

ارایه روشی نوین جهت انتخاب بهینه شاخص‌های مرتبط با پوشش زمین به منظور شناسایی جزایر حرارتی شهری، با بکارگیری داده‌های سنجش از دور

حسین آقامحمدی زنجیرآباد^۲

مهدی رضانی^۴

نیکروز مستوفی^۱

علیرضا وفایی نژاد^۳

امیرهومن حمصی^۵

تاریخ پذیرش مقاله: ۹۸/۱۲/۰۹

تاریخ دریافت مقاله: ۹۸/۰۵/۱۳

چکیده

در تحقیقات اخیر، دانشمندان توجه ویژه‌ای به مسئله گرمایش جهانی داشته‌اند، زیرا دمای سطح زمین در طول قرن گذشته به طور قابل توجهی افزایش یافته است. جزایر حرارتی شهری به پدیده‌ای ناشی از آثار شهرنشینی اشاره دارد که درجه حرارت در محیط شهری از مناطق اطراف آن بالاتر می‌رود. بررسی این دما توسط سنسورها دارای مشکلاتی همچون هزینه و گسسته بودن نقاط اندازه‌گیری را دارد. بنابراین تحقیق حاضر تلاش می‌کند، با تکنیک سنجش از دور مدلی کمی و پیوسته را برای پوشش این مشکلات در شهر تهران ارائه دهد. لذا با استفاده از تصاویر لندست ۸^۶، داده‌های سنجنده مودیس، فاکتورهای تولید و بررسی می‌شوند که در تولید جزایر حرارتی شهری مؤثر هستند. به منظور تولید این فاکتورها ابتدا با انجام تصحیحات لازم بر روی تصاویر مورد نیاز، تعداد چهارده شاخص انتخاب و در سه سناریو مختلف محاسباتی شامل روش رگرسیون خطی، رگرسیون بردار پشتیبان و با استفاده از الگوریتم ژنتیک بکارگرفته شد. به منظور مدل‌سازی رویکردهای بیان شده، مجموعاً ۲۴۰۰ نقطه دارای دما به عنوان داده میدانی از منطقه مورد مطالعه (شهر تهران) جمع‌آوری شده است. برای ارزیابی کارایی سناریوهای مورد استفاده، ۳۰٪ داده‌ها (جمعاً ۷۲۰ نقطه) به صورت اتفاقی انتخاب شده و بعنوان داده‌های آموزشی در نظر گرفته و مابقی ۷۰٪ داده‌ها (جمعاً ۱۶۸۰ نقطه) به عنوان داده‌های تست مورد ارزیابی قرار گرفت. براساس نتایج بدست آمده، ترکیب مدل رگرسیون بردار پشتیبان و الگوریتم ژنتیک بهترین تطابق را (میانگین خطای مربعی ۰/۹۳۲۴، نرمال شده میانگین خطای مربعی ۰/۲۶۹۵ و ضریب همبستگی ۰/۹۳۱۵) با داده‌های زمینی مورد استفاده دارند.

واژه‌های کلیدی: جزایر حرارتی شهری، رگرسیون خطی، رگرسیون بردار پشتیبان، الگوریتم ژنتیک، تصاویر لندست ۸.

۱- دانشجوی دکتری تخصصی، گروه سنجش از دور و سیستم اطلاعات مکانی، دانشکده منابع طبیعی و محیط زیست، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران n-mostofi@azad.ac.ir

۲- استادیار عضو هیأت علمی، گروه سنجش از دور و سیستم اطلاعات مکانی، دانشکده منابع طبیعی و محیط زیست، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران (نویسنده مسئول) aghamohammadi@srbiau.ac.ir

۳- استادیار دانشکده عمران، گروه حمل و نقل، دانشکده عمران، آب و محیط زیست، دانشگاه شهید بهشتی، تهران، ایران a-vafaei@sbu.ac.ir

۴- استادیار عضو هیأت علمی، گروه سنجش از دور و سیستم اطلاعات مکانی، دانشکده منابع طبیعی و محیط زیست، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران dr.mramezani@yahoo.com

۵- استاد عضو هیأت علمی، گروه علوم و صنایع چوب و کاغذ، دانشکده منابع طبیعی و محیط زیست، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران h-hemmasi@srbiau.ac.ir

6- Urban Heat Islands

7- Landsat 8

۱- مقدمه

امروزه با بکارگیری علوم ثابت شده که تغییرات آب و هوایی زیادی در طول حیات کره زمین، صورت گرفته است و این تغییرات با مداخله بشر در طبیعت تسریع شده است. بر اساس این تغییرات مشکلات فراوانی در زمینه وقوع بلایای طبیعی از جمله گرم‌شدگی جهانی، بالا آمدن سطح آب‌ها، تخریب جنگل‌ها، سوراخ شدن لایه ازن، باران‌های اسیدی و کاهش تنوع گونه‌های زیستی بوجود آمده است (Zhu et al., 2019). یکی از نمودهای مهم این تغییرات و پیامدهای آن، بحث جزایر گرمایی است (Dihkan et al., 2015).

شهرها را از نظر دمایی می‌توان به جزیره‌ای تشبیه کرد که دمای هوای آن نسبت به چشم‌انداز اطراف آن بیشتر است و این امر، بیان‌کننده ایجاد جزیره حرارتی است (Mathew et al., 2018). جزیره گرمایی با شیب تند دمایی بین شهر و حومه آن مشخص می‌شود. تغییرات دمایی در این بخش از شهر از چهار درجه در هر کیلومتر نیز می‌تواند فراتر رود، در حالی که داخل شهر غالباً با شیب آرام و نامحسوس دما مشخص می‌گردد (Actionbioscience, 2015). به طور کلی شدت جزیره حرارتی هر شهر تابعی از عوامل مختلف مانند وسعت، جمعیت شهر، درجه ابرناکی و میزان وزش باد است (Streutker, 2002; Xian & Crane, 2006). بررسی تغییرات روزانه دمای هوا نشان دهنده‌ی این است که کاهش دما در ساعات بعدازظهر و نزدیک شب، در نواحی شهری با شدت کمتری نسبت به مناطق حومه‌ی شهر انجام می‌گیرد و این موضوعی است که باعث می‌شود درجه حداقل دمای هوا در شهر بالاتر از حومه آن باشد (Fabrizi et al., 2010; Huang & Wang, 2019).

مشکلات متعددی در سنجش دمای نقاط شهری بصورت نقطه‌ای وجود دارد. از جمله این مشکلات می‌توان به هزینه قابل توجه تهیه سنسورهای مورد نیاز سنجش حرارت شهری اشاره کرد.

از سوی دیگر، با فرض امکان اندازه‌گیری نقطه‌ای درجه حرارت مقادیر حاصل از این نوع اندازه‌گیری همواره ماهیت

گسسته دارند (Sanchez & Reames, 2019). لذا در این روش اندازه‌گیری، برای ایجاد مدلی پیوسته برای محدوده شهر همواره باید از روش‌های درون‌یابی استفاده نمود که این امر با محدودیت دقت همراه خواهد بود. با بکارگیری سامانه سنجش از دور^۱ و استفاده از تصاویر تکرارپذیر ماهواره‌ای امکان استخراج جزایر حرارتی و مناطق مستعد در یک دوره زمانی وجود دارد (Voogt & Oke, 2003).

از سوی دیگر، استخراج مناطق گرم در یک شهر به منظور نظارت و پایش محیط زیست کلان شهرها صورت می‌گیرد. تعیین اثر نوع پوشش و کاربری زمین بر دمای شهر و یافتن شاخص‌های مؤثر در این پدیده امری ضروری است (Wang et al., 2019). در این تحقیق، برای رسیدن به این شاخص‌ها، طبقه‌بندی پوشش اراضی شهری بر اساس تغییر در خصوصیات بیوفیزیکی (ویژگی‌های بیولوژیکی و خواص فیزیکی) آن، در سطح طبقه بندی پیکسل، پوشش گیاهی و سطوح نفوذناپذیر با استفاده از تصاویر لندست ۸ انجام خواهد شد (Abbassi et al., 2020). در نتیجه انتظار می‌رود با استفاده از این تکنیک بتوان روشی کارآمد جهت شناسایی جزایر گرمایی و همچنین تغییرات حرارتی کلان شهرها مانند شهر تهران معرفی نمود.

اخیراً، الگوریتم‌های کمی برای بررسی جزایر حرارتی شهری و فاکتورهای وابسته آن، مورد بررسی قرار گرفته است. برای مثال، رابطه این پدیده با انواع پوشش اراضی و مدل رگرسیون مربوط به آن در تحقیقات لیو و همکارانش (Liu et al., 2015)(Odindi et al., 2015) و (Xian & Crane, 2006) مورد بررسی قرار گرفته است. همچنین در تحقیقات مشابه مدل‌هایی از رابطه بین شاخص‌های مختلف پوشش گیاهی و دمای سطح ایجاد شده است (Xiong et al., 2012).

در این تحقیقات سعی شده است از یک روش کمی برای ردیابی رابطه بین شاخص‌های پوشش زمین و انتخاب شاخص‌های مناسب از جمله شاخص پوشش گیاهی NDVI (Kriegler et al., 1969) و شاخص بهبود پوشش گیاهی (EVI)

جدول ۱: مشخصات داده‌های ماهواره لندست ۸

سنجنده	باندها	طول موج (میکرومتر)	توان تفکیک مکانی (متر)	
OLI	باندها ۱	Coastal aerosol	0.43-0.45	30
	باندها ۲	Blue	0.45-0.51	30
	باندها ۳	Green	0.53-0.59	30
	باندها ۴	Red	0.64-0.67	30
	باندها ۵	(Near Infrared (NIR	0.85-0.88	30
	باندها ۶	SWIR 1	1.57-1.65	30
	باندها ۷	SWIR 2	2.11-2.29	30
	باندها ۸	Panchromatic	0.50-0.68	15
	باندها ۹	Cirrus	1.36-1.38	30
TIRS	باندها ۱۰	Thermal Infrared 1	1.36-1.38	100
	باندها ۱۱	Thermal Infrared 2	11.50-12.51	100

صورت مرحله به مرحله توضیح داده می‌شود.

