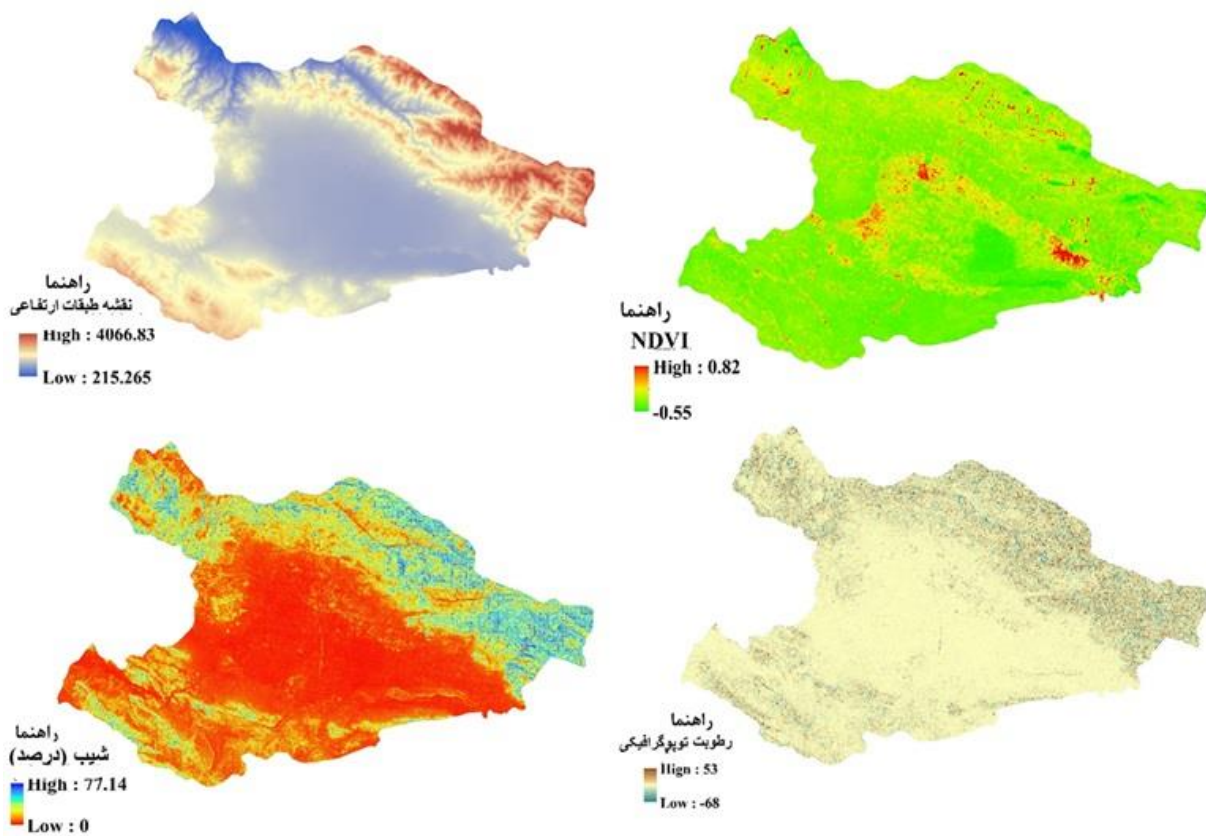
$$\text{FAR} = \frac{FP}{TP+FP} \quad \text{رابطه ۶}$$

که در اینجا برخوردها یا مثبت‌های درست، هشدار نادرست یا مثبت‌های نادرست، خطاها یا منفی‌های نادرست و منفی‌های درست در جدول توافقی به ترتیب TP، FP، FN و TN می‌باشند. مساحت زیر منحنی ROC دارای مقادیر بین ۰/۵ تا ۱ می‌باشد که در بهترین حالت سطح زیر منحنی نزدیک به ۱ است در حالی که مقادیر نزدیک به ۰/۵ نشان دهنده عدم دقت کافی در مدل می‌باشد [۲۳، ۱۲، ۴]. در این مرحله از ۳۰ درصد داده‌هایی که جهت اعتبارسنجی مدل‌ها به صورت تصادفی انتخاب شده بودند، استفاده شد.

