

جنگل و فرآورده‌های چوب، مجله منابع طبیعی ایران  
دوره ۶۹، شماره ۴، زمستان ۱۳۹۵

تاریخ دریافت: ۱۳۹۳/۱۲/۲۲

تاریخ پذیرش: ۱۳۹۴/۰۲/۲۲

ص ۶۶۷-۶۶۷

## طبقه‌بندی تراکم توده‌های جنگلی با استفاده از تصویر

### ماهواره IRS و الگوریتم ناپارامتریک kNN

- ❖ **رؤیا عابدی\***؛ دکتری جنگلداری، دانشکده منابع طبیعی، دانشگاه گیلان، صومعه‌سرا
- ❖ **امیر اسلام بنیاد**؛ استاد، دانشکده منابع طبیعی، دانشگاه گیلان، صومعه‌سرا
- ❖ **اسدالله شاه بهرامی**؛ دانشیار، دانشکده فنی، دانشگاه گیلان، رشت

#### چکیده

برآوردهای کمی و دقیق از مشخصه‌های توده‌های جنگلی لازمه مدیریت صحیح آنها است. تصاویر سنجش از دور با توجه به اطلاعات مکانی دقیق و وسیع، همواره ابزاری مقرون به صرفه در مدیریت جنگل است و یکی از متداول‌ترین کاربردهای تصاویر در علم جنگلداری، طبقه‌بندی مشخصه‌های توده‌های جنگلی و تهیه نقشه‌های موضوعی آنها است. هدف این پژوهش بهینه‌سازی طبقه‌بندی تراکم (تعداد درختان در هکتار) در توده‌های جنگلی با استفاده از روش ناپارامتریک kNN بود. بررسی در منطقه پیلمبرا، در حوضه آبخیز شفارود در غرب استان گیلان انجام گرفت. نتایج نشان داد که روش kNN با تعداد بهینه نزدیک‌ترین همسایه‌ها در نقطه  $k=6$  و معیار فاصله بهینه اقلیدسی، نتایج مناسبی برای تهیه نقشه پهنه‌بندی مشخصه تراکم (تعداد درختان در هکتار) در جنگل‌های مورد بررسی دارد. برآورد مشخصه تراکم در این روش، با مقادیر  $RMSE=228/58$ ،  $relative\ RMSE=78/9$  درصد و مقدار همبستگی  $r=0/50$  محاسبه شد. مقدار صحت کل و ضریب کاپای نقشه حاصل از طبقه‌بندی به ترتیب برابر ۸۵/۱۹ درصد و ۰/۵۶ به دست آمد. بنابراین الگوریتم ناپارامتریک kNN می‌تواند روش به نسبت مناسبی برای برآورد مشخصه تراکم جنگل روی تصویر ماهواره IRS سنجنده P6-LIS III است.

واژگان کلیدی: تراکم جنگل، شفارود، تصویر ماهواره ای، kNN.

## مقدمه

استفاده از روش‌های ناپارامتریک اطلاعات ارزنده‌ای را در زمینه کمک به مدیریت منابع جنگلی و برای برنامه‌ریزی‌های بهتر ارائه می‌دهد، چراکه با توجه به ماهیت خود نیاز به پیش‌فرضی مبنی بر نرمال بودن مجموعه داده‌های اولیه ندارد [۷]. اطلاعات حاصل از نمونه‌برداری در جنگل اغلب به دلیل ساختار طبیعی توده‌های جنگلی پراکنش نرمال ندارند، از این رو استفاده از الگوریتم‌های ناپارامتریک بسیار کارآمد خواهد بود. یکی از این الگوریتم‌ها، kNN (k Nearest Neighbor) یا k نزدیک‌ترین همسایه) است. الگوریتم kNN توسط کیلیکی و پایوینن (۱۹۸۷) در اندازه‌گیری جنگل به عنوان روشی برای طبقه‌بندی براساس نزدیک‌ترین نمونه‌ها در فضای طیفی معرفی شد و تاکنون به صورت گسترده در علوم جنگل در کشورهای شمال اروپا و آمریکا به منظور برآورد و تهیه نقشه متغیرهای مختلف بر روی تصاویر ماهواره‌ای گوناگون استفاده شده است [۸]. فرانکو-لوپز و همکاران (۲۰۰۱) در برآورد حجم، تراکم و تیپ توده‌های جنگلی در منطقه جنگلی لویس در مینه‌سوتا و بر روی تصاویر لندست TM اقدام کردند و  $k=1$  و معیار فاصله اقلیدسی را گزارش و بیان کردند که روش kNN به دلیل سادگی و سهولت استفاده، ابزاری مناسب برای تهیه نقشه‌های متغیرهای جنگل است [۹]. لورنزو و همکاران (۲۰۰۵) سطح مقطع توده‌های جنگلی را با استفاده از داده‌های تصویر ETM+ لندست ۷ به روش kNN برآورد و نقشه سطح مقطع را با مقدار بهینه  $k=6$  و معیار ماهالانویس تهیه کردند [۱۰]. ریس و همکاران (۲۰۰۳) در برآورد مشخصه‌های حجم و سن در توده‌های جنگلی و بر روی تصاویر لندست TM و SPOT HRV در سطح کل کشور سوئد از روش kNN استفاده کردند و مقدار خطای این روش را برای

