نشریه علمی علوم و فنون نقشه برداری، دوره دهم، شماره ۴، خرداد ماه ۱۴۰۰ - مقاله پژوهشی

شناسایی اتوماتیک تغییرات کاربری/پوشش اراضی از تصاویر سنجش از دور چندزمانه و نقشه‌های قدیمی با پالایش نمونه‌های آموزشی مبتنی بر آزمون خی-دو و خوشه‌بندی k-means

وحید صادقی[۱]، حمید عبادی[۲]، آرمین مقیمی*[۳]، علیرضا نوزادی[۴]

[۱] استادیار گروه مهندسی نقشه‌برداری - دانشکده مهندسی عمران - دانشگاه تبریز
vahid.sadeghi.1985@gmail.com

[۲] استاد دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
ebadi.h@gmail.com

[۳] دانشجوی دکتری فتوگرامتری - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
moghimi.armin@gmail.com

[۴] کارشناس ارشد فتوگرامتری - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی
surveying14@gmail.com

(تاریخ دریافت اسفند ۱۳۹۸، تاریخ تصویب فروردین ۱۴۰۰)

چکیده

انتخاب نمونه‌های آموزشی، مرحله‌ای بسیار مهم و تأثیرگذار در نتایج طبقه‌بندی و شناسایی تغییرات از تصاویر سنجش از دور بوده و لازم است با حساسیت بالایی انجام گیرد. این نمونه‌ها اغلب توسط عامل انسانی تعیین می‌شوند که فرآیندی زمانبر بوده و مستعد خطای بالایی است. نقشه‌های قدیمی می‌توانند منبع اطلاعاتی ارزشمندی برای انتخاب و تهیه نمونه‌های آموزشی باشند. در صورتی که این نمونه‌ها بطور صحیح پالایش شوند، می‌توانند سبب تسریع، تسهیل و همچنین افزایش صحت فرآیند شناسایی تغییرات شوند. نوآوری اصلی مقاله حاضر، اهتمام در فرآیند پالایش نمونه‌ها است که با پیشنهاد مدلی مبتنی بر آزمون آماری خی‌دو و خوشه‌بندی k-means میسر شده است. این روش ضمن اینکه با بکارگیری آزمون آماری خی-دو، سعی در انتخاب نمونه‌های آموزشی خالص دارد، با خوشه‌بندی چندگانه نمونه‌های آموزشی با تکنیک k-means و انتخاب نمونه‌های نزدیک به مراکز خوشه‌های داخلی هر کلاس، تنوع طیفی داخلی کلاسها را نیز لحاظ می‌نماید. در روش پیشنهادی؛ جهت بهبود صحت طبقه‌بندی و تشخیص تغییرات، ویژگی‌های طیفی و بافتی مرتبه اول و دوم ماتریس رخداد همزمان، استخراج و در فرآیند طبقه‌بندی به روش ماشین بردار پشتیبان (SVM) مورد استفاده قرار می‌گیرد. لازم به ذکر است به منظور ارتقای صحت طبقه بندی و شناسایی تغییرات، فرآیند انتخاب مجموعه ویژگی‌های بهینه و پارامترهای طبقه‌بندی‌کننده SVM با الگوریتم ژنتیک بهینه شده‌اند. جهت پیاده‌سازی و ارزیابی روش پیشنهادی، از تصاویر دوزمانه در بازه زمانی ۲۰۱۱ و ۲۰۱۵ و نقشه کاربری اراضی سال ۲۰۰۹ مربوط به شهر شیراز استفاده شد. با بکارگیری روش ارائه شده در این تحقیق، نقشه‌های موضوعی منطقه مورد مطالعه با صحت کلی ۹۸/۷۲٪ و ۹۴/۵۷٪ بهنگام گردیده و با مقایسه آنها، نقشه ماهیت تغییرات حاصل گردید. ارزیابی نتایج نشان داد؛ فرآیند پالایش نمونه‌های آموزشی سبب بهبود نتایج طبقه‌بندی تصویر سال ۲۰۱۱ (افزایش ضریب کاپا از ۶۵٪ به ۸۷٪ و افزایش صحت کلی از ۷۳٪ به ۹۱٪) و همچنین تصویر سال ۲۰۱۵ (افزایش صحت کلی از ۶۹٪ به ۸۶/۳۲٪ و افزایش ضریب کاپا از ۵۹٪ به ۸۰/۴۸٪) شده است.

واژگان کلیدی: شناسایی تغییرات، بهنگام‌رسانی، پالایش نمونه‌های آموزشی، آزمون آماری خی-دو، خوشه‌بندی k-means

* نویسنده رابط

۱- مقدمه

شناسایی تغییرات زمین یکی از کاربردهای فتوگرامتری و سنجش‌ازدور است که در آن به پردازش و معرفی تغییرات در یک منطقه جغرافیایی در تصاویر هوایی و ماهواره‌ای چندزمانه پرداخته می‌شود [۱]. با توجه به رشد و گسترش روزافزون شهرها و روستاها، دسترسی به اطلاعات دقیق، جامع و به‌هنگام از چگونگی و میزان تغییرات سطح زمین، برای مدیریت صحیح و کارآمد، امری ضروری است. این تغییرات می‌توانند شامل تغییرات در کاربری و پوشش اراضی، پوشش گیاهی، رشد و توسعه شهری، جنگل‌زدایی و غیره باشد. بنابراین شناسایی تغییرات و به‌تبع آن بهنگام‌رسانی نقشه‌ها جهت مدیریت صحیح منابع، یکی از چالش‌های اساسی در حوزه فتوگرامتری و سنجش‌ازدور است. نقشه‌های موضوعی یک منبع باارزش جهت مدیریت منابع طبیعی و انسانی است که حاوی اطلاعات مهمی در مورد پوشش اراضی، کاربری اراضی، محصولات کشاورزی و غیره می‌باشد. تحلیل تصاویر ماهواره‌ای چندزمانه، امکان شناسایی و بررسی پدیده‌های متغیر و پویا در سطح زمین را فراهم نموده و افقی مناسب را جهت بهنگام‌رسانی نقشه‌های موضوعی و همچنین کاربردهای مذکورگشوده است [۲-۴]. توسعه روش‌های شناسایی تغییرات در سنجش‌ازدور، موضوعی چالش‌برانگیز با ماهیتی پویا است [۵]. در بکارگیری داده‌های سنجش از دور برای شناسایی تغییرات سطح زمین، فرض بر این است که تغییر در عارضه موردنظر، سبب تغییر رفتار طیفی (مقدار بازتاب یا بافت) آن عارضه شده و این تغییرات طیفی از تغییرات ایجاد شده توسط عوامل دیگر همچون شرایط اتسمفری، روشنایی، زاویه دید و رطوبت خاک قابل تفکیک است [۱, ۶, ۷]. بنابراین شناسایی تغییرات از داده‌های سنجش‌ازدوری تحت تاثیر عوامل مختلفی نظیر قیود مکانی، طیفی، زمانی، قدرت تفکیک رادیومتریکی، شرایط اتمسفری و شرایط رطوبت خاک است [۷, ۸]. با در نظر گرفتن این قیود، روش‌های مختلفی از گذشته تاکنون برای شناسایی تغییرات بسته به نیاز و شرایط مسئله توسعه داده شده است. با این‌حال انتخاب یک الگوریتم و روش مناسب همیشه چالش‌برانگیز بوده است [۸].

روش‌های شناسایی تغییرات مبتنی بر طبقه‌بندی عمدتاً به‌صورت یک فرآیند نظارت‌شده بر روی تصاویر چندزمانه اعمال می‌شوند [۸]. بنابراین، این روش‌ها نیازمند نمونه‌های آموزشی بوده و نقشه تغییرات حاصل از این روش‌ها، به‌صورت نقشه ماهیت تغییرات می‌باشد که تغییرات را به‌صورت «از-به» نشان می‌دهد [۵, ۸]. روش «مقایسه پس از طبقه‌بندی» متداول‌ترین رویکرد در زمینه شناسایی تغییرات است که در آن، تصاویر به طور مستقل طبقه‌بندی شده و سپس پیکسل به پیکسل با یکدیگر مقایسه می‌شوند [۵].

مطالعات صورت گرفته در زمینه شناسایی تغییرات با استفاده از روش مقایسه پس از طبقه‌بندی بیشتر مبتنی بر توسعه یا ارتقاء روش طبقه‌بندی می‌باشند. در حالیکه کیفیت و نحوه جمع‌آوری نمونه‌های آموزشی به عنوان یک اصل مهم در موفقیت و عدم موفقیت یک روش طبقه‌بندی نظارت‌شده، صرف نظر از نوع آن (پارامتریک و غیرپارامتریک) تاثیرگذار است. به عنوان مثال، صادقی و همکاران در سال ۲۰۱۶، [۹]، به منظور بهنگام‌رسانی پایگاه داده مکانی یک سیستم خبره دانش‌پایه توسعه داده‌اند که در آن نحوه گردآوری نمونه‌های آموزشی مبتنی بر تفسیر بصری تصاویر سنجش از دور و همچنین مبتنی بر همپوشانی تصاویر سنجش از دور چندزمانه و نقشه موضوعی منطقه بوده است. قید هم‌زمان بودن تصویر قدیمی و نقشه قدیمی منطقه، از محدودیت‌های تکنیک پیشنهادی ایشان است. سعیدزاده در سال ۱۳۹۴[۱۰]، جهت شناسایی تغییرات از روش مقایسه پس از طبقه‌بندی شی‌مبنا استفاده کرد که نمونه‌های آموزشی بکار رفته در این مطالعه از تفسیر بصری تصاویر ماهواره‌ای بدست آمده است. عدم استخراج اتوماتیک نمونه‌های آموزشی از نقشه-های موجود و همچنین عدم پالایش این نمونه‌ها از جمله محدودیت‌های روش‌های ارائه شده در این زمینه است که موجب کاهش سطح اتوماسیون بهنگام‌رسانی و همچنین کاهش صحت این فرآیند شده است.

استفاده از نقشه به عنوان داده اضافی برای ایجاد نمونه‌های آموزشی در طبقه‌بندی تصاویر چندزمانه یک منطقه، می‌تواند سبب تسهیل فرآیند جمع‌آوری نمونه‌های آموزشی شده و در صورتی که کیفیت و کمیت نمونه‌های آموزشی کافی باشد، می‌تواند بهبود صحت نتایج طبقه‌بندی و شناسایی تغییرات و همچنین سطح اتوماسیون را نیز به دنبال داشته باشد. در این راستا تحقیقات زیادی انجام شده است که در ادامه به مهمترین آن‌ها پرداخته می‌شود.

Xian و همکارانش در سال ۲۰۰۹، [۱۱]، جهت بهنگام‌رسانی نقشه کاربری اراضی ملی ایالات متحده که به طور اسمی در سال ۲۰۰۱ تهیه شده بود، از نتایج شناسایی تغییرات تصاویر چندزمانه ماهواره لندست استفاده کردند. در این مطالعه از نقشه پوششی/ کاربری اراضی سال ۲۰۰۱ به عنوان مبنای انتخاب داده‌های آموزشی و بهنگام‌رسانی پایگاه داده اطلاعات مکانی استفاده گردید. اساس کار به این صورت بود که ابتدا مناطق تغییریافته از تصاویر ماهواره‌ای لندست مربوط به سال ۲۰۰۱ و ۲۰۰۶ جدا شدند و داده‌های آموزشی در این مناطق با تطابق با نقشه سال ۲۰۰۱ جمع‌آوری شد. سپس با استفاده از این داده‌ها مناطق تغییریافته سال ۲۰۰۶ طبقه‌بندی گردید و از نتایج این طبقه‌بندی در بهنگام‌رسانی نقشه پوششی سال ۲۰۰۱ استفاده شد.

Chen و همکارانش در سال ۲۰۱۲، [۱۲]، به منظور کاهش تعامل انسانی در بهنگام‌رسانی نقشه کاربری اراضی، یک رویکرد یکپارچه و اتوماتیک را ارائه دادند که مبنای آن تشخیص مناطق تغییریافته، مدل تصادفی مارکوف[۱] (MRFs) و یک روش انتخاب نمونه آموزشی تکراری بود. نتایج این روش بر روی تصاویر ماهواره‌ای چندزمانه نشان داد که روش مذکور پتانسیل بالایی در بهنگام‌رسانی نقشه‌ها دارد. در ادامه این رویکرد، Wessels و همکارانش در سال ۲۰۱۶، [۱۳]، در جهت بهنگام‌رسانی پایگاه داده نقشه کاربری اراضی آفریقا (تهیه شده در سال ۲۰۰۸)، از روش IRMAD به منظور تفکیک مناطق تغییریافته از تصاویر ماهواره‌ای چندزمانه لندست ۲۰۰۸ و ۲۰۱۱ استفاده کردند. در ادامه از این مناطق به منظور تهیه نمونه‌های آموزشی برای تولید نقشه موضوعی از تصویر ماهواره‌ای سال ۲۰۱۱ با استفاده از طبقه‌بندی جنگل تصادفی استفاده شد.