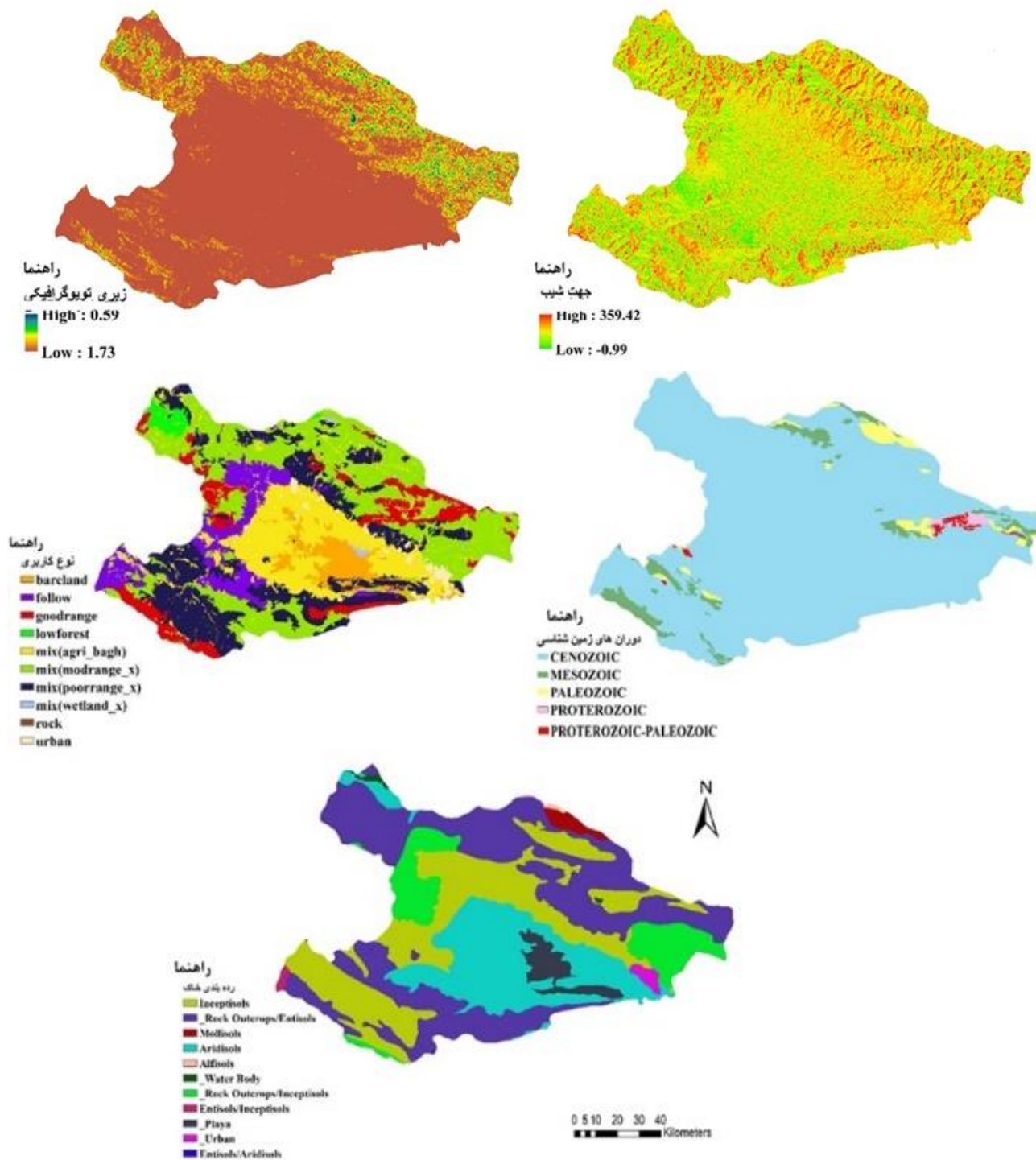
∩ Hits or True positive
 √ False Alarms/False positive
 √ Misses/False negative
 ✕ Correct Negative/True negative