۲-۱- داده‌های سنجنش از دوری مورد نیاز

در این پژوهش از دو دسته داده ماهواره‌ای لندست ۸ و مودیس به منظور استخراج و نظارت بر جزایر حرارتی در شهر تهران استفاده شده است. ماهواره لندست ۸ در فوریه سال ۲۰۱۳ در مدار زمین قرار گرفت، این هفتمین ماهواره لندست است که با موفقیت در مدار زمین قرار گرفت. لندست ۸ دارای ۱۱ باندها در طول موج مرئی، مادون قرمز نزدیک، مادون قرمز کوتاه و مادون قرمز حرارتی می‌باشد، که به طور متوسط معمولاً ۴۰۰ تصویر در هر روز از کره زمین برداشت می‌کند. توان تفکیک مکانی ۳۰ متر برای محدوده مرئی، فروسرخ نزدیک، فروسرخ کوتاه و توان تفکیک مکانی ۱۰۰ متر برای فروسرخ حرارتی و یک باندها پانکروماتیک با توان تفکیک مکانی ۱۵ متر دارد (USGS, 2015) (جدول ۱). لذا به علت توان تفکیک مناسب در باندها حرارتی، تصاویر این سنجنده در تحقیق حاضر مورد استفاده قرار گرفت.

باندهای حرارتی حاصل از سنجنده فروسرخ حرارتی آن (TIRS^۱) ۱۰ و ۱۱ در ارائه درجه حرارت دقیق‌تر از سطح زمین مفید هستند و در ابعاد ۱۰۰ متری بر روی زمین جمع

بهره گرفته شود ولی در هیچیک از مطالعات مذکور میزان تأثیر و اهمیت شاخص‌های مورد بررسی با توجه به ساختار منطقه مورد مطالعه ارزیابی نشده است.

هدف اصلی این مقاله ایجاد چارچوبی برای محاسبه جزایر حرارتی شهری بهینه با استفاده از شاخص‌های مناسب برای پوشش سطح زمین از تصاویر Landsat 8 است. در تحقیقات اشاره شده آثار شاخص‌های متعدد مطرح، مورد بررسی قرار گرفته ولی اثر کلی این شاخص‌ها مورد بررسی قرار نگرفته است. لذا در این تحقیق سه رویکرد مختلف مورد بررسی قرار گرفته و پیاده‌سازی شده است: الف) ترکیب مدل رگرسیون خطی با مجموعه کامل ویژگی‌ها و بدون انتخاب ویژگی. ب) استفاده از مدل رگرسیون بردار پشتیبان با مجموعه ویژگی‌های کامل و ج) ترکیب ویژگی‌های مناسب انتخاب شده به روش بهینه‌سازی ژنتیک با روش رگرسیون بردار پشتیبان. از نتایج این مطالعه می‌توان به افزایش عملکرد خروجی برآورد جزایر حرارتی شهری با استفاده از تصاویر لندست ۸ اشاره کرد.

۲- مبانی نظری

در این بخش مراحل محاسبه دمای سطح با استفاده از باندهای حرارتی سنجنده TIRS در ماهواره لندست ۸ به

1- Thermal Infrared Sensor

دست آید (Jimenez-Munoz et al., 2014; Mathew et al., 2019).

۲-۲-۱- دمای روشنایی (حرارت ظاهری)

پس از انجام تصحیحات هندسی و رادیومتریک، با اعمال تصحیحات اتمسفری، برای محاسبه دمای روشنایی می‌توان از رابطه‌ی (۱) استفاده کرد، در این رابطه T_R دمای روشنایی بر حسب کلون و مقادیر K_1 و K_2 در جدول شماره ۲ ارائه شده است (رابطه ۱).

$$T_R = \frac{K_2}{\log(K_1 / C_R + 1)} \quad \text{رابطه ۱}$$

جدول ۲: ضریب ثابت محاسبه دما در باندهای حرارتی

Landsat-8

شماره باند	K_1	K_2
باند ۱۰	۷۷۴/۸۹	۱۳۲۱/۰۸
باند ۱۱	۴۸۰/۸۹	۱۲۰۱/۱۴

سنجنده‌ی TIRS دو باند حرارتی (باند ۱۰ و باند ۱۱) را برای محاسبه دمای سطح در اختیار قرار می‌دهد. از این رو در تحقیق حاضر جهت برآورد بهینه و مناسب دمای سطح از روش Split-Window استفاده می‌شود. این روش به صورت رابطه ۲ است (Skokovic et al., 2014). با توجه به تحقیقات اخیر انجام شده در زمینه تولید مستقیم دمای سطح زمین^۱ (LST) با استفاده از تصاویر لندست ۸ توسط خیمنز و همکاران و با بکارگیری دو باند حرارتی ۱۰/۹ و ۱۲/۰ میکرومتر این ماهواره در روش Split-Window دمای سطح زمین به صورت مستقیم و بدون نیاز به مقدار رطوبت منطقه، استخراج می‌شود (Jimenez-Munoz et al., 2014).

رابطه (۲)

$$LST = C_0 + T_{b10} + C_1(T_{b10} - T_{b11}) + C_1(T_{b10} - T_{b11})^2 + (C_3 - C_4\omega)(1 - \varepsilon) + (C_5 - C_6\omega)\Delta\varepsilon$$

در رابطه (۲) ضرایب C_i به صورت جدول شماره ۳ معین می‌شود. در رابطه فوق T_{b10} و T_{b11} حرارت ظاهری بر حسب کلون و اندازه‌گیری شده در سنجنده است. بعلاوه مقدار ω برابر با مقدار آب محتوی اتمسفر با واحد ($g \cdot cm^{-2}$)

آوری می‌شوند ولی عرضه آنها از سایت‌های مرتبط با ابعاد ۳۰ متری در باندهای حرارتی می‌باشد. اندازه هر تصویر برداشت شده بر روی زمین تقریباً ۱۷۰ کیلومتر از شمال به جنوب، ۱۸۳ کیلومتر از شرق به غرب است. توان تفکیک زمانی این ماهواره برای هر منطقه دقیقاً ۱۶ روز می‌باشد، این بدین معنا است که اگر قرار به مطالعه دوره‌ای منطقه باشد، می‌توان هر ۱۶ روز یکبار از منطقه مورد نظر تصویر داشت و به مطالعه مورد نظر پرداخت. با بررسی به عمل آمده مشخص شده است که تهران در $Path=164$ و $Row=35$ قرار می‌گیرد. لذا تمامی تصاویر اخذ شده در این منطقه مورد مطالعه با شرط عدم وجود ابر، دانلود و مورد مطالعه قرار گرفت.

مودیس سنجنده‌ای است که بر روی ماهواره ترا EOS AM-1 قرار داده شده است. در هر کدام از ماهواره‌های EOS PM و EOS AM یک سنجنده مودیس قرار داده شده است (MOD11A2, 2015). سنجنده مودیس ماهواره ترا، هر دو روز یکبار از تمام سطح زمین تصویربرداری می‌کند و داده‌هایی در ۳۶ باند طیفی اخذ می‌نماید.

این داده‌ها، همراه با داده‌هایی که توسط مودیس دوم (که بر روی ماهواره آکوا) قرار داده شده است، اخذ می‌شوند. با توجه به اینکه داده‌های این سنجنده، یک منبع مناسب جهت مطالعات در مناطق بزرگ می‌باشند، لذا در این تحقیق از داده‌ها و تصاویر هم زمان این سنجنده با تصاویر لندست استفاده شده است، با این تفاوت که توان تفکیک زمانی سنجنده مودیس هر روزه می‌باشد، یعنی از یک مکان جغرافیایی می‌توان نظارت هر روزه با استفاده از تصاویر سنجنده مودیس داشت (Imhoff et al., 2010).

۲-۲-۲- شاخص‌های مورد نیاز

در این بخش ابتدا عوامل مختلف تأثیرگذار بر افزایش یا کاهش درجه حرارت سطح زمین بررسی می‌شود. سپس با در نظر گرفتن عواملی که بیشترین تأثیر را بر جزیره گرمایی دارند، تلاش می‌شود رابطه بین آنها و دمای سطح زمین به

1- Land Surface Temperature

فصلنامه علمی - پژوهشی اطلاعات جغرافیایی (SID)
 ارایه روشی نوین جهت انتخاب بهینه شاخص‌های... / ۶۱

حرارتی در مناطق کوچک است. عوارض گوناگون سطح و اینرسی حرارتی زمین ویژگی‌های فیزیکی خاصی دارند که به شدت بر دمای آنها اثرگذار است (Zohary & Sharifi, 2011). در شهرهایی که تغییرات توپوگرافی زیادی ندارند، پوشش سطح زمین بیشترین اثر را در ایجاد اختلاف دمای شهر نسبت به محیط اطراف آن می‌گذارد (Odindi et al., 2015). بنابراین می‌توان رابطه‌ای بین دما و پوشش سطح زمین به دست آورد که کمک می‌کند نقش انواع عوارض در ایجاد جزیره گرمایی بهتر بررسی شود (Mijani et al., 2019). از آنجا که پوشش سطح زمین نقش زیادی در ایجاد تنوع حرارتی دارد، اهمیت تفکیک‌پذیری مکانی سنجنده‌ها در دقت برآورد دما بیشتر می‌شود. هرچه قابلیت سنجنده در شناسایی عوارض کوچکتر بیشتر باشد، بهتر می‌توان رابطه دما و پوشش سطح زمین را تعیین کرد (Jiang & Tian, 2010). یکی از مهمترین روش‌های تخمین جزایر حرارتی شهری استفاده از شاخص‌های گیاهی و شهری به منظور مرتبط ساختن تغییرات این شاخص‌ها با دمای سطح بخصوص جزیره حرارتی از نوع SHI است. لذا در این تحقیق سعی شده است مهمترین شاخص‌های گیاهی و شهری مورد بررسی قرار گیرد (Henao et al., 2020, p.).

که بنا به تحقیقی که توسط خیمنز و همکارانش انجام شده است مقدار آن برابر با سه در نظر گرفته می‌شود (Jimenez Munoz et al. 2014). همچنین ϵ مقدار گسیلمندی میانگین و $\Delta\epsilon$ نیز مقدار تفاضلی گسیلمندی با استفاده از این دو باند حرارتی می‌باشند (رابطه ۷) (Latif, 2014).

جدول ۳: ضرایب معادله Split-window

ضرایب	c_0	c_1
مقدار	- ۰/۲۶۸	۱/۳۷۸

همچنین در رابطه فوق مقادیر ϵ و $\Delta\epsilon$ از به ترتیب از رابطه شماره ۳ و ۴ محاسبه می‌شوند (Skokovic et al, 2014; Sobrino et al, 1996; 2003; Shaouhua Zhao et al, 2009)

$$\epsilon = 0.5(\epsilon_{b10} + \epsilon_{b11})$$

رابطه ۳

$$\Delta\epsilon = (\epsilon_{b10} - \epsilon_{b11})$$

رابطه ۴

لازم به ذکر است که در این تحقیق، برای تخمین دو مقدار گسیلمندی یعنی مقادیر ϵ_{b10} و ϵ_{b11} از داده‌های همزمان اخذ شده توسط سنجنده مودیس استفاده شده است.

۲-۲-۲- عوامل تأثیرگذار بر دمای سطح زمین

پوشش سطح زمین مهمترین عامل به وجود آورنده تنوع

جدول ۴: شاخص‌های شهری قابل استخراج از تصاویر لندست ۸

شاخص شهری	رابطه
Normalized Difference Bareness Index (NDBaI)	$NDBaI = \frac{SWIR1 - TIRS1}{SWIR1 + TIRS1}$
Normalized Difference Build-up Index (NDBI)	$NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR}$
Bare Soil Index (BI)	$BI = \frac{(SWIR1 + RED) - (NIR + BLUE)}{(SWIR1 + RED) + (NIR + BLUE)}$
Urban Index (UI)	$UI = \frac{SWIR2 - NIR}{SWIR2 + NIR}$
Index-based Built-Up Index (IBI)	$IBI = \frac{\frac{2 \times SWIR1}{SWIR1 + NIR} - (\frac{NIR}{NIR + RED} - \frac{GREEN}{GREEN + SWIR1})}{\frac{2 \times SWIR1}{SWIR1 + NIR} + (\frac{NIR}{NIR + RED} - \frac{GREEN}{GREEN + SWIR1})}$
Enhanced Built-Up and Bareness Index (EBBI)	$EBBI = \frac{SWIR1 - NIR}{10\sqrt{SWIR1 + TIRS1}}$

جدول ۵: شاخص‌های گیاهی قابل استخراج از تصاویر لندست ۸

شاخص گیاهی	رابطه
Normalized Difference Vegetation Index (NDVI)	$NDVI = \frac{NIR - RED}{NIR + RED}$
Enhanced Vegetation Index (EVI)	$EVI = G \times \frac{NIR - RED}{NIR + C_1 \times RED - C_2 \times BLUE + L}$ $L = 1; C_1 = 6; C_2 = 7.5; G = 2.5$
Soil Adjusted Vegetation Index (SAVI)	$SAVI = \frac{NIR - RED}{NIR + RED + L} \times (L + 1)$ $0 < L < 1 \Rightarrow L = 0.5$
Normalized Difference Water Index (NDWI)	$NDWI = \frac{NIR - SWIR1}{NIR + SWIR1}$
Modified Normalized Difference Water Index (MNDWI)	$MNDWI = \frac{GREEN - NIR}{GREEN + NIR}$
Tasseled Cap Transformation (TCT)	Brightness
Tasseled Cap Transformation (TCT)	Greenness
Tasseled Cap Transformation (TCT)	Wetness

رفلکتانس مرتبط که در هر دو صورت قابل محاسبه خواهد بود (Gao, 1996). لیست مهمترین شاخص‌های گیاهی استفاده شده در این تحقیق در جدول ۵ ارائه شده است.

۳- روش شناسی تحقیق

در این بخش مفاهیم ریاضی مدل‌های استفاده شده در پژوهش به ترتیب تشریح می‌گردد.

۳-۱- مدل رگرسیون خطی

با فرض داشتن داده‌هایی آماری با فرمت $\{y_i, x_{i1}, \dots, x_{ip}\}_{i=1}^n$ در مدل رگرسیون خطی (LRM¹) فرض بر این است که ارتباطی بین متغیر مستقل y_i و بردارهای x_i پتابی به صورت خطی است. این ارتباط با استفاده از توزیع متغیر خطا (که ناشی از متغیر تصادفی بعلاوه نویز موجود در داده‌ها) با ارتباط خطی بین متغیر مستقل (در اینجا SHI) و عوارض و شاخص‌های استخراج

۲-۲-۱- شاخص‌های شهری

شاخص‌های شهری با بهره‌گیری از خصوصیات مختلف طیفی و نیز ساختاری سعی بر شناسایی محل‌های ساخته شده و به طور کلی دستخوش تغییر در مناطق شهری دارد. بعضی از این شاخص‌ها از مقادیر سطوح خاکستری برای محاسبه مقادیر خود استفاده می‌نمایند و بعضی دیگر از مقادیر رادیانس یا رفلکتانس مرتبط که در هر دو صورت قابل محاسبه خواهد بود. مهمترین شاخص‌های شهری استفاده شده در این تحقیق در جدول ۴ ارائه شده است (As-syakur et al., 2012).

۲-۲-۲- شاخص‌های گیاهی

شاخص‌های گیاهی با بهره‌گیری از خصوصیات مختلف طیفی و نیز ساختاری سعی بر شناسایی سطوح با پوشش گیاهی مختلف در مناطق شهری دارد. بعضی از این شاخص‌ها از مقادیر سطوح خاکستری برای محاسبه مقادیر خود استفاده می‌نمایند و بعضی دیگر از مقادیر رادیانس یا

1- Linear Regression Model

فصلنامه علمی - پژوهشی اطلاعات جغرافیایی (۱۳۸۳)
 ارایه روشی نوین جهت انتخاب بهینه شاخص‌های... / ۶۳

می‌گیریم:
 رابطه ۸ $\{(x_1, y_1), \dots, (x_l, y_l)\}$, $x_i \in \mathbb{R}^n, y \in \mathbb{R}$

تابع رگرسیون به وسیله تابع رابطه ۹ تخمین زده می‌شود.

رابطه ۹ $f(x) = \langle w, x \rangle + b$

که $\langle \cdot \rangle$ ضرب داخلی می‌باشد. تابع بهینه رگرسیون به وسیله کمینه تابع رابطه ۱۰ بیان می‌گردد.

$$\min_{\omega, b, \xi, \xi^*} \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$

رابطه ۱۰ subject to $\begin{cases} \langle w, x \rangle + b - y_i \leq \varepsilon + \xi_i \\ \langle w, x \rangle + b - y_i \geq -\varepsilon - \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{cases}$

که C یک مقدار از پیش تعیین شده بوده ξ_i و ξ_i^* متغیرهای شل هستند که قیود بالا و پایین خروجی سیستم را تعیین می‌کنند. اگر داده‌ها به صورت خطی مجزا از هم باشند، یک سطح بهینه که داده‌ها را بدون خطا و با حداکثر فاصله میان صفحه و نزدیکترین نقاط آموزشی (بردارهای پشتیبان) تفکیک می‌نماید، آموزش می‌دهد (Drucker et al., 1997). اگر نقاط آموزشی را به صورت $[x_i, y_i]$ و بردار ورودی $x_i \in \mathbb{R}^n$ تعریف کنیم، در حالتی که داده‌ها به صورت خطی قابل تفکیک باشند، معادله به صورت رابطه ۱۱ است.

رابطه ۱۱ $y = f(x) = \text{sign} \left[\sum_{i=1}^N y_i a_i \langle x_i, x \rangle + b \right]$

که در آن y خروجی معادله و yi ارزش طبقه نمونه آزمایشی xi می‌باشد. بردار $x = (x_1, x_2, \dots, x_n)$ نشان دهنده یک داده ورودی و بردارهای $x_i, i=1, 2, \dots, N$ بردارهای پشتیبان هستند (Moser & Serpico, 2009).

۳-۳- الگوریتم ژنتیک

الگوریتم ژنتیک یکی از زیرمجموعه‌های محاسبات تکامل یافته می‌باشد که رابطه مستقیمی با مبحث هوش مصنوعی دارد (Goldberg & Holland, 1988). الگوریتم ژنتیک را می‌توان یک روش جستجوی کلی نامید که از قوانین تکامل بیولوژیک طبیعی تقلید می‌کند. الگوریتم ژنتیک بر

شده (x_i) انجام می‌شود (Kutner et al., 2004). این مدل به شکل رابطه ۵ است.

رابطه ۵ $y_i = \beta_0 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i = X_i^T \beta + \varepsilon_i, i = 1, \dots, n$

که در رابطه ۵، T علامت ترانپوز و مقدار $X_i^T \beta$ ترکیب خطی بین متغیرهای ورودی x_i و ضرایب β می‌باشد. اگر تمامی فرمول‌ها با هم نوشته شوند، شکل برداری و ماتریسی آن به شکل رابطه ۶ خواهد شد.

رابطه ۶ $y = X\beta + \varepsilon$

در این رابطه ماتریس X را در بعضی از مراجع ماتریس طراحی نیز می‌نامند. صحبت از اینکه کدامیک از متغیرها برای مدل‌سازی متغیر مستقل (در اینجا SHI) مناسب است، بستگی به تأثیر متغیرهای وابسته به یکدیگر و تأثیر مستقیم آن بر متغیر مستقل دارد. با حل رابطه ۸ با روش کمترین مربعات^۱ ضرایب رگرسیون (متغیر β) تخمین زده خواهد شد. تخمین آماری و استنتاج در رگرسیون خطی وابسته به مقادیر متغیر β است. مؤلفه‌های این بردار به مقادیر مشتقات جزئی از متغیر وابسته تفسیر خواهد شد. متغیرهای ε_i به ترم خطا یا نویز مشهور است.

۳-۲- مدل رگرسیون بردار پشتیبان^۲

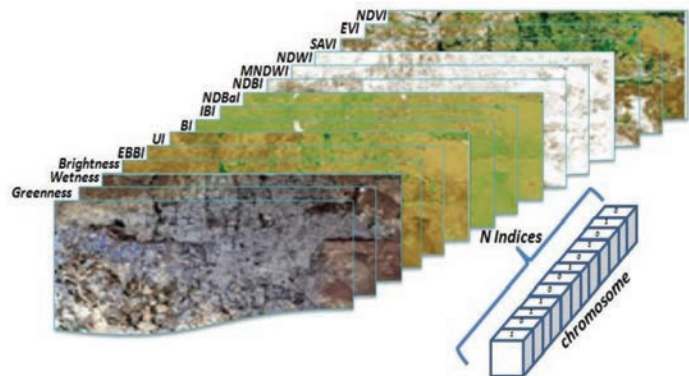
مدل رگرسیون ماشین بردار پشتیبان در حل مسائل پیش‌بینی و مدل‌سازی کاربرد دارد. ماشین بردار پشتیبان مبتنی بر کمینه کردن ساختاری ریسک^۳ می‌باشد که از تئوری آموزش آماری گرفته شده است (Drucker et al., 1997). این تابع به صورت رابطه ۷ تعریف می‌شود.

$$L(y, f(x, a)) =$$

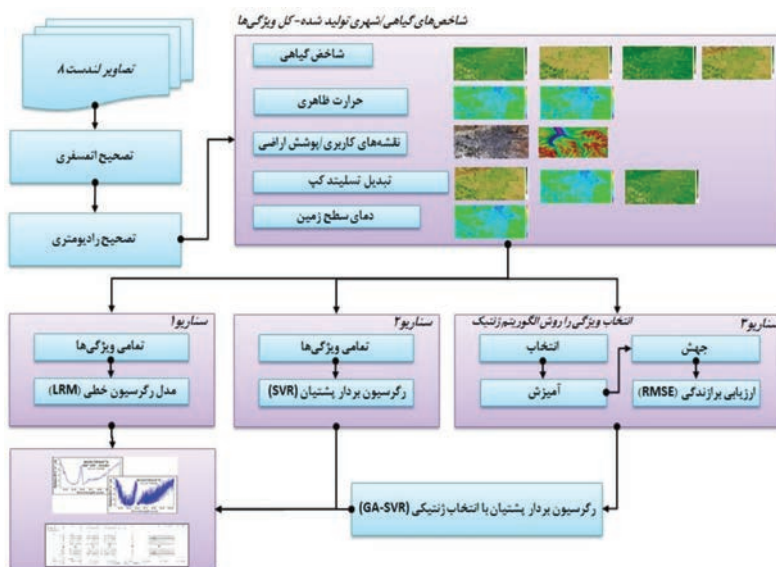
رابطه ۷ $|y - f(x, a)|_\varepsilon = \begin{cases} 0 & \text{for } |y - f(x, a)| \leq \varepsilon \\ |y - f(x, a)| - \varepsilon & \text{for } |y - f(x, a)| > \varepsilon \end{cases}$

این تابع خطا، مقادیر خطای کمتر از ε را در نظر نمی‌گیرد. مسأله تقریب یک مجموعه از داده‌های رابطه ۸ را در نظر

- 1- Least square
- 2- Support Vector Regression (SVR)
- 3- Structural Risk Minimization (SRM)



نگاره ۱: نمایش کروموزم ژنتیکی به منظور انتخاب شاخص‌های مناسب با استفاده از الگوریتم ژنتیک



نگاره ۲: روندنمای کلی انجام تحقیق

بار تکرار می‌شود تا نسل بعدی جمعیت تولید شود. سپس این جمعیت بررسی می‌شود و در صورتی که ضوابط همگرایی برآورده شوند، فرآیند فوق خاتمه می‌یابد. با استفاده از الگوریتم ژنتیک در مسأله بهینه‌سازی انتخاب شاخص‌های مناسب به منظور مدل‌سازی SHI بدین صورت فرض می‌شود که تعداد N عارضه یا شاخص قرار است در مدل‌سازی شرکت داشته باشد (نگاره ۱). در این حالت یک کروموزم دودویی تشکیل شده است و وجود عدد یک به این معنا است که شاخص مورد نظر در محاسبات مدل‌سازی شرکت داده می‌شود و اگر عدد صفر بود شرکت داده نمی‌شود. در تحقیق حاضر، با استفاده از معیار خطای مربعی میانگین کمینه (RMSE^۵) به ارزیابی هریک از کروموزم‌ها پرداخته خواهد شد.

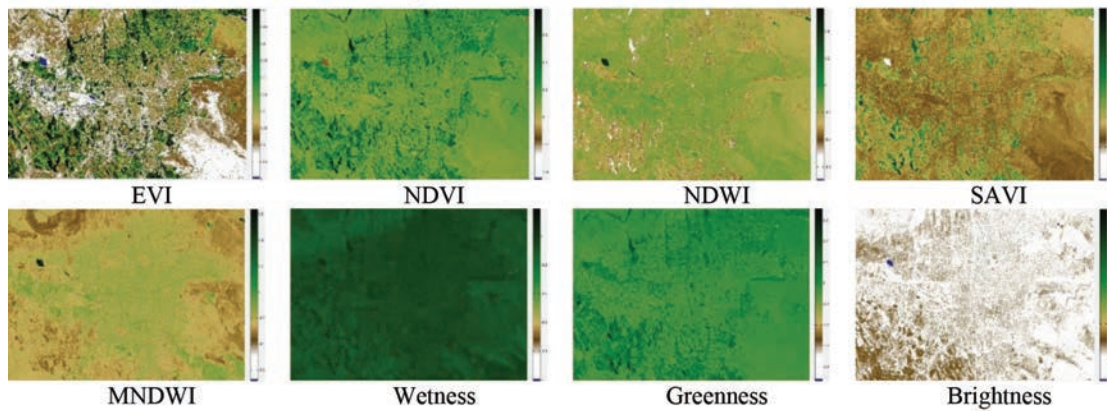
۴- پیاده سازی

5- Root Mean Square Error (RMSE)

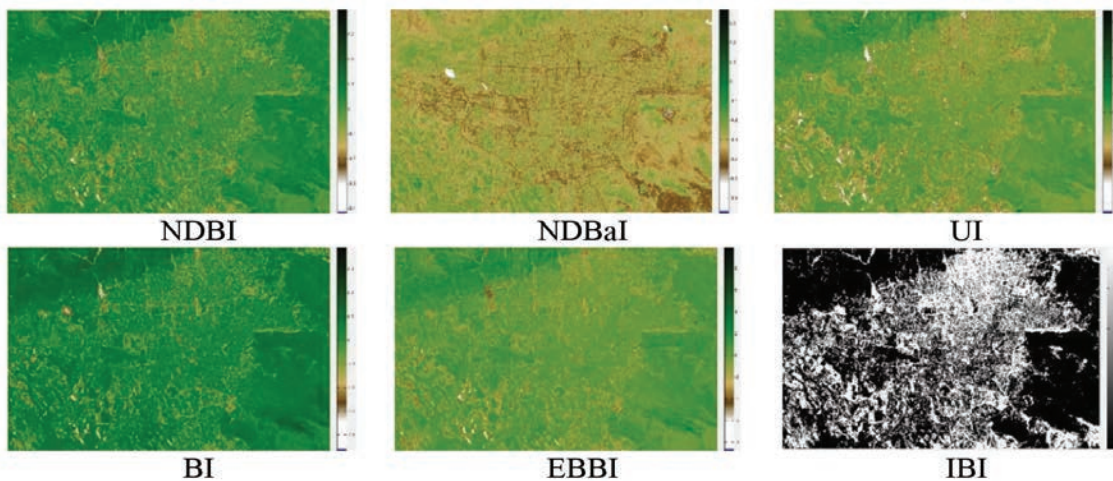
روی یکسری از جواب‌های مسأله به امید بدست آوردن جواب‌های بهتر قانون بقای بهترین را اعمال می‌کند. قبل از این که یک الگوریتم ژنتیک بتواند اجرا شود، ابتدا باید کدگذاری^۱ (یا نمایش) مناسبی برای مسئله مورد نظر پیدا شود. هر متغیر تصمیم‌گیری به صورت دودویی در آمده و سپس با کنار هم قرار گرفتن این متغیرها کروموزوم ایجاد می‌شود. در این روش لازم است یک تابع برازندگی^۲ ابداع شود تا به هر راه‌حل کدگذاری شده ارزشی را نسبت دهد. در طی اجرا، والدین برای تولید مثل انتخاب می‌شوند و با استفاده از عملگرهای آمیزش^۳ و جهش^۴ با هم ترکیب می‌شوند تا فرزندان جدیدی تولید کنند. این فرآیند چندین

- 1- Coding
- 2- Fitness Function
- 3- Cross Over
- 4- Mutation

فصلنامه علمی - پژوهشی اطلاعات جغرافیایی (SID)
 آرایه روشی نوین جهت انتخاب بهینه شاخص‌های... / ۶۵



نگاره ۳: شاخص‌های گیاهی تولید شده از تصاویر لندست ۸



نگاره ۴: شاخص‌های شهری تولید شده از تصاویر لندست ۸

تصاویر همزمان اخذ شده لندست ۸ و مودیس، پردازش‌هایی جهت تولید محصولات رادیومتریکی (رادینانس، رفلکتانس و حرارت ظاهری)، تولید انواع شاخص‌های گیاهی، تولید انواع شاخص‌های شهری و پردازش‌های دیگر بر روی آنها انجام گیرد. این رویه برای کلیه داده‌ها انجام شده و تعداد بسیار زیادی از محصولات مختلف (شاخص‌ها گیاهی و شهری، باندهای طیفی، باندهای رادینانسی، باندهای رفلکتانسی، دمای سطح زمین) تولید گردید.

با بکارگیری تصاویر همزمان سنجنده مودیس، امکان تخمین گسیلمندی دو باند ۱۰ و ۱۱ برای سنجنده TIRS ماهواره لندست ۸ (E_{b10} و E_{b11}) فراهم می‌گردد. با بکارگیری روش‌های اندازه‌گیری شاخص گیاهی که قبلاً در

در این مطالعه از تصاویر لندست ۸ اخذ شده از شهر تهران به عنوان مجموعه داده پایه محاسبات استفاده شده است. این مجموعه داده متعلق به نزدیکی فصل تابستان است، (بیست و پنجم خردادماه سال ۹۷) که دمای هوا نزدیک به ۴۰ درجه سانتیگراد بوده است. محاسبه درجه حرارت سطح از طریق تصویر لندست ۸ و تصویر همزمان سنجنده مودیس، جهت تخمین میزان انتشار دو باند حرارتی صورت می‌گیرد. با توجه به مفاهیم ریاضی E_{b11} و E_{b10} تبیین شده و نیز اهداف پژوهش با بکارگیری سه رویکرد محاسباتی، مطابق روند نمای نگاره ۲ اقدام گردید.

همانطور که گفته شده طبق نگاره شماره ۲ لازم است بعد از انجام تصحیحات اتمسفری و رادیومتری بر روی

$$SHI = \beta_1 + \beta_2 EVI + \beta_3 NDVI + \beta_4 NDWI + \beta_5 NDBI + \beta_6 NDBaI + \beta_7 MNDWI + \beta_8 BI + \beta_9 UI + \beta_{10} IBI + \beta_{11} EBBI + \beta_{12} SAVI + \beta_{13} Wetness + \beta_{14} Greenness + \beta_{15} Brightness$$

تخمین این ضرایب مجهول با استفاده از روش کمترین مربعات در جدول شماره ۷ ارائه شده است.

جدول ۶: مدل سازی داده‌ها با استفاده از روش رگرسیون خطی با ۱۴ شاخص (رویکرد اول)

	β_i	مقدار	خطای استاندارد	tStudent
Intercept	β_1	-۵۴/۴۴۸۶	۱۳/۸۲۲۸	-۳/۹۳۹
EVI	β_2	-۸/۲۷۹۷	۱/۵۰۰۴	-۵/۵۱۸۳
NDVI	β_3	-۲۹/۶۸۷۴	۳/۹۶۱۱	-۷/۴۹۴۷
NDWI	β_4	-۱۵/۹۵۳۲	۵/۷۰۵۶	-۲/۷۹۶۱
NDBI	β_5	-۲۳/۳۵۵۳	۱۱/۵۲۷۵	-۲/۰۲۶۱
NDBaI	β_6	-۵۵/۸۱۱۳	۰/۹۷۲۸	-۵۷/۳۷۴۶
MNDWI	β_7	-۹/۰۷۴۴	۶/۲۱۹	-۱/۴۵۹۱
BI	β_8	۰/۷۸۶۷	۲/۰۳۰۵	۰/۳۸۷۴
UI	β_9	۲۲/۰۳۵۶	۴/۴۰۰۱	۵/۰۰۸
IBI	β_{10}	-۰/۰۵۲۲	۰/۰۹۹۱	-۰/۵۲۶۷
EBBI	β_{11}	۷۹/۴۴۰۹	۱۳/۱۲۷۷	۶/۰۵۱۴
SAVI	β_{12}	۵۴/۹۱۱	۵/۶۴۵۱	۹/۷۲۷۲
Wetness	β_{13}	۵۶/۰۲۴۷	۸/۴۹۱	۶/۵۹۸۱
Greenness	β_{14}	۰/۶۹۹۵	۶/۰۱۵۲	۰/۱۱۶۳
Brightness	β_{15}	۴۶/۴۵۷۹	۳/۰۵۴۳	۱۵/۲۱۰۶

همانطور که از جدول ۶ مشخص است، مقادیر تخمین زده شده برای ضرایب (β_i) در ستون دوم، خطای استاندارد تخمین در ستون سوم، در ستون چهارم مقدار تخمین توزیع t-student آمده است. همچنین با بکارگیری ضرایب تخمین زده شده و ۷۰٪ مابقی داده‌ها (داده‌های تست) مدل را مورد ارزیابی قرار داده‌ایم. نتایج دقت مدلی روی داده‌های تست در جدول ۷ ارائه شده است.

جدول ۷: کارایی روش رگرسیون خطی (رویکرد اول)

مورد آنان صحبت شد، تعداد هشت شاخص گیاهی تولید گردید که در نگاره شماره ۳ نشان داده شده است.

همچنین با بکارگیری روابط مربوط به شاخص‌های شهری، تعداد شش شاخص شهری مطابق نگاره ۴ تولید گردید. به منظور مدل‌سازی در سه رویکرد پیش‌رو، مجموعاً ۲۴۰۰ نقطه به عنوان داده میدانی از منطقه مورد مطالعه (شهر تهران) جمع‌آوری شده است.

بنابراین برای اینکه بتوان کارایی الگوریتم مورد محاسبه را تخمین زد و نیز مقدار خطای احتمالی را مورد محاسبه قرار داد، ۳۰٪ داده‌ها (جمعاً ۷۲۰ نقطه) به صورت اتفاقی انتخاب شده و به عنوان داده‌های آموزشی در نظر گرفته می‌شود و مابقی ۷۰٪ داده‌ها (جمعاً ۱۶۸۰ نقطه) به عنوان داده‌های تست یا کنترل مورد ارزیابی قرار خواهد گرفت. همچنین در سه رویکرد پیش‌رو، از چهار معیار مقایسه مانند میانگین خطای مربعی^۱ (MSE)، ریشه میانگین خطای مربعی^۲ (RMSE)، ریشه نرمال شده میانگین خطای مربعی^۳ (NRMSE) و ضریب تعیین^۴ (R^2) برای بررسی نتایج بدست آمده استفاده شده است (Durbha et al., 2007).

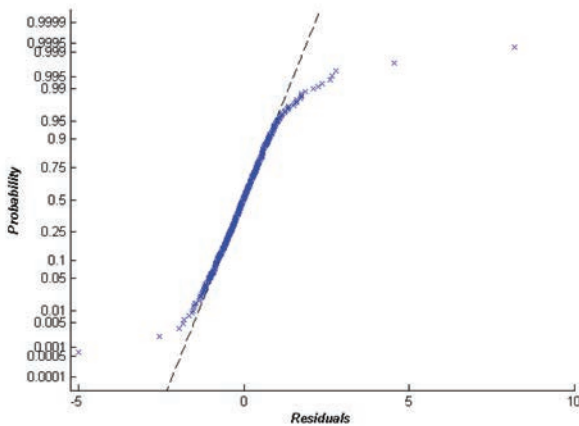
۴-۱- رویکرد اول

همانطور که از فلوچارت کلی الگوریتم در نگاره شماره ۲ نیز مشخص است، در این رویکرد با استفاده از مدل رگرسیون خطی (LRM) چهارده شاخص/ویژگی شامل EVI, NDVI, NDWI, NDBI, NDBaI, MNDWI, BI, UI, IBI, EBBI, SAVI, Wetness, Greenness, Brightness را در قالب رابطه ۱۲ بکارگرفته شده است. ضرایب مدل (مجهولات) با استفاده از داده‌های آموزشی (۳۰٪ داده‌ها) قابل تخمین خواهد بود.

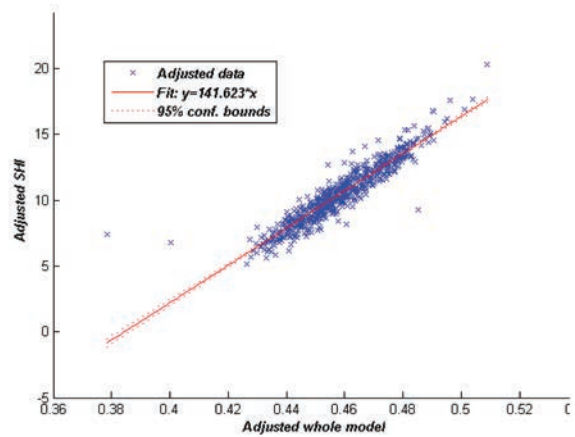
- 1- Mean Square Error
- 2- Root Mean Square Error
- 3- Normalized Root Mean Square Error
- 4- R-squared

رابطه ۱۲

فصلنامه علمی - پژوهشی اطلاعات جغرافیایی (SID)
 آرایه روشی نوین جهت انتخاب بهینه شاخص‌های... / ۶۷



الف



ب

نگاره ۵: الف) نمایش باقیمانده احتمال نرمال شده و ب) نمایش متغیر اضافه شده به تمامی مدل

در این رویکرد همانند رویکرد اول، از تمامی چهارده شاخص به عنوان ورودی استفاده می‌شود. ولی با این تفاوت که بجای مدل رگرسیون خطی از مدل رگرسیون بردار پشتیبان^۱ (SVR) استفاده می‌گردد. این روش سعی دارد تا بهترین تابع تابع با شرایط رابطه ۱۳ را به داده‌ها برازش دهد. رابطه ۱۳

$$SHI = f(NDVI, EVI, SAVI, NDWI, MNDWI, Brightness, Greenness, Wetness, NDBaI, NDBI, BI, UI, IBI, EBBI)$$

برای انجام محاسبات در روش SVR همانطور که قبلاً بیان شد، لازم است یکسری پارامتر حساس و مهم برای مدل تخمین زده شود تا بتوان به کارایی بالا دست یافت. پارامتر C یک موازنه بین صافی سطح و درجه تغییرات بزرگ ϵ به منظور بهینه‌سازی بهتر فرمول استفاده شده بکار گرفته می‌شود (رابطه ۱۴). در این صورت، اگر مقدار C بسیار بزرگ تخمین زده شود، هدف کمینه کردن ریسک تجربی^۲ بدون در نظر گرفتن قسمت صافی در فرمول بهینه‌سازی می‌باشد. مقدار بزرگ ϵ ، باعث می‌شود که بردارهای پشتیبان کمتری را شامل شود. بنابراین تخمین صافی زیاد، نتیجه اتخاذ مقادیر بزرگ برای ϵ است. در واقع دو متغیر C و ϵ در صافی مدل

نوع داده‌ها	MSE	RMSE	NRMSE	R2
داده‌های آموزشی	۰/۵۶۷۲	۰/۷۵۳۱	۰/۳۴۸۳	۰/۸۷۶۲
داده‌های تست	۰/۵۴۸۶	۰/۷۴۰۷	۰/۳۴۱۹	۰/۸۷۸۵

از جدول ۷ مشخص است که تطابق و کارایی بالایی در تخمین مدل وجود داشته است و ارتباط بین SHI و چهارده شاخص بکار گرفته شده با استفاده از داده‌های تست به RMSE حدود ۰/۷۴ درجه رسیده است. همچنین ضریب تعیین این مدل R^2 انطباق بسیار بالای ۰/۸۸ را نشان می‌دهد. از سوی دیگر، نمایش احتمال نرمال باقیمانده‌ها با استفاده از مدل برازش داده شده در نگاره ۵-الف و متغیرهای اضافه شده برای کل مدل برازش داده شده در نگاره ۵-ب ارائه شده است. نگاره ۵-ب نشان دهنده تأثیر پاسخ یکی از ورودی‌ها با حذف تأثیرات همه ورودی‌های دیگر است. شیب خط برازش داده شده، ضریب ترکیب خطی ورودی‌های مشخص منعکس شده در بهترین جهت برازش است. همچنین از نگاره ۵-ب مشخص است که مدل یک خط افقی بین محدوده اطمینان خود برازش نداده ولی با توجه به خروجی‌های استخراج شده از جدول ۷ نتایج قابل قبول است.

۴-۲- رویکرد دوم

1- Support Vector Regression

2- Empirical risk

جدول ۸: تخمین پارامترهای بهینه روش SVR با $C=22.40$ (رویکرد دوم)

	$\epsilon = 0$	1	2	3	4	5
$\gamma = 2^{-7}$	۸/۹۲۴۸	۸/۹۶۴۷	۹/۰۴۴	۹/۲۶۱۵	۹/۴۹۱۲	۹/۸۸۲۶
2^{-6}	۸/۱۹۳۱	۸/۳۰۲	۸/۴۳۶۵	۹/۸۲۵۱	۹/۰۹۵۹	۹/۶۶۰۱
2^{-5}	۷/۲۲۶۷	۷/۳۲۷۶	۷/۶۷۲	۸/۱۶۲۲	۸/۵۸۳۱	۹/۲۷۰۸
2^{-4}	۵/۸۵۲۲	۵/۹۹۴۲	۶/۶۱۹۵	۷/۲۶۶۸	۷/۹۷۷۸	۸/۷۶
2^{-3}	۳/۹۹۴۹	۴/۴۵۳۲	۵/۲۸۹۵	۶/۲۳۹۹	۷/۲۸۵۴	۸/۱۰۵
2^{-2}	۲/۳۷۲	۲/۹۳۹۸	۳/۹۷۴۲	۵/۲۴۲۷	۶/۳۹۷	۷/۴۰۹۹
2^{-1}	۱/۵۴۷۳	۲/۱۱۰۴	۳/۱۷۴	۴/۴۶۷۹	۵/۷۲۹۷	۶/۸۹۴۶
2^0	۱/۴۰۱۳	۱/۸۰۱	۲/۸۴۳۷	۴/۰۵۰۲	۵/۴۵۷۸	۶/۵۹۸
2^1	۱/۳۸۳۳	۱/۶۸۳۶	۲/۵۵۲	۳/۷۷۰۹	۵/۱۷۹۴	۶/۵۷۸۷
2^2	۱/۵۰۹۲	۱/۶۲۰۵	۲/۴۶۸۶	۳/۷۱۱۹	۵/۲۰۳۵	۶/۷۵۹
2^3	۱/۷۲۶۴	۱/۸۲۵۸	۲/۶۸۵۸	۳/۹۷۳۷	۵/۳۶۰۵	۷/۱۲۲۳
2^4	۱/۹۶۳۱	۲/۲۳۳۲	۳/۲۸۸۳	۴/۴۶۸	۵/۹۲۹۷	۷/۶۰۵۹
2^5	۲/۴۸۸۵	۲/۹۵۷۱	۴/۲۲۸۳	۵/۵۶۱۹	۶/۹۷۴۴	۸/۲۲۰۱
2^6	۳/۵۵۴	۴/۱۰۰۱	۵/۴۸۷۲	۶/۸۱۱۹	۸/۱۶۰۲	۹/۲۱۶۵
2^7	۵/۱۸۹۷	۵/۹۴	۷/۱۳۱	۸/۲۵۷۵	۹/۳۷۱۳	۱۰/۳۸۵۵

تخمین پارامتر ϵ بین ۰ تا ۵ و برای پارامتر هسته گوسی (γ) RBF بین 2^{-7} تا 2^7 است. لذا جدول ۸ به منظور انتخاب پارامترهای بهینه SVR با استفاده از تصویر و شاخص‌های مختلف استخراج شده است. از جدول ۸ مشخص است که پارامترهای بهینه برای تصویر لندست استفاده شده با مقادیر کمینه RMSE مقادیر $\epsilon = 0$ و $\gamma = 2$ و $C = 22/40$ می‌باشد. همچنین با استفاده از تخمین سه پارامتر مهم (ϵ و γ و C) با کمترین مقدار RMSE و استفاده از داده‌های آموزشی می‌توان روش SVR را آموزش داد تا مدل مناسبی تولید شود. کارایی و مدل نهایی SVR با استفاده از چهارده داده ورودی در جدول ۹ ارائه شده است.

جدول ۹: کارایی روش SVR (رویکرد دوم)

نوع داده‌ها	MSE	RMSE	NRMSE	R ²
داده‌های آموزشی	۰/۷۵۰۷	۰/۸۶۶۴	۰/۲۴۲۴	۰/۹۴۴۲
داده‌های تست	۱/۱۱۵۵	۱/۰۵۶۲	۰/۳۰۵۳	۰/۹۱۰۰

همانطور که مشخص است، انطباق بالایی بین اطلاعات

(پیچیدگی مدل) مؤثر است. در این تحقیق مقدار پارامتر C با استفاده از رابطه ۱۴ بر مبنای روابط مطالعات چرکاسکی و همکاران (Cherkassky and Ma 2004) محاسبه می‌شود.

$$C = \max(\text{Trainingdata}) - \min(\text{Trainingdata})$$

قبل از تخمین رگرسیون لازم است یک تغییر مقیاس/ نرمال‌سازی بر روی داده‌های آموزشی انجام شود. مهمترین مزیت استفاده از نرمال‌سازی، دوری از محاسبات با اعداد بسیار بزرگ در مقابل اعداد بسیار کوچک و کاهش پیچیدگی محاسبات است. مرحله بعد در روش SVR، مرحله آموزش است که قبل از این مرحله باید پارامترهای حیاتی روش SVR شامل پارامتر ϵ و γ به عنوان پارامتر هسته RBF تخمین زده شود. این کار با استفاده از روش جستجوی شبکه‌ای برای تخمین پارامتر ϵ و γ ، و با استفاده از روش خطای اعتبارسنجی متقابل^۱ (CV) با استفاده از معیار RMSE با پنج فولد انجام می‌شود. محدودیت جستجوی شبکه‌ای برای

1- Cross-Validation

فصلنامه علمی - پژوهشی اطلاعات جغرافیایی (SID)
 ارایه روشی نوین جهت انتخاب بهینه شاخص‌های... / ۶۹

پارامترها	مقادیر
Population size	۲۰
Crossover rate	۰/۷
Elitism Ratio	۱
Mutation Ratio	۰/۰۵
Crossover Method	دو نقطه ای
Max Iterations	۵۰
Elite count	۱

ب

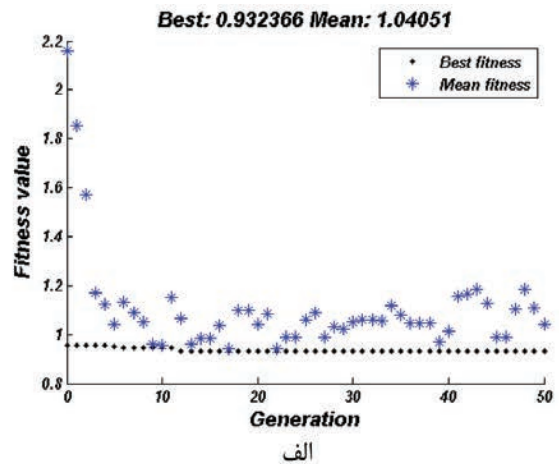
پارامترهای روش انتخاب ویژگی ژنتیک (GA-SVR، ب) که به عنوان ورودی به روش رگرسیون SVR داده می‌شود، با استفاده از الگوریتم ژنتیک و با روش کدگذاری که قبلاً بیان شد، بهترین ورودی‌ها برای ورود به روش مدل‌سازی SVR انتخاب شود و سپس فضای ورودی کاهش یافته برای تخمین درجه حرارت سطح استفاده گردد. نگاره ۶، نشان دهنده تمامی پارامترهای مهم استفاده شده در روش ژنتیک به منظور انتخاب ویژگی است. تعداد نمونه‌ها به صورت اتفاقی انتخاب می‌شوند، هرچند که کلیه نمونه‌ها تأثیر مستقیمی بر کارایی روش دارند. به منظور مطمئن شدن از قابلیت اطمینان نتایج بدست آمده از روش GA-SVR، این الگوریتم ۱۰ بار تکرار شده و ویژگی‌هایی که بیشتر تکرار شده‌اند انتخاب شده است.

خروجی روش GA-SVR، پنج شاخص بهینه از بین چهارده شاخص شامل شاخص‌های NDBI، NDBaI، BI، Greenness، Brightness می‌باشد که به عنوان ورودی‌های روش SVR در مدل‌سازی استفاده می‌شود (رابطه ۱۵).

رابطه (۱۵)

$$SHI = f(NDBI, NDBaI, BI, Greenness, Brightness)$$

با استفاده از شاخص‌های انتخاب شده، روند آموزش و تست در روش SVR آغاز می‌گردد. نتایج و کارایی این رویکرد (GA-SVR) با استفاده از داده‌های لندست ۸ از منطقه تهران در جدول ۱۰ ارائه شده است. از این جدول



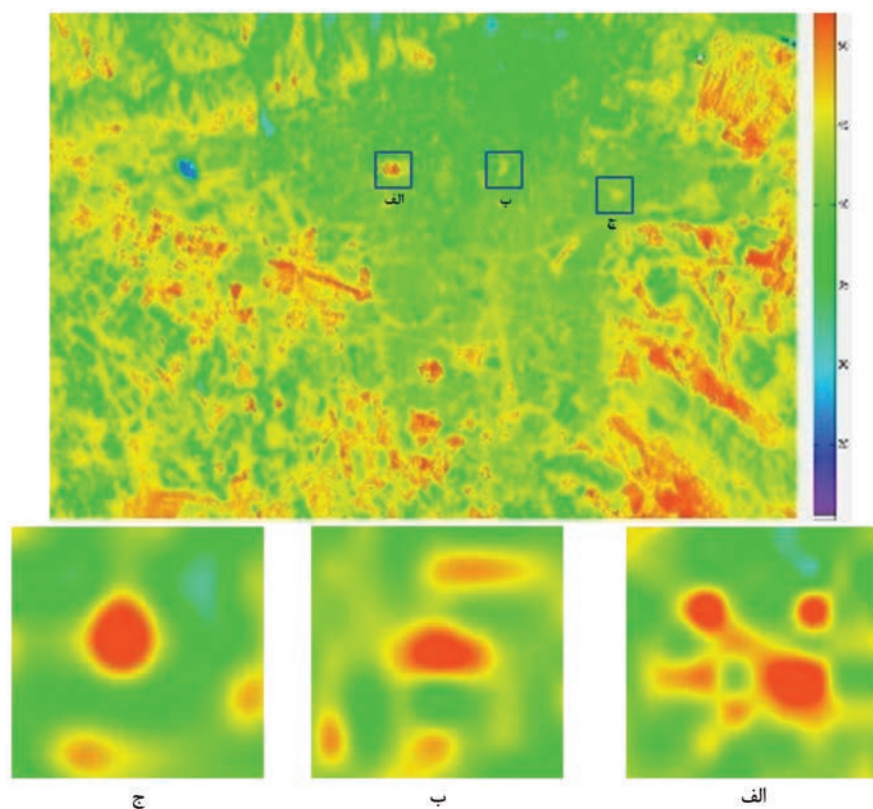
الف

نگاره ۶: الف) مقادیر تابع برازندگی با استفاده از روش استفاده شده در روش هسته مبنا و تخمین داده‌های SHI وجود دارد. برای مثال در رویکرد دوم، مقدار ضریب تعیین $R^2 = ۰/۹۴$ و بهبود مقدار NRMSE نسبت به رویکرد اول است.

۴-۳- رویکرد سوم

جهت کاهش هزینه محاسباتی مربوط به بررسی تمامی چهارده شاخص مؤثر ذکر شده و همچنین، حذف شاخص‌های کم اثر در بررسی جزایر حرارتی شهری، در این رویکرد از روش بهینه‌سازی ژنتیک استفاده می‌گردد. با استفاده از الگوریتم ژنتیک در مسأله بهینه‌سازی انتخاب شاخص‌های مناسب به منظور مدل‌سازی جزایر حرارتی بدین صورت فرض می‌شود که تعداد N عارضه یا شاخص قرار است در مدل‌سازی شرکت داشته باشد (نگاره شماره ۱). در این حالت یک کروموزم دودویی تشکیل شده است و وجود عدد یک بدین معنا است که شاخص مورد نظر در محاسبات مدل‌سازی شرکت داده می‌شود و اگر عدد صفر بود شرکت داده نمی‌شود. در این تحقیق با استفاده از معیار خطای مربعی میانگین کمینه ($RMSE^1$) به ارزیابی هریک از کروموزم‌ها پرداخته خواهد شد. در این رویکرد، از ترکیب مدل SVR و الگوریتم ژنتیک استفاده گردید (GA-SVR). در این مرحله، هدف تعیین مقادیر ورودی بهینه برای مسئله است، بدین معنا که از چهارده شاخص گیاهی و شهری

1- Root Mean Square Error (RMSE)



نگاره ۷: دمای سطح شهر تهران حاصل از رویکرد سوم و سه نقطه نمونه انتخابی الف، ب و ج از جزایر حرارتی

با دقت بیشتری انجام گیرد. بعلاوه اینکه بکارگیری معیار NRMSE به منظور مقایسه بین روش‌ها، تفاوت نتایج حاصل از سه رویکرد مورد استفاده را بهتر نمایان کرده است. با توجه به نتایج حاصل مشخص می‌گردد، رویکرد سوم بیان شده، کارایی بالاتری با توجه به تعداد شاخص‌های ورودی به منظور مدل‌سازی درجه حرارت سطح داشته است. همچنین استفاده از روش انتخاب ویژگی ژنتیک باعث کاهش ابعاد فضای ورودی به مدل‌سازی شده که خود تأثیر مستقیمی در افزایش اعتمادپذیری و کاهش پیچیدگی محاسباتی دارد. ضمناً این روش باعث شده است، شاخص‌های تأثیرگذارتر در مدل‌سازی درجه حرارت سطح استفاده شود. با بکارگیری روش تخمین دمای سطح زمین با استفاده از داده‌های لندست ۸ محصول درجه حرارت سطح منطقه مورد مطالعه تولید شده که در نگاره شماره ۷ نشان داده شده است. محله‌هایی در تصویر با رنگ‌های گرم‌تر، در داخل شهر مشخص شده که از مناطق دیگر داغ‌تر است.

۵- نتیجه‌گیری

مشخص است که کمترین مقدار NRMSE بوسیله این رویکرد با مقدار ۰/۲۶ کسب شده است. کوچک بودن مقدار NRMSE به مفهوم کوچک بودن واریانس بین داده‌هاست. همچنین مقدار ضریب تعیین این رویکرد R^2 (مقدار ۰/۹۳)، نشان دهنده انطباق حداکثری بین شاخص‌های انتخاب شده بوسیله روش ژنتیک و داده‌های SHI می‌باشد.

جدول ۱۰: کارایی روش GA-SVR (رویکرد سوم)

نوع داده‌ها	MSE	RMSE	NRMSE	R2
داده‌های آموزشی	۰/۸۸۵۲	۰/۹۴۰۹	۰/۲۶۳۳	۰/۹۳۳۸
داده‌های تست	۰/۸۶۹۳	۰/۹۳۲۴	۰/۲۶۹۵	۰/۹۳۱۵

۴-۴- تحلیل نتایج

به طور خلاصه، با به کارگیری سه رویکرد مختلف با روش‌های رگرسیون و بهینه‌سازی بیان شده، این امکان به وجود می‌آید که روند تخمین جزایر حرارتی در شهر تهران

منابع و مأخذ

1. Abbassi, Y., Ahmadikia, H., & Baniasadi, E. (2020). Prediction of pollution dispersion under urban heat island circulation for different atmospheric stratification. *Building and Environment*, 168, 106374. <https://doi.org/10.1016/j.buildenv.2019.106374>
2. Actionbioscience. (2015). Urban Heat Islands: Hotter Cities. <http://www.actionbioscience.org/environment/voogt.html>
3. As-syakur, A. R., Adnyana, I. W. S., Arthana, I. W., Nuarsa, I. W., As-syakur, A. R., Adnyana, I. W. S., Arthana, I. W., & Nuarsa, I. W. (2012). Enhanced Built-Up and Bareness Index (EBBI) for Mapping Built-Up and Bare Land in an Urban Area. *Remote Sensing*, 4(10), 2957–2970. <https://doi.org/10.3390/rs4102957>
4. Dihkan, M., Karsli, F., Guneroglu, A., & Guneroglu, N. (2015). Evaluation of surface urban heat island (SUHI) effect on coastal zone: The case of Istanbul Megacity. *Ocean & Coastal Management*, 118, 309–316. <https://doi.org/10.1016/j.ocecoaman.2015.03.008>
5. Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support Vector Regression Machines. *Advances in Neural Information Processing Systems* 9, 155–161.
6. Durbha, S. S., King, R. L., & Younan, N. H. (2007). Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. *Remote Sensing of Environment*, 107(1), 348–361. <https://doi.org/10.1016/j.rse.2006.09.031>
7. Fabrizi, R., Bonafoni, S., Biondi, R., Fabrizi, R., Bonafoni, S., & Biondi, R. (2010). Satellite and Ground-Based Sensors for the Urban Heat Island Analysis in the City of Rome. *Remote Sensing*, 2(5), 1400–1415. <https://doi.org/10.3390/rs2051400>
8. Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58(3), 257–266. [https://doi.org/10.1016/S0034-4257\(96\)00067-3](https://doi.org/10.1016/S0034-4257(96)00067-3)
9. Goldberg, D. E., & Holland, J. H. (1988). Genetic Algorithms and Machine Learning. *Machine Learning*, 3(2–3), 95–99. <https://doi.org/10.1023/A:1022602019183>
10. Henao, J. J., Rendón, A. M., & Salazar, J. F. (2020). Trade-off between urban heat island mitigation and air quality in urban valleys. *Urban Climate*, 31, 100542. <https://doi.org/10.1016/j.uclim.2019.100542>
11. Huang, X., & Wang, Y. (2019). Investigating the effects of 3D urban morphology on the surface urban heat island effect in urban functional zones by using high-resolution remote sensing data: A case study of Wuhan,

با در نظر گرفتن مراحل اجرایی این پژوهش نتایج حاصل را می‌توان بصورت کلی بیان نمود. در این تحقیق، از تمامی باندهای طیفی ماهواره لندست ۸ بخصوص باندهای حرارتی استفاده شده است و این امر، کارایی بالای این ماهواره را در مدل‌سازی کمیت‌های فیزیکی و بیولوژیکی نشان می‌دهد. همچنین در این تحقیق از ظرفیت سنجنده مودیس نیز برای تخمین گسیلمندی در مناطق مختلف با فرض هر روزه بوده این تصاویر و استفاده همزمان آن تصاویر لندست ۸ بهره گرفته شده است. این تحقیق نشان داد که استفاده از شاخص‌های شهری و گیاهی مدل‌سازی پدیده جزایر حرارتی شهری را میسر می‌سازد. نتایج حاصل از رویکرد سوم (جدول ۱۰)، نشان دهنده کارایی بالای این رویکرد در تخمین پدیده جزیره گرمایی در شهر تهران است. همچنین به عنوان بخشی از این تحقیق شاخص‌های مناسب و تأثیرگذار در تخمین پدیده جزایر حرارتی سطحی مشخص شده است که سه شاخص شهری و دو شاخص گیاهی از مهمترین عوامل تأثیرگذار در این تخمین هستند (NDBI, NDBaI, BI, Greenness, Brightness).

پیشنهاد می‌گردد در مطالعات آتی پیرامون مسئله شناسایی پدیده جزیره حرارتی شهری، به منظور بهبود روش‌های مدل‌سازی از داده‌های میدانی بیشتر که با روش زمینی و ایستگاه‌های هواشناسی استخراج شده است استفاده گردد. ضمناً به منظور بررسی سنجنده‌های جدید سنجش‌ازدوری مخصوصاً داده‌های ماهواره‌ای Sentinel-3 لازم است توانایی آنها نیز شناسایی شود. ضمناً همانگونه که اشاره شد، شاخص‌های متعدد دیگری نیز مانند عرض معابر، کانیون‌های حرکت باد و غیره نیز در درجه حرارت شهر مؤثر است که لازم است در تحقیقات بعدی و با بهره‌گیری از روش‌های دیگر بهینه‌سازی، مورد بررسی قرار گیرند. بعلاوه، لازم است به منظور مطالعات آتی، مکانی از مناطق شهری انتخاب شده و با استفاده از روش‌های کاهش اثرات پدیده جزایر حرارتی و استفاده از تصاویر ماهواره‌ای نظارت و کنترل مستقیم انجام شود.

- and Services. https://lpdaac.usgs.gov/products/modis_products_table/mod11a2
22. Moser, G., & Serpico, S. B. (2009). Automatic Parameter Optimization for Support Vector Regression for Land and Sea Surface Temperature Estimation From Remote Sensing Data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(3), 909–921. <https://doi.org/10.1109/TGRS.2008.2005993>
23. Odindi, J. O., Bangamwabo, V., & Mutanga, O. (2015). Assessing the Value of Urban Green Spaces in Mitigating Multi-Seasonal Urban Heat using MODIS Land Surface Temperature (LST) and Landsat 8 data. *International Journal of Environmental Research*, 9(1), 9–18. <https://doi.org/10.22059/ijer.2015.868>
24. Sanchez, L., & Reames, T. G. (2019). Cooling Detroit: A socio-spatial analysis of equity in green roofs as an urban heat island mitigation strategy. *Urban Forestry & Urban Greening*, 44, 126331. <https://doi.org/10.1016/j.ufug.2019.04.014>
25. Streutker, D. R. (2002). A remote sensing study of the urban heat island of Houston, Texas. *International Journal of Remote Sensing*, 23(13), 2595–2608. <https://doi.org/10.1080/01431160110115023>
26. Voogt, J. A., & Oke, T. R. (2003). Thermal remote sensing of urban climates. *Remote Sensing of Environment*, 86(3), 370–384. [https://doi.org/10.1016/S0034-4257\(03\)00079-8](https://doi.org/10.1016/S0034-4257(03)00079-8)
27. Wang, R., Cai, M., Ren, C., Bechtel, B., Xu, Y., & Ng, E. (2019). Detecting multi-temporal land cover change and land surface temperature in Pearl River Delta by adopting local climate zone. *Urban Climate*, 28, 100455. <https://doi.org/10.1016/j.uclim.2019.100455>
28. Xian, G., & Crane, M. (2006). An analysis of urban thermal characteristics and associated land cover in Tampa Bay and Las Vegas using Landsat satellite data. *Remote Sensing of Environment*, 104(2), 147–156. <https://doi.org/10.1016/j.rse.2005.09.023>
29. Xiong, Y., Huang, S., Chen, F., Ye, H., Wang, C., & Zhu, C. (2012). The Impacts of Rapid Urbanization on the Thermal Environment: A Remote Sensing Study of Guangzhou, South China. *Remote Sensing*, 4(7), 2033–2056. <https://doi.org/10.3390/rs4072033>
30. Zhu, Z., Zhou, Y., C Seto, K., Stokes, E., Deng, C., S.T.A., P., & Taubenböck, H. (2019). Understanding an Urbanizing Planet: Strategic Directions for Remote Sensing (Vol. 228). <https://doi.org/10.1016/j.rse.2019.04.020>
31. Zohary and Sharifi, 2011, Thermal Anomalies Detection in Birjand Using Landsat Data, *Geomatics 90* (National Conference & Exhibition).
- Central China. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 119–131. <https://doi.org/10.1016/j.isprsjprs.2019.04.010>
12. Imhoff, M. L., Zhang, P., Wolfe, R. E., & Bounoua, L. (2010). Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sensing of Environment*, 114(3), 504–513. <https://doi.org/10.1016/j.rse.2009.10.008>
13. Jiang, J., & Tian, G. (2010). Analysis of the impact of Land use/Land cover change on Land Surface Temperature with Remote Sensing. *Procedia Environmental Sciences*, 2, 571–575. <https://doi.org/10.1016/j.proenv.2010.10.062>
14. Jimenez-Munoz, J. C., Sobrino, J. A., Skokovic, D., Mattar, C., & Cristobal, J. (2014). Land Surface Temperature Retrieval Methods From Landsat-8 Thermal Infrared Sensor Data. *IEEE Geoscience and Remote Sensing Letters*, 11(10), 1840–1843. <https://doi.org/10.1109/LGRS.2014.2312032>
15. Krieglner, F. J., Malila, W. A., Nalepka, R. F., & Richardson, W. (1969). Preprocessing Transformations and Their Effects on Multispectral Recognition. 97. <http://adsabs.harvard.edu/abs/1969rse.conf...97K>
16. Kutner, M., Nachtsheim, C., results, search, & Li, W. (2004). *Applied Linear Statistical Models* (5th edition). McGraw-Hill/Irwin.
17. Liu, K., Su, H., Zhang, L., Yang, H., Zhang, R., Li, X., Liu, K., Su, H., Zhang, L., Yang, H., Zhang, R., & Li, X. (2015). Analysis of the Urban Heat Island Effect in Shijiazhuang, China Using Satellite and Airborne Data. *Remote Sensing*, 7(4), 4804–4833. <https://doi.org/10.3390/rs70404804>
18. Mathew, A., Khandelwal, S., Kaul, N., & Chauhan, S. (2018). Analyzing the diurnal variations of land surface temperatures for surface urban heat island studies: Is time of observation of remote sensing data important? *Sustainable Cities and Society*, 40, 194–213. <https://doi.org/10.1016/j.scs.2018.03.032>
19. Mathew, A., Sreekumar, S., Khandelwal, S., & Kumar, R. (2019). Prediction of land surface temperatures for surface urban heat island assessment over Chandigarh city using support vector regression model (Vol. 186). <https://doi.org/10.1016/j.solener.2019.04.001>
20. Mijani, N., Alavipanah, S. K., Hamzeh, S., Firozjaei, M. K., & Arsanjani, J. J. (2019). Modeling thermal comfort in different condition of mind using satellite images: An Ordered Weighted Averaging approach and a case study. *Ecological Indicators*, 104, 1–12. <https://doi.org/10.1016/j.ecolind.2019.04.069>
21. MOD11A2. (2015). NASA Land Data Products