حاج احمدی و همکارانش در سال ۲۰۱۶، [۲۳]، روشی نوین را جهت تولید نقشه تغییرات از تصاویر ماهواره‌ای چندزمانه و به کمک نقشه‌های قدیمی ارایه کردند. در این روش، نمونه‌های آموزشی با استفاده از هم-پوشانی با نقشه‌های قدیمی و تصاویر صورت گرفت. در راستای خالص‌سازی و پالایش این نمونه‌ها از خوشه‌بندی k-means و روش k نزدیکترین همسایه استفاده شد. نتایج نشان داد که استفاده از رویکرد به کار رفته در پالایش نمونه‌های آموزشی موجب افزایش ۱۵ درصدی دقت طبقه‌بندی شده است. این استراتژی هر چند نمونه‌های خالص را بر می‌گزیند، ولی تنوع طیفی داخلی کلاسها را لحاظ نمی‌کند.

تمیمی و همکارانش در سال ۲۰۱۸، [۲۴]، روشی نوین با سطح اتوماسیون بالا را جهت بهینه‌سازی طبقه-بندی SVM شیءمبنا با الگوریتم‌های فرابتکاری جهت بهنگام‌رسانی نقشه‌های بزرگ مقیاس ارایه کردند. در این تحقیق نمونه‌های آموزشی با استفاده از روشی مبتنی بر فیلتر مساحت به صورت نیمه خودکار از سطح تصاویر جمع‌آوری گردیدند که موجب افزایش سطح خودکارسازی روند بهنگام‌رسانی نقشه‌ها گردید.

بررسی تحقیقات انجام شده در زمینه بهنگام‌رسانی نقشه‌های موضوعی با تصاویر سنجش از دور نشان می‌دهد، سه مسئله اصلی شامل؛ ۱- فرض تطابق زمانی تصاویر و نقشه‌های موضوعی موجود، ۲- عدم پالایش نمونه‌های آموزشی و ۳- عدم توجه کافی به مسئله تنوع طیفی نمونه‌های آموزشی پالایش‌شده از محدودیت‌های تحقیقات قبلی است. در این مقاله سعی شده است به منظور رفع محدودیت‌های تکنیک‌های موجود، یک روش با سطح اتوماسیون بالا و کارآمد جهت بهنگام‌رسانی نقشه‌های موضوعی طراحی شود که مشکلات موجود در هر یک از مراحل؛ ۱- استخراج و پالایش نمونه‌های آموزشی و ۲- استخراج و انتخاب مجموعه ویژگی‌های طیفی و بافت بهینه را مرتفع نماید به طوری که محدودیت تطابق زمانی تصاویر چندزمانه و نقشه‌های موضوعی موجود با بکارگیری فرآیند پالایش الگوهای آموزشی برطرف شده و با استخراج و انتخاب ویژگی طیفی و مکانی بهینه، صحت فرآیند بهنگام‌رسانی نقشه‌های موضوعی بهبود یابد.

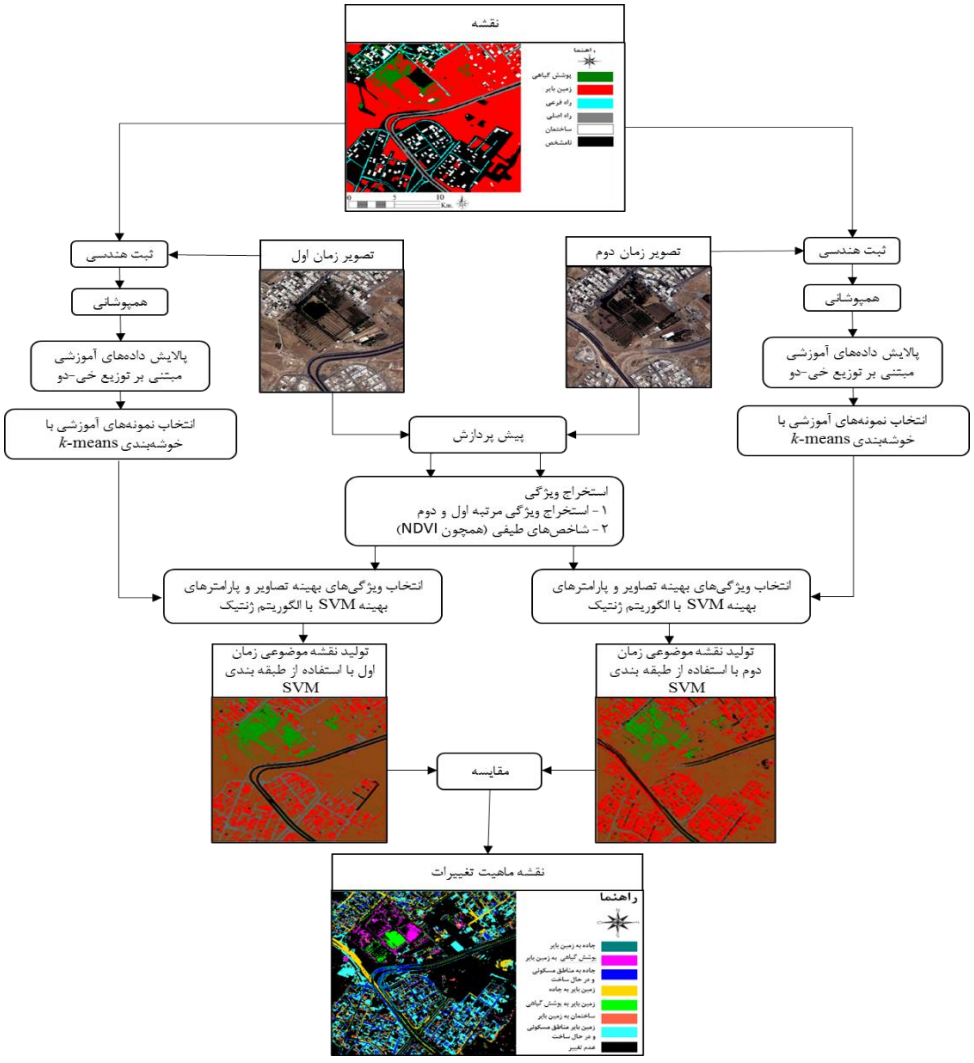
نوآوری اصلی مقاله حاضر، اهتمام در فرآیند پالایش نمونه‌های آموزشی است که با پیشنهاد مدلی مبتنی بر آزمون آماری خی‌دو و خوشه‌بندی k-means میسر شده است. این روش ضمن اینکه با بکارگیری آزمون آماری خی-دو، سعی در انتخاب نمونه‌های آموزشی خالص دارد، با خوشه‌بندی چندگانه نمونه‌های آموزشی با تکنیک K-means و انتخاب نمونه‌های نزدیک به مراکز خوشه‌های داخلی هر کلاس، تنوع طیفی داخلی کلاسها را نیز لحاظ می‌نماید. لازم به ذکر است، تنوع طیفی نمونه‌های آموزشی

[۱] Markov random field

کلاسهای موضوعی، مسئله‌ای مغفول در تحقیقات پیشین بوده است که برای اولین بار در این تحقیق پیشنهاد شده و مورد بررسی و ارزیابی قرار گرفته است. در بخش انتخاب ویژگی نیز از الگوریتم بهینه‌یابی ژنتیک استفاده شده و طبقه‌بندی با الگوریتم ماشین بردار پشتیبان بهینه‌شده با الگوریتم ژنتیک انجام می‌گیرد. هر چند به ظاهر، در بخش انتخاب ویژگی و طبقه‌بندی نوآوری خاصی مشاهده نمی‌گردد، ولی تلفیق این تکنیک‌ها در کاربرد بهنگام‌رسانی نقشه‌های موضوعی و شناسایی تغییرات، اولین بار در مقاله حاضر بطور عملی مورد ارزیابی قرار گرفته است.

۲- روش پیشنهادی

هدف اصلی تحقیق حاضر ارائه یک روش نوین و کارآمد جهت شناسایی تغییرات و بهنگام‌رسانی نقشه‌های موضوعی، مبتنی بر تصاویر سنجش از دور چندزمانه است. روش پیشنهادی از چهار مرحله اصلی تشکیل شده است: ۱) پالایش و انتخاب اتوماتیک نمونه‌های آموزشی برای هر یک از تصاویر با بکارگیری آزمون آماری خی-دو و خوشه‌بندی k-means، ۲) استخراج و انتخاب ویژگی‌های طیفی و بافتی برای هر یک از تصاویر ورودی، ۳) تولید نقشه‌های موضوعی دوزمانه (مربوط به هر یک از تصاویر ورودی) با بکارگیری ویژگی‌ها و پارامترهای بهینه تعیین‌شده توسط الگوریتم ژنتیک و طبقه‌بندی‌کننده SVM و ۴) تولید نقشه ماهیت تغییرات با مقایسه نقشه‌های موضوعی دوزمانه. شکل (۱)، فلوچارت روش پیشنهادی به منظور شناسایی تغییرات و بهنگام‌سازی نقشه‌های موضوعی را نشان می‌دهد. جزئیات روش پیشنهادی در ادامه تشریح می‌گردد. لازم به ذکر است در مرحله پیش‌پردازش، فرآیندهای مهمی از جمله؛ تبدیل فرمت برداری به فرمت رستری (و بالعکس)، تصحیح هندسی، تصحیح رادیومتریکی و اثرات ابر در صورت نیاز باید انجام گیرد.



شکل ۱- فلوچارت روش پیشنهادی به منظور شناسایی تغییرات و بهنگام‌رسانی نقشه‌های موضوعی

۲-۱- مرحله اول: انتخاب و پالایش نمونه‌های آموزشی بهینه مبتنی بر نقشه

پالایش نمونه‌های آموزشی فرآیندی است که به منظور انتخاب نمونه‌های آموزشی خالص برای کلاس‌های مختلف صورت میگرد. لذا قبل از فرآیند پالایش، لازم است تا نمونهای آموزشی خام برای کلاسهای مختلف تهیه گردد. بدین منظور؛ ابتدا نقشه پوشش/کاربری اراضی و تصاویر ورودی از نظر هندسی هم‌مرجع شده و با همپوشانی آنها، برای هر پیکسل؛ یک برچسب اختصاص داده می‌شود (کلاس موضوعی هر یک از پیکسل‌ها تعیین می‌گردد). در ادامه فرآیند پالایش نمونه‌ها آغاز می‌شود.

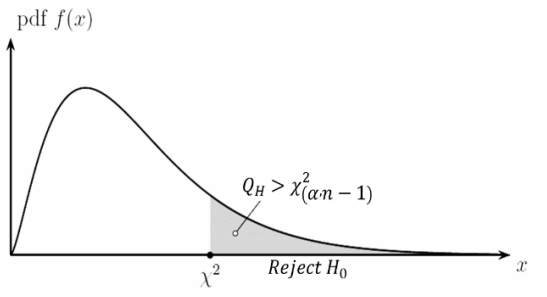
متداول‌ترین روش پالایش نمونه‌های آموزشی استفاده از تکنیک‌های خوشه‌بندی است [۲۳]. فارغ از تفاوت بین الگوریتم‌های خوشه‌بندی مورد استفاده در تحقیقات پیشین، استراتژی بکاررفته تقریبا مشابه است؛ خوشه‌بندی نمونه‌های آموزشی و انتخاب نمونه‌های نزدیک به مرکز به عنوان نمونه‌های خالص. راهکار مذکور، هر چند نمونه‌های خالص را بر می‌گزیند؛ ولی تنوع طیفی داخلی کلاسها را لحاظ نمی‌کند. در واقع تنوع طیفی نمونه‌های آموزشی کلاسهای موضوعی، مسئله‌ای مغفول در تحقیقات پیشین است که در تحقیق حاضر مدنظر قرار گرفته و راهکار مناسب و کارآمدی پیشنهاد شده است.

در روش پیشنهادی هرچند از روشهای خوشه‌بندی برای پالایش نمونه‌های آموزشی استفاده می‌شود؛ ولی روند پالایش نمونه‌ها، متفاوت با روشهای پیشین است. در روشهای پیشین پالایش در یک مرحله و با انتخاب نمونه های نزدیک به مرکز خوشه هر کلاس صورت گرفته است ولی در روش پیشنهادی، پالایش نمونه‌ها در دو مرحله متوالی صورت می‌گیرد. به این شرح که در مرحله اول؛ از آزمون آماری خی-دو برای پالایش اولیه نمونه‌ها استفاده می‌شود و سپس پالایش ثانویه نمونه‌ها با خوشه‌بندی k-means صورت می‌گیرد. این روش دو مرحله‌ای ضمن انتخاب نمونه‌های خالص، تنوع طیفی داخلی کلاسها را نیز لحاظ می‌نماید. جزئیات روش پیشنهادی در ادامه تشریح می‌گردد.

در بسیاری از کاربردها، محقق علاقه‌مند به شمارش افراد، اشیا و یا پاسخ‌هایی است که در طبقه‌های خاصی قرار می‌گیرند. در کاربرد خاص نظیر طبقه‌بندی نظارت‌شده، پاسخ‌ها می‌تواند به صورت فرض صفر (H_0) «تعلق داشتن نمونه‌ها به یک کلاس» و فرض یک (H_1) «عدم تعلق نمونه‌ها به آن کلاس» باشد. با این تفسیر، محقق می‌تواند این فرضیه را که پاسخ‌های ذکرشده از نظر فراوانی با یکدیگر تفاوت معناداری خواهند داشت، آزمون نماید. در این صورت آزمون تطابق توزیع خی-دو یا خی-مربع می‌تواند کمک شایانی در انتخاب نمونه‌های آموزشی صحیح برای فرآیند طبقه‌بندی تصاویر سنجش از دور کند. آماره این آزمون دارای توزیع خی-دو با n-1 درجه آزادی است (n تعداد کلاس‌ها است) که مقادیر آن در سطح معناداری α در جدولی به نام جدول توزیع خی-دو ارائه می‌گردد. استراتژی ردّ فرض صفر به این صورت است که هرچه مقادیر مشاهده‌شده دور از انتظار باشند، مقدار این آماره بزرگ و در نتیجه فرض صفر به مخاطره می‌افتد و برعکس؛ تفاوت کمتر بین مقادیر مشاهداتی و مقادیر مورد انتظار، آماره این آزمون را کوچک کرده و فرض صفر را نمی‌توان رد کرد [۱۴]. لازم به ذکر است که در کاربرد حاضر، داده‌های آموزشی استخراج شده از هم‌پوشانی نقشه و تصویر به عنوان مقادیر مشاهداتی و میانگین آن‌ها (متاثر از کواریانس آن‌ها در باندهای طیفی) به عنوان مقادیر مورد انتظار در نظر گرفته می شود. استراتژی ردّ فرض صفر به صورت رابطه (۱) بیان می‌شود:

$$Q_H > \chi^2_{(\alpha,n-1)} \qquad (۱)$$

در رابطه (۱)؛ Q_H بیانگر مقادیر مشاهداتی و $\chi^2_{(\alpha,n-1)}$، آماره آزمون یا شاخص آماری محاسبه شده می‌باشد که دارای توزیع خی-دو با درجه آزادی $n-1$ است و اصطلاحاً «خی-دو بحرانی» نامیده می‌شود. رابطه (۱) نشان می‌دهد در صورتی‌که مقدار Q_H از مقدار خی-دو بحرانی بیشتر باشد؛ فرض صفر ردّ شده و مقدار مشاهده‌شده نماینده خوبی برای انتخاب شدن به عنوان نمونه آموزشی مناسب نیست و در غیر اینصورت فرض صفر را نمی‌توان ردّ کرد و آن نمونه می‌تواند به عنوان داده آموزشی مناسب انتخاب گردد. جهت تشریح بیشتر این موضوع، نحوه پذیرش یا ردّ فرض صفر (H_0) در شکل (۲) ارائه داده شده است.



شکل ۲- چگونگی پذیرش یا ردّ فرض H_0 در توزیع خی-دو

به طور کلی اینگونه استدلال می‌شود که در یک جامعه آماری اعضای متعلق به یک گروه خاص باید دارای رفتارهای مشابه باشند. با تعمیم این استدلال به طبقه‌بندی تصاویر سنجش از دور چندطیفی، می‌توان این ایده را در نظر گرفت که داده‌های آموزشی مستخرج از نقشه در هر کلاس، دارای رفتار مشابهی هستند و از توزیع نرمال چندمتغیره تبعیت می‌کنند. بنابراین برای متغیر تصادفی x که بیانگر درجه خاکستری نمونه‌های آموزشی در هر کلاس است، توزیع فوق به صورت رابطه (۲) قابل تعریف است.

$$\phi(x) = \frac{1}{2\pi|\Sigma|^{0.5}} e^{-0.5\left((X-\mu)\Sigma^{-1}(X-\mu)'\right)} \qquad (۲)$$

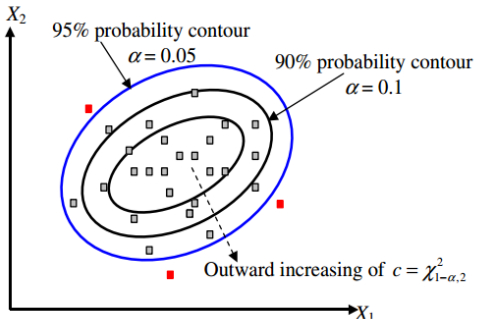
در رابطه (۲)، X بیانگر بردار طیفی مربوط به نمونه‌های آموزشی در باندهای طیفی مختلف، Σ بیانگر ماتریس کواریانس بین‌باندی داده‌های آموزشی و μ بردار میانگین داده‌ها در باندهای طیفی مختلف است. با توجه به این که $((X-\mu)\Sigma^{-1}(X-\mu)')$ از توزیع خی-دو با درجه آزادی ۲ پیروی می‌کند و بیانگر فاصله ماهالانوبیس[۱] می‌باشد. بنابراین منحنی‌های احتمال توزیع چندمتغیره[۲]، برای هر مقدار مثبت C می‌تواند به صورت رابطه (۳) تخمین زده شود:

$$(X-\mu)\Sigma^{-1}(X-\mu)' = C \qquad (۳)$$

با توضیحات ارائه‌شده از رابطه فوق می‌توان از Q_H یا معیار آزمون خی-دو به عنوان معیارپالایش داده‌های آموزشی هر کلاس استفاده کرد که به صورت رابطه (۴) قابل‌تعریف است [۱۴].

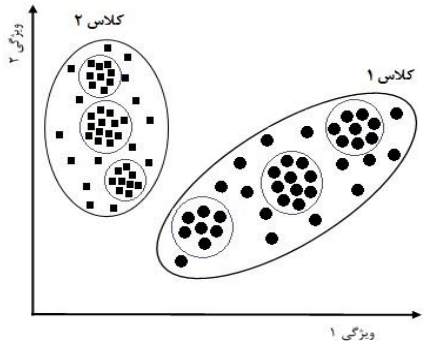
$$Q_H = P[(X-\mu)\Sigma^{-1}(X-\mu)'] \le \chi^2_{(1-\alpha),2} \qquad (۴)$$

پالایش داده‌های آموزشی می‌تواند در سطوح اطمینان مختلفی انجام پذیرد. شکل (۳) سطوح اطمینان مختلف در پالایش نمونه‌های آموزشی را نشان می‌دهد. با نگاه اجمالی به این شکل می‌توان دریافت که با افزایش سطح اطمینان، عدد خی-دو (در جدول مربوطه) افزایش یافته و احتمال ورود داده‌های پرت[۳] به محاسبات افزایش می‌یابد که در نتیجه می‌تواند منجر به کاهش صحت طبقه‌بندی گردد. در واقع میزان سخت‌گیری فرآیند پالایش کاهش یافته و تعداد بیشتری از نمونه‌های آموزشی در ناحیه پذیرش فرض صفر قرار می‌گیرند (شکل ۳).



شکل ۳- سطوح اطمینان مختلف جهت پالایش نمونه‌های آموزشی با آزمون آماری خی-دو [۱۴]

پس از پالایش داده‌های آموزشی با آزمون آماری خی-دو، نمونه‌های آموزشی اولیه برای هر یک از کلاس‌های موضوعی فراهم می‌شود. نمونه‌های آموزشی در هر کلاس باید از تعداد و توزیع مناسبی برخوردار باشند و تنوع طیفی داخلی کلاسها نیز مد نظر قرار گیرد. برای این منظور در روش پیشنهادی، نمونه‌های پالایش‌شده از مرحله قبل، به تعداد خوشه‌های متعددی، دسته‌بندی می گردند. این فرآیند با تکنیک k-means با تعداد ۳ خوشه برای هر کلاس موضوعی صورت می‌گیرد. سپس نمونه‌های نزدیک به مراکز خوشه‌های داخلی هر کلاس، به عنوان نمونه‌های نهایی انتخاب می‌شوند (شکل ۴).



شکل ۴- مرحله دوم پالایش، خوشه‌بندی نمونه‌های آموزشی پالایش شده از مرحله قبل به خوشه‌های مجزا در هر یک از کلاسهای موضوعی

۱ Mahalanobis distance
۲ Multivariante Normal Distribution
۳ Outlier

در این روش، ابتدا نمونه‌های پالایش‌شده در هر کلاس که به صورت مجموعه $(x_{1,c}, x_{2,c}, \dots, x_{j,c})$ هستند؛ به عنوان کاندیداهای اولیه نمونه‌های آموزشی در آن کلاس در نظر گرفته شده و سپس با استفاده از الگوریتم خوشه-بندی K-means به سه گروه براساس تابع هدف زیر (روابط ۵ و ۶) تقسیم می‌گردند.

$$J_C = \sum_{i=1}^{3}\sum_{j=1}^{n_c}\left\|x_{j,c} - v_{i,c}\right\|^2 \qquad (۵)$$

$$v_{i,c} = \sum_{j=1}^{n_c} x_j \Big/ n_c \qquad (۶)$$

که در روابط فوق $\|.\|$ بیانگر معیار فاصله بوده، x_j درجه خاکستری داده‌های آموزشی jام در کلاس cام، و v_i مرکز خوشه iام در کلاس cام و n_c بیانگر تعداد نمونه‌های آموزشی در کلاس cام است [۱۵]. پس از تقسیم هر یک از نمونه‌های آموزشی پالایش‌شده به سه خوشه، تعداد مشخصی از آن‌ها که دارای کیفیت بالاتری نسبت به سایر داده‌ها در آن خوشه هستند، به عنوان داده‌های آموزشی پالایش‌شده نهایی در هر کلاس انتخاب می‌گردند. در این راستا به داده‌های پالایش‌شده متعلق به هر یک از سه خوشه در هر کلاس، یک امتیاز بر مبنای کمترین فاصله تا مرکز خوشه تعلق می‌گیرد (رابطه ۷).

$$S_{j,c} = \left((x_{j,c} - v_{i,c})\Sigma^{-1}(x_{j,c} - v_{i,c})'\right), \qquad (۷)$$

پس از محاسبه امتیازات داده‌های پالایش‌شده در هر سه خوشه از هر کلاس، ۱۰٪ از نمونه‌های آموزشی پالایش‌شده در هر خوشه که دارای بیشترین امتیاز (کمترین فاصله تا مرکز هر خوشه) باشند، به عنوان کاندیداهای نهایی داده‌های آموزشی برای هر کلاس در نظر گرفته می‌شوند. راهکار پیشنهادی سبب می‌شود که نمونه‌های آموزشی پالایش‌شده برای هر یک از کلاس‌های موضوعی، از توزیع و تراکم و تنوع طیفی مناسبی برخوردار باشند. همچنین علت در نظر گرفتن ۱۰٪ از نمونه‌های پالایش شده در هر خوشه، تسریع در فرآیند-های پردازشی بعدی (طبقه‌بندی SVM) می‌باشد.

۲-۲- مرحله دوم: استخراج ویژگی از تصاویر و طبقه‌بندی آن‌ها با استفاده از روش GASVM

استفاده صرف از اطلاعات طیفی جهت طبقه‌بندی مناطق همگن در تصاویر ماهواره‌ای با قدرت تفکیک مکانی پایین و متوسط، می‌تواند نتایج رضایت‌بخشی را به همراه داشته باشد؛ با این حال، در طبقه‌بندی تصاویر ماهواره‌ای با توان تفکیک مکانی بالا و دارای مناطق غیرهمگون، استفاده از ویژگی‌های صرفاً طیفی کافی نیست. بنابراین در جهت بهبود نتایج طبقه‌بندی می‌توان از ویژگی‌های طیفی و مکانی (بافت) در کنار باندهای طیفی اصلی سنجنده‌ها استفاده کرد. در تحقیق حاضر به منظور بهبود صحت فرآیند بهنگام‌رسانی نقشه‌های موضوعی، در مرحله طبقه‌بندی تصاویر، از شاخص‌های طیفی و ویژگی‌های مکانی (بافت) استفاده می‌شود. در این راستا، از شاخص گیاهی نرمالیزه شده (NDVI)[۱] به عنوان متداول‌ترین ویژگی طیفی و با هدف برجسته نمودن پوشش گیاهی سطح زمین استفاده شده است که با استفاده از رابطه (۸) محاسبه می شود[۱۶].

$$NDVI = \frac{b_{NIR} - b_{Red}}{b_{NIR} + b_{Red}} \qquad (۸)$$

در رابطه (۸)، b_{NIR} و b_{Red} به ترتیب بیانگر مقادیر درجه خاکستری در باند مادون قرمز نزدیک و باند قرمز است. استفاده از اطلاعات مکانی (بافت) و ترکیب آن‌ها با اطلاعات طیفی در کلاس‌هایی با پاسخ‌های طیفی مشابه ولی با ساختار/بافت گوناگون (مانند مناطق گیاهی کم تراکم و مناطق شهری) می‌تواند سبب افزایش صحت طبقه‌بندی شود. تکنیک‌های متنوعی برای این منظور توسعه داده شده‌اند. در تحقیق حاضر به منظور استخراج ویژگی‌های مکانی (بافت)، از ویژگی‌های آماری مرتبه اول و دوم مستخرج از ماتریس هم‌رخداد GLCM[۲] [۱۷] استفاده شد. از آنجا که هر یک از عوارض موجود در تصاویر دارای جهات مختلفی با توجه به زاویه تصویربرداری سنجنده‌ها هستند، ویژگی‌های انتخابی در این مطالعه با میانگین‌گیری چهار جهته در توصیفگرهای آماری مرتبه دوم با ابعاد پنجره ۳*۳ محاسبه گردیدند تا جهت غالب هر یک از عوارض موجود در سطح تصاویر مشخص گردند. از میان توصیف‌گرهای مرتبه اول و دوم ماترس GLCM موارد فهرست شده در جدول (۱) در این مطالعه مورد استفاده قرار گرفت.