جنگل‌ها تقریباً ۳۱ درصد از کل سطح زمین، ۶۶ درصد از کل کربن جهان و ۹۰ درصد گونه‌های خشکی را تشکیل می‌دهند [۱]. برآوردهای کمی و دقیق از مشخصه‌های توده‌های جنگلی لازمه بر مدیریت صحیح جنگل است [۲]. در گذشته معمول‌ترین روش دستیابی به اطلاعات، نمونه‌برداری به وسیله قطعه نمونه‌های زمینی و تنها با حضور مستقیم در عرصه جنگل امکان‌پذیر بود. اما امروزه روش‌های تلفیق اطلاعات سنجش از دور و اطلاعات نمونه‌برداری زمینی، محاسبات را راحت‌تر کرده است. این ارتباط نیازمند به‌کارگیری الگوریتم‌هایی است تا برآورد متغیرهای مورد نظر را از اعداد طیفی تصاویر و اطلاعات حاصل از نمونه‌برداری توده‌های جنگلی انجام دهد [۳، ۴]. تصاویر سنجش از دور با توجه به اطلاعات مکانی دقیق و گستره وسیع می‌تواند ابزاری مقرون به صرفه در مدیریت جنگل بوده و یکی از عمومی‌ترین کاربردهای تصاویر در علم جنگل، طبقه‌بندی مشخصه‌های توده‌های جنگلی و تهیه نقشه‌های موضوعی آنها بوده است [۵]. ارتباط بین متغیرهای اندازه‌گیری شده از قطعه نمونه‌های زمینی در جنگل و داده‌های سنجش از دور منتج به تولید نقشه‌های موضوعی می‌شود. سنجش از دور می‌تواند به تهیه نقشه‌های به‌روز از توده‌های جنگلی و تغییرات آنها برای تشخیص تفاوت‌ها در مشخصه‌های توده‌های وسیع و مکان این تغییرات به منظور اعمال فعالیت‌های مدیریتی کمک کند. ۵۰ تا ۸۰ درصد تغییرات مشخصه‌های جنگل را می‌توان به وسیله داده‌های سنجش از دور تشریح کرد، چراکه مقدار انعکاس باندها در تصویر ماهواره به مشخصه‌های جنگل وابسته است [۶].

آگاهی از وضعیت تعداد در هکتار درختان در توده‌های جنگلی، یکی از اولویت‌هایی است که برای مدیران جنگل به‌منظور برنامه‌ریزی و ارزیابی منابع جنگلی لازم است [۱۵]. از این‌رو در ایران بررسی‌های مختلفی بر روی این مشخصه انجام گرفته است. پیرباوقار (۲۰۱۱) امکان برآورد تراکم جنگل‌های زاگرس را با استفاده از تحلیل رگرسیونی چندگانه بر روی تصاویر ماهواره IRS-P6 در جنگل‌های شهرستان بانه ارزیابی و بیان کرد که نتایج حاصل قابلیت متوسط داده‌های مورد استفاده را نشان می‌دهد [۱۶]. پیرباوقار و همکاران (۲۰۱۱) متغیر تراکم توده را به‌منظور بررسی قابلیت داده‌های SPOT 5 در جنگل‌های شهرستان بانه با استفاده از روش رگرسیون چندمتغیره بررسی کردند. این محققان ضمن گزارش مقدار خطای ۵۷/۴ درصد، بیان کردند که به‌منظور برآوردهای دقیق‌تر این مشخصه باید راهکارهای بهتری را پی گرفت [۱۷]. محمدی و همکاران (۲۰۰۸) برآورد تعداد درختان در هکتار بر روی تصاویر ماهواره‌ای ETM+ را با روش زمین‌آمار در جنگل‌های بلوط لوه مقایسه و گزارش کردند که سنجش از دور نسبت به زمین‌آمار این مشخصه را با میانگین مجذور خطای مربعات کمتری (۱۷۰/۳) برآورد کرد [۱۵]. محمدی و شتایی (۲۰۰۷) برآورد تراکم درختان را در جنگل لوه و تهیه نقشه این مشخصه را به روش آنالیز رگرسیون چندمتغیره بر روی تصاویر ETM+ بررسی کردند و نقشه‌های حاصل را به‌منظور استفاده در برنامه‌های مدیریتی، مناسب و کارآمد ارزیابی کردند [۶]. با وجود استفاده از این الگوریتم در کشورهای دیگر، تاکنون پژوهشی در ایران، برای برآورد تراکم درخت در توده‌های جنگلی با استفاده از این الگوریتم بر روی تصاویر سنجش از دور در علوم جنگل منتشر نشده