شکل ۳. پراکندگی منشأ گرد و غبار



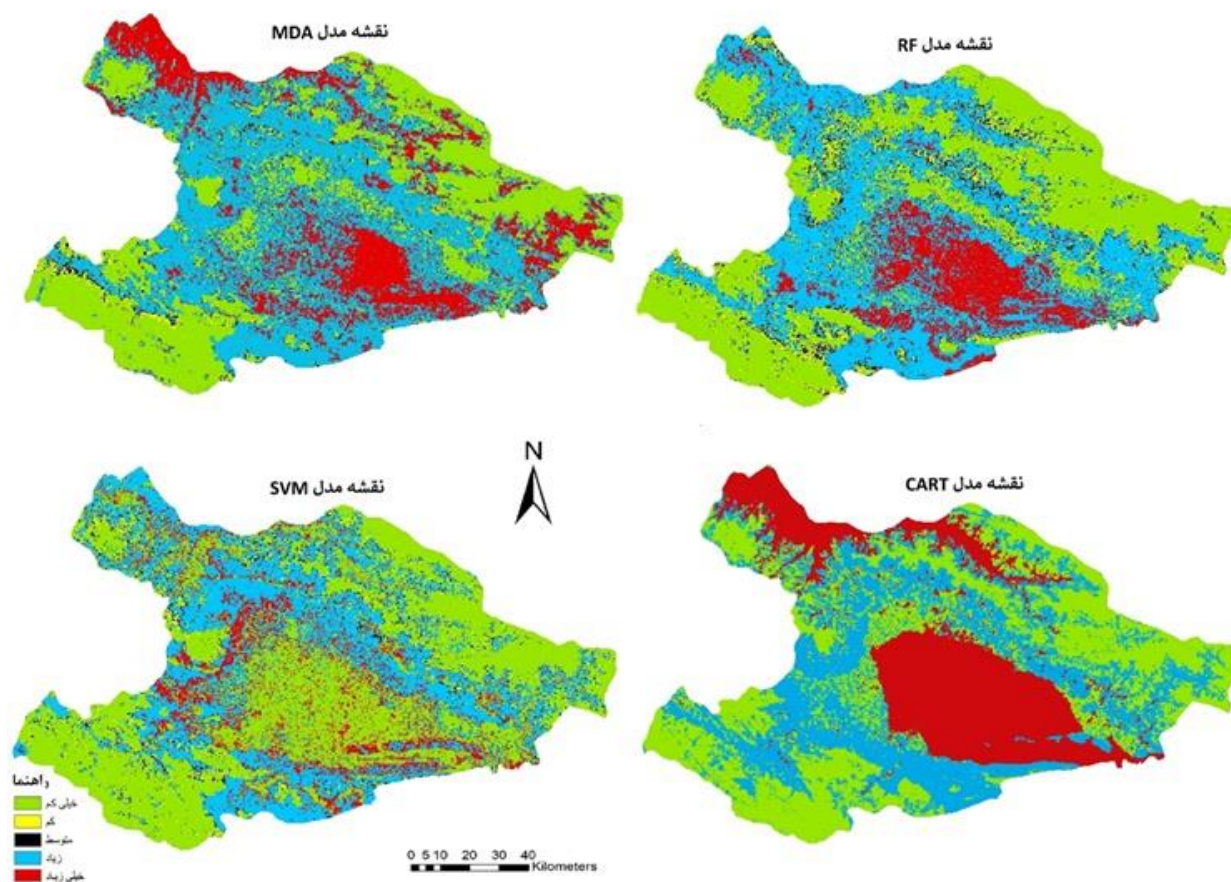
شکل ۴. نقشه لایه‌های اطلاعاتی



ادامه شکل ۴

استفاده از متغیرهای تأثیرگذار در ایجاد گرد و غبار با استفاده از مدل‌های SVM، RF، CART و MDA در شکل (۵) نمایش داده شده‌اند.

۳.۲. مدل‌سازی پهنه‌بندی کانون‌های گرد و غبار
نقشه‌های اولویت‌بندی کانون‌های تولید گرد و غبار با



شکل ۵. نقشه‌های پهنه‌بندی مدل‌های MDA، RF، SVM، CART

مربوط به مناطقی با میزان تولید گرد و غبار زیاد و کمترین مساحت مربوط به مناطقی با میزان تولید گرد و غبار کم است که از این نظر با مدل RF شباهت دارد اما ترتیب سایر مساحت‌ها نسبت به مدل RF متفاوت است. اما در مدل CART بر اساس پارامترهای تعریف شده مناطق مختلف تولید گرد و غبار به سه دسته طبقه‌بندی شد و دو طبقه کم و متوسط از طبقه‌بندی کانون‌های تولید گرد و غبار حذف گردید و بر اساس طبقه‌بندی مدل مناطقی با میزان تولید گرد و غبار خیلی کم بیشترین مساحت و مناطقی با میزان تولید گرد و غبار خیلی زیاد کمترین مساحت را به خود اختصاص دادند. مساحت هر کدام از این طبقات در نقشه‌های تولید شده توسط مدل‌های استفاده شده در جدول (۱) ارائه شده است.

نقشه‌های اولویت‌بندی کانون‌های تولید گرد و غبار نشان می‌دهد که مناطق کم ارتفاع و اراضی بایر که در مرکز و شمال شرق منطقه مورد مطالعه قرار دارد در طبقه زیاد تا بسیار زیاد تولید گرد و غبار قرار می‌گیرند. طبق نقشه‌های به دست آمده در مدل‌سازی پهنه‌بندی کانون‌های گرد و غبار هر چهار مدل دارای طبقه‌بندی متفاوتی می‌باشند. به طوری که در مدل SVM بیشترین مساحت مربوط به مناطقی با میزان تولید گرد و غبار خیلی کم و کمترین مساحت مربوط به مناطقی با میزان تولید گرد و غبار کم است. برای مدل RF بیشترین مساحت مربوط به مناطقی با میزان تولید گرد و غبار زیاد و کمترین مساحت مربوط به مناطقی با میزان تولید گرد و غبار کم است. در مدل MDA نیز بیشترین مساحت

جدول ۱. مساحت (هکتار) کلاس‌های پهنه‌بندی کانون‌های تولید گرد و غبار

| ردیف | کلاس‌های پهنه‌بندی | SVM | RF | CART | MDA |
|------|--------------------|---------|--------|--------|--------|
| ۱ | خیلی کم | ۱۰۱۷۲۷۹ | ۷۶۳۳۶۴ | ۸۰۵۳۸۹ | ۶۲۶۶۹۱ |
| ۲ | کم | ۶۷۴۷۲ | ۱۰۹۴۲۲ | - | ۵۶۲۴۷ |
| ۳ | متوسط | ۶۸۰۶۹ | ۱۲۲۸۹۰ | - | ۶۴۱۶۳ |
| ۴ | زیاد | ۶۹۳۴۳۹ | ۸۳۹۹۶۶ | ۷۲۲۴۰۴ | ۹۵۰۵۵۸ |
| ۵ | خیلی زیاد | ۲۲۷۲۳۹ | ۲۳۸۱۱۶ | ۵۴۵۷۶۵ | ۳۷۳۹۹۲ |

اما در مقدار اریبی^۵ (۱/۰۳۵)، نسبت به مدل SVM کمتر می‌باشند. با توجه به این جدول، عملکرد مدل‌های RF و MDA نسبت به سایر مدل‌ها بهتر بوده و پس از آن‌ها به ترتیب مدل‌های SVM و CART قرار دارند.

۴.۳. آنالیز حساسیت

به منظور آنالیز حساسیت در این پژوهش ۹ پارامتر با استفاده از آزمون جک‌نایف در مدل‌سازی پهنه‌بندی کانون‌های گرد و غبار مورد بررسی قرار گرفت. به منظور کاهش مقدار تخمین‌گر اریبی که هنگام استفاده از چارچوب‌های پیچیده به وجود می‌آید و همچنین به خاطر سادگی، آزمون جک‌نایف از یک مشتق جزئی بهره می‌برد که روشی بسیار مناسب برای آنالیز حساسیت است. شکل‌های (۷) اهمیت نسبی پارامترهای مورد استفاده در مدل‌سازی پهنه‌بندی کانون‌های گرد و غبار برای هر یک از مدل‌های مورد استفاده در این پژوهش را نشان داده شده است.

۳.۳. نتایج اعتبارسنجی نقشه‌های پهنه‌بندی

کانون‌های گرد و غبار

به منظور اعتبارسنجی شبیه‌سازی مدل‌ها مقدار AUC در منحنی ROC و پنج آماره خطا برای هر مدل محاسبه شد. منحنی ROC صحت شبیه‌سازی هر مدل را با رسم مقادیر شبیه‌سازی در مقابل مقادیر مشاهداتی ارزیابی می‌کند. نتایج منحنی ROC نشان می‌دهد که مدل RF دارای صحت پیش‌بینی بیشتری (۰/۸۹) نسبت به سایر مدل‌ها می‌باشد. به دنبال آن مدل‌های MDA، CART و SVM به ترتیب با مقادیر AUC ۰/۸۷، ۰/۸۳ و ۰/۸۱ قرار دارند. شکل (۶) نتایج منحنی ROC را برای مدل‌های مورد استفاده نشان می‌دهد. برای اعتبارسنجی مدل‌ها در منحنی ROC مقادیر AUC که در فرآیند آزمون مدل به دست می‌آید، مد نظر قرار می‌گیرد.

طبق مقادیر آماره‌های خطا در جدول (۲) مدل‌های RF و MDA دارای بیشترین مقادیر صحت^۱ (۰/۹۶)، دقت^۲ (۰/۹۴)، احتمال آشکارسازی^۳ (۰/۹۸) و کمترین نرخ هشدار اشتباه^۴ (۰/۰۵۱) نسبت به بقیه مدل‌ها است؛