۱ Normalized Difference Vegetation Index
۲ Gray-Level Co-Occurrence Matrix (GLCM)

جدول ۱- ویژگی‌های بافت استخراج شده از GLCM

ماتریس رخداد همزمان	پارامتر استخراج شده	رابطه آماری
مرتبه اول	میانگین	$\mu_1 = \left(\sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} I_{i,j}\right)/n$
	انتروپی	$E = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} P(i,j)\,ln(P(i,j))$
	واریانس	$Var = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} I_{i,j} \times (I_{i,j} - \mu)^2$
	چولگی	$S = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} I_{i,j} \times (I_{i,j} - \mu)^3$
	رنج داده	$D_R = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} ln(I_{i,j})$
مرتبه دوم	میانگین	$\mu_i = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} i \times P(i,j)$
	واریانس	$\delta^2 = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} (i-\mu_i)^2 \times P(i,j)$
	تجانس	$Cont = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} (i-j)^2 \times P(i,j)$
	عدم شباهت	$Diss = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} \lvert i-j\rvert \times P(i,j)$
	انتروپی	$Ent = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} P(i,j)\,log(P(i,j))$
	همگنی	$Hom = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} \frac{P(i,j)}{1+(i-j)^2}$
	وابستگی	$Corr = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} \frac{(i-\mu_i)(i-\mu_i)P(i,j)}{\sigma_i\sigma_j}$
	گشتاور دوم زاویه‌ای	$ASM = \sum_{i=0}^{N_{g-1}}\sum_{j=0}^{N_{g-1}} (P_d(i,j))^2$

پس از تهیه نمونه‌های آموزشی مناسب برای هر یک از کلاس‌های موضوعی، فرآیند طبقه‌بندی تصاویر سنجش از دوری بر روی هر یک از تصاویر زمان اول و دوم صورت می‌گیرد تا نقشه‌های موضوعی متناظر آن‌ها تهیه گردد. طبقه‌بندی تصاویر می‌تواند با روش‌های متعددی از جمله، تکنیک‌های کمترین‌فاصله، بیشترین‌شباهت، درخت تصمیم‌گیری، شبکه‌های عصبی و ماشین بردار پشتیبان صورت گیرد. در این مقاله به منظور طبقه‌بندی تصاویر چندطیفی و تولید نقشه‌های موضوعی چندزمانه از الگوریتم طبقه‌بندی ماشین‌های بردار پشتیبان [۱] (SVM) استفاده شد [۱۸]. بررسی مطالعات پیشین نشان می‌دهد این طبقه‌بندی‌کننده در مواقعی که داده‌ها از توزیع آماری خاصی پیروی نمی‌کنند و همچنین در مواردی که داده‌های آموزشی هر کلاس ناهمگن باشند، عملکرد بهتری نسبت به سایر روش‌ها دارد [۱۹]. ایده اصلی SVM، تعیین ابرصفحه بهینه، به منظور جداسازی دو کلاس با بیشترین حاشیه جداسازی است [۱۸]. برای داده‌های که به صورت خطی قابل تفکیک نیستند، از طریق یک تابع کرنل فضایی، ورودی به یک فضای ویژگی با ابعاد بالاتر که در آن یک فوق صفحه خطی یافت می‌شود، نگاشت می‌گردد. در صورت استفاده از توابع کرنل، تابع تصمیم‌گیری SVM را می‌توان از رابطه ۹ محاسبه کرد [۱۸]:

$$f(x) = \sum_{i=1}^{N} y_i \alpha_i k(x, x_i) + b \qquad (۹)$$

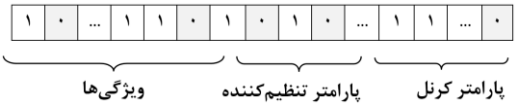
که در رابطه (۹)، N تعداد نمونه‌های آموزشی مربوط به دو کلاس، $\alpha_i, y_i \in \{-1, +1\}$ ضرایب لاگرانژ و $k(x, x_i)$ تابع کرنل و b ضریب بایاس می‌باشد.

از متداول‌ترین توابع کرنل می‌توان به کرنل چندجمله‌ای، سیگموئیدی و توابع پایه شعاعی (RBF) اشاره کرد. برای استفاده از SVM در حالت چند کلاسه دو استراتژی یک در برابر بقیه (OAA) و یک در برابر یک (OAO) وجود دارد. در روش یک در برابر یک، برای هر زوج کلاس ممکن، از یک SVM باینری استفاده می‌شود، بنابراین برای C کلاس $C(C-1)/2$ طبقه‌بندی کننده‌ی باینری خواهیم داشت. در روش یک در برابر بقیه هر SVM، داده‌های یک کلاس را از داده‌های کلاس دیگر جدا می‌کند. در این روش برای هر C کلاس C طبقه‌بندی کننده‌ی باینری خواهیم داشت. در هر دو روش برچسب نهایی داده از طریق روش رأی‌گیری حداکثر تعیین می‌شود [۲۰].

دو چالش اصلی که منجر به کاهش احتمالی صحت طبقه‌بندی مورد نظر در راستای تولید نقشه موضوعی می‌شود عبارتند از: ۱- حجم بالای فضای ویژگی و وابستگی‌های موجود بین آن‌ها و۲- تعیین مقادیر پارامترهای کرنل

[۱] Support vector machine: SVM

مورد استفاده. در مورد چالش اول، استفاده از همه ویژگی‌ها در یک الگوریتم به دلیل مشکلات ناشی از بالا بودن تعداد ویژگی‌ها و وابستگی بین آنها، لزوماً منجر به بهبود نتایج نخواهد شد [۲۱]. انتخاب زیرمجموعه ای از ویژگی‌ها، منجر به کاهش نویز ایجاد شده در هنگام استخراج ویژگی‌ها، کاهش شدید حجم محاسبات و زمان نسبت به حالت استفاده از تمامی ویژگیها و افزایش باورپذیری نسبت به استفاده از فضای کوچک تر ویژگی می‌شود [۲۱]. بنابراین انتخاب ویژگی‌های بهینه ضروری می‌باشد. در مورد چالش دوم، در تعیین مقادیر پارامترهای مورد استفاده در کرنل طبقه‌بندی‌کننده SVM، معمولاً از روش زمان‌بر سعی و خطا استفاده می‌شود. بنابراین، به منظور استفاده از طبقه‌بندی‌کننده SVM که بالاترین صحت و بیشترین کارآیی را داشته باشد، تعیین مقادیر بهینه پارامترهای کرنل، امری ضروری می‌باشد. تاکنون روش‌های کلاسیک و همچنین تکنیک‌های هوشمند متعددی برای رفع این مسئله ارائه شده است. از آن جمله می‌توان به روش جستجوی شبکه‌ای و روش‌های گرادیان‌مبنا اشاره نمود. عمده روش‌هایی که به این طریق اقدام به استخراج ویژگی و تعیین پارامتر می‌نمایند، پیچیدگی‌های محاسباتی بالایی داشته و زمان‌بر هستند و در نهایت مجبور به لحاظ فرضیات محدودکننده در خصوص پارامترهای فوق هستند. در سال‌های اخیر استفاده از الگوریتم‌های بهینه‌سازی فراابتکاری به سبب توانایی بالا در انتخاب و ترکیب راه‌حل-های امیدبخش و جهت‌گیری به سمت جواب بهینه مورد توجه قرارگرفته‌اند. در این تحقیق با الهام گرفتن از مقاله [۲۴]، از الگوریتم ژنتیک به عنوان یک الگوریتم فراابتکاری جهت برآورد مقادیر پارامترهای SVM و انتخاب ویژگی‌های بهینه طبقه‌بندی مورد استفاده قرار گرفته است. انتظار می‌رود، انتخاب ویژگی‌های بهینه و تنظیم مناسب مقادیر پارامترهای کرنل SVM با الگوریتم ژنتیک سبب بهبود نتایج طبقه‌بندی و بهنگام‌رسانی نقشه‌های موضوعی شود. در الگوریتم ژنتیک بکار رفته در این مقاله، هر یک از کروموزوم‌ها از سه بخش مجزا تشکیل شده‌اندکه ساختار آن در شکل (۵) ارائه شده است.



شکل ۵- ساختار کروموزوم‌های بکار رفته در این تحقیق

مطابق شکل (۵)، بخش اول، تعیین‌کننده ویژگی‌های مورد استفاده در الگوریتم است که طولی برابر با تمامی ویژگی‌های مورد استفاده است. رمزگشایی این بخش به این صورت می‌باشد که عدد یک نشانگر حضور ویژگی موردنظر در طبقه‌بندی SVM بوده و صفر بودن آن به معنای عدم حضور آن در طبقه‌بندی است. بخش دوم و سوم نیز به ترتیب مربوط به پارامترهای کرنل است. رمزگشایی این بخش بصورت تبدیل اعداد از مبنای ۲ به ۱۰ می‌باشد.

الگوریتم ژنتیک پیاده‌سازی شده در این تحقیق، براساس کمینه کردن مقدار تابع هزینه تعریف شده است، لذا مقدار شایستگی براساس کمینه کردن هزینه از رابطه (۱۰)، محاسبه می‌گردد که با کمینه شدن این مقدار، صحت طبقه‌بندی بیشینه می‌گردد.

$$Fitness\ Value = 1 - \ kappa\ coefficient \qquad (۱۰)$$

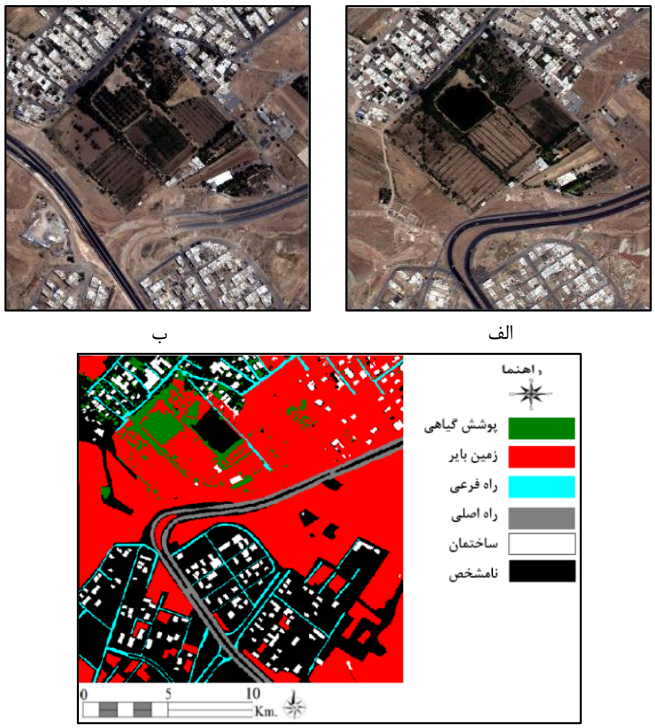
پس از طبقه‌بندی جداگانه تصاویر چندزمانه با نمونه‌های آموزشی پالایش شده، نقشه‌های موضوعی چندزمانه فراهم شده و با مقایسه این نقشه‌ها، نقشه تغییرات که بیانگر ماهیت تغییرات است، تولید می‌شود.

۳- داده‌ها و منطقه مورد مطالعه

داده‌های مورد استفاده در این تحقیق شامل دو تصویر ماهواره‌ای (از سال ۲۰۱۱ و ۲۰۱۵) و یک نقشه موضوعی قدیمی (سال ۲۰۰۹) از بخشی از شهر شیراز است که جزییات آن در جدول (۲) ارائه شده است. شکل (۶) نیز نقشه موجود منطقه (شامل ۵ کلاس موضوعی) و ترکیب رنگی طبیعی تصاویر سنجش از دور مورد استفاده در این تحقیق را نشان می‌دهد.