مشخصه‌های مورد بررسی به‌ترتیب ۳۳ و ۲۳ درصد به‌دست آوردند و در نهایت این روش را روشی بهینه برای سطوح وسیع معرفی کردند [۱۱]. ماکلا و پکارینن (۲۰۰۴) برآورد حجم بر روی تصاویر لندست TM را به روش kNN انجام دادند و مقدار  $k=5$  را با استفاده از معیار اقلیدسی به‌عنوان مقادیر بهینه معرفی کردند [۱۲]. هوایان و همکاران (۲۰۰۶) برای برآورد حجم در منطقه‌ای جنگلی در چین از روش kNN بر روی تصاویر لندست TM استفاده کردند و مقدار خطا برای این مشخصه را ۴۴/۲ درصد با استفاده از  $k=7$  و معیار فاصله معکوس مربع اقلیدسی به‌دست آوردند و بیان کردند که این روش برای مناطقی که تنها اطلاعات آماربرداری زمینی دارد، مناسب خواهد بود [۱۳]. لابرکو و همکاران (۲۰۰۶) برآورد و تهیه نقشه زی‌توده در منطقه جنگلی در غرب کانادا بر روی تصاویر لندست TM را با مقدار  $k=5$  و فاصله معکوس مربع اقلیدسی انجام دادند و مقدار خطای  $RMSE = 59$  را گزارش کردند. آنها بیان کردند در حالتی که داده‌های زمینی در اختیار باشد و هدف رسیدن به دقت مناسبی در برآوردها باشد، روش kNN روش مناسبی است [۱۴]. لیمی و تمسگن (۲۰۰۵) روش kNN را در بررسی تراکم و سطح مقطع توده‌های جنگلی در غرب بریتیش کلمبیا در کانادا بر روی عکس‌های هوایی با مقایسه سه معیار فاصله مربع اقلیدسی<sup>۱</sup>، شبیه‌ترین همسایه<sup>۲</sup> و فاصله مطلق وزنی<sup>۳</sup> استفاده کردند. آنها مقدار  $k=3$  و فاصله شبیه‌ترین همسایه را بهترین نتیجه گزارش کرده و نقشه این مشخصه‌ها را با استفاده از این مقادیر بهینه تهیه کردند [۷].

1. Squared Euclidean distance
2. Most similar neighbor distance (MSN)
3. Weighted absolute distance

سفیدپلت (*Populus caspica* Bornm.) است. منطقه تحقیق از نظر پستی و بلندی، مسطح و تقریباً بدون شیب است (شکل ۱).

در بررسی حاضر از تصویر ماهواره IRS سنجنده P6-LISSIII مربوط به مسیر ۶۷ و ردیف ۴۳ در تاریخ ۱۸ جولای ۲۰۰۸ (۲۸ تیر ۱۳۸۷) استفاده شد (جدول ۱).

### روش کار

داده‌های آماربرداری زمینی و داده‌های سنجنش از دور: نمونه برداری زمینی در شبکه بندی به ابعاد ۱۵۰×۲۰۰ متر و در ۴۰۸ قطعه نمونه به ابعاد ۱۰۰۰ متر مربع صورت گرفت. در هر قطعه نمونه نوع گونه‌ها شناسایی و تعداد درختان در هر قطعه نمونه شمارش شد. سپس تراکم درختان در سطح قطعه نمونه و در سطح هکتار (تعداد درختان در هکتار) محاسبه شد. سپس این شبکه آماربرداری بر روی تصویر ماهواره IRS سنجنده P6-LISSIII که با تصحیح هندسی و انطباق بر نقشه توپوگرافی ۱:۲۵۰۰۰ منطقه از نظر همپوشانی با جاده‌ها و رودها تأیید شده بود، قرار گرفت.

است. هدف اصلی از این تحقیق بهینه سازی طبقه بندی تراکم توده های جنگلی بر روی تصویر سنجنده P6-LISS III ماهواره IRS با استفاده از الگوریتم ناپارامتریک kNN است.

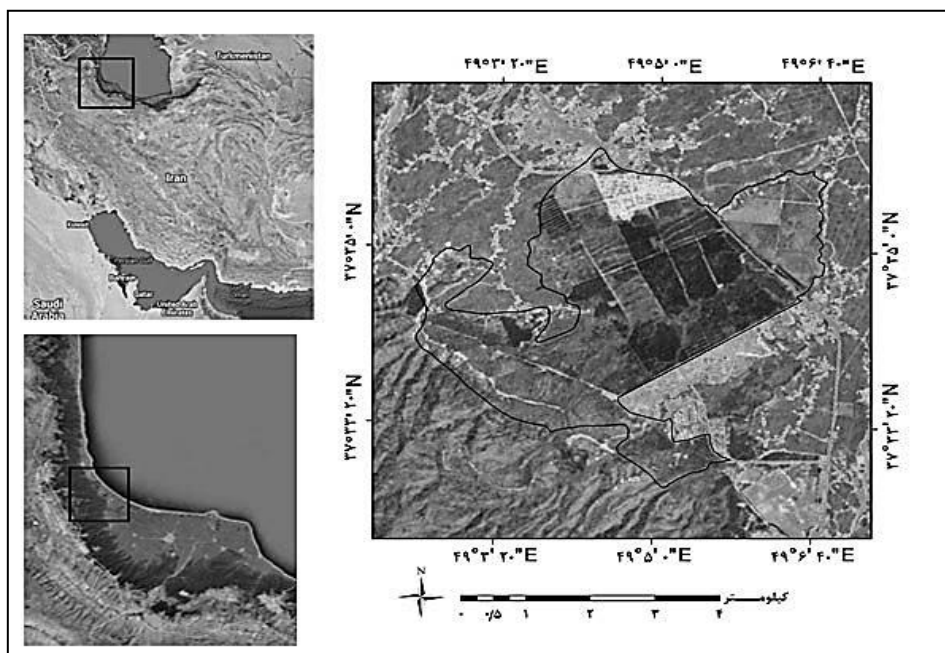
### مواد و روش ها

#### منطقه تحقیق

بررسی حاضر در منطقه پیلمبرا، در حوضه آبخیز شفارود در غرب استان گیلان انجام گرفت. این منطقه در موقعیت جغرافیایی "۳۷° ۳۲' ۴۰" تا "۳۷° ۳۶' ۲۰" عرض شمالی و "۴۹° ۲' ۲۰" و "۴۹° ۷' ۲۰" طول شرقی قرار دارد. مساحت منطقه ۱۸۵۰/۴۴ هکتار و دارای گونه های غالب کاج تدا (*Pinus taeda* L.)، توسکای قشلاقی (*Alnus glutinosa* (L.) Gaertn.)، توسکای بیلاقی (*Alnus subcordata* C. A. Mey.)، صنوبر (*Populus* sp.) و توده های آمیخته طبیعی از دیگر گونه های پهن برگ شامل بلند مازو (*Quercus castaneaefolia* C. A. Mey.)، ممرز (*Carpinus betulus* L.)، پلت (*Acer velutinum* Boiss.)، لرگ (*Pterocarya fraxinifolia* (Lam.) Spach.) و

جدول ۱. مشخصات تصویر ماهواره ای مورد استفاده در بررسی حاضر

ماهواره	ماهواره
سنجنده	IRS
	P6 LISS III
	میکرومتر
	باند ۲- سبز ۵۲۰-۵۹۰
	باند ۳- قرمز ۶۲۰-۶۸۰
باندها و محدوده طیفی (نانومتر)	باند ۴- مادون قرمز نزدیک ۷۷۰-۸۶۰
	باند ۵- مادون قرمز طول موج کوتاه ۱۷۰۰-۱۵۵۰
قدرت تفکیک زمینی (متر)	۲۳/۵
عرض تصویربرداری (کیلومتر)	۱۴۰
قدرت تفکیک رادیومتری (بیت)	۷



شکل ۱. موقعیت منطقه تحقیق

در واقع الگوریتم kNN این دو مجموعه داده را همزمان و به‌طور توأم به‌کار برد، به‌طوری که براساس ساختار برچسب‌دهی در الگوریتم kNN ارتباط بین اطلاعات حاصل از اندازه‌گیری قطعه نمونه‌های زمینی و ارزش‌های طیفی تصویر ماهواره‌ای متناظر با مکان جغرافیایی قطعه نمونه در قالب ماتریس داده برقرار می‌شود، به‌طوری که هر قطعه نمونه زمینی دارای برچسب اطلاعات طیفی متناظر و برعکس هر پیکسل تصویر دارای برچسب اطلاعات اندازه‌گیری زمینی خواهد بود. به این ترتیب یک پایگاه اطلاعاتی مرجع براساس این ارتباط، از منطقه تحقیق تشکیل خواهد شد. با استفاده از این پایگاه اطلاعاتی، پیش‌بینی‌ها و در نهایت برآورد مشخصه‌های مورد نظر برای دیگر پیکسل‌های تصویر که دارای اطلاعات قطعه نمونه‌ها نیستند و مجموعه هدف نامیده می‌شوند، انجام گرفت (جدول ۲).

روش kNN دارای الگوریتمی ساده و قابل فهم

## روش تجزیه و تحلیل

در این بررسی، پایگاه اطلاعاتی مرجع به‌صورت ماتریس داده‌های ورودی که حاوی اطلاعات حاصل از قطعه نمونه‌های زمینی و ارزش‌های طیفی استخراج‌شده از پیکسل‌های متناظر با قطعه نمونه‌ها بود، تدوین شد. الگوریتم kNN نیازمند ماتریس مرجع به‌عنوان داده ورودی است، که دارای اطلاعات داده اندازه‌گیری زمینی و تصاویر ماهواره‌ای است. داده زمینی شامل اندازه‌گیری مشخصه مورد نظر در قطعه نمونه (تعداد درختان در قطعه نمونه و سپس تعداد در هکتار درختان) و مختصات طول و عرض جغرافیایی مرکز قطعه نمونه و داده تصویر ماهواره‌ای نیز شامل ارزش‌های طیفی پیکسل‌ها و موقعیت جغرافیایی این پیکسل‌ها است. این اطلاعات در ابتدا مجزا هستند و هر یک در جدول‌هایی جداگانه تشکیل می‌شوند که باید با هم ادغام شوند و مجموعه‌ای به‌صورت ماتریس مرجع را تشکیل دهند،

که  $y_i$  مقدار مشاهده شده  $i$  امین قطعه نمونه و  $\hat{y}_i$  مقدار برآورد kNN برای  $i$  امین قطعه نمونه و  $n$  تعداد کل قطعه نمونه‌ها است.

$$\text{relative RMSE} = \frac{\text{RMSE}}{\bar{y}} \quad (2)$$

$\bar{y}$  میانگین برآورد kNN برای  $i$  امین قطعه نمونه است.

بهینه‌سازی فرایندی است برای اصلاح ورودی‌ها تا در نهایت، حداقل‌ها و حداکثرهای خروجی، به‌عنوان نتایج مطلوب پیدا شود [۱۹]. پیدا کردن بهینه‌ترین مقدار  $k$  و معیار فاصله، هدف بهینه‌سازی الگوریتم kNN در این پژوهش است. کلیه محاسبه‌های مربوط به بهینه‌سازی تعداد همسایه‌های  $k$  و تعیین معیار فاصله بهینه با استفاده از کدها و دستورالعمل‌های این الگوریتم در پکیج `yaImpute` در محیط نرم‌افزار R توسط نگارندگان برنامه‌نویسی شد. در پایان نقشه حاصل از تجزیه و تحلیل تصاویر با نقشه واقعیت زمینی (اطلاعات آماربرداری زمینی) مقایسه و معیارهای صحت کلی و ضریب کاپا محاسبه شد.

است. تعداد  $k$  از نزدیک‌ترین نمونه‌های تعلیمی در مجموعه مرجع به نقطه هدف (در این بررسی منظور پیکسل‌هایی از تصویر که بدون اطلاعات قطعه نمونه زمینی بوده و در نتیجه دارای طبقه نامعلوم هستند و قصد طبقه‌بندی آنها را داریم)، استفاده می‌شود [۱۸]. بنابراین دو عامل مهم در بهینه‌سازی این الگوریتم وجود دارد: ۱. معیار فاصله بهینه برای پیدا کردن همسایگی و ۲. مقدار  $k$  برای پیدا کردن تعداد همسایه‌ها. برای تعیین معیارهای فاصله در این الگوریتم روش‌های مختلفی وجود دارد، مانند فاصله اقلیدسی، منهن، ماهالانویس، فازی و غیره. انتخاب مقادیر بهینه برای دو عامل ذکر شده در روش kNN، براساس یک مقدار آستانه‌ای است که دارای کمترین مقدار خطای مربع میانگین (RMSE) یا مقدار نسبی این شاخص (relative RMSE) باشد. برای تعیین مقدار خطای مربع میانگین (RMSE) و مقدار نسبی این شاخص (relative RMSE) از رابطه‌های ۱ و ۲ استفاده شد [۹]:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

جدول ۲. بخشی از ماتریس اطلاعات مرجع که حاوی اطلاعات قطعه نمونه‌های زمینی و ارزش‌های طیفی پیکسل‌های متناظر با قطعه نمونه‌ها است در محیط نرم‌افزار R

Plot	X	Y	Density	Band 2	Band 3	Band 4	Band 5
۱	۳۳۲۸۶۱	۴۱۶۱۱۹۳	۵۰۰	۴۹	۲۵	۱۱۳	۳۳
۲	۳۳۲۹۹۱	۴۱۶۰۹۸۰	۳۰۰	۵۱	۲۵	۱۰۷	۴۰
۳	۳۳۲۶۹۳	۴۱۶۱۰۰۹	۲۶۰	۵۲	۲۷	۱۱۲	۳۹
۴	۳۳۲۸۴۱	۴۱۶۰۹۹۴	۲۹۰	۵۱	۲۵	۱۱۱	۳۹
۵	۳۳۳۱۴۰	۴۱۶۰۹۶۵	۲۳۰	۵۰	۲۵	۱۰۶	۳۹
۶	۳۳۳۲۸۹	۴۱۶۰۹۵۱	۳۷۰	۵۴	۲۸	۱۲۰	۴۴
۷	۳۳۲۷۱۱	۴۱۶۱۲۰۸	۴۹۰	۴۹	۲۵	۱۱۱	۳۱
۸	۳۳۳۱۵۹	۴۱۶۱۱۶۵	۲۹۰	۵۱	۲۶	۱۱۰	۳۹
۹	۳۳۳۰۱۰	۴۱۶۱۱۷۹	۳۷۰	۵۱	۲۶	۱۰۹	۳۹
۱۰	۳۳۲۸۹۳	۴۱۶۱۵۹۲	۶۳۰	۵۱	۲۵	۱۱۴	۴۱
⋮							
⋮							
⋮							

## نتایج و بحث

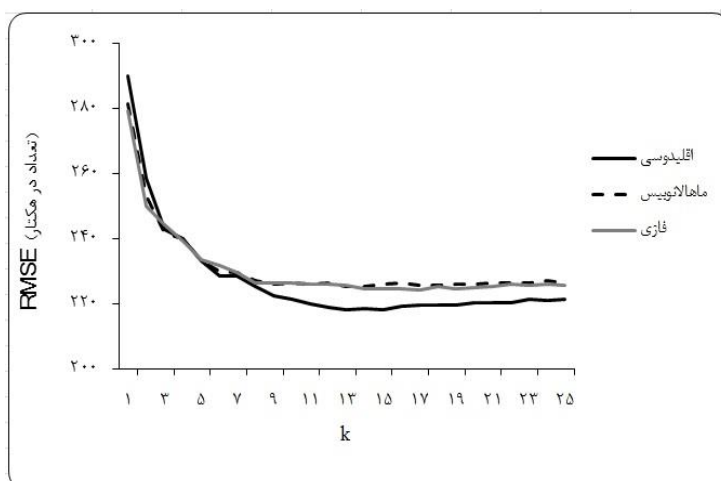
پس از تصحیح هندسی تصویر، با قرار دادن لایه‌برداری جاده و رودخانه بر روی آن، از صحت هندسی آن اطمینان حاصل شد. همخوانی دقیق لایه‌های جاده و آبراهه‌ها روی تصویر تصحیح‌شده بیانگر دقت تصحیح‌های هندسی انجام‌گرفته است. بر پایه بررسی‌های انجام‌گرفته، هیچ‌گونه خطای رادیومتریکی مشاهده نشد.

در این بررسی تعیین معیار فاصله که لازمه اجرای روش ناپارامتریک kNN است، برای سه معیار شامل فاصله اقلیدسی، فاصله ماهالانویس و فازی محاسبه شد که مقدار خطا به ترتیب برابر با  $۲۲۸/۵۸$ ،  $۲۳۰/۱۳$  و  $۲۳۱/۷۹$  به دست آمد. براساس نتایج معیار فاصله اقلیدسی نسبت به فاصله ماهالانویس و فازی در سطح پایین‌تری از خطا قرار دارد. بنابراین در اجرای روش ناپارامتریک kNN برای طبقه‌بندی به منظور تعیین تراکم (تعداد درختان در هکتار) از معیار فاصله اقلیدسی استفاده شد. بنابراین، کمترین مقدار میانگین مربعات خطا با استفاده از تصویر و روش kNN در تحقیق حاضر برابر با  $relative\ RMSE=۷۸/۹$  درصد به دست آمد. محمدی و همکاران (۲۰۰۸) با استفاده از مقایسه زمین‌آمار و داده‌های سنجش از دور در برآورد تعداد در هکتار، میانگین مربعات خطای  $۱۷۰/۳$  را با روش رگرسیونی در سنجش از دور و محمدی و شتایی (۲۰۰۷) میانگین مربعات خطای  $۱۷۰/۱۳$  را برای مشخصه تراکم با استفاده از تکنیک رگرسیون چندمتغیره در توده‌های مورد بررسی خود به دست آوردند [۶، ۱۵]. در بررسی‌هایی که به آنها اشاره شد، به دلیل ماهیت نوع روش‌های به‌کاررفته، هیچ نوع بهینه‌سازی در مقدار

خطا و جست‌وجویی برای یافتن مقدار خطای کمتر انجام نگرفته بود. در بررسی حاضر، پس از محاسبات و ترسیم شکل ۲ و دستیابی به  $k=۱$  بر روی این منحنی به‌عنوان اولین همسایه نزدیک، نتیجه نشان داد که مقدار خطا در این نقطه (با یک همسایه نزدیک) زیاد است. بنابراین روی این منحنی نقطه بهینه برای  $k$  (تعداد نزدیک‌ترین همسایه‌های مناسب) برابر با  $k=۶$  به دست آمد، چراکه پس از این نقطه روند کاهش خطا بسیار ناچیز بود. بنابراین با محاسبه نقطه  $k=۶$ ، هدف بهینه‌سازی در بررسی حاضر به دست آمد.

نتایج این بررسی نشان داد که تعداد بهینه  $k$  به‌عنوان تعداد نزدیک‌ترین همسایه‌ها به منظور طبقه‌بندی مشخصه تراکم (تعداد درختان در هکتار) برابر با ۶ نمونه مرجع بود، زیرا کمترین مقدار خطای RMSE که بعد از آن کاهش مقدار خطا ناچیز است، براساس معیار بهینه فاصله اقلیدسی برابر با  $۲۲۸/۵۸$  و  $relative\ RMSE=۷۸/۹$  درصد در نقطه  $k=۶$  به دست آمد (شکل ۲). مطابق شکل ۲، منحنی خطای محاسبه‌شده براساس معیار فاصله اقلیدسی در سطح پایین‌تری نسبت به دو معیار دیگر (معیار فاصله ماهالانویس و فازی) قرار دارد و مقدار خطا از ابتدا تا نقطه  $k=۶$  بیشترین کاهش را داشته و از آن پس، روند کاهش خطا کم شده است، به طوری که در نقطه  $k=۱۳$  که منحنی به مقدار ثابت رسیده، مقدار خطا نسبت به مقدار بهینه  $k=۶$  فقط  $۳/۵۷$  درصد کاهش داشته است.

نتایج نشان داد که درصد خطای  $relative\ RMSE$  در  $k=۱$  برابر با  $۱۰۰/۱$  درصد بود، درحالی که این مقدار در  $k=۶$  به  $۷۸/۹۶$  درصد رسید، یعنی در حدود  $۲۱/۱۸$  درصد کاهش در مقدار خطا، که بسیار شایان توجه است. از سوی دیگر، مقدار این خطا برای نقطه



شکل ۲. مقدار RMSE بر حسب معیارهای بهینه فاصله و تعداد k

دانست؛ چراکه یکی از مشکلات روش kNN زمان زیاد پردازش اطلاعات و جست‌وجو است، به‌خصوص زمانی که مجموعه مرجع دارای اطلاعات زیادی باشد، بنابراین یک معیار فاصله پیچیده‌تر که به چندین مرحله آزمون نیاز داشته باشد، کارایی این روش را کم خواهد کرد [۱۹، ۲۲]. فرانکو لویز و همکاران (۲۰۰۱)، سوزا و همکاران (۲۰۱۳) و رحمان (۲۰۰۶) نیز در بررسی‌های خود معیار اقلیدسی را به‌عنوان معیار بهینه فاصله برای برآورد مشخصه‌های مختلف جنگل به‌دست آوردند [۹، ۱۸، ۲۳].

نتایج بررسی لیمی و تمسگان (۲۰۰۵) در محاسبه تعداد درختان در هکتار با استفاده از الگوریتم ناپارامتریک kNN نشان داد در حالتی که از مقدار بهینه  $k=3$  استفاده شود، مقدار RMSE کمتر می‌شود و نتایج مطلوب‌تری نسبت به حالتی که  $k=1$  باشد، به‌دست خواهد آمد [۷]. فرانکو لویز و همکاران (۲۰۰۱) در برآورد تراکم نشان دادند که اگر تعداد همسایه‌ها از ۱ به ۵ افزایش یابد، مقدار RMSE حدود ۱۴ درصد کاهش خواهد یافت، اما بعد از این تعداد، مقدار کاهش خطا بسیار ناچیز خواهد بود [۹]. در این بررسی مقدار

بعدی یعنی  $k=7$  برابر با  $78/98$  درصد است، یعنی فقط  $0/16$  درصد کاهش در مقدار خطا اتفاق افتاد که معنی‌دار نیست. همچنین مقدار این خطا در  $k=25$  (به‌عنوان بیشترین مقدار k در این بررسی) به  $76/51$  درصد رسید، یعنی فقط  $2/45$  درصد نسبت به نقطه  $k=6$  کاهش در مقدار خطا اتفاق افتاد.

براساس شکل ۳ مقدار همبستگی معیار فاصله اقلیدسی نیز در سطح بالاتری نسبت به فاصله ماهالانویس و فازی قرار داشت ( $r=0/50$ ). بنابراین، این نتیجه نیز تأیید می‌کند که  $k=6$  بهینه‌ترین مقدار در روش ناپارامتریک kNN برای مشخصه تراکم جنگل در منطقه تحقیق است.

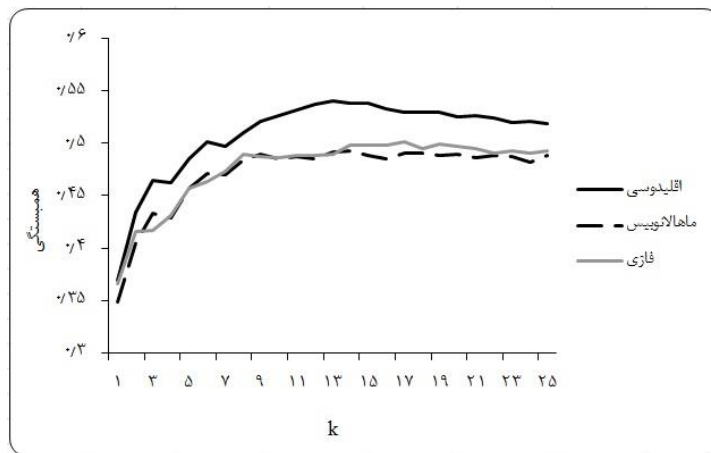
معیار فاصله اقلیدسی در میان دیگر معیارهای فاصله عمومیت بیشتری دارد و به‌عنوان معیار بهینه جایگاه مناسبی در تحقیقات مختلف کسب کرده است [۱۸، ۱۹، ۲۰، ۲۱]. در پژوهش حاضر نیز سطح خطای برآورد با استفاده از این معیار فاصله، کمترین بود که می‌توان براساس نظر قائمی و فیضی درخشی (۲۰۱۴) و شوهر (۲۰۱۰) دلیل این آن را ساده بودن محاسبه این معیار و همچنین هماهنگی بیشتر آن با مشخصه مورد بررسی



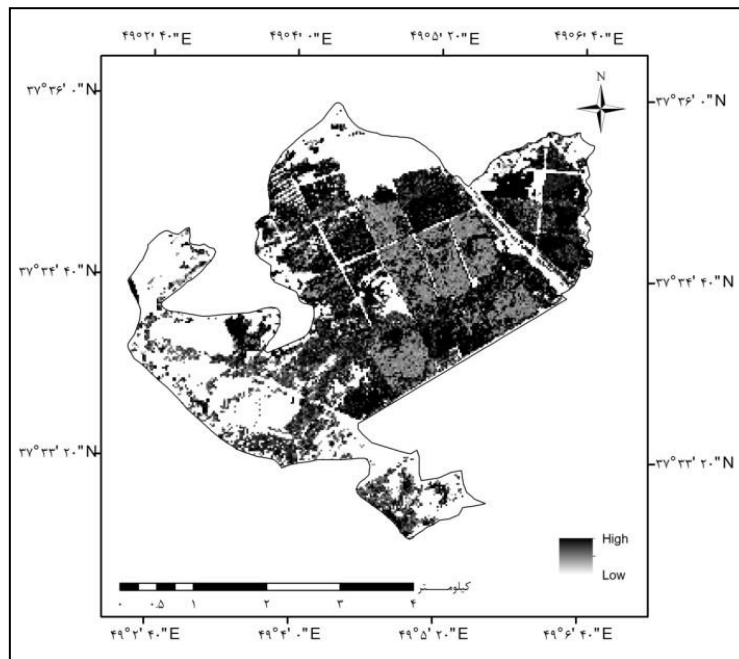
با برنامه‌نویسی در محیط نرم‌افزار R تصویر ماهواره ای طبقه‌بندی و نقشه تعداد در هکتار تهیه شد (شکل ۴). صحت کلی و ضریب کