

پهنه‌بندی حساسیت مناطق برداشت گرد و غبار با روش‌های داده کاوی در استان خراسان رضوی

مهدی بروغنی*: استادیار مرکز پژوهشی علوم جغرافیا و مطالعات اجتماعی، دانشگاه حکیم سبزواری

سیما پورهاشمی: دکتری ژئومورفولوژی دانشکده‌ی جغرافیا و علوم محیطی، دانشگاه حکیم سبزواری

تاریخ پذیرش: ۱۳۹۸/۱۰/۲۴

تاریخچه مقاله (تاریخ دریافت: ۱۳۹۸/۰۶/۲۵)

چکیده

توفان گرد و غبار از خطرات طبیعی است که بر شرایط جوی، سلامت انسان و اکوسیستم اثر می‌گذارد. استان خراسان رضوی دارای آب و هوای خشک و نیمه‌خشک است که سالانه توفان‌های گرد و غبار متعددی در سطح آن اتفاق می‌افتد. هدف از این تحقیق، ارزیابی عوامل مؤثر بر وقوع گرد و غبار و تهیه نقشه حساسیت آن در استان خراسان رضوی با استفاده از مدل‌های رگرسیون لجستیک و جنگل تصادفی است. در ابتدا ۶۵ کانون برداشت در منطقه‌ی مورد مطالعه و برای دوره‌ی زمانی ۲۰۱۶-۲۰۰۵ شناسایی شد. از مجموع کانون‌های شناسایی شده، ۷۰ درصد (۴۶ کانون) برای آموزش و ۳۰ درصد (۱۹ کانون) برای آزمون در نظر گرفته شد. سپس ۷ عامل شامل خاک، سنگ‌شناسی، شیب توپوگرافی، شاخص اختلاف پوشش گیاهی نرمال شده، فاصله از رودخانه، واحدهای ژئومورفولوژی و کاربری اراضی به عنوان متغیر مستقل و مؤثر بر برداشت گرد و غبار بررسی شد. سپس با استفاده از روش‌های رگرسیون لجستیک و جنگل تصادفی، نقشه‌های حساسیت خطر گرد و غبار تهیه شد. نتایج به دست آمده از هر دو مدل نشان داد که شیب و کاربری اراضی با ضریب معنی‌داری ۰/۸۵ و ۰/۶۷، بیشترین تأثیر را در ایجاد کانون‌های برداشت گرد و غبار داشته‌است. نتایج همچنین بیانگر آن است که بیشترین نقاط برداشت گرد و غبار (۵۷/۹ درصد)، در منطقه‌ی با حساسیت زیاد قرار دارد. همچنین حدود ۸۴ درصد نقاط برداشت گرد و غبار، در منطقه‌ی با حساسیت زیاد و خیلی زیاد واقع شده‌است. ارزیابی مدل‌ها با استفاده از مشخصه‌ی عملکرد سیستم نیز نشان داد که در زمینه‌ی میزان نرخ موفقیت، مدل‌های رگرسیون لجستیک و جنگل تصادفی به ترتیب دارای دقت ۷۸ و ۹۰ درصد است. هر دو مدل نیز همبستگی بالایی را بین نقشه‌های حساسیت و پراکنش کانون‌های گرد و غبار نشان می‌دهد. واژگان کلیدی: استان خراسان رضوی، پهنه‌بندی، جنگل تصادفی، کانون برداشت گرد و غبار، رگرسیون لجستیک.

۱- مقدمه

خطرات طبیعی هر ساله خسارت‌های زیادی در نقاط مختلف کره‌ی زمین به همراه دارد. توفان گرد و غبار از خطرات طبیعی است که بر شرایط جوی، سلامت انسان و اکوسیستم اثر می‌گذارد (Hahnenberger and Kathleen, 2014). افزایش توفان گرد و غبار در دهه‌ی گذشته سبب شده‌است بسیاری از محققان به این امر توجه کنند (Silva et al, 2016). مناطق اصلی گرد و غبار در جهان، مناطق خشک با توپوگرافی هموار است که خاک حساس به فرسایش و پوشش گیاهی ضعیفی دارد و به آسانی توسط باد فرسایش می‌یابد (Pourhashemi et al, 2016 & Prospero et al, 2002). نتایج نشان می‌دهد شدت و تعداد توفان گرد و غبار در مناطق خشک و نیمه خشک - که بیش از ۳۰ درصد جهان را شامل می‌شود - در حال افزایش است (Namdari et al, 2018). انتقال گرد و غبار، مواد غذایی و ذرات خاک را از نواحی منبع، حذف و در مناطق ترسیب پخش می‌کند که پیامدهای زیست-محیطی زیانباری دارد (Hasselquist et al, 2011 & Sankey et al, 2012). فرایندهای جوی به طور مستقیم در اثر جذب و پراکندگی تابش به وسیله‌ی گرد و غبار تحت تأثیر قرار می‌گیرد و گرد و غبار در ابرها به عنوان هسته‌ی تراکم عمل می‌کند (Danyali et al, 2018 & Wang and Niu, 2013). تغییر در انتقال گرد و غبار، بر آب و هوا در مقیاس محلی و جهانی تأثیر می‌گذارد (Martinez-Garcia et al, 2011 & Prospero et al, 2002).

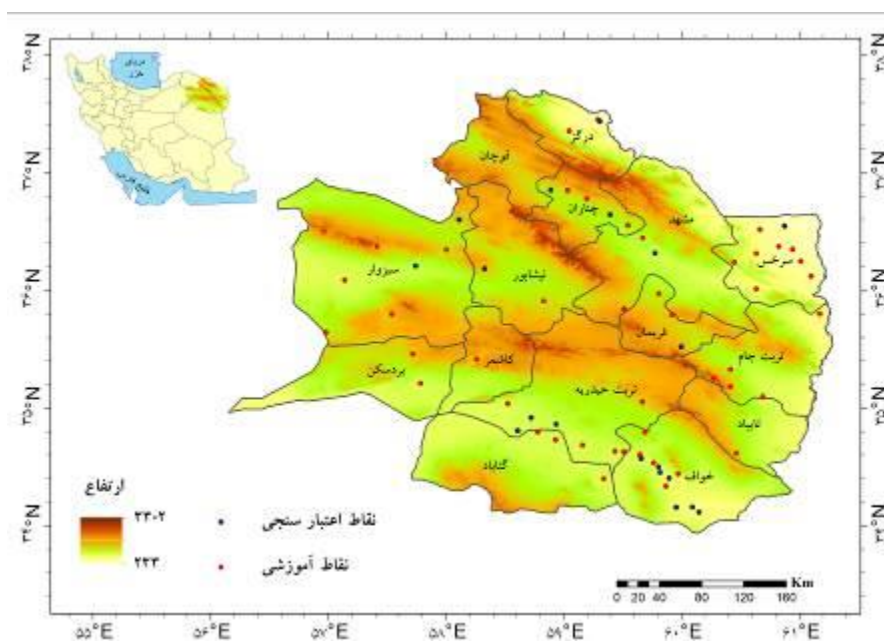
ایران به علت قرار گرفتن در کمربند خشک و نیمه خشک جهان، همواره در معرض سیستم‌های متعدد گرد و غبار محلی و فرا منطقه‌ای قرار دارد (Goudie, 2014 & Rashki et al, 2012). مشاهدات سال‌های اخیر، افزایش وقوع این پدیده را در کشور و منطقه‌ی مورد مطالعه نشان می‌دهد. استان خراسان رضوی دارای آب و هوای خشک و نیمه خشک است که سالانه توفان‌های گرد و غبار متعددی در سطح آن رخ می‌دهد. در بعضی سال‌ها بیش از ۵۰ رخدادهای گرد و غبار در این استان ثبت شده‌است. میزان زیاد گرد و غبار در این منطقه، برداشت زیاد از مناطق سطح استان را نشان می‌دهد که بیانگر اهمیت تحقیق در این زمینه می‌باشد (Lashkari and Keykhosravi, 2009). یکی از راه‌های کاهش خسارت‌های ناشی از خطرات محیطی، شناسایی و بررسی خصوصیات آنهاست. همین امر بر لزوم استفاده از پهنه‌بندی در بحث کانون‌های گرد و غبار تأکید می‌کند. در واقع، با استفاده از نقشه‌ی پهنه‌بندی خطر و حساسیت و اختصاص اقدامات مدیریتی مناسب در نواحی با پتانسیل بالا، می‌توان خسارت‌های وارد شده را به حداقل رساند. برای پهنه‌بندی نیز می‌توان از تکنیک‌های آماری و روش‌های مختلف داده‌کاوی استفاده کرد. تکنیک‌های آماری شامل مقادیر زیادی از داده‌هاست که برای به دست آوردن نتایج قابل اعتماد می‌باشد و استفاده از آنها معمولاً برای مطالعات در مقیاس منطقه‌ای مناسب است (Ylmaz, 2009). در زمینه‌ی به کارگیری روش‌های مختلف داده‌کاوی در انتخاب متغیرهای مؤثر و تهیه‌ی نقشه‌های حساسیت، مطالعات گسترده‌ای بر روی سیل (Al-Abadi et al, 2016)، زمین‌لغزش (Hong et al, 2016)، چشمه‌های آب زیرزمینی (Dube et al, 2014 & Golkarian et al, 2018)، بهمن (Bühler et al, 2013) و آب‌کند (Manap et al, 2014)، گرد و غبار (Boroughani et al, 2019) و ... انجام شده‌است.

داده‌کاوی، استخراج دانش در پایگاه داده‌ها نامیده می‌شود و روشی برای کشف اطلاعات سودمند جدید و بالقوه از بین حجم انبوهی از اطلاعات است (Fayyad et al, 1996). مفهوم داده‌کاوی شامل الگوریتم‌ها و روش‌هایی است که به استخراج اطلاعات از داده‌ها منجر می‌شود (Akbari et al, 2017). Can و همکاران (۲۰۰۵) و Garcia-Rodriguez و همکاران (۲۰۰۸)، طی تحقیقی

دو نکته‌ی مهمی را در زمینه‌ی اعتبارسنجی نقشه‌های حساسیت ذکر کرده‌اند: ۱- کلاس‌های حساسیت زیاد و خیلی زیاد باید مساحت کمتری را نسبت به سایر کلاس‌های حساسیت پوشش دهند. ۲- درصد بالایی از داده‌های اعتبارسنجی باید در کلاس‌های زیاد و خیلی زیاد واقع شود. دو نمونه از مهم‌ترین الگوریتم‌های داده‌کاوی، جنگل تصادفی و رگرسیون لجستیک است. با توجه به اینکه هیچ‌گونه مطالعه‌ای در زمینه‌ی پهنه‌بندی حساسیت کانون‌های گرد و غبار با استفاده از الگوریتم رگرسیون لجستیک و جنگل تصادفی صورت نگرفته؛ از این رو، هدف از این تحقیق پهنه‌بندی خطر مناطق برداشت گرد و غبار با استفاده از این الگوریتم‌ها در استان خراسان رضوی است.