جدول ۲- مشخصات تصاویر مورداستفاده

سکو	Worldview2	Worldview3
باندهای طیفی	Red, Green, Blue, Coastal, Yellow, Red Edge, Near-IR1, Near-IR2	Red, Green, Blue, Coastal, Yellow, Red Edge, Near-IR1, Near-IR2
تاریخ اخذ تصویر	۲۰۱۱	۲۰۱۵
قدرت تفکیک مکانی	۱/۶۰ متر	۱/۲۰ متر
ارتفاع مداری	۷۰۰ کیلومتر	۶۱۷ کیلومتر
قدرت تفکیک رادیومتریکی	۱۱ بیت	۱۱ بیت
ابعاد مکانی (بعد از یکسان‌سازی)	۶۰۰×۶۰۰ پیکسل	۶۰۰×۶۰۰ پیکسل



شکل ۶- الف) تصویر ماهواره‌ای Worldview2 (سال ۲۰۱۱)، ب) تصویر ماهواره‌ای Worldview3 (سال ۲۰۱۵)، ج) نقشه موجود منطقه (سال ۲۰۰۹) ازمنطقه ای در شهر شیراز

۴- پارامترهای ارزیابی

روش استاندارد ارزیابی صحت طبقه‌بندی؛ انتخاب یک مجموعه نمونه‌های ارزیابی است که با عناوین نمونه‌های واقعیت زمینی[۱] یا داده‌های مرجع[۲] نیز معروف هستند. در تحقیق حاضر نمونه‌های ارزیابی با استفاده از تفسیر بصری تصاویر با قدرت تفکیک بالا و نقشه‌های موجود فراهم گردید که یک روش رایج در این زمینه می‌باشد. مقایسه برچسب نمونه‌های واقعیت زمینی با نتایج حاصل از طبقه‌بندی، منجر به ایجاد ماتریس خطا[۳] می‌شود. با استفاده از این ماتریس می‌توان معیارهای ارزیابی صحت متعددی از جمله صحت کلی[۴] ، صحت تولیدکننده[۵] ، صحت کاربر[۶] و ضریب کاپا [۷] را محاسبه نمود که در این مطالعه مورد استفاده قرار گرفته است.

۱ Ground truth
۲ Reference data
۳ Error Matrix
٤ Overall accuracy
٥ Producer’s accuracy
٦ User’s accuracy
۷ Kappa coefficient

۵- پیاده‌سازی و ارزیابی نتایج

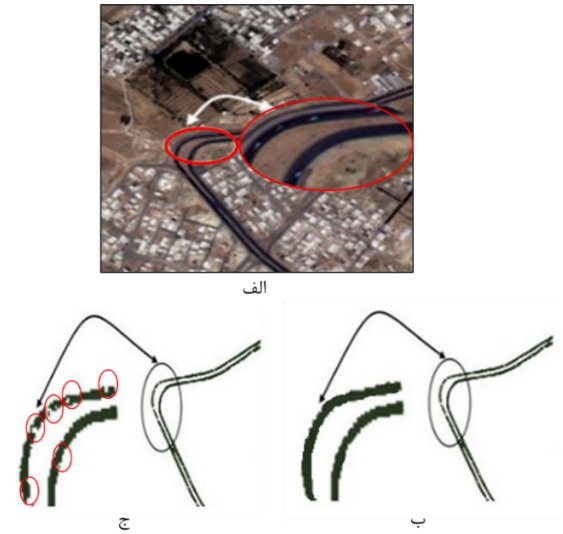
در این بخش جزئیات پیاده‌سازی و نتایج حاصله، بیان شده و مورد بحث قرار می‌گیرد. جهت پیاده‌سازی روش پیشنهادی، از نرم‌افزار MATLAB 2013b استفاده شده است.

در مطالعه موردی انجام گرفته در تحقیق حاضر، تصحیحات رادیومتریک و اثرات ابر ضروری نبوده و لذا از آنها صرفنظر شده است. از آنجاییکه که نقشه قدیمی مورد استفاده، زمین‌مرجع بوده و تصاویر نیز دارای اطلاعات زمینی دقیق بودند؛ ابتدا نقشه از فرمت برداری به فرمت رستری تبدیل گردید و سپس یکی از تصاویر به عنوان مبنای ثبت هندسی دقیق بین نقشه و تصاویر قرار گرفت. جهت ثبت هندسی دقیق نقشه رستری از تبدیل افاین با تعداد نقطه کنترل مناسب اخذ شده از گوشه‌های ساختمان استفاده شد. در این مطالعه، کلاس‌های پوششی و کاربری اراضی از منطقه مورد نظر شامل راه اصلی، راه فرعی، مناطق مسکونی و در حال ساخت، پوشش گیاهی و زمین بایر می‌باشد. همانطور که در بخش قبل اشاره شد، از الگوریتم SVM با کرنل شعاعی (RBF) به عنوان طبقه‌بندی‌کننده استفاده شده و از الگوریتم ژنتیک برای انتخاب ویژگی‌های بهینه و همچنین بهینه‌سازی پارامترهای کرنل طبقه‌بندی

کننده SVM استفاده شده است. در این راستا، طول کروموزوم‌ها معادل با تعداد ویژگی‌های هر یک از تصاویر ماهواره‌ای (شامل ۸ ویژگی) در نظر گرفته شد. همچنین با توجه به محدوده در نظر گرفته شده برای تغییر پارامترهای C و γ در طبقه‌بندی‌کننده SVM، طول کروموزوم‌ها به ۲۸ افزایش یافت که شامل ۸ ژن برای تعداد ویژگی‌ها و ۱۰ ژن برای مقادیر هر کدام از پارامترهای کرنل است. روش مورد استفاده برای انتخاب والدین روش تورنمنت[1] بوده و تعداد نسل‌ها برابر با ۶۰۰ و تعداد جمعیت موجود در هر نسل ۵۰ عضو است. احتمال جهش ۱۰٪ بوده و در هر نسل ۲۰٪ آن به نسل آتی منتقل می‌شوند.

۵-۱- پالایش و انتخاب نمونه‌های آموزشی بهینه

در این قسمت ابتدا، نمونه‌های آموزشی اولیه هر یک از کلاس‌های مذکور با هم‌پوشانی نقشه و تصاویر زمان‌های اول و دوم تعیین می‌شوند و مطابق با روش پیشنهادی جهت پالایش، آزمون خی-دو با سطح معناداری $\alpha = 0.05$ انجام می‌گردد. در شکل (۷)، تأثیر پالایش داده‌ها با استفاده از روش پیشنهادی نسبت به استفاده از داده‌های ویرایش نشده ارائه شده است.



الف ب ج

شکل ۷- الف) ناخالصی کلاس راه در تصویر ب) داده‌های کلاس راه قبل از پالایش ج) داده‌های کلاس راه بعد از پالایش. (دایره‌های قرمز بیانگر حفره‌های ایجاد شده (نمونه‌های ناخالص) بعد از پالایش کلاس راه است

همان‌طور که در شکل (۷) مشاهده می‌شود در کلاس راه اصلی، عوارض غیرمرتبط مثل خودروها پس از پالایش به صورت دقیق از داده‌های آموزشی حاصل از نقشه حذف شده‌اند. جهت تعیین تأثیر سطوح اطمینان مختلف، بکارگیری سطوح اطمینان ۹۹٪، ۹۷/۵٪ و ۹۵٪ در پالایش نمونه‌های آموزشی مبتنی بر آزمون آماری خی-دو مورد بررسی قرار گرفت. بدین منظور نمونه‌های پالایش‌شده در هر یک از سطوح اطمینان مذکور به صورت مجزا در طبقه‌بندی پارامتریک بیشترین شباهت مورد استفاده قرار می‌گیرند (جدول (۳)). همچنین نقشه‌های موضوعی تولید شده برای تصاویر اخذ شده در سال ۲۰۱۱ و ۲۰۱۵ را در سطوح اطمینان مختلف در شکل (۸) نشان داده شده‌اند.

جدول ۳- نتایج طبقه‌بندی تصاویر به ازای سطوح اطمینان مختلف آزمون آماری خی-دو در پالایش تصاویر

معیارهای ارزیابی صحت						
ردیف	سطح اطمینان	ضریب کاپا (درصد)		صحت کلی (درصد)		
		تصویر سال ۲۰۱۵	تصویر سال ۲۰۱۱	تصویر سال ۲۰۱۵	تصویر سال ۲۰۱۱
1	99%	85.33	88.88	79.04	83.47
2	97.5%	77.75	74.81	69.68	67.219
3	**95%**	**86.32**	**90.32**	**80.84**	**87.87**

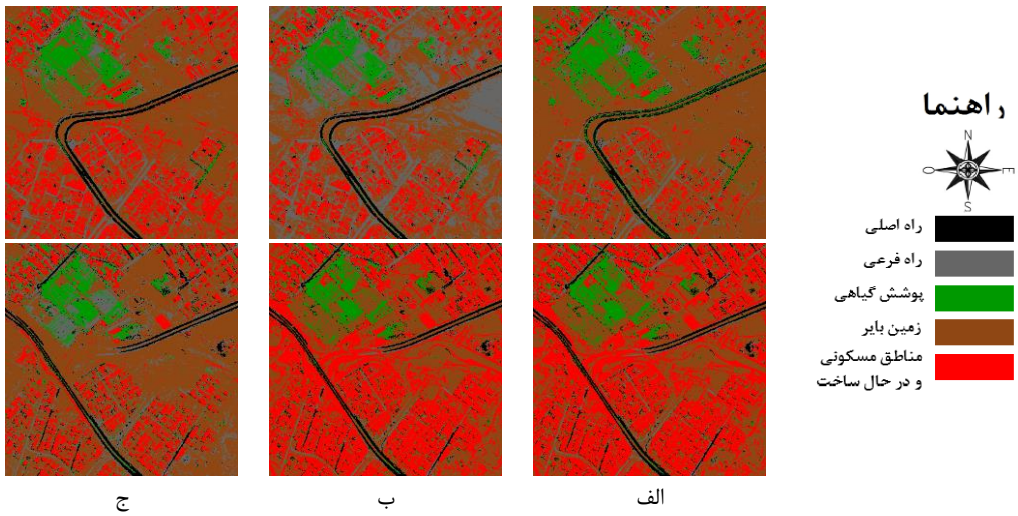
همانطور که در شکل (۸)، مشاهده می‌شود هر چه سطح اطمینان افزایش یابد، میزان خطا افزایش می‌یابد. در این راستا، انتظار می‌رود با در نظر گرفتن سطح اطمینان ۹۰٪ (افزایش سختگیری) صحت کلی نسبت به سطح اطمینان ۹۹٪ بیشتر شود که تا حدودی در مورد نتایج حاصل از داده‌های این تحقیق صادق است، جاییکه صحت کلی بدست آمده برای سطح اطمینان ۹۵٪ بیشتر از دیگر سطوح اطمینان است. اما باید به این مهم نیز توجه کرد که بر خلاف انتظار صحت کلی و ضریب کاپا برای سطح اطمینان ۹۹٪ بیشتر از سطح اطمینان ۹۷/۵٪ می‌باشد. چنین نتیجه‌ای نشان می‌دهد که، انتخاب سطح اطمینان مناسب کاملا وابسته به نوع عوارض موجود در داده‌های بکار رفته می‌باشد. در واقع، توزیع مختلف عوارض در سطح تصویر و همچنین گوناگونی فراوانی آن‌ها، موجب شده که تاثیر پالایش در سطوح مختلف اطمینان دارای رفتار سیستماتیک (افزایشی/کاهشی) نباشد.

همچنین مطابق با جدول (۳)، ضریب کاپای حاصل برای تصویر سال ۲۰۱۵ کمتر از سال ۲۰۱۱ می‌باشد، علت این امر می‌تواند ناشی از عوامل مختلفی نظیر کیفیت و وضوح تصویر مربوط به سال ۲۰۱۵، اختلاف زمانی نسبتا زیاد بین نقشه و تصویر سال ۲۰۱۵ باشد. با توجه به نتایج بدست آمده می‌توان اینطور استنباط کرد که پالایش نمونه

[1] Tournament

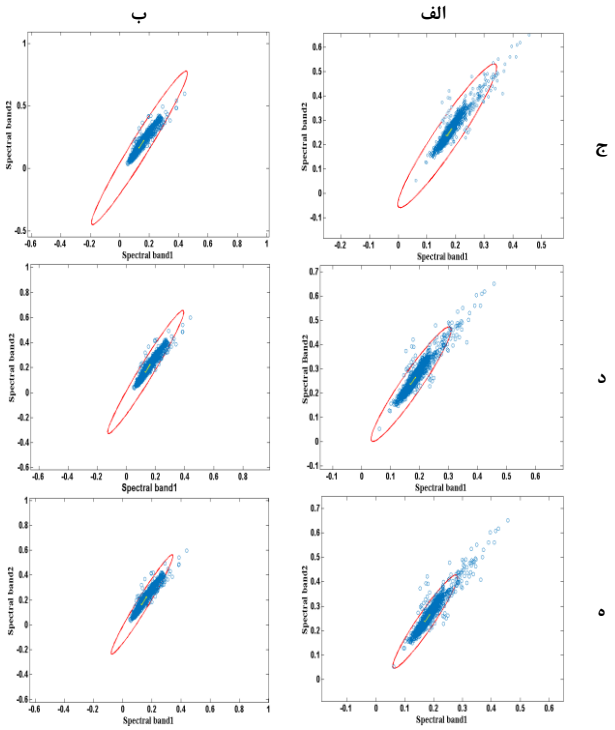
های آموزشی وابستگی بالایی به سطح اطمینان انتخاب شده دارد. همچنین عدم سیستماتیک نبودن نتایج به ازای سطوح اطمینان مختلف را می‌توان ناشی از عدم توزیع یکنواخت داده‌ها در سطح تصاویر و همچنین عدم وجود داده‌های پرت با اندازه یکسان در کلاس‌ها دانست. در نمونه‌های پالایش‌شده با سطح اطمینان ۹۵٪، صحت بالاتری حاصل شده است. استفاده از آزمون خی-دو سبب شده است که پیکسل‌های اشتباه (پرت) که به کلاس‌های مورد نظر تعلق نداشته‌اند، در جریان پالایش حذف گردند.

شکل (۹) منحنی توزیع خی-دو، تحت سطوح اطمینان مختلف برای هر کلاس که محدوده پذیرش یا ردّ نمونه‌های آموزشی در فرآیند پالایش می‌باشد را نشان می‌دهد. این شکل با پارامترهای نیم قطر بزرگ (a)، نیم قطر کوچک (b) و زاویه بین محور x با بزرگ‌ترین بردار ویژه مشخص شده است.

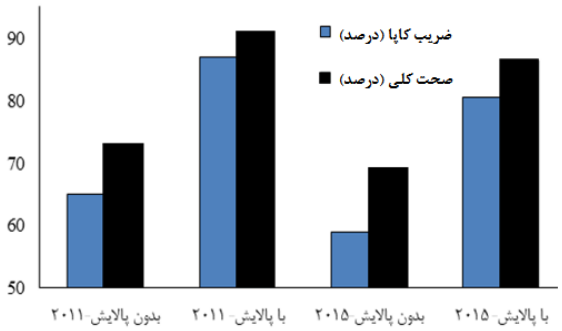


شکل ۸- تاثیر سطوح اطمینان آزمون خی-دو در پالایش نمونه‌های آموزشی در نتیجه طبقه‌بندی تصاویر مورد استفاده. نقشه کاربری/پوششی اراضی تهیه شده از تصویر سال ۲۰۱۱ (ردیف بالا) و ۲۰۱۵ (ردیف پایین) تحت سطوح اطمینان مختلف: الف) سطح اطمینان۹۹٪، ب) سطح اطمینان ۹۷/۵ ٪ ج) سطح اطمینان ۹۵٪

پس از همپوشانی نقشه با هر یک از تصاویر، پیکسل‌های مربوط به هر کلاس تعیین شده و با آزمون خی-دو پالایش می‌شوند و سپس به منظور انتخاب نمونه‌های آموزشی بهینه از الگوریتم *k*-means برای خوشه‌بندی استفاده می‌شود. ولی در یک حالت به‌منظور بررسی تأثیر پالایش بر روی صحت نتایج طبقه‌بندی، قبل از انجام خوشه‌بندی از فرآیند پالایش صرف نظر گردید و نمونه‌های آموزشی فقط با استفاده از الگوریتم خوشه‌بندی k-means انتخاب شده و وارد طبقه‌بندی به روش بیشترین‌شباهت شده و نتایج در شکل (۱۰) ارائه شده است. در این حالت صحت کلی و ضریب کاپای بدست آمده برای تصویر سال ۲۰۱۱ به ترتیب ۷۳٪ و ۶۵٪ می‌باشند. در حالی که در صورت استفاده از آزمون آماری خی-دو برای پالایش (قبل از مرحله خوشه‌بندی)، صحت کلی و ضریب کاپا به ترتیب ۹۱٪ و ۸۷٪ بوده است. برای تصویر سال ۲۰۱۵ صحت کلی و ضریب کاپا در حالت بدون استفاده از آزمون خی-دو ۶۹٪و ۵۹٪ می‌باشد که در صورت استفاده از آزمون آماری خی-دو صحت کلی و ضریب کاپا به ترتیب به ۸۶/۳۲٪ و ۸۰/۴۸٪ افزایش پیدا می‌کند.

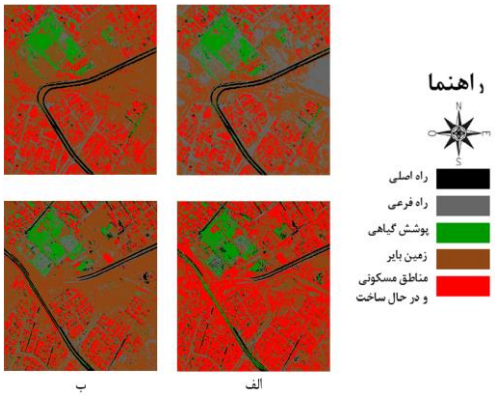


شکل ۹- منحنی توزیع خی-دو در پالایش نمونه‌های آموزشی کلاس راه اصلی؛ ستون الف نتایج مربوط به تصویر ۲۰۱۱ و ستون (ب) نتایج مربوط به تصویر ۲۰۱۵ بوده و سطرهای (ج)، (د) و (ه) به ترتیب بیانگر سطوح اطمینان ۹۹٪، ۹۷/۵٪ و ۹۵٪ می‌باشد



شکل ۱۰- تأثیر پالایش نمونه‌های آموزشی با آزمون آماری خی-دو در نتایج طبقه‌بندی

شکل (۱۱)، نتیجه طبقه‌بندی هر یک از تصاویر ۲۰۱۱ و ۲۰۱۵ در صورت استفاده و عدم استفاده از آزمون آماری خی-دو را نشان می‌دهند.

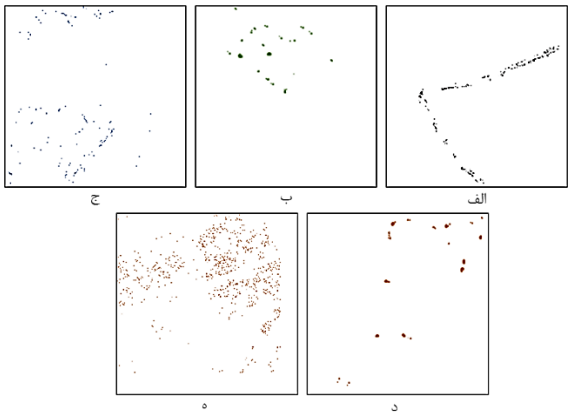


شکل ۱۱- نتایج طبقه‌بندی سال ۲۰۱۱ (ردیف بالا) و ۲۰۱۵ (ردیف پایین)، الف) بدون اعمال پالایش، ب) با اعمال پالایش مبتنی بر آزمون آماری خی-دو

جدول (۴) تعداد نمونه‌های آموزشی هر کلاس قبل و بعد از پالایش برای تصاویر سال ۲۰۱۱ و ۲۰۱۵ را نشان می‌دهد. همچنین شکل (۱۲)، تراکم و توزیع مکانی نمونه های آموزشی پالایش‌شده به روش پیشنهادی در تصویر سال ۲۰۱۱ را نشان می‌دهد.

جدول ۴- تعداد نمونه‌های آموزشی هر کلاس قبل و بعد از پالایش در تصاویر سالهای ۲۰۱۱ و ۲۰۱۵ در سطح اطمینان ٪۹۵

کلاس	تعداد نمونه‌های آموزشی (در واحد پیکسل)					
	تصویر ۲۰۱۱			تصویر ۲۰۱۵		
	قبل از پالایش	پس از پالایش	پس از خوشه‌بندی	قبل از پالایش	پس از پالایش	پس از خوشه‌بندی
راه اصلی	۱۱۳۷۸	۱۰۵۲۰	۱۰۸۷	۱۱۳۷۸	۱۰۲۶۴	۲۰۷۶
درخت	۱۵۹۵۹	۱۴۰۴۰	۹۲۳	۱۵۹۵۹	۱۴۵۱۲	۱۵۸۶
راه فرعی	۱۵۱۱۵	۱۳۵۱۰	۴۷۷	۱۵۱۱۵	۱۳۳۴۰	۱۱۰۶
زمین بایر	۱۷۱۵۱۷	۱۵۱۷۵۳	۲۲۷۶	۱۷۱۵۱۷	۱۵۱۱۷۹	۲۶۳۱
مسکونی	۱۸۳۵۹	۱۶۵۹۶	۹۴۲	۱۸۳۵۹	۱۶۱۴۶	۱۱۲۲



شکل ۱۲- نمونه‌های آموزشی در تصویر ۲۰۱۱؛ الف) کلاس راه اصلی ب) کلاس درخت ج) کلاس راه فرعی د) کلاس مسکونی ه) کلاس زمین بایر

۵-۲- تولید نقشه موضوعی با استفاده از الگوریتم GASVM

همانطور که در بخش قبل بیان شد، در این تحقیق جهت طبقه‌بندی تصاویر سنجش از دور و تولید نقشه موضوعی، از GASVM استفاده شده است. در روش GASVM، نمونه‌های آموزشی و ویژگی‌های طیفی (NDVI) و بافتی ماتریس GLCM (درجه اول و دوم) توسط الگوریتم ژنتیک انتخاب شده و به‌عنوان ورودی، وارد فاز طبقه‌بندی می‌شوند. همان-طور که اشاره شد، در این تحقیق بهینه‌سازی پارامترهای SVM، همزمان با انتخاب ویژگی‌ها صورت می‌گیرد. در جدول (۵) مقادیر بهینه پارامترهای GASVM به ازای ورودی‌های مختلف (S: صرفاً باندهای طیفی؛ S+NDVI: باند-های طیفی به همراه شاخص NDVI؛ S+NDVI+C-1st: باندهای طیفی به همراه شاخص طیفی NDVI و ویژگی بافت مرتبه ۱؛ S+NDVI+C-2nd: باندهای طیفی به همراه شاخص طیفی NDVI و ویژگی بافت مرتبه ۲؛ S+NDVI+C-1st+C-2nd: باندهای طیفی به همراه شاخص طیفی NDVI و ویژگی بافت مرتبه ۱ و ۲ ارائه شده است.

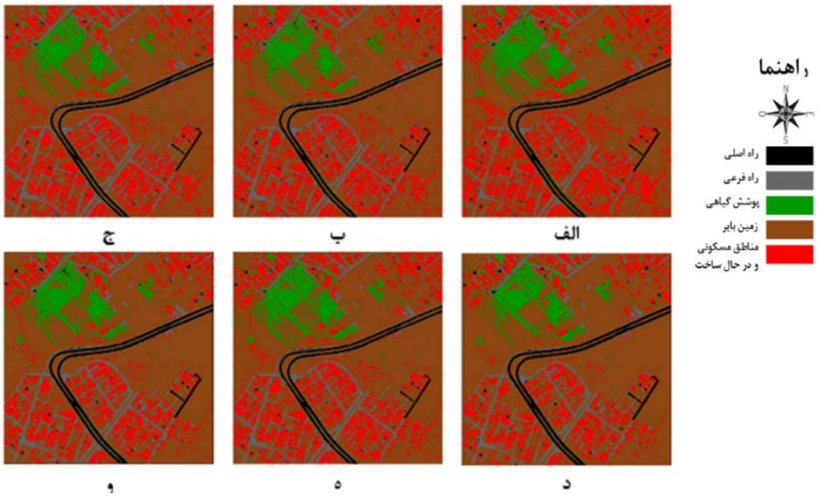
جدول ۵- مقادیر بهینه پارامترهای GASVM به ازای ورودی‌های مختلف

نوع داده ورودی	تصویر	پارامتر تنظیم (C)	پارامتر کرنل (γ)
S	2011	476	0.0104
	2015	1016	0.0652
S+NDVI	2011	286	0.3121
	2015	118	0.002
S+NDVI+C-1st	2011	470	0.0509
	2015	1016	0.0652
S+NDVI+C-2nd	2011	958	0.0048
	2015	175	0.0572
S+NDVI+C-1st+C-2nd	2011	621	0.0814
	2015	3	1.7601

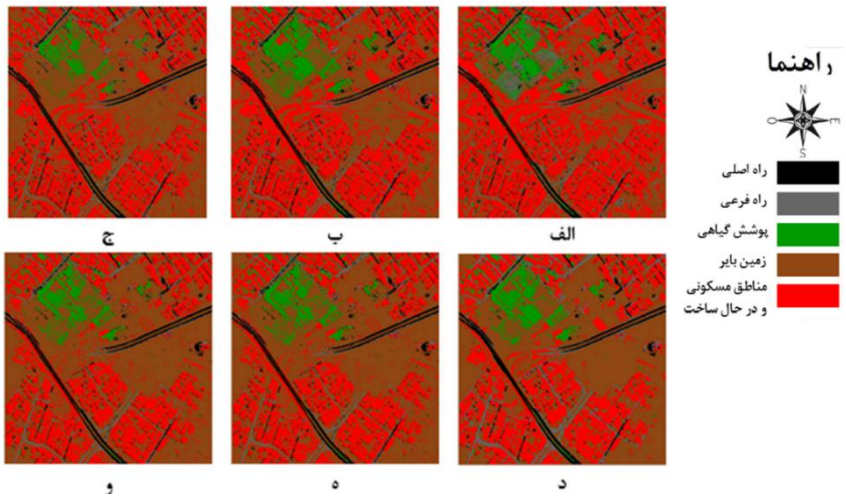
شکل‌های (۱۳) و (۱۴) به ترتیب نقشه طبقه‌بندی‌شده مربوط به سال ۲۰۱۱ و ۲۰۱۵ را که از اعمال الگوریتم SVM و GASVM بر ورودی‌های مختلف بدست آورده شده‌اند را نشان می‌دهد. همچنین، در جدول (۶) نتایج پارامترهای ارزیابی صحت نقشه کاربری اراضی سال ۲۰۱۱ و ۲۰۱۵ بر مبنای روشهای نامبرده ارائه شده است. همچنین نمودارهای ارائه شده در شکل (۱۵)، میزان تاثیر اطلاعات بافتی و طیفی در نتایج طبقه‌بندی را به صورت کیفی و کمی نشان می‌دهد.

بررسی پارامترهای ارزیابی صحت طبقه‌بندی این تصاویر نشان می‌دهد که استخراج ویژگی‌های بافت و همچنین استفاده از الگوریتم ژنتیک در انتخاب مجموعه ویژگی‌های بهینه سبب بهبود قابل ملاحظه نتایج طبقه‌بندی شده است. طبق جدول (۶)، این نتیجه حاصل می‌شود که استفاده از ویژگی‌های بهینه موجب افزایش ضریب کاپا از ۸۹/۹۴ به ۹۵/۳۴ و صحت کلی از ۹۳/۲۰٪ به ۹۶/۸۸٪ شده است. در مورد استفاده از شاخص طیفی می‌توان گفت، استفاده از شاخص گیاهی نرمالیزه شده (NDVI) باعث بهبود صحت کلی و ضریب کاپا و صحت تولیدکننده در تصویر سال ۲۰۱۱ شده است. البته با توجه به میزان کم پوشش گیاهی در تصاویر ممکن است تأثیر شاخص گیاهی نرمالیزه شده در پارامتر صحت کلی یا ضریب کاپای کل تصویر محسوس نباشد.

بهترین نتایج طبقه‌بندی مربوط به حالت استفاده از ویژگی‌های بافت مراتب اول و دوم می‌باشد. همچنین تأثیر ویژگی‌های بافت مرتبه اول بیشتر از ویژگی‌های بافت مرتبه دوم می‌باشد. این امر می‌تواند عمدتاً به این دلیل باشد که بافت مرتبه دوم باعث از بین رفتن جزئیات و نرم شدن لبه برخی از عوارض شده است. بر این اساس و طبق جدول (۶) صحت تولیدکننده راه اصلی از ۹۵/۶۲٪ به ۹۹/۲۳٪ و صحت تولیدکننده مسکونی از ۹۷/۳۲٪ به ۹۹/۴۵٪ افزایش یافته است.



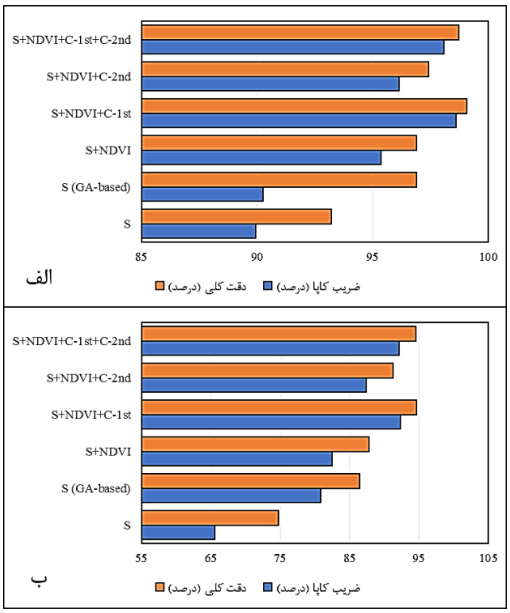
شکل ۱۳- نقشه پوشش/کاربری اراضی سال ۲۰۱۱ با اعمال الف) روش SVM روی S، ب) روش GASVM روی S، ج) روش GASVM روی S+NDVI، د) روش GASVM روی S+NDVI+C-1st، ه) روش GASVM روی S+NDVI+C-2nd، و) روش GASVM روی S+NDVI+C-1st+C-2nd



شکل ۱۴- نقشه پوشش/کاربری اراضی سال ۲۰۱۵ با اعمال الف) روش SVM روی S، ب) روش GASVM روی S، ج) روش GASVM روی S+NDVI، د) روش GASVM روی S+NDVI+C-1st، ه) روش GASVM روی S+NDVI+C-2nd، و) روش GASVM روی S+NDVI+C-1st+C-2nd

جدول ۶- پارامترهای ارزیابی صحت تهیه نقشه کاربری/پوشش اراضی سال ۲۰۱۱ و سال ۲۰۱۵ با استفاده از ورودی‌های مختلف

داده‌ها	ویژگی‌ها	روش طبقه‌بندی	صحت تولیدکننده «راه اصلی»	صحت کاربر «راه اصلی»	صحت تولیدکننده «پوشش گیاهی»	صحت کاربر «پوشش گیاهی»	صحت تولیدکننده «راه فرعی»	صحت کاربر «راه فرعی»	صحت تولیدکننده «زمین بایر»	صحت کاربر «زمین بایر»	صحت تولیدکننده «مسکونی»	صحت کاربر «مسکونی»	ضریب کاپا	صحت کلی
تصویر ۲۰۱۱	S	SVM	96.72	99.37	99.63	98.09	97.10	81.22	91.37	96.66	88.11	86.48	89.94	93.20
	S	GASVM	96.40	99.18	97.93	99.44	97.61	93.33	96.69	98.84	97.09	90.26	95.34	96.8
	S+NDVI	GASVM	95.62	99.22	99.67	96.91	98.14	92.21	96.57	99.13	97.32	91.12	95.36	96.8
	S+NDVI+C-1^{st}	GASVM	99.23	99.25	99.95	99.90	98.31	98.22	98.31	99.14	99.45	98.95	98.60	99.07
	S+NDVI+C-2^{nd}	GASVM	96.33	99.69	99.12	97.24	97.92	92.47	97.42	99.02	97.53	93.82	96.14	97.42
	S+NDVI+C-1^{st}+C-2^{nd}	GASVM	98.80	98.77	98.25	99.76	98.63	96.70	98.8	99.40	98.50	97.47	98.08	98.72
تصویر ۲۰۱۵	S	SVM	91.84	77.88	98.92	98.80	42.15	79.47	70.75	86.46	97.62	49.92	65.55	74.78
	S	GASVM	93.30	79.61	98.29	99.55	53.40	89.50	90.60	93.56	99.46	71.16	80.88	86.45
	S+NDVI	GASVM	95.96	78.57	83.43	100	51.26	89.70	95.46	90.73	98.98	85.15	82.48	87.82
	S+NDVI+C-1^{st}	GASVM	95.77	82.89	99.05	100	75.30	94.74	99.03	99.21	99.89	92.39	92.34	94.61
	S+NDVI+C-2^{nd}	GASVM	93.71	82.12	95.46	100	62.31	91.49	98.17	92.65	98.39	94.05	87.39	91.24
	S+NDVI+C-1^{st}+C-2^{nd}	GASVM	95.58	84.19	98.55	100	75.28	94.69	98.99	98.020	99.81	93.56	92.18	94.52



شکل ۱۵- ضریب کاپا و صحت کلی نقشه کاربری اراضی الف) سال ۲۰۱۱ و ب) ۲۰۱۵ با استفاده از صرفاً اطلاعات طیفی و سایر داده‌های ورودی

پایین‌ترین صحت طبقه‌بندی مربوط به حالتی است که تنها از اطلاعات طیفی و بدون حضور الگوریتم ژنتیک، طبقه بندی صورت گرفته است. استفاده از الگوریتم ژنتیک موجب افزایش ضریب کاپا از ۶۵/۵۵٪ به ۸۰/۸۸٪ شده است. در صورت استفاده از شاخص گیاهی نرمالیزه شده (NDVI) نیز ضریب کاپا و صحت کلی به ترتیب از ۸۶/۴۵٪ و ۸۰/۸۸٪ به ۸۷/۸۲٪ و ۸۲/۴۸٪ افزایش یافته است. در کل صحت کلی و ضریب کاپای طبقه‌بندی تصویر سال ۲۰۱۵ کمتر از تصویر سال ۲۰۱۱ می‌باشد که علت عمده آن می‌تواند مربوط به فاصله زمانی زیاد بین نقشه و تصویر موردنظر و همچنین کیفیت و وضوح پایین‌تر این تصویر باشد.

۵-۳- تهیه نقشه تغییرات به روش مقایسه پس‌ازطبقه‌بندی

در این مرحله، تصاویر طبقه‌بندی شده مربوط به سال‌های ۲۰۱۱ و ۲۰۱۵ با یکدیگر مقایسه شده و نقشه ماهیت تغییرات تولید می‌گردد. به منظور حصول بالاترین صحت ممکن در شناسایی تغییرات، از نقشه‌های کاربری اراضی حاصل از ورودی S+NDVI+C-1^{st}+C-2^{nd} که دارای بالاترین صحت طبقه‌بندی بودند، استفاده شد. برای ارزیابی صحت شناسایی تغییرات، نمونه‌های ارزیابی به صورت بصری از سطح تصویر و در هفت کلاس تغییر کاربری شامل: «جاده به زمین بایر»، «پوشش گیاهی به زمین بایر»، «جاده به مناطق مسکونی و در حال ساخت»، «زمین بایر به جاده»، «زمین بایر به پوشش گیاهی»، «زمین بایر به مناطق مسکونی و در حال ساخت» و «مسکونی به زمین بایر» جمع‌آوری شد. کلاس عدم تغییر نیز با رنگ مشکی مشخص شده است. شایان ذکر است؛ خطای موجود در طبقه‌بندی هر یک از تصاویر در نقشه تغییرات نهایی، انتشاریافته و موجب کاهش صحت نقشه ماهیت تغییرات می‌شود [۲۲]. از آنجا که نقشه ماهیت تغییرات، اطلاعات کاملی از تغییرات را ارائه می‌دهد؛ بنابراین می‌توان نوع تغییرات رخداده در کاربری‌ها و پوشش اراضی را به خوبی در این نقشه نمایش داد. لازم به ذکر است در این مرحله کلاس مربوط به راه اصلی و راه فرعی باهم ادغام شده و به عنوان کلاس کلی تر «جاده» در نقشه ماهیت تغییرات نشان داده شده است. همچنین کلاس مسکونی با عنوان

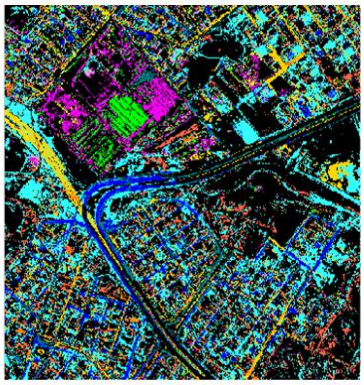
مناطق مسکونی و در حال ساخت مشخص گردیده‌اند. شکل (۱۶) نقشه ماهیت تغییرات حاصل از تکنیک مقایسه‌پس‌از طبقه‌بندی را نشان می‌دهد. از جمله تغییرات مشهود در این نقشه، تغییر کاربری اراضی از کلاس «بایر» به «مناطق مسکونی» می‌باشد که با رنگ فیروزه‌ای مشخص گردیده است. در درجه دوم، تغییر کاربری اراضی از کلاس زمین «بایر» به «جاده» از دیگر تغییرات بارزتر می‌باشد.

شکل ۱۶- نقشه ماهیت تغییرات مستخرج از روش پیشنهادی

نتایج ارزیابی صحت نشان می‌دهد، صحت کلی و ضریب کاپای نقشه ماهیت تغییرات به ترتیب ۸۳/۳۸٪ و ۷۵/۳۳٪ می‌باشد. جدول (۷) سایر پارامترهای ارزیابی صحت نقشه ماهیت تغییرات را در سه حالت نشان می‌دهد که شامل، حالت اول؛ بدون انجام فرآیند پالایش نمونه‌های آموزشی، حالت دوم؛ اعمال پالایش نمونه‌های آموزشی بدون انتخاب ویژگی‌های بهینه و حالت سوم؛ اعمال پالایش نمونه‌های آموزشی و انتخاب ویژگی‌های بهینه (روش پیشنهادی) می‌باشد. همچنین شکل (۱۷) نقشه ماهیت تغییرات حاصل از سه روش مذکور را نشان می‌دهد.