^۱ Accuracy

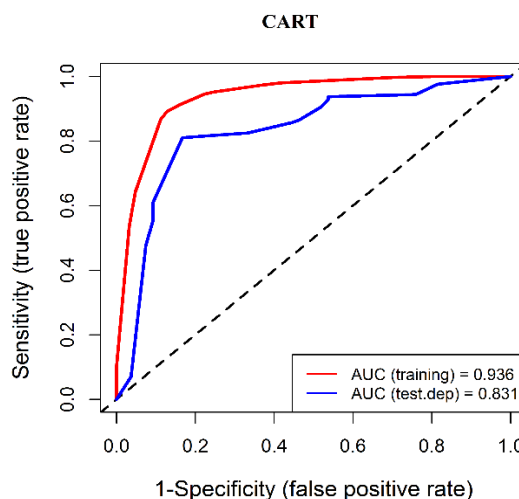
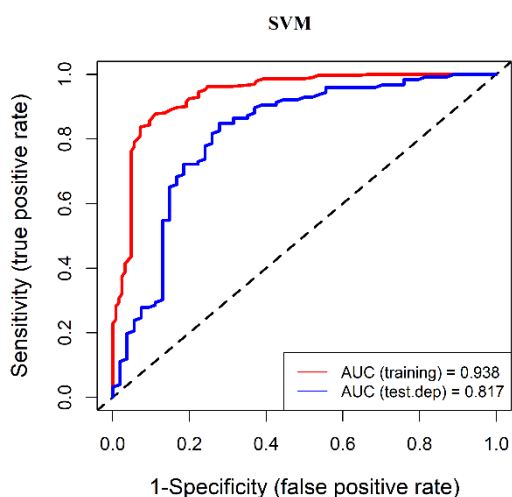
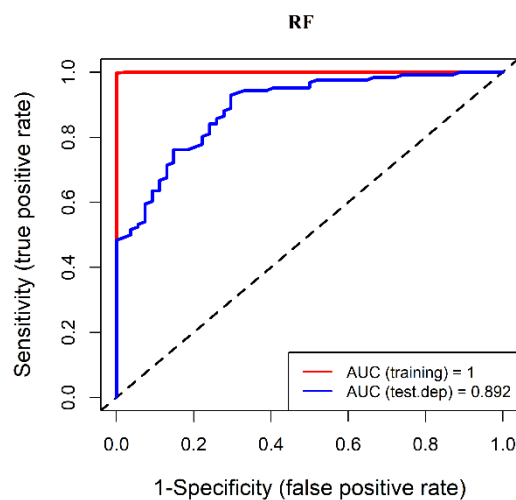
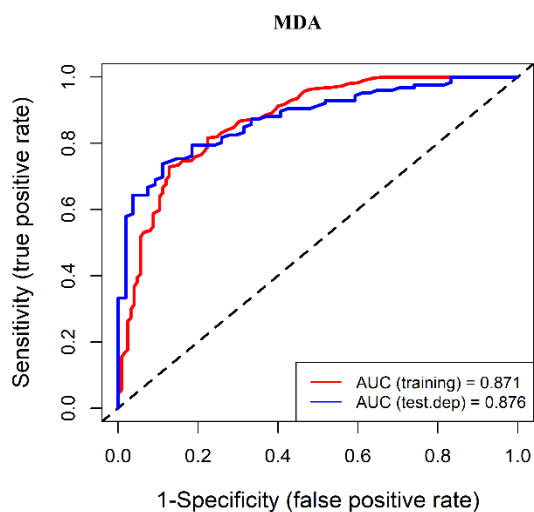
^۲ Precision

^۳ Probability of Detection (POD)

^۴ False Alarm Ratio (FAR)

^۵ Bias

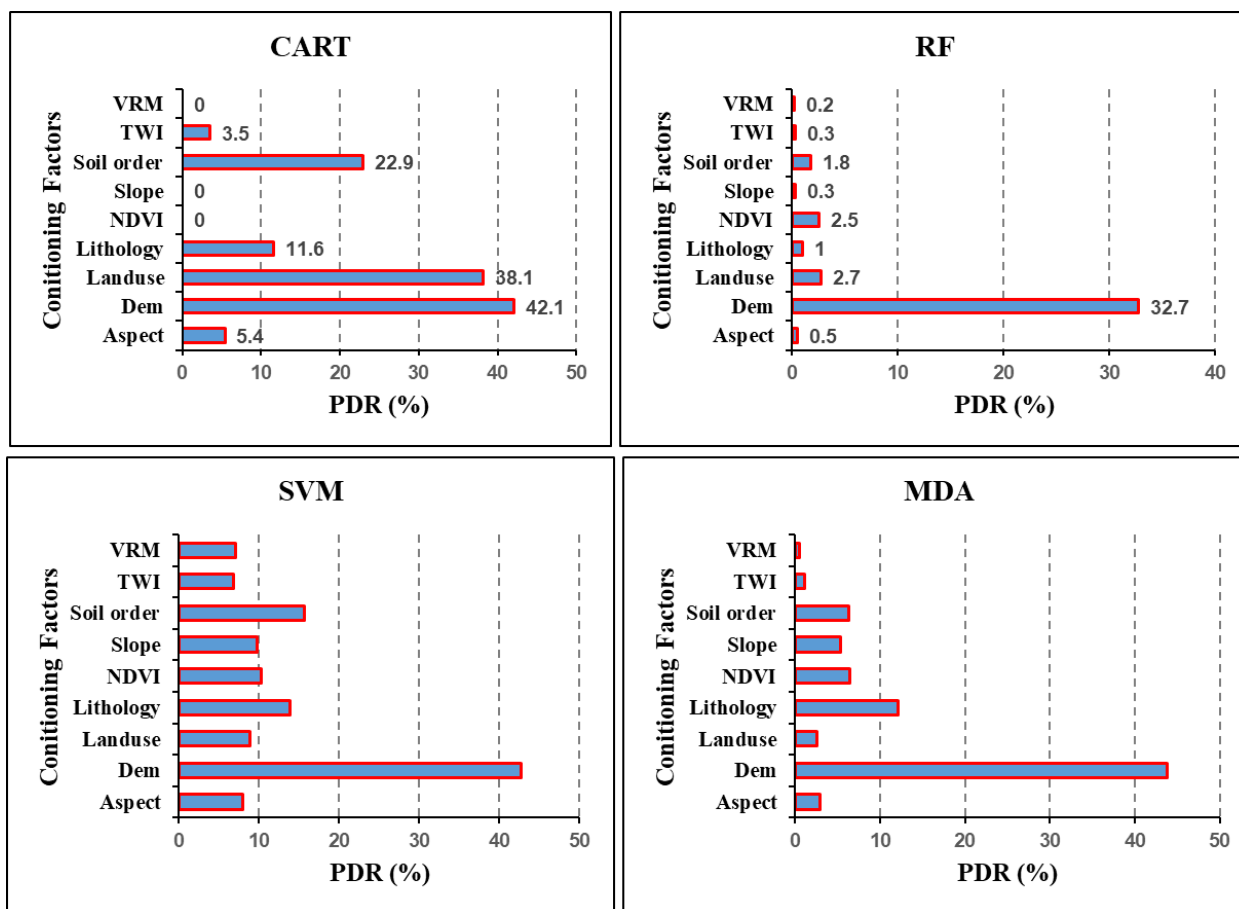
^۶ Partial - derivative



شکل ۶. مقدار سطح زیر منحنی (AUC) در منحنی مشخصه عملکرد سیستم (ROC) برای نقشه‌های پهنه‌بندی کانون‌های گرد و غبار

جدول ۲. آماره‌های خطا در عملکرد مدل‌ها جهت تهیه نقشه‌های پهنه‌بندی کانون گرد و غبار

| ردیف | آماره‌های خطا | SVM | RF | CART | MDA |
|------|---------------|-------|-------|------|-------|
| ۱ | Accuracy | ۰/۷۹۵ | ۰/۹۶۳ | ۰/۶۵ | ۰/۹۶۳ |
| ۲ | Precision | ۰/۷۴۵ | ۰/۹۴۸ | ۰/۶۷ | ۰/۹۴۸ |
| ۳ | Bias | ۱/۲ | ۱/۰۳۵ | ۰/۸۵ | ۱/۰۳۵ |
| ۴ | POD | ۰/۸۹ | ۰/۹۸۱ | ۰/۵۷ | ۰/۹۸۱ |
| ۵ | FAR | ۰/۲۵ | ۰/۰۵۱ | ۰/۳۲ | ۰/۰۵۱ |



شکل ۷. نتایج آنالیز حساسیت آزمون جک‌نایف در مدل RF, MDA, CART, SVM

۴. بحث و نتیجه‌گیری

در این تحقیق به مقایسه مدل‌های یادگیری ماشینی جهت اولویت‌بندی مناطق مستعد تولید گرد و غبار پرداخته شد. جهت شناسایی کانون‌های گرد و غبار از تصاویر ماهواره‌ای MODIS برای سالهای ۲۰۲۰-۲۰۰۵ استفاده شد و نقشه پراکنش کانون‌های گرد و غبار تهیه گردید. سپس لایه‌های اطلاعاتی مربوط به مهم‌ترین عوامل مؤثر در وقوع گرد و غبار شامل کاربری اراضی، خاک‌شناسی، شیب، جهت، ارتفاع، پوشش گیاهی، رطوبت سطح توپوگرافیکی، نسبت سطح توپوگرافیکی و زمین‌شناسی تهیه شد. با اعمال مدل‌های RF، MDA، CART و SVM عوامل مؤثر بر وقوع گرد و غبار رده‌بندی و نقشه‌های نهایی پهنه‌بندی در ۵ کلاس حاصل شد.

طبق نتایج حاصل شده عامل ارتفاع در تمامی مدل‌ها نسبت به سایر پارامترهای مورد استفاده در مدل از اهمیت بیشتری برخوردار است و در مدل‌های RF، MDA، CART و SVM سهم هر کدام از عامل‌های تأثیرگذار بر گرد و غبار در نمودار آنالیز حساسیت با PDR نشان داده شده است به ترتیب ۳۲/۷، ۴۳/۸، ۴۲/۱ و ۴۲/۷ درصد می‌باشد. همچنین در مدل RF به عنوان بهترین مدل بر اساس اعتبارسنجی صورت گرفته پارامترهای کاربری اراضی و NDVI با مقدار PDR ۲/۷ و ۲/۵ درصد پس از عامل ارتفاع از اهمیت بیشتری نسبت به سایر عوامل برخوردارند. در مدل MDA نیز پس از عامل ارتفاع، پارامترهای زمین‌شناسی و NDVI با مقدار آماره PDR ۱۲/۱ و ۶/۴ درصد بیشترین اهمیت را دارند، در مدل

منفرد، نتیجه بهتری به دست می‌دهد [۲۹]. در جنگل تصادفی از چندین درخت تصمیم استفاده شده و واریانس کمتری در طبقه‌بندی تولید می‌شود. مدل جنگل تصادفی یک مجموعه داده را به صورت تصادفی بین طبقه‌بندها توزیع کرده و با میانگین‌گیری از نتایج طبقه‌بندها، طبقه‌بندی نهایی را ایجاد می‌کند. با توجه به نتایج آنالیز حساسیت در این مطالعه در میان متغیرهای توصیفی، ارتفاع اهمیت زیاد و قابل توجهی نسبت به بقیه پارامترها دارد. این بدین معنی است که در بین مجموعه ویژگی‌ها یک ویژگی قوی وجود دارد که می‌تواند باعث بیش‌برازش طبقه‌بندی شود. از آنجایی یادگیری ماشینی در زمینه‌های دیگر استفاده شده است و کمتر به اولویت‌بندی کانون‌های گرد و غبار با استفاده از این مدل‌های پرداخته شده است. پیشنهاد می‌شود از این مدل‌ها به‌خصوص مدل جنگل تصادفی که به عنوان بهترین مدل انتخاب شد، جهت اولویت‌بندی کانون‌های گرد و غبار استفاده شود.

SVM پارامترهای رده خاک و زمین‌شناسی با آماره PDR ۱۵/۷ و ۱۳/۸ پس از عامل ارتفاع از مهم‌ترین عوامل تأثیرگذار بر نتایج مدل می‌باشند و در مدل CART پس از عامل ارتفاع دو عامل کاربری اراضی و رده‌های خاک بیشترین اهمیت را در مدل‌سازی دارند.

همچنین نتایج حاصل از مقایسه و اعتبارسنجی مدل‌های اولویت‌بندی مناطق مستعد تولید گرد و غبار با استفاده از منحنی راک، آماره‌های خطا در هر مدل نشان داد که نقشه‌های تولیدی توسط مدل RF نتیجه بهتری نسبت به سایر مدل‌ها دارد. این نتایج با یافته‌های [۱۸، ۳۳، ۷، ۱۵] که در مطالعات خود عملکرد مدل جنگل تصادفی را بهتر از سایر مدل‌ها معرفی می‌کنند مطابقت دارد. این به این علت است که درخت تصمیم با استفاده از چندین الگوریتم طبقه‌بندی، نتایج بهتری نسبت به یک الگوریتم طبقه‌بندی منفرد دارد و در مقابل مدلی مانند RF به عنوان یک مدل ترکیبی، از چندین درخت تصمیم استفاده می‌کند و در نتیجه نسبت به یک درخت تصمیم

References

- [1] Abdolshahnejad, M., Khosravi, H., Nazari Samani, A. A., Zehtabian, G. R. and Alambaigi, A. (2020). Determining the Conceptual Framework of Dust Risk Based on Evaluating Resilience (Case Study: Southwest of Iran). *Strategic Research Journal of Agricultural Sciences and Natural Resources*, 5(1), 33-44.
- [2] Alamdarloo, E. H., Behrang Manesh, M. and Khosravi, H. (2018). Probability assessment of vegetation vulnerability to drought based on remote sensing data. *Environmental monitoring and assessment*, 190(12), 1-11.
- [3] Albugami, S., Palmer, S., Meersmans, J. and Waive, T. (2018). Evaluating MODIS dust-detection indices over the Arabian Peninsula. *Remote Sensing*, 10(12), 1993.
- [4] Boroughani, M. and Pourhashemi, S. (2019). Susceptibility Zoning of Dust Source Areas by Data Mining Methods over Khorasan Razavi Province. *Environmental Erosion Research Journal*, 9(3), 1-22.
- [5] Breiman L. (2001), Random forests. *Machine Learning*, 45 (1): pp.5-32
- [6] Cao, H., Amiraslani, F., Liu, J. and Zhou, N. (2015). Identification of dust storm source areas in West Asia using multiple environmental datasets. *Science of the Total Environment*, 5: 224-235.
- [7] Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S. and Ghamisi, P. (2020). Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. *Science of The Total Environment*, 701, 134474.
- [8] Damizadeh, M., Mahdavi, R., Noroozi, A. A., Hollisaz, A. and Gholami, H. (2021). Dust Storm Analysis and Detection in Hormozgan Province. *Watershed Engineering and Management*, 13(1), 111-124.

- [9] Darvishi, A., Mobarghaee Dinan, N., Barghjelveh, S., & Yousefi, M. (2020). Assessment and Spatial Planning of Landscape Ecological Connectivity for Biodiversity Management (Case Study: Qazvin Province). *Iranian Journal of Applied Ecology*, 9(1), 15-29.
- [10] Ebrahimi-Khusfi, Z., Taghizadeh-Mehrjardi, R. and Mirakbari, M. (2021). Evaluation of machine learning models for predicting the temporal variations of dust storm index in arid regions of Iran. *Atmospheric Pollution Research*, 12(1), 134-147.
- [11] Faridi, S., Rahmani, S., Hashemi, N., Ghobadian, S. and Zokaie, M. S. (2021). The Economic Effects of Dust Storm. *Journal of Health*, 11(5), 699-713.
- [12] Floyd, K.W., and Gill, T.E. (2011). The association of land covers with aeolian sediment production at Jornada Basin, New Mexico, USA. *Aeolian Research*, 3, 55-66.
- [13] Gholami, H., Mohamadifar, A., Sorooshian, A. and Jansen, J. D. (2020). Machine-learning algorithms for predicting land susceptibility to dust emissions: The case of the Jazmuriyan Basin, Iran. *Atmospheric Pollution Research*, 11(8), ۱۳۰۳-۱۳۱۵.
- [14] Goudie, A. (2014). Review Desert dust and human health disorders. *Environment International*, 3: 101-113.
- [15] Grinand, C., Vieilledent, G., Razafimbelo, T., Rakotoarijaona, J. R., Nourtier, M. and Bernoux, M. (2019). Landscape-scale spatial modelling of deforestation, land degradation and regeneration using machine learning tools. *Land Degradation & Development*.
- [16] Hosseini, C. F., Farokhkhani, F. A. and Amerkhan, H. (2019). Difference Vegetation Index (NDVI), land surface temperature (LST) and normalized moisture (NDMI) indices.
- [17] Isazadeh, M., Biazar, S., Ashrafzadeh, A. and Khanjani, R. (2019). Estimation of aquifer qualitative parameters in Guilans plain using gamma test and support vector machine and artificial neural network models. *Journal of Environmental Science and Technology*, 21(2), 1-21.
- [18] Jafari, M., Zehtabian, G., Ahmadi, H., Mesbahzadeh, T. and Norouzi, A. A. (2019). Detection of dust storm paths using numerical models and satellite images (case study: Isfahan province). *Iranian Journal of Range and Desert Research*, 26(1).
- [19] Lee, J., Shi, Y. R., Cai, C., Ciren, P., Wang, J., Gangopadhyay, A. and Zhang, Z. (2021). Machine Learning Based Algorithms for Global Dust Aerosol Detection from Satellite Images: Inter-Comparisons and Evaluation. *Remote Sensing*, 13(3), 456.
- [20] Moradi, E., Abdolshanejad, M., Borji, M., Ghohestani, G., da Silva, A. M., Khosravi, H. and Cerda, A. (2021). Machine learning approach to predict susceptible growth regions of *Moringa peregrina* (Forssk). *Ecological Informatics*, 101267.
- [21] Moradi, S., Yousefi, H., Noorollahi, Y. and Rosso, D. (2020). Multi-criteria decision support system for wind farm site selection and sensitivity analysis: Case study of Alborz Province, Iran. *Energy Strategy Reviews*, 29, 100478.
- [22] Nandi, A., and Shakoor, A. (2009). A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Engineering Geology*, 110, 11-20.
- [23] Nazari, S., Kermani, M., Fazlzadeh, M., Matboo, S.A. and Yari, A.R. (2016). The origins and sources of dust particles, their effects on environment and health, and control strategies: a review. *J. Air Pollut. Health* 1 (2), 137-152.
- [24] Qaderi Nasab, F. and Rahnama, M. B. (2018). Detection of dust storms in Jazmoriyan drainage basin using multispectral techniques and MODIS image. *Physical Geography Research Quarterly*, 50(3), 545-562.
- [25] Rahimi, M., Damavandi, A. A. and Jafarian, V. (2014). Investigating remote sensing applications in evaluating and monitoring land degradation and desertification. *Scientific-Research Quarterly of Geographical Data (SEPEHR)*, 22(88), 115-128.

- [26] Rahmati, O., Mohammadi, F., Ghiasi, S. S., Tiefenbacher, J., Moghaddam, D. D., Coulon, F. and Bui, D. T. (2020). Identifying sources of dust aerosol using a new framework based on remote sensing and modelling. *Science of The Total Environment*, 737, 139508.
- [27] Rahmati, O., Panahi, M., Ghiasi, S. S., Deo, R. C., Tiefenbacher, J. P., Pradhan, B. and Bui, D. T. (2020). Hybridized neural fuzzy ensembles for dust source modeling and prediction. *Atmospheric Environment*, 224, 117320.
- [28] Rayegani, B., Barati, S., Goshtasb, H., Gachpaz, S., Ramezani, J. and Sarkheil, H. (2020). Sand and dust storm sources identification: A remote sensing approach. *Ecological Indicators*, 112, 106099.
- [29] Seni, G. and Elder, J. F. (2010). Ensemble methods in data mining: improving accuracy through combining predictions. *Synthesis lectures on data mining and knowledge discovery*, 2(1), 1-126.
- [30] Sobhani, B., Safarian Zengir, V. and Faizollahzadeh, S. (2020). Modeling and prediction of dust in western Iran. *Physical Geography Research Quarterly*, 52(1), 17-35.
- [31] Taghavi, F., Oulad, E. and Ackerman, S. A. (2017). Enhancement and identification of dust events in the south-west region of Iran using satellite observations. *Journal of Earth System Science*, 126(2), 28.
- [32] Wang, Y., Stein, A. F., Draxler, R. R., Jesús, D. and Zhang, X. (2011). Global sand and dust storms in 2008: Observation and HYSPLIT model verification. *Atmospheric environment*, 45(35), 6368-6381.
- [33] Zhao, G., Pang, B., Xu, Z., Yue, J. and Tu, T. (2018). Mapping flood susceptibility in mountainous areas on a national scale in China. *Science of The Total Environment*, 615, 1133-1142.