۲- منطقه‌ی مورد مطالعه

استان خراسان رضوی با مساحت ۱۴۳۸۶۴/۱۱۸ کیلومتر مربع، چهارمین استان کشور از نظر وسعت است (شکل ۱). بلندترین نقطه‌ی استان، در ارتفاعات بینالود در شمال نیشابور با ارتفاع ۳۲۱۱ متر از سطح دریا و پست‌ترین نقطه‌ی استان، دشت سرخس با ارتفاع ۳۰۰ متر از سطح دریا و در حاشیه‌ی مرز ایران و ترکمنستان قرار دارد. میانگین بارش سالیانه‌ی مناطق مختلف این استان، از ۱۱۶ میلی‌متر در شهرستان خواف تا ۳۱۲ میلی‌متر در شهرستان قوچان متغیر است (Mosavi Baygi and Ashraf, 2011). این منطقه دارای آب و هوای خشک و نیمه خشک است، شرایط مطلوبی برای ایجاد گرد و غبار دارد و مستعد فرسایش بادی است (Alizadeh Choobari et al, 2014). مراتع در سطح منطقه، بیشترین گسترش را به خود اختصاص داده‌است و دشت‌های پهناور متشکل از سازندهای آبرفتی و بادرفتی کواترنری و رشته کوه‌های مرکب از سازندهای رسوبی و آتشفشانی هوازده متعلق به مزوزوئیک و سنوزوئیک، از دیگر ویژگی‌های بارز این استان است که در زمینه‌ی تولید گرد و غبار مستعد می‌باشد (Mosavi Baygi and Ashraf, 2011).



شکل ۱: موقعیت استان خراسان رضوی در کشور

۳- مواد و روش

۳-۱- شناسایی مناطق برداشت گرد و غبار

در این تحقیق، تصاویر سنجنده MODIS برای روزهایی که توسط ایستگاه‌های سینوپتیک به عنوان روز وقوع توفان گرد و غبار مشخص شده‌است، از ماهواره‌های ترا و آکوا جمع‌آوری شد (Vickery and Eckardt, 2013). در ابتدا روزهای وقوع گرد و غبار در سال‌های ۲۰۰۵ تا ۲۰۱۶ با استفاده از داده‌های هواشناسی مانند قابلیت دید کمتر از ۲۰۰۰ متر، سرعت باد بالاتر از ۷ متر بر ثانیه و وضعیت ابرناکی تعیین شد. سپس روزهایی که وقوع گرد و غبار هم‌زمان با تصویربرداری بود، تعیین شد و در نهایت از ۲۳ تصویر سنجنده MODIS ماهواره‌ی ترا مربوط به روزهای گرد و غباری انتخابی از سال‌های ۲۰۰۵ تا ۲۰۱۶ استفاده شد (<https://ladsweb.modaps.eosdis.nasa.gov/>). این تصاویر برای تهیه‌ی شاخص‌های تعیین مناطق برداشت گرد و غبار فراهم شد و تصحیح هندسی و تصحیحات اولیه بر آن صورت گرفت. خصوصیات زمانی این تصاویر در جدول (۱) ارائه شده‌است.

۳-۲- شاخص‌های مورد استفاده برای بارزسازی و تشخیص مناطق

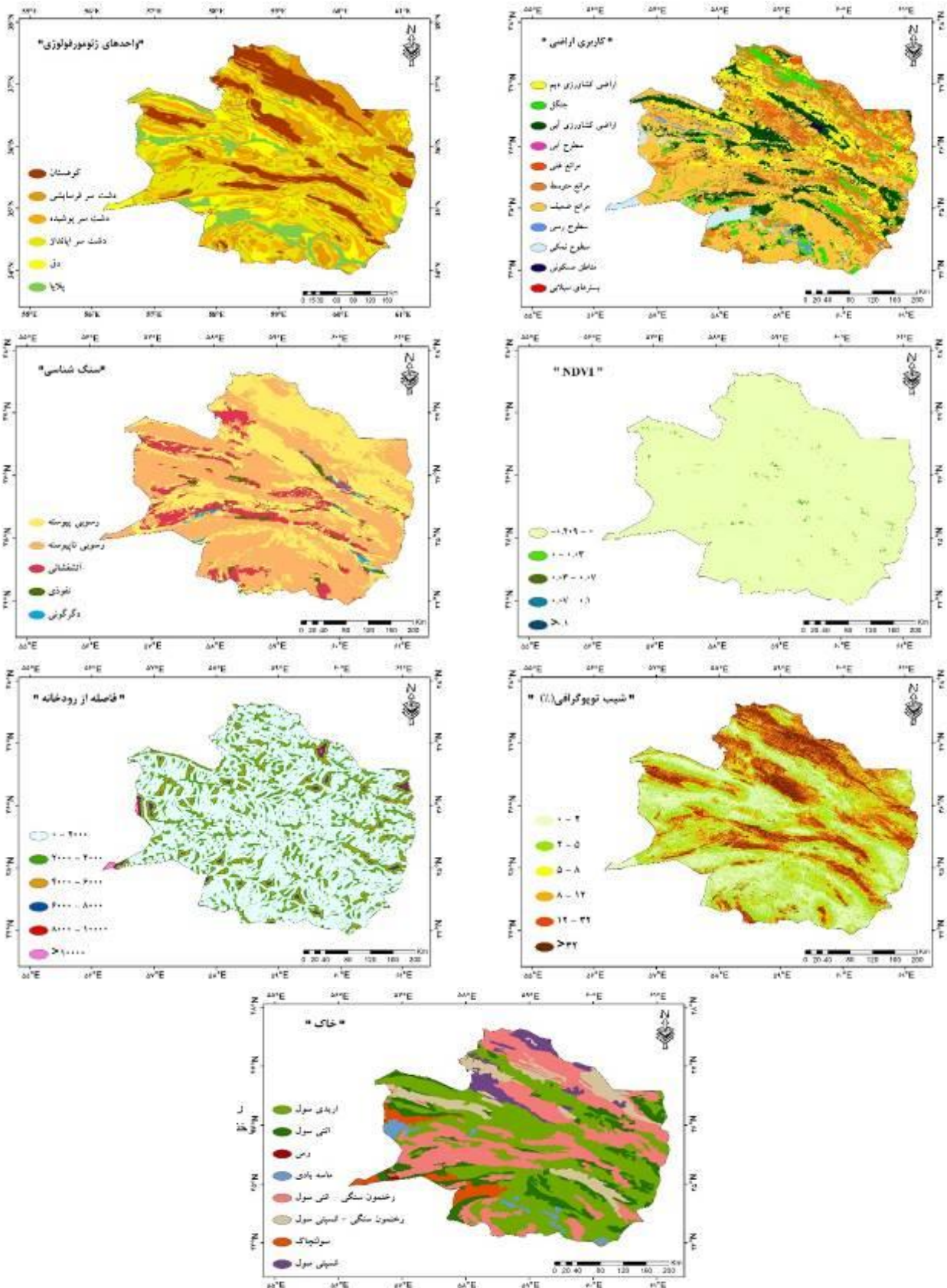
شاخص‌های برداشت گرد و غبار شامل Brightness Temperature Difference₃₁₃₂، Brightness Temperature، Normalized Difference Dust Index، Difference₂₉₃₁ و D می‌باشد که برای تمام تصاویر ماهواره‌ای محاسبه شد. سپس از چهار روش ترکیب رنگی کاذب که شامل R: D, G: R: D, G: BT₂₉₃₁, B: NDDI، R: NDDI, G: B4, B: B3 و BT₃₁₃₂, B: NDDI، R: BT₂₉₃₁, G: B4, B: B3 است، برای تعیین بهترین روش در شناسایی کانون‌های برداشت گرد و غبار استفاده شد (Roscovensky and Liou, 2005 & Qu et al, 2006 & Akerman, 1997 & Boroughani et al, 2007).

جدول ۱: مشخصات زمانی تصاویر سنجنده MODIS

ردیف	ماهواره	تاریخ	ردیف	ماهواره	تاریخ
1	Terra	17 may 2004	13	Terra	19 June 2010
2	Terra	23 June 2005	14	Terra	10 April 2011
3	Terra	8 July 2005	15	Terra	6 march 2012
4	Terra	1 August 2006	16	Terra	17 may 2013
5	terra	26 December 2006	17	Terra	1 November 2013
6	Terra	24 June 2007	18	Terra	1 July 2014
7	Terra	13 November 2007	19	Terra	5 July 2015
8	Terra	5 June 2008	20	Terra	17 July 2015
9	Terra	3 August 2008	21	Terra	23 April 2016
10	Terra	9 August 2008	22	Terra	13 June 2016
11	Terra	31 may 2009	23	Terra	2 October 2016
12	Terra	11 August 2009			

بعد از تهیه نقشه‌ی پراکنش مناطق برداشت گرد و غبار، عوامل مؤثر بر این مناطق شامل نقشه‌های خاک، لیتولوژی، شیب، شاخص پوشش گیاهی^۱، فاصله از رودخانه، واحدهای ژئومورفولوژی و کاربری اراضی بررسی شد. برای تهیه نقشه‌ی کاربری اراضی از تصاویر ماهواره‌ای Land sat 8 برای بازه‌ی زمانی اردیبهشت تا شهریور ۱۳۹۵ استفاده شد. دلیل استفاده از این بازه‌ی زمانی، همزمانی با توفان‌های گرد و غبار در منطقه‌ی مورد مطالعه است. این تصاویر از سایت USGS دانلود شد. با توجه به اینکه بیشتر توفان‌های گرد و غبار در سطح استان در ماه‌های اردیبهشت و خرداد رخ داده‌است، تصاویر این ماه‌ها دانلود و با استفاده از نرم‌افزار ENVI، نقشه‌ی کاربری اراضی برای سال ۱۳۹۵ تهیه شد. با استفاده از DEM ۳۰ متر، نقشه‌ی شیب در محیط GIS بر اساس درصد تهیه شد. کمترین و بیشترین شیب در منطقه‌ی مورد مطالعه به ترتیب ۰ تا ۲ و بیش از ۳۲ درصد می‌باشد. نقشه‌ی زمین‌شناسی و خاک‌شناسی با مقیاس ۱:۲۵۰۰۰۰ برای استان خراسان رضوی تهیه شد. سپس نقشه‌ی فاصله از رودخانه‌ها از نقشه‌ی توپوگرافی منطقه در محیط GIS استخراج شد. برای تهیه نقشه‌ی پوشش گیاهی منطقه، از شاخص تفاضلی پوشش گیاهی نرمال شده (NDVI) استفاده شد؛ بدین منظور، از تصاویر ETM⁺ ماهواره لندست ۸ (مربوط به سال ۲۰۱۶) استفاده شد. در نهایت، برای تهیه نقشه‌ی واحدهای ژئومورفولوژی ابتدا با بهره‌جویی از نقشه‌های شیب و توپوگرافی با دقت ۱:۵۰۰۰۰ و زمین‌شناسی محدوده-ی مورد مطالعه، نقشه‌ی مورد نیاز برای عملیات صحرایی آماده شد. در مرحله‌ی بعد، از تفسیر تصاویر ماهواره‌ای Google Earth و لندست ۸ در سال ۲۰۱۶، اطلاعات جزئی‌تر استخراج و بر روی نقشه‌ی اولیه منتقل شد و در ادامه، با تکمیل اطلاعات صحرایی و کنترل نهایی (در تاریخ ۲۱ تا ۲۶ اردیبهشت ۱۳۹۷)، نقشه‌ی واحدهای ژئومورفولوژی منطقه تهیه شد. شکل (۲) عوامل مؤثر بر ایجاد کانون‌های برداشت گرد و غبار را نشان می‌دهد.

¹ Normalized Difference Vegetation Index



شکل ۲: نقشه‌ی عوامل مؤثر بر کانون‌های برداشت گرد و غبار

۳-۳- مدل‌سازی

اولین گام در روش مدل‌سازی، آماده‌سازی مجموعه داده‌های آموزشی و اعتبارسنجی است. در مرحله‌ی بعد، داده‌های مربوط به عوامل مؤثر بر گرد و غبار و نقشه‌ی پراکنش کانون‌ها در نرم‌افزار R وارد و در آن، الگوریتم‌های رگرسیون لجستیک و جنگل تصادفی با استفاده از بسته‌های مختلف اجرا شد. پس از آن، حساسیت منبع گرد و غبار برای هر پیکسل منطقه‌ی مورد مطالعه محاسبه شد. در مرحله‌ی نهایی، نتایج به محیط ARC MAP 10.2 منتقل شد تا نقشه‌ی حساسیت گرد و غبار را برای منطقه‌ی مورد مطالعه تولید کند. شرح الگوریتم‌های مورد استفاده در زیر بیان شده‌است.

۳-۴- رگرسیون لجستیک

رگرسیون لجستیک چندگانه، تکنیکی چند متغیره است که چند پارامتر فیزیکی را - که ممکن است بر احتمال تأثیر بگذارد - قابل توجه قرار می‌دهد. در مدل رگرسیون بر اساس داده‌های موجود، معادله‌ای برای مقادیر متغیر وابسته (Y) بر اساس یک یا چند متغیر مستقل Xi به دست می‌آید. این مدل، وجود و نبود متغیر وابسته را در ارتباط با متغیر مستقل تحلیل می‌کند (رابطه‌ی ۱) که متغیر وابسته (وجود یا عدم وجود نقطه برداشت گرد و غبار) متغیر کیفی دو حالتی است و فقط ارزش ۱ و ۰ دارد. این رگرسیون، روشی کاملاً کمی است که میزان تأثیر هر یک از متغیرهای مستقل را به صورت کمی و از طریق ضرایب و آنتی لگاریتم ضرایب مشخص می‌کند (Moghim et al, 2011 & Garcia-Rodriguez et al, 2008 & Lee and Sambath, 2006). در این تحقیق با استفاده از رگرسیون لجستیک، به بررسی مهم‌ترین عوامل مؤثر بر کانون‌های برداشت گرد و غبار پرداخته می‌شود. در تهیه‌ی مناطق برداشت گرد و غبار، هدف رگرسیون لجستیک یافتن بهترین مدل تشریح روابط میان حضور یا عدم حضور متغیر وابسته (مناطق برداشت گرد و غبار) و مجموعه گروه‌هایی از متغیر مستقل (عوامل مؤثر) است. خروجی مدل، ضریب‌هایی بین ۰ و ۱ خواهد داشت که از طریق تئوری لجیت به احتمالات بیش از ۰/۵ ارزش ۱ (مؤثر بر منطقه برداشت) و کمتر از ۰/۵ ارزش صفر (عدم تأثیر منطقه برداشت) می‌دهد (Pontius and Schneider, 2001 & Martinez-Garcia et al, 2011 & Chung and Fabbri, 2003). مدل لجستیک در ساده‌ترین شکل می‌تواند به صورت زیر بیان شود:

$$P = \frac{1}{1 + e^{-z}} \quad \text{رابطه‌ی ۱}$$

که در آن P احتمال وقوع یک رخداد (گرد و غبار) است که مقدار آن از ۰ تا ۱ در یک منحنی S شکل در نوسان می‌باشد و Z به صورت معادله‌ای (مدل لجستیک خطی) تعریف می‌شود که مقدار (ارزش) آن از $-\infty$ تا $+\infty$ در نوسان است (Schicker and Moon, 2012). در مورد مطالعه‌ی حساسیت نقاط برداشت گرد و غبار، هدف رگرسیون لجستیک یافتن بهترین (در عین حال معقول) مدل برازش برای بیان رابطه بین وجود یا فقدان گرد و غبار (متغیر وابسته) و مجموعه‌ای از پارامترهای مستقل است و بر اساس رابطه‌ی زیر بیان می‌شود (Akhlq et al, 2012):

رابطه‌ی ۲

$$Y = \text{Logit}(p) = \ln\left(\frac{p}{1-p}\right) = C_0 + C_1X_1 + C_2X_2 + \dots + C_nX_n$$

که در این رابطه، P : احتمال وقوع حادثه (در این تحقیق احتمال وقوع نقاط برداشت گرد و غبار)، $\left(\frac{p}{1-p}\right)$: اصطلاحاً شانس یا نسبت درست‌نمایی است، C_0 : عرض از مبدأ یا ضریب ثابت و $C_1, C_2, \dots, C_n$ ضرایب متغیرهای مستقل $(X_1, X_2, \dots, X_n)$ هستند که در این مدل، وجود یا فقدان متغیر وابسته (نقاط برداشت گرد و غبار) را در زمینه‌ی متغیرهای مستقل تحلیل و بررسی می‌کند (Garcia-Rodriguez et al, 2008).

۳-۵- مدل جنگل تصادفی^۲

جنگل تصادفی از مجموعه‌ای از طبقه‌بندی‌کننده‌های با ساختار درختی و طبقه‌بندی‌کننده‌ی مجموعه‌ای تشکیل می‌شود که شامل بسیاری از درخت‌های تصمیم‌گیری است (Ebrahim khani et al, 2012). این الگوریتم را نخستین بار لئو برایمن و آدل کاتلر ایجاد کردند و توسعه دادند (Farshad and Sadeh, 2014). الگوریتم جنگل تصادفی، روش یادگیری مبتنی بر دسته‌ای از درخت‌های تصمیم است (Breiman, 2001). در این روش به جای رشد یک درخت، صدها یا هزاران درخت طبقه‌بندی تولید می‌شود (Akbari et al, 2017). اینک این الگوریتم از بهترین الگوریتم‌های یادگیری است و برای بسیاری از مجموعه داده‌ها، طبقه‌بندی با صحت بالایی دارد (Calderoni et al, 2015). ویژگی مثبت دیگر این طبقه‌بندی‌کننده این است که روی مجموعه داده‌های بزرگ، بسیار خوب عمل می‌کند (Ebrahim Khani et al, 2012 & Zabihi et al, 2016). مهم‌ترین ویژگی جنگل‌های تصادفی، عملکرد بالای آنها در اندازه‌گیری اهمیت متغیرهاست که مشخص می‌کند هر متغیر چه نقشی در پیش‌بینی پاسخ دارد. اساس کار روش جنگل تصادفی (RF)، ترکیبی از چندین درخت تصمیم است که چندین نمونه‌ی بوت استرپ از داده‌ها در ساخت آن شرکت دارد و در ساخت هر درخت نیز به طور تصادفی، تعدادی از متغیرهای ورودی شرکت می‌کنند. روش بوت استرپ، روش نمونه‌گیری با جای‌گذاری است؛ از این رو با تکرار عملیات نمونه‌گیری، تعدادی مجموعه داده‌ی خارج از کیسه^۳ از مجموعه‌ی آموزشی به وجود می‌آید که می‌توان برای هر مجموعه، داده‌ی آموزشی یک درخت تصمیم را محاسبه کرد (Catani et al, 2013). روش جنگل تصادفی دارای پارامترهای $mtry$ و $ntree$ است که به ترتیب تعداد متغیرهای مورد استفاده در ساخت هر درخت - که قدرت هر درخت مستقل را بیان می‌کند و تعداد درخت می‌باشد - قابل تغییر بوده و توسط کاربر تعیین می‌شود. کلاس‌بندی جنگل تصادفی، ابتدا بر روی داده‌ی آموزشی سپس بر داده‌های اعتبارسنجی اجرا می‌شود. در نهایت، مدلی انتخاب می‌شود که دارای کمترین خطای OOB باشد. در این مدل، از دو عامل میانگین کاهش دقت^۴ و میانگین کاهش جینی^۵ برای تعیین اولویت تأثیر هر یک از عوامل مؤثر استفاده می‌شود.

^۲ Random Forest

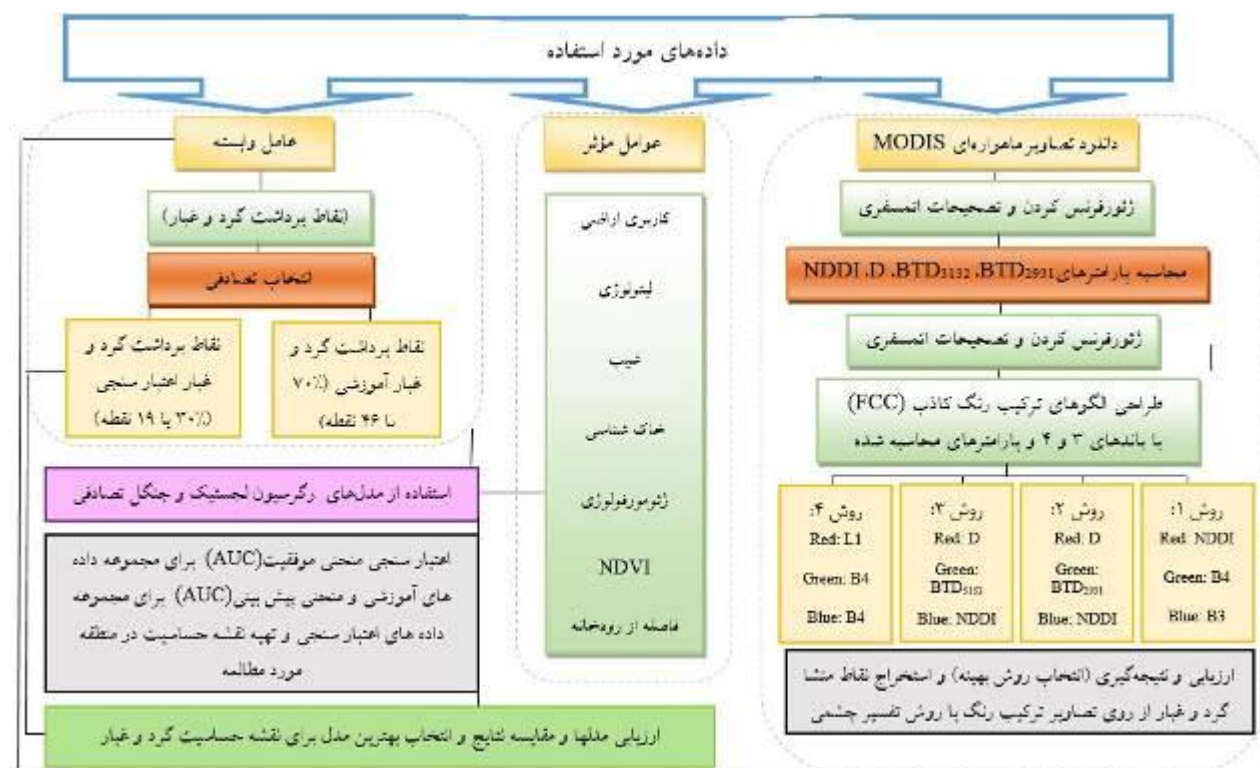
^۳ Out Of Bag

^۴ Mean Decrease Accuracy

^۵ Mean Decrease Gini

۳-۶- ارزیابی نقشه‌ی حساسیت گرد و غبار

یکی از مهم‌ترین بخش‌های مربوط به مدل‌سازی، اعتبارسنجی نتایج پیش‌بینی شده‌است (Chung and Fabbri, 2003). در این مطالعه، تمامی کانون‌های گرد و غبار شناسایی شده به دو بخش تقسیم شد؛ یکی برای آموزش و دیگری برای اعتبارسنجی. به این صورت که از ۷۰٪ کانون‌های گرد و غبار برای آموزش مدل و ۳۰٪ آن برای اعتبارسنجی استفاده شد. فرض اصلی در انتخاب کانون‌های گرد و غبار برای آموزش و اعتبارسنجی مدل‌ها، بر اساس منطق تصادفی در منطقه‌ی مورد مطالعه بود. برای ارزیابی صحت مدل آماری، از شاخص ROC استفاده شد. منحنی ROC یکی از مفیدترین و کارآمدترین روش‌ها در ارائه‌ی خصوصیت تعیینی، شناسایی احتمالی و پیش‌بینی سیستم‌هاست (Catani et al, 2013). این منحنی، مقدار پیش‌بینی سیستم را از طریق توصیف توانایی آن در تخمین درست وقایع رخ داده (مناطق برداشت گرد و غبار) و عدم تخمین آن نشان می‌دهد. ایده‌آل‌ترین مدل، بیشترین سطح زیر منحنی را دارد و مقدار آن بین ۰/۵ تا ۱ است. همبستگی کیفی - کمی سطح زیر منحنی و ارزیابی تخمین به صورت (۱-۰/۹ عالی، ۰/۸-۰/۹ خیلی خوب، ۰/۷-۰/۸ خوب، ۰/۶-۰/۷ متوسط و ۰/۵-۰/۶ ضعیف) می‌باشد (Can et al, 2005). از منحنی ROC (ویژگی عامل گیرنده) برای ارزیابی عملکرد نقشه‌ی حساسیت ایجاد شده در این مطالعه استفاده شد. این منحنی، نمایشی گرافیکی از موازنه‌ی بین میزان خطای منفی و مثبت برای هر مقدار احتمالی از برش‌هاست و از کارآمدترین روش‌ها در ارائه‌ی خصوصیت تعیینی، شناسایی احتمالی و پیش‌بینی سیستم‌ها به شمار می‌رود که میزان دقت را به صورت کمی برآورد می‌کند (Pradhan and Lee, 2010). در روش ROC، مساحت زیر منحنی دارای مقادیری بین ۰/۵ تا ۱ است و از آن برای ارزیابی دقت مدل استفاده می‌شود (Nandi and Shakoor, 2009). روش اولویت‌بندی عوامل مؤثر بر ایجاد کانون گرد و غبار و تهیه‌ی نقشه‌ی حساسیت آن با استفاده از مدل‌های رگرسیون لجستیک و جنگل تصادفی، به صورت نمودار جریان در شکل (۳) ارائه شده‌است.

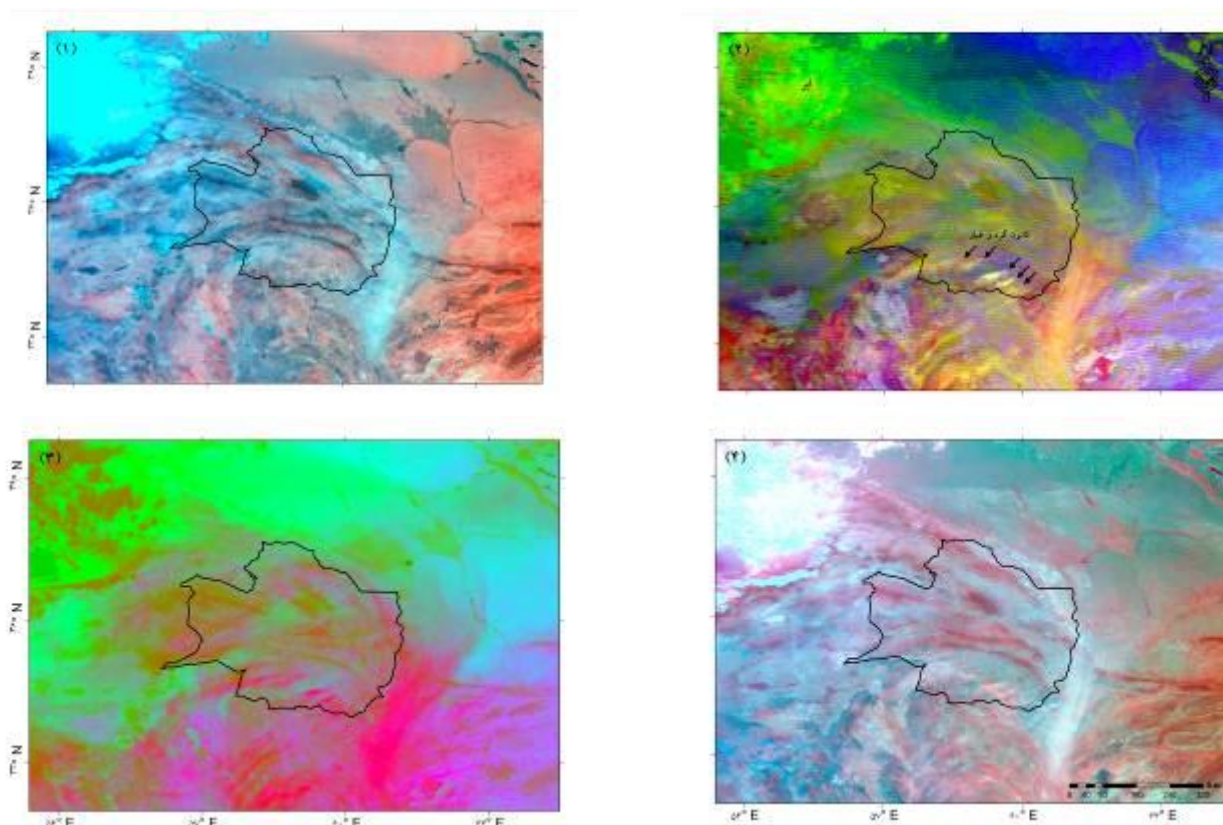


شکل ۳: نمودار جریانی تهیه نقشه‌های حساسیت گرد و غبار با استفاده از مدل‌های رگرسیون لجستیک و جنگل تصادفی

۴- یافته‌ها (نتایج)

۴-۱- شناسایی کانون‌های برداشت گرد و غبار

برای آشکارسازی مناطق گرد و غباری در تمامی تصاویر، از چهار روش مختلف ترکیب رنگی کاذب استفاده شد. هر کدام از این روش‌ها، توده‌ی گرد و غباری را به گونه‌ای آشکارسازی کردند. بهترین روشی که برای آشکارسازی گرد و غبار و تعیین مناطق برداشت گرد و غبار انتخاب شد، روش دوم (R: D, G: BTD₂₉₃₁, B: NDDI) می‌باشد؛ زیرا مقایسه‌ی چشمی تصاویر رنگی شده در چهار روش نشان داد که این روش نسبت به سایر روش‌ها در آشکارسازی مناطق برداشت گرد و غبار، قابلیت بهتری دارد و به طور بارزتری مناطق گرد و غباری را از سایر عوارض و مناظر متمایز می‌کند. یک تصویر به عنوان مثال از اجرای هر چهار روش در شکل (۴) نشان داده شده‌است.



شکل ۴: تصاویر چهار روش الگوی ترکیب رنگی کاذب برای تصویر سنجنده MODIS در تاریخ ۱ جولای ۲۰۱۴

در نهایت، با کمک گرفتن از این روش‌ها و ایجاد ترکیب رنگی کاذب و اعمال آن بر تصاویر موجود و تفسیر چشمی آنها، ۶۵ نقطه‌ی برداشت گرد و غبار در کل منطقه شناسایی شد.

۲-۴- مدل‌سازی حساسیت کانون‌های گرد و غبار

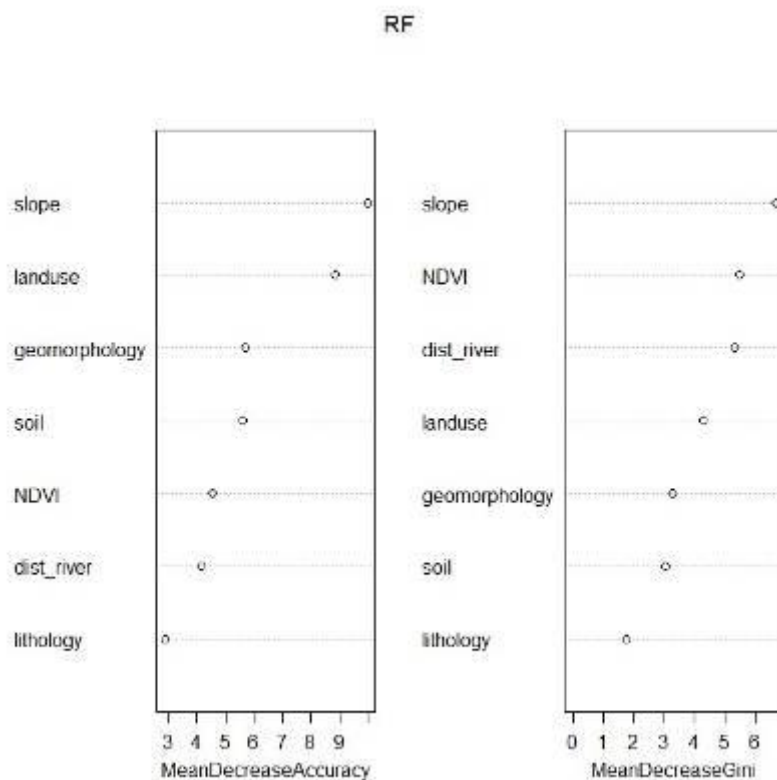
برای ارزیابی عوامل مؤثر بر مناطق برداشت گرد و غبار (متغیر وابسته) در استان خراسان رضوی، از مدل رگرسیون لجستیک با ۷ متغیر مستقل استفاده شد. پهنه‌بندی حساسیت این رویداد در محدوده‌ی مطالعاتی نیز بر اساس ضرایب به دست آمده برای متغیرهای مستقل از طریق اجرای مدل صورت گرفت. نتایج حاصل از اجرای مدل رگرسیون لجستیک، برای بررسی مهم‌ترین عامل مؤثر بر ایجاد مناطق برداشت گرد و غبار استان خراسان رضوی در جدول (۲) ارائه شده‌است. بر اساس نتایج، شیب بیشترین تأثیر را در ایجاد مناطق برداشت گرد و غبار داشته‌است. ژئومورفولوژی، لیتولوژی، شاخص پوشش گیاهی، خاک و کاربری اراضی در رده‌های بعدی نیز بر کانون‌های برداشت گرد و غبار تأثیر داشته‌اند. ضرایب منفی بیانگر رابطه‌ی معکوس متغیر وابسته (کانون گرد و غبار) با متغیرهای مستقل است. در واقع ضرایب منفی، فقدان همبستگی را نشان نمی‌دهد، بلکه به معنای همبستگی ضعیف‌تر نسبت به عوامل دیگر است. عامل فاصله از رودخانه، کمترین تأثیر را در ایجاد کانون‌های گرد و غبار دارد. میزان

Pseudo R-Square، ۰/۳۸۱۲ است که در محدوده‌ی مورد قبول قرار دارد.

جدول ۲: ضرایب متغیرهای مستقل و آزمون والد از تجزیه و تحلیل آماری رگرسیون لجستیک

متغیرهای مستقل	ضریب متغیر	Wald	Sig	Exp(B)
شیب	۱/۷۰۵	۰/۰۳۵	۰/۰۰۰	۵/۴۹۹
ژئومورفولوژی	۰/۱۴۳	۰/۱۸۰	۰/۰۰۵	۰/۸۶۷
کاربری اراضی	-۰/۰۲۰	۰/۰۵۸	۰/۰۰۰	۰/۹۸۰
لیتولوژی	-۰/۱۳۰	۰/۲۹۸	۰/۰۰۰	۰/۸۷۸
خاک	۰/۰۴۹	۰/۱۷۴	۰/۰۰۱	۱/۰۵۰
شاخص پوشش گیاهی	-۰/۰۷۹	۳/۰۹۷	۰/۰۱۵	۰/۹۲۴
فاصله از رودخانه	۰	۲/۱۷۹	۰/۰۳۸	۱
مقدار ثابت	۱/۱۲۲	۰/۹۴۶	۰/۰۱۱	۳/۰۷۱

نتایج حاصل از رابطه‌ی بین کانون‌های گرد و غبار و عوامل مؤثر همچنین پیش‌بینی‌های OOB کل (ماتریس درهم ریختگی) با استفاده از مدل RF، در جدول ۳ و شکل ۴ نشان داده شده‌است. نتایج OOB بیانگر میزان خطای پیش‌بینی شده در حدود ۳۵٪ می‌باشد؛ به عبارت دیگر، دقت مدل را می‌توان ۶۴٪ در نظر گرفت. بر طبق نتایج حاصل از جدول ۳، مدل در قسمت فقدان گرد و غبار (ردیف ۱)، ۱۵ مورد را در قسمت وجود گرد و غبار به صورت غلط پیش‌بینی کرد (خطای ۱) و در بخش وجود گرد و غبار، ۱۶ مورد را در ستون وجود گرد و غبار پیش‌بینی کرد (خطای ۲). از سوی دیگر، مدل به درستی فقدان کانون گرد و غبار را برای ۳۱ مورد و وجود کانون گرد و غبار را برای ۳۰ مورد پیش‌بینی کرد. شکل ۵، نتایج اولویت‌بندی متغیرهای مؤثر را برای مدل RF با استفاده از دو معیار میانگین کاهشی دقت و شاخص اهمیت جینی نشان می‌دهد. میانگین کاهشی دقت، معیار مناسب‌تری برای اولویت‌بندی فاکتورهاست و وزن بالاتر متغیر نیز نشان می‌دهد که آن عامل در نسبت وقوع پدیده نقش مهمی دارد (Mohamadi and Pourghasemi, 2018 & Nicodemous, 2011). نتایج میانگین کاهشی دقت نشان می‌دهد که شیب و کاربری اراضی، جزء مهم‌ترین عوامل مؤثر بر وقوع کانون‌های گرد و غبار است. شیب و شاخص پوشش گیاهی نیز با توجه به شاخص اهمیت جینی از بیشترین اهمیت برخوردار است. اولویت‌بندی عوامل، با توجه به شاخص اهمیت جینی نسبت به میانگین کاهشی دقت متفاوت است. با این وجود هر دو شاخص، شیب را به عنوان مؤثرترین عامل وقوع گرد و غبار در منطقه‌ی مورد مطالعه شناسایی کردند.



شکل ۵: میانگین کاهش دقت و میانگین کاهش جینی با استفاده از مدل جنگل تصادفی

جدول ۳: ماتریس سردرگمی برای مدل جنگل تصادفی: ۰ = عدم وجود کانون گرد و غبار و ۱ = وجود کانون گرد و غبار

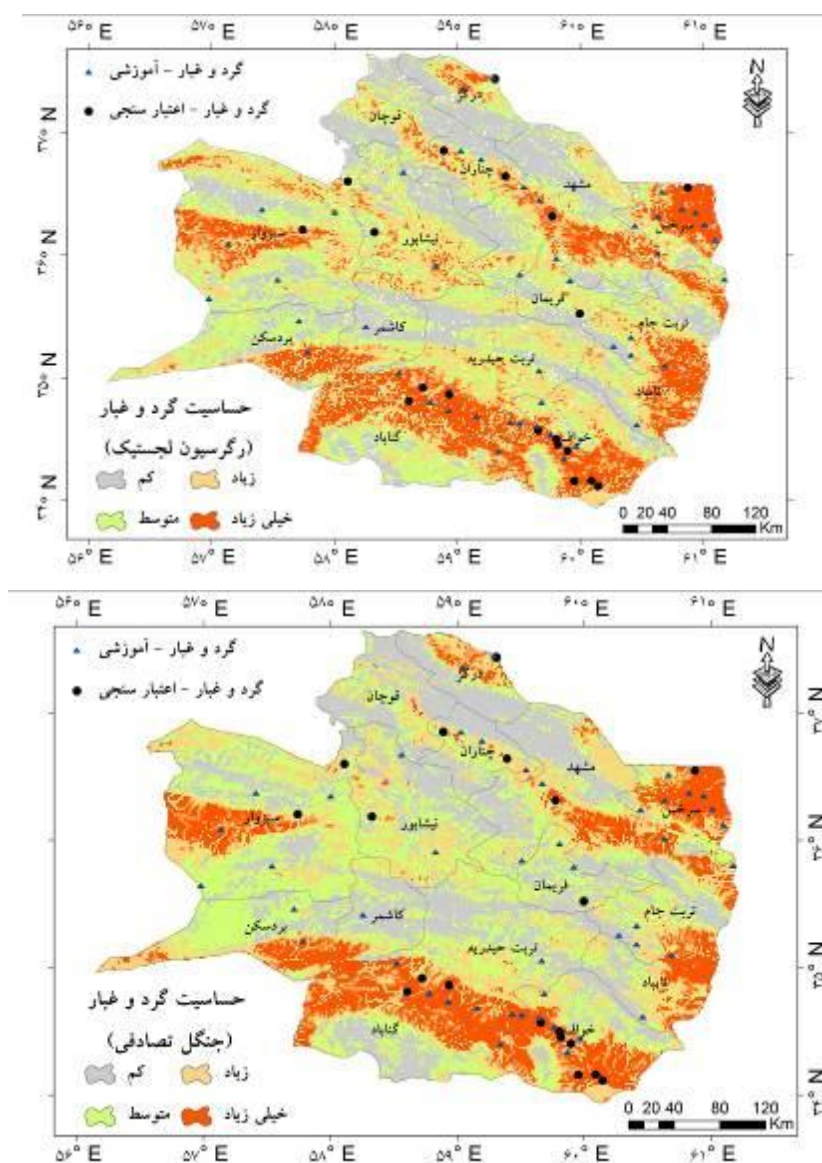
	Predicted		Class error
	عدم حضور	حضور	
Actual			
عدم حضور	۳۱	۱۵	۰/۳۴
حضور	۱۶	۳۰	۰/۳۵

۳-۴. پهنه‌بندی مناطق حساس به برداشت گرد و غبار

پس از اجرای مدل‌ها در نرم‌افزار R برای داده‌های مشاهده‌ای و تعمیم فرآیند مدل‌سازی به کل منطقه‌ی مورد مطالعه، خروجی به صورت یک فایل با دو ستون اعداد به صورت صفر و یک می‌باشد. ستون صفر به معنای فقدان نقاط گرد و غبار برای هر پیکسل و ستون یک بیانگر نقاط گرد و غبار برای هر پیکسل از کل منطقه‌ی مورد مطالعه است. به منظور پهنه‌بندی حساسیت گرد و غبار نتایج حاصل از مدل‌سازی بر اساس مختصات هر یک از پیکسل‌ها، در قالب یک فایل نقطه‌ای به محیط GIS وارد شد. سپس از مسیر فرمت نقطه به رستر، نقشه‌ی نهایی برای منطقه‌ی مورد مطالعه تهیه و بر اساس شکست‌های طبیعی^۶ در چهار

^۶ Naturals Breake

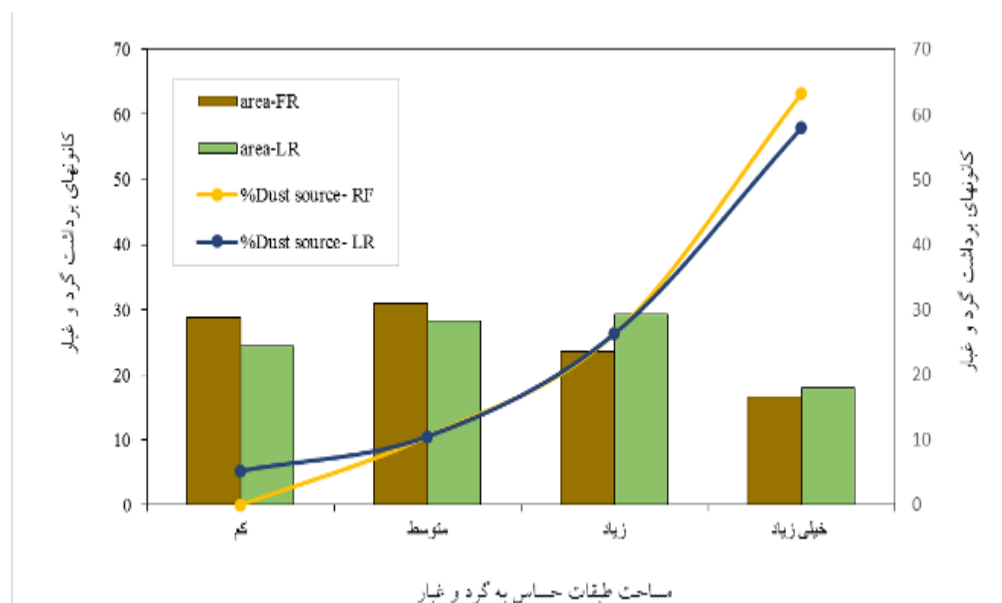
طبقه‌ی حساسیت کم، متوسط، زیاد و خیلی زیاد طبقه‌بندی شد (شکل ۶). سپس به منظور مشخص شدن وضعیت کلی منطقه‌ی مورد مطالعه، مساحت طبقات حساسیت گرد و غبار نیز محاسبه شد. نتایج حاصل از الگوریتم رگرسیون لجستیک نشان داد که ۲۴/۵ درصد (۲۸۲۲۱/۹ کیلومترمربع) از مساحت استان خراسان رضوی در منطقه‌ی با حساسیت کم، ۲۸/۲ (۳۲۳۸۱/۹ کیلومترمربع) درصد در منطقه‌ی با حساسیت متوسط، ۲۹/۴ (۳۳۶۴۷/۶ کیلومترمربع) درصد در منطقه‌ی با حساسیت زیاد و ۱۷/۹ درصد (۲۰۹۸۹/۹ کیلومترمربع) در منطقه‌ی با حساسیت خیلی زیاد قرار دارد. نتایج حاکی از آن است که بیشترین نقاط برداشت گرد و غبار (۵۷/۹ درصد) در منطقه‌ی با حساسیت زیاد و حدود ۸۴/۲ درصد از نقاط در منطقه‌ی با حساسیت زیاد و خیلی زیاد واقع شده‌اند. بررسی مساحت طبقات نقشه‌ی حساسیت گرد و غبار با استفاده از الگوریتم جنگل تصادفی نشان داد که منطقه‌ی با حساسیت خیلی زیاد، کمترین مساحت (۱۹۱۸۸/۹ کیلومترمربع) را در بر گرفته و مساحت طبقات زیاد، متوسط و کم به ترتیب ۲۷۰۹۶/۵، ۳۵۷۲۲/۵ و ۳۳۲۳۳/۳ کیلومترمربع برآورد شده‌است. حدود ۶۳/۲ درصد کانون‌های برداشت، در منطقه‌ی با حساسیت خیلی زیاد واقع شده‌است و در منطقه‌ی با حساسیت کم، هیچ کانون برداشتی وجود ندارد.



شکل ۶: نقشه‌ی پهنه‌بندی حساسیت مناطق برداشت گرد و غبار در استان خراسان رضوی

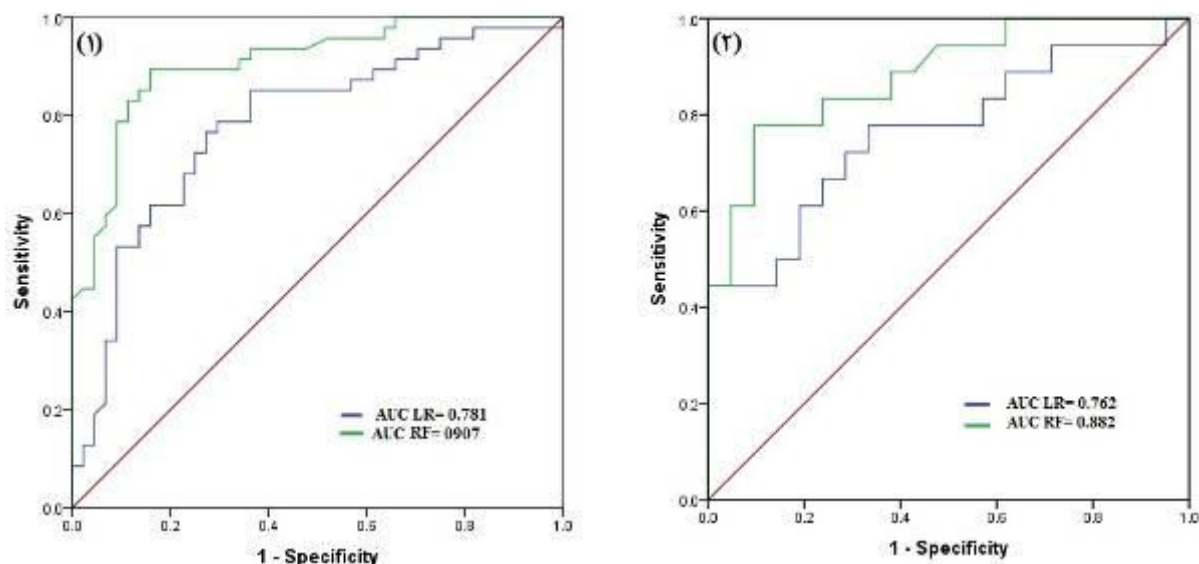
۴-۴- ارزیابی صحت مدل آماری رگرسیون لجستیک و جنگل تصادفی

همان طور که ملاحظه می‌شود، طبقات با حساسیت خیلی زیاد به طور کلی درصد کمی از مساحت منطقه‌ی مورد مطالعه را در بر گرفته‌است؛ در حالی که بیشترین درصد نقاط برداشت گرد و غبار در این طبقه‌ی حساسیت قرار گرفته و این امر دقت بالای مدل‌ها را نشان می‌دهد. این نتایج، صحت اجرای هر دو مدل آماری را بیان می‌کند. شکل (۷) مساحت هر کلاس حساسیت و درصد تعداد گرد و غبارهای اعتبارسنجی را در نقشه‌ی حساسیت به وقوع گرد و غبار نشان می‌دهد (Carcia-Rodriguez et al, 2008 & Can et al, 2005).



شکل ۷: نمودار مساحت طبقات حساس به گرد و غبار و تعداد گرد و غبارهای اعتبارسنجی در نقشه‌ی حساسیت به گرد و غبار به درصد

در تحقیق حاضر از ۳۰ درصد از نقاط برداشت گرد و غبار برای مرحله‌ی ارزیابی (تحت عنوان نرخ پیش‌بینی) و از ۷۰ درصد از نقاط برای مرحله‌ی مدل‌سازی (تحت عنوان میزان موفقیت) استفاده شد. براساس نتایج منحنی ROC (نرخ پیش‌بینی)، مقدار سطح زیر منحنی مدل‌های رگرسیون لجستیک و جنگل تصادفی به ترتیب ۰/۷۶۲ و ۰/۸۸۲ محاسبه شد. با توجه به طبقه‌بندی ارائه شده برای سطح زیر منحنی، دقت مدل رگرسیون لجستیک در طبقه‌ی خوب و مدل جنگل تصادفی در طبقه‌ی خیلی خوب قرار گرفت که این امر عملکرد بهتر مدل جنگل تصادفی را در منطقه‌ی مورد مطالعه نشان می‌دهد. شکل (۸)، نتایج ارزیابی مدل‌ها را نشان می‌دهد.



شکل ۸: منحنی ROC و سطح زیر منحنی مربوط به نقشه‌ی حساسیت گرد و غبار با استفاده از مدل‌های رگرسیون لجستیک و جنگل تصادفی.

(۱) منحنی میزان موفقیت و (۲) منحنی میزان پیش‌بینی

۵- بحث و نتیجه گیری

در این پژوهش به اولویت بندی عوامل مؤثر و پهنه بندی حساسیت برداشت گرد و غبار در استان خراسان رضوی پرداخته شد. این امر با استفاده از روش های رگرسیون لجستیک و جنگل تصادفی و تعیین اهمیت هر یک از عوامل محیطی مؤثر بر برداشت گرد و غبار صورت گرفت؛ بدین منظور از ۷ متغیر خاک، لیتولوژی، شیب، شاخص پوشش گیاهی، فاصله از رودخانه، واحدهای ژئومورفولوژی و کاربری اراضی به عنوان عوامل مؤثر بر وقوع کانون های گرد و غبار استفاده شد. نتایج حاصل از روش رگرسیون لجستیک نشان داد که به ترتیب عوامل شیب و ژئومورفولوژی بیشترین نقش را در وقوع گرد و غبار داشته است. پهنه های حساسیت به نقاط برداشت با استفاده از این مدل نشان داد که حدود ۱۸ درصد منطقه ی مورد مطالعه در رده ی حساسیت خیلی زیاد قرار گرفته است که بیش از ۵۷ درصد از کل نقاط برداشت گرد و غبار نیز در این محدوده قرار دارد. در هر دو مدل، بیشترین درصد نقاط برداشت گرد و غبار در طبقه ی حساسیت خیلی زیاد قرار گرفته است که مساحت کمی از منطقه را به خصوص در مدل جنگل تصادفی شامل می شود. طبق نتایج حاصل از روش جنگل تصادفی نیز به ترتیب عوامل شیب و کاربری اراضی بیشترین و لیتولوژی کمترین تأثیر را در ایجاد کانون های برداشت گرد و غبار داشته است. با توجه به اینکه بیشتر کانون های برداشت گرد و غبار در مناطق با شیب کم وجود دارد و در منطقه ی مورد مطالعه به علت رشد جمعیت، تخریب و تغییرات زیاد کاربری اراضی شرایط برای ایجاد کانون برداشت فراهم شده است؛ از این رو، این دو پارامتر بیشترین نقش را در ایجاد کانون برداشت گرد و غبار داشته اند. در زمینه ی رده های خطر در این مدل نیز ۱۶/۷ درصد محدوده، در رده ی خطر خیلی زیاد قرار دارد که بیش از ۶۳ درصد از کل نقاط برداشت گرد و غبار در این رده واقع می باشد. نتایج ارزیابی مدل ها با استفاده از روش ROC نشان داد که مدل رگرسیون لجستیک در رده ی خوب و مدل جنگل تصادفی در رده ی بسیار خوب قرار می گیرد. جنگل تصادفی با سطح زیر منحنی AUC برابر با ۰/۸۸۲ در مقایسه با روش رگرسیون لجستیک با سطح زیر منحنی برابر با ۰/۷۶۲، کارایی بالاتری برای ارزیابی دارد. نقشه ی پهنه بندی حساسیت گرد و غبار، مناطق با حساسیت های مختلف را در استان خراسان رضوی نشان می دهد؛ از این رو، نتایج این مدل در کنترل و کاهش گرد و غبار می تواند برای برنامه ریزان و مدیران مفید باشد.

منابع

1. Ackerman, S. A., 1997. Remote sensing aerosols using satellite infrared observations, *Journal of Geophysical Research*, 102, 17069-17080.
2. Akbari, M.; Bashiri, M. ; & A. Rangavar, 2017. Application of Data Mining Algorithms to Appreciate Sensitivity and Spatial Zoning Prone to Floating View in Khorasan Razavi Display Basins, *Journal of Environmental Erosion Research*, 7(26), 16-42. (in Persian)
3. Akhlaq, M.; Sheltami, T. R.; & H. T. Mouftah, 2012. A review of techniques and technologies for sand and dust storm detection, *Reviews in Environmental Science and Biotechnology*, 11(3), 305-322.
4. Al-Abadi, A. M.; Shahid, S.; & A. K. Al-Ali, 2016. A GIS-based integration of catastrophe theory and analytical hierarchy process for mapping flood susceptibility: a case study of Teeb area, Southern Iraq, *Environmental Earth Science*, 75, 687. <http://dx.doi.org/10.1007/s12665-016-5523-7>.

5. Alizadeh Choobari, O.; Zawar-Reza, P.; & A. Sturman, 2014. The global distribution of mineral dust and its impacts on the climate system: A review, *Atmospheric Research*, 138(1), 152-165.
6. Ayalew, L., & H. Yamagishi., (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*. 65, 15-31.
7. Bai, S. B.; Wang, J.; Lu, G. N.; Zhou, P. G.; Hou, S. S.; & S. N. Xu, 2010. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China, *Geomorphology*, 115, 23-31.
8. Boroghani, M.; Pourhashemi, S.; Zangane Asadi, M. A.; & H. Moradi, 2017. Dust source identification in the Middle East by using remote sensing, *Natural hazards environment magazine*, 66 (11), 101-118. (in Persian)
9. Breiman, L., 2001. Random forests, *Machine learning*, 45(1), 5-32.
10. Bühler, Y.; Kumar, S.; Veitinger, J.; Christen, M.; & A. Stoffel, 2013. Automated identification of potential snow avalanche release areas based on digital elevation models, *Natural Hazards Earth System Science*, 13, 13-21.
11. Calderoni, L.; Ferrara, M.; Franco, A.; & D. Maio, 2015. Indoor localization in a hospital environment using random Forest classifiers, *Expert System Apply*, 42 (1), 125-134.
12. Can, T.; Nefeslioglu, H. A.; Gokceoglu, C.; Sonmez, H.; & T. Y. Duman, 2005. Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses, *Geomorphology*, 72, 250-271.
13. Catani, F.; Lagomarsino, D.; Segoni, S.; & V. Tofani, 2013. Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues, *Natural Hazard Earth System Science*, 13, 2815-2831.
14. Chung, C. J. F., & A. G. Fabbri., (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards*. 30(3), 451-472.
15. Danyali, M.; MohamadNejad, B.; & N. Karimi, 2018. Spatial Analysis of Dust in Khuzestan Province Using Satellite Images, *Remote sensing and GIS in natural resources*, 9(1), 58-73.
16. Dube, F.; Nhapi, I.; Murwira, A.; Gumindoga, W.; Goldin, J.; & D. A. Mashauri, 2014. Potential of Weight of Evidence Modelling for Gully Erosion Hazard Assessment in Mbire District -Zimbabwe, *Physics and Chemistry of the Earth*, 67, 145-152.
17. Ebrahim Khani, S.; Afzali, M.; & A. Shokohi, 2012. Forecasting and studying the causes of road accidents using data mining algorithms, *Journal of Zanjan Law Enforcement*, 1(1), 111-127. (in Persian)
18. Farshad, M., & J. Sadeh., (2014). Short-circuit fault location in high voltage direct current transmission lines using neural networks, generalized regression and Random Forest algorithm. *Intelligent Systems in Electrical Engineering*. 4(2), 1-14. (in Persian)
19. Fayyad, U.; Piatetsky-Shapiro, G.; & P. Smyth, 1996. From data mining to knowledge discovery in databases, *Artificial Intelligence magazine*, 17(3), 37-54.
20. Flagg, C.; Neff, J.; Reynolds, R.; & J. Belnap, 2013. Spatial and temporal patterns of dust emissions (2004–2012) in semi-arid landscapes, southeastern Utah, USA, *Aeolian research*, 1-13. <http://dx.doi.org/10.1016/j.aeolia.2013.10.002>.
21. Garcia-Rodriguez, M. J.; Malpica, J. A.; Benito, B.; & M. Diaz, 2008. Susceptibility assessment of earthquake etriggered landslides in El Salvador using logistic regression, *Geomorphology*, 95, 172-193.

22. Golkarian, A.; Naghibi, S. A.; Kalantar, B.; & B. Pradhan, 2018. Groundwater potential mapping using C5. 0, random forest, and multivariate adaptive regression spline models in GIS, *Environmental monitoring and assessment*, 190(3), 149.
23. Goudie, A., 2014. Review Desert dust and human health disorders, *Environment International*, 63(3), 101-113.
24. Hahnenberger, M., & N. Kathleen., (2014). Geomorphic and land cover identification of dust sources in the eastern Great Basin of Utah, U.S.A. *Geomorphology*. 204(2), 657-672.
25. Hasselquist, N. J.; Germino, M. J.; Sankey, J. B.; Ingram, L. J.; & N. F. Glenn, 2011. Aeolian nutrient fluxes following wildfire in sagebrush steppe: implications for soil carbon storage, *Biogeosciences*, 8(11), 3649-3659.
26. Hong, H.; Naghibi, S. A.; Pourghasemi, H. R.; & B. Pradhan, 2016. GIS-based landslide spatial modeling in Ganzhou City, China, *Arabian Journal of Geoscience*, 9, 112. <http://dx.doi.org/10.1007/s12517-015-2094-y>.
27. Lashkari, H., & G. H. Keykhosravi., (2009). Synoptic analysis of dust storms Razavi Khorasan Province in the Period of (1993-2005). *Physical Geography Research*. 65(3), 17-33.
28. Lee, S., & T. Sambath., (2006). Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology*. 50, 847-855.
29. Manap, M. A.; Nampak, H.; Pradhan, B.; Lee, S.; Sulaiman, W. N. A.; & M. F. Ramli, 2014. Application of Probabilistic based Frequency Ratio Model in Groundwater Potential Mapping Using Remote Sensing Data and GIS, *Arabian Journal of Geosciences*, 7(2), 711-724.
30. Martinez-Garcia, A.; Rosell-Mele, A.; Jaccard, S. L.; Geibert, W.; Sigman, D. M.; & G. H. Haug, 2011. Southern Ocean dust-climate coupling over the past four million years, *Nature*, 476(7360), 312-315.
31. Moghimi, E.; Alavi panah, K.; & T. Jafari, 2011. Assessment and zonation factors on the northern of landslides Aladagh, *Geographical research*, 40, 53-75. (in Persian)
32. Mohamadi, M., & H. R. Pourghasemi., (2018). Prioritizing Factors Affecting Landslide and Preparing Its Sensitivity Map Using Random Forest Algorithm (Case Study: Part of Golestan Province). *Watershed Management Research*. 8(15), 161-170. (in Persian)
33. Moradi, H. R.; Dashti Marvili, M.; & A. Ildromi, 2012. Prioritize factors affecting landslides using logistic regression (Case Study: Black Canyon watershed Hamadan), *Iranian Journal of Watershed Management Science and Engineering*, 19(6), 65-70.
34. Mosavi Baygi, M., & B. Ashraf., (2011). Study of leading to drought of autumn and winter synoptic patterns in Khorasan Razavi Province. *Journal of soil and water conservation*. 18(4), 184-167. (in Persian).
35. Namdari, S.; Karimi, N.; Sorooshian, A.; Mohamadi, G. H.; & S. Sehatkashani, 2018. Impacts of climate and synoptic fluctuations on dust storm activity over the Middle East, *Atmospheric Environment*, 173, 265-276. <https://dx.doi.org/10.1016%2Fj.atmosenv.2017.11.016>.
36. Nandi, A., & A. Shakoor., (2009). A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Engineering Geology*. 110, 11-20.
37. Nicodemous, K. K., 2011. Letter to the Editor: On the stability and ranking of predictions from random forest variable importance measures. *Briefings in Bioinformatics*. 12, 369-373.
38. Pontius, R. G., & L. Schneider., (2001). Land use change model validation by a ROC method for the Ipswichwatershed Massachusetts, USA. *Journal of Agriculture, Ecosystems and Environment*. 85(30), 239-248.
39. Pourhashemi, S.; Boroghani, M.; Zangane Asadi, M. A.; & A. Amirahmadi, 2016. Analysis relation of vegetation cover on the number of dust event in khorasan razavi using geographic

- information system and remote sensing, *Remote sensing and GIS in natural resources*, 6(4), 33-45. (in Persion)
40. Pradhan, B., & S. Lee., (2010). Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environmental Earth Science*. 60, 1037-1054. <http://dx.doi.org/10.1007/s12665-009-0245-8>.
41. Prospero, J. M.; Ginoux, P.; Torres, O.; Nicholson, S. E.; & T. E. Gill, 2002. Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus 7 total ozone mapping spectrometer absorbing aerosol product, *Reviews of Geophysics*, 40(1), 2-31.
42. Qu, J.; Hao, X. J.; Kafatos, M.; & L. Wang, 2006. Asian Dust Storm Monitoring Combining Terra and Aqua MODIS SRB Measurements, *IEEE Geoscience and Remote Sensing Letters*, 3(4), 484-486.
43. Rashki, A.; Kaskaoutis, D. G.; Rautenbach, C.; Eriksson, P. G.; Qiang, M.; & P. Gupta, 2012. Dust storms and their horizontal dust loading in the Sistan region, Iran, *Aeolian Research*, 5(3), 51-62.
44. RazaviTermeh, S. V., & K. Shirani., (2018). Landslide hazard zoning using frequency ratio, entropy and TOPSIS decision making methods (Case Study: Fahlian Basin, Fars). *Remote sensing and GIS in natural resources*. 9(4), 119-138.
45. Roscovensky, J. K., & K. N. Liou., (2005). Differentiating airborne dust from cirrus clouds using MODIS data. *Geophysical Research Letters*. 32, L12809. Doi: 10. 1029/2005GL022798.
46. Sankey, J. B.; Germino, M. J.; Benner, S. G.; Glenn, N. F.; & A. N. Hoover, 2012. Transport of biologically important nutrients by wind in an eroding cold desert, *Aeolian Research*, 7(4), 17-27.
47. Schicker, R., & V. Moon., (2012). Comparison of bivariate and multivariate statistical approaches in landslide susceptibility mapping at a regional scale. *Geomorphology*. 40-57.
48. Silva, H. G.; Lopes, F. M.; Pereira, S.; Nicoll, K.; Barbosa, S. M.; Conceicao, R.; Neves, S.; Harrison, R. G.; & M. Collares Pereira, 2016. Saharan dust electrification perceived by a triangle of atmospheric electricity stations in Southern Portugal, *Journal of Electrostatics*, 84, 106-120. <https://doi.org/10.1016/j.elstat.2016.10.002>.
49. Vickery, K., & F. Eckardt., (2013). Dust emission controls on the lower Kuiseb River valley, Central Namib. *Aeolian Research*. 10(3), 125-133.
50. Wang, H., & T. Niu., (2013). Sensitivity studies of aerosol data assimilation and direct radiative feedbacks in modeling dust aerosols. *Atmospheric Environment*. 64(1), 208-218.
51. Ylmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey), *Computer & Geosciences*, 35, 1125-1138.
52. Zabihi, M.; Pourghasemi, H. R.; & M. Behzadfar, 2016. Preparation of Zizhmini Water Potential Map Using Shannon Entropy Models and Random Forest in Bojnourd Plain, *Journal of Ecohydrology*, 2(2), 221-233.
53. Zhang, B.; Tsunekawa, A.; & M. Tsubo, 2008. Contributions of sandy lands and stony deserts to long-distance dust emission in China and Mongolia during 2000-2006, *Global and Planetary Change*, 60(6), 487-504.

Susceptibility Zoning of Dust Source Areas by Data Mining Methods over Khorasan Razavi Province

Mahdi Boroughaniⁱ: Assistant Professor of Research Center for Geoscience and Social Studies, Hakim Sabzevari University

Sima Pourhashemi: Geomorphology PhD in Geographic Faculty of Hakim Sabzevari University

Article History (Received: 2016.9.16

Accepted: 2020.01.14)

Extended abstract

1- Introduction

Dust storms are natural hazards that effect on weather conditions, human health and ecosystem. Atmospheric processes are directly affected by the absorption and diffusion of radiation by dust, and dust in the cloud acts as a nucleus of congestion. The main dust areas in the world are flat topographically dry areas with erosion-sensitive soil and poor vegetation that is easily eroded by wind. Atmospheric processes are directly affected by the absorption and scattering of radiation by dust, and the dust in the clouds acts as a condensation nucleus. Iran is exposed to numerous local and trans-regional dust systems due to being in the arid and semi-arid belt of the world. Recent observations indicate that this phenomenon is increasing in the country and region under study. Khorasan Razavi province has arid and semi-arid climates with numerous dust storms occurring annually. In some years, more than 50 dust events have been recorded in the province. The large amount of dust in this area indicates large harvesting areas across the province, indicating the importance of research in this area. The purpose of this study was to evaluate the factors affecting the occurrence of dust and to prepare its susceptibility map in Khorasan Razavi province using logistic regression and random forest regression models.

2- Methodology

Khorasan Razavi Province with an area of 143864.118 square kilometers and with a population of 5.999529 is the fourth largest province in the country. The highest point of the province is in the highlands of Binalood, north of Neyshabur, at an altitude of 3211 meters above sea level, and the lowest point of the province is 300 meters above sea level, on the border of Iran and Turkmenistan. Dust indices include BTD_{2931} , D, NDDI and BTD_{3132} , which were calculated for all satellite images of these indices. Then there are four false color combination methods including 1- R: NDDI, G: $B4$, B: $B3$, 2- R: D, G: BTD_{2931} , B: NDDI, 3- R: D, G: BTD_{3132} , B: NDDI and 4- R: BTD_{2931} , G: $B4$, B: $B3$, were used to determine the best method for identifying dust source. Initially, 65 dust sources were identified in the study area and for the period 2005-2016. Of the identified focal points, 70% (46 dust source) were used for modeling and 30% (19 dust source) were used for evaluation. Then, 7 factors including soil, lithology, slope, vegetation index (NDVI), distance from the river, geomorphic units and land use as an independent and effective variable on dust were prepared. Then using logistic regression and random forest regression models, the weight of each of the factors in R software was calculated, finally, the dust hazard sensitivity maps were prepared.

3- Results

Four different false color combinations were used to detect dust areas in all images. Each of these methods revealed a dusty mass. The second method (R: D, G: BTD_{2931} , B: NDDI) is the best method for detecting dust and determining dust collection zones. Finally, by employing these techniques and creating false color combinations and applying them to existing images and their visual interpretation,

ⁱ Corresponding Author: m.boroughani@hsu.ac.ir

65 dust collection points were identified throughout the area. Based on the results, slope had the greatest impact on creating dust source zones. Geomorphology, lithology, vegetation index, soil and land use are the next influential categories of dust source area. Negative coefficients indicate the inverse relationship of the dependent variable (dust source) with the independent variables; in fact, negative coefficients do not indicate a lack of correlation, but mean a weaker correlation than other factors. The distance factor from the river has the least impact on the formation of these dust source. The OOB results show a predicted error rate of about 35.16%. In other words, the accuracy of the model can be considered 64.84%. The results of logistic regression algorithm showed that 24.5% of Khorasan Razavi province was in low sensitivity area, 28.2% in medium sensitivity area, 29.4% in high sensitivity area and 17.9% in high sensitivity area. It is very sensitive. The results show that the highest dust source points (57.9%) are located in high sensitivity area. The results indicate that about 84.2% of the dust collection points are located in the area of high and high sensitivity. Surveying the area of dust susceptibility map using random forest algorithm showed that the area with the highest sensitivity had the lowest area (16.7) and the areas of high, medium and low were 23.5, 31 and 8.8, respectively. 28 estimated. About 63.2% of source are located in the area of high sensitivity and there is no source for the area of low sensitivity.

4- Discussion & Conclusions

The results of both models show that slope and land use with the significant coefficients of 0.852 and 0.673 had the most effect on the creation of dust source area. The results also indicate that most dust removal sites (57.9%) are located in high sensitivity area. The results indicate that about 84.2% of the dust collection points are located in the area of high and high sensitivity. The evaluation of models using the ROC curve showed that in terms of success rate, logistic regression models and random forest had a precision of 781 and 907 respectively. Each of the two high correlation models between the sensitivity maps and the distribution of dust source areas can be observed. But in general, the random forest model has a higher accuracy than the other logistic regression models.

Key Words: Khorasan Razavi Province, Zoning, Random Forest, Dust source, Logistic Regression.