همانطور که در شکل (۱۷) مشخص است؛ پالایش داده-های آموزشی تأثیر بسزایی در نتیجه نقشه ماهیت تغییرات داشته است. در شکل نامبرده مشاهده می‌شود؛ زمانی که از داده‌های آموزشی پالایش نشده استفاده می‌شود، اکثر مناطق تغییرنیافته به صورت تغییر کاربری مشخص شده‌اند که به معنای خطای کاذب بالا در نقشه تغییرات است. در واقع عدم پالایش نمونه‌های آموزشی ضمن افزایش خطای کاذب، صحت نقشه تغییرات بین سایر کلاس‌ها را نیز کاهش می دهد. همچنین با مقایسه شکل ۱۷ (ب) با شکل ۱۷ (ج) مشاهده می‌شود که تنها استفاده از الگوریتم پالایش نمونه-های آموزشی نمی‌تواند منجر به نتایج مناسب شود و انتخاب ویژگی‌های بهینه در کنار پالایش داده‌های آموزشی امری ضروری است. به عبارت دیگر می‌توان گفت، اگرچه استفاده از الگوریتم پیشنهادی در پالایش نمونه‌های آموزشی توانسته تاثیر بسزایی در بهبود دقت نقشه ماهیت تغییرات از نظر کمی و کیفی داشته است، اما کارآیی آن زمانی بیشتر خواهد شد که همراه با فرآیند انتخاب ویژگی با استفاده از یک الگوریتم معتبر نظیر الگوریتم ژنتیک باشد. بطور کلی؛ مطابق با جدول (۷)، ضریب کاپا و صحت کلی در حالت پالایش نمونه‌های آموزشی (حالت ب) نسبت به حالت بدون پالایش (حالت الف) به اندازه ۴۱/۹۳٪ و ۳۳/۰۷٪ افزایش یافته است. همچنین استفاده از پالایش و انتخاب ویژگی مناسب (حالت ج) موجب افزایش صحت کلی به اندازه ۳۵/۷۷٪ و ۲/۷٪ نسبت به حالات (الف) و (ب) شده است.

الف



ب

ج

شکل ۱۷- نقشه ماهیت تغییرات در سه حالت منظور شده: الف) بدون انجام فرآیند پالایش، ب) با انجام فرآیند پالایش و بدون انتخاب ویژگی‌های بهینه و ج) با انجام فرآیند پالایش و انتخاب ویژگی‌های بهینه (روش پیشنهادی)

جدول ۷- پارامترهای ارزیابی صحت نقشه ماهیت تغییرات در سه حالت منظور شده: حالت اول: بدون انجام فرآیند پالایش، حالت دوم: با انجام فرآیند پالایش (بدون انتخاب ویژگی‌های بهینه) و حالت سوم: با انجام فرآیند پالایش و انتخاب ویژگی‌های بهینه (روش پیشنهادی)

روش	دقت تولیدکننده «جاده به زمین بایر»	دقت کاربر «جاده به زمین بایر»	دقت تولیدکننده «درخت به زمین بایر»	دقت کاربر «درخت به زمین بایر»	دقت تولیدکننده «جاده به مناطق مسکونی و در حال ساخت»	دقت کاربر «جاده به مناطق مسکونی و در حال ساخت»	دقت تولیدکننده «زمین بایر به جاده»	دقت کاربر «زمین بایر به جاده»	دقت تولیدکننده «زمین بایر به درخت»	دقت کاربر «زمین بایر به درخت»	دقت تولیدکننده «زمین بایر به مناطق مسکونی و در حال ساخت»	دقت کاربر «زمین بایر به مناطق مسکونی و در حال ساخت»	دقت تولیدکننده «ساختمان به زمین بایر»	دقت کاربر «ساختمان به زمین بایر»	صحت کلی	ضریب کاپا
حالت اول	19.57	0.33	0	0.25	94.20	22.34	61.90	100	**100**	65.31	38.40	92.94	68.57	85.71	47.62	29.23
حالت دوم	37.63	**85.37**	6.25	**100**	**94.37**	50	**100**	**67.63**	**100**	**100**	**80.36**	95.74	88.64	**100**	80.69	71.16
حالت سوم	**89.36**	66.66	**92**	**100**	70.42	**52.63**	**100**	64.49	**100**	**100**	80.05	**98.94**	**91.48**	**100**	**83.39**	**75.33**

۶- نتیجه گیری

هدف از تحقیق حاضر، افزایش سطح اتوماسیون و ارتقاء کارآیی تکنیک‌های شناسایی تغییرات و بهنگام رسانی نقشه‌های پوششی و کاربری اراضی مبتنی بر تصاویر سنجش‌ازدور و نقشه‌های موجود بود که با توسعه روش‌های نوین و کارآمد در مراحل؛ انتخاب اتوماتیک نمونه‌های آموزشی و پالایش اتوماتیک آن‌ها، محقق گردید. با استفاده از روش پیشنهادی در مرحله انتخاب و پالایش نمونه‌های آموزشی، «شرط مطابقت زمانی تصاویر و نقشه موضوعی موجود منطقه» برای جمع‌آوری صحیح و اتوماتیک نمونه‌های آموزشی که یکی از محدودیت‌های روش‌های قبلی در این زمینه بوده است، تا حد زیادی برطرف گردید. نتایج حاصل از پالایش داده‌های آموزشی نشان داد که بکارگیری آزمون آماری خی-دو در سطوح مختلف اطمینان، و همچنین استفاده از خوشه‌بندی k-means در انتخاب داده‌های آموزشی با تنوع طیفی بالا، نسبت به عدم استفاده از آن‌ها بسیار امیدبخش بوده و باعث افزایش چشمگیر صحت کلی و ضریب کاپای طبقه‌بندی تصویر سال ۲۰۱۱ (افزایش ضریب کاپا از ۶۵٪ به ۸۷٪ و افزایش صحت کلی از ۷۳٪ به ۹۱٪) و همچنین سبب بهبود صحت کلی و ضریب کاپای طبقه‌بندی تصویر سال ۲۰۱۵ (افزایش صحت کلی از ۶۹٪ به ۸۶/۳۲٪ و افزایش ضریب کاپا از ۵۹٪ به ۸۰/۴۸٪) شده است. همچنین از آنجا که صرفاً، استفاده از نمونه‌های آموزشی پالایش شده نمی‌تواند منجر به افزایش صحت طبقه‌بندی تصاویر مورد نظر گردد؛ از الگوریتم ژنتیک به عنوان روش جستجو برای انتخاب ویژگی‌های بهینه و در عین حال پارامترهای بهینه طبقه‌بندی SVM استفاده شد. نتایج نشان داد که فرآیند انتخاب ویژگی با استفاده از الگوریتم ژنتیک، سبب بهبود قابل توجه صحت طبقه‌بندی و شناسایی تغییرات شده است. همچنین، معیارهای ارزیابی صحت نشان می‌دهد استفاده از ویژگی‌های بافت مرتبه اول تاثیر بیشتری از ویژگی‌های بافت مرتبه دوم در بهبود صحت طبقه-بندی داشته است.

با توجه به صحت نسبتاً بالای روش پیشنهادی در بهنگام‌رسانی نقشه‌های پوششی و کاربری اراضی، می‌توان از این روش برای بهنگام‌رسانی طیف وسیعی از نقشه‌های موضوعی استفاده کرد. روش پیشنهادی ضمن ارتقاء صحت طبقه‌بندی و تشخیص تغییرات، تسهیل و تسریع فرآیند بهنگام‌رسانی را به همراه دارد. لازم بذکر است؛ روش ارائه شده در این مقاله، مختص شناسایی تغییرات کاربری/پوشش اراضی است که دارای ماهیت تغییرات تدریجی هستند و در مواردی که تغییرات زمانی (فصلی) مد نظر باشد (تغییرات فنولوژیکی و موارد مشابه)، تضمینی بر حصول نتایج دقیق وجود ندارد.

مراجع

[1] A. Singh, "Review article digital change detection techniques using remotely-sensed data," International journal of remote sensing, vol. 10, no. 6, pp. 989-1003, 1989.

[2] M. Khoshlahjeh, B. Ranjbar, A. Moghimi, S. Beheshtifar, Y. Maghsudi, and A. Mohammadzade, "An Overview of the Methods and Models Used to Identify Land use Changes Based on Remote Sensing and GIS (with Emphasis on Studies in Iran)," (in eng), Journal of Geomatics Science and Technology, Tarviji vol. 9, no. 2, pp. 225-242, 2019. [Online]. Available: http://jgst.issge.ir/article-1-909-en.html.

[3] A. Moghimi, H. Ebadi, and V. Sadeghi, "Review of Change Detection Methods from Multitempolar Satellite Images by Pixel-Based and Object-Based Approach," (in eng), Geospatial Engineering Journal, Research vol. 7, no. 2, pp. 99-110, 2016. [Online]. Available: http://gej.issge.ir/article-1-170-en.html.

[4] E. Ghanvati, P, Ziaeian, M.Sardashti, A. Jangi, "Detecting Morphodynamical Changes Using Remote Sensing Data and Analysis Principal Components (PCA) and Fuzzy Logic (LOGIC-FUZZY), Case Study: (Taleghan Watershed)", Volume 40, Number 1 - Sequential Issue 1849. 2009. (Persian).

[5] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," ISPRS Journal of photogrammetry and remote sensing, vol. 80, pp. 91-106, 2013.

[6] P. Deer, "Digital change detection techniques in remote sensing," DEFENCE SCIENCE AND TECHNOLOGY ORGANIZATION CANBERRA (AUSTRALIA), 1995 .

[7] J. R. Jensen, Introductory digital image processing: a remote sensing perspective (no. Ed. 2). Prentice-Hall Inc., 1996.

[8] D. Lu, P. Mausel, E. Brondizio, and E. Moran, "Change detection techniques," International journal of remote sensing, vol. 25, no. 12, pp. 2365-2401, 2004.

[9] V. Sadeghi, F. F. Ahmadi, and H. Ebadi, "Design and implementation of an expert system for updating thematic maps using satellite imagery (case study: changes of Lake Urmia)," Arabian Journal of Geosciences, vol. 9, no. 4, p. 257, 2016.

[10] F.Saeidzadeh, M. Sahebi, H. Ebadi, V. Sadeghi, "Investigating and Analyzing of Object-Based Change Detection method in Urban Management and Planning", First National Conference on Urban Development, Urban Management and Sustainable Development, Iran, 2017. (Persian).

[11] G. Xian, C. Homer, and J. Fry, "Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods," Remote Sensing of Environment, vol. 113, no. 6, pp. 1133-1147, 2009.

[12] X. Chen, J. Chen, Y. Shi, and Y. Yamaguchi, "An automated approach for updating land cover maps based on integrated change detection and classification methods," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 71, pp. 86-95, 2012.

[13] K. J. Wessels et al., "Rapid land cover map updates using change detection and robust random forest classifiers," Remote sensing, vol. 8, no. 11, p. 888, 2016.

[14] S. Teng, Y. Chen, K. Cheng, and H. Lo, "Hypothesis-test-based landcover change detection using multi-temporal satellite images–A comparative study," Advances in Space Research, vol. 41, no. 1 ,۱pp. 1744-1754, 2008.

[15] M. Paul and T. Brandt, "Classification Methods for Remotely Sensed Data," ed: Taylor & Francis, New York, 2001.

[16] G. T. Yengoh, D. Dent, L. Olsson, A. E. Tengberg, and C. J. Tucker III, Use of the Normalized Difference Vegetation Index (NDVI) to Assess Land Degradation at Multiple Scales: Current Status, Future Trends, and Practical Considerations. Springer, 2015.

[17] M. Hall-Beyer, "GLCM texture: A tutorial," National Council on Geographic Information and Analysis Remote Sensing Core Curriculum, vol. 3, 2000.

[18] V. Vapnik and S. Mukherjee, "Support vector method for multivariate density estimation," in Advances in neural information processing systems, 2000, pp. 659-665 .

[19] C. Lardeux et al., "Support vector machine for multifrequency SAR polarimetric data classification," IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 12, pp. 4143-4152, 2009.

[20] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," IEEE transactions on Neural Networks, vol. 13, no. 2, pp. 415-425, 2002.

[21] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," IEEE transactions on pattern analysis and machine intelligence, vol. 19, no. 2, pp. 153-158 ,1997.

[22] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection," IEEE Transactions on Geoscience and Remote sensing, vol. 36, no. 5, pp. 1566-1577, 1998

[23] S. Hajahmadi, M. Mokhtarzedeh, A. Mohammadzadeh, M. J. Valadanzoej. Mapping Changes in Urban Areas Using Satellite Imagery with an Emphasis on Maximum Use of Old Maps. GEJ. 2015; 6 (1) :31-40

[24] Tamimi E, Ebadi H, Kiani A. Developing a New Method in Object Based Classification to Updating Large Scale Maps with Emphasis on Building Feature. JGST. 2019; 8 (4) :203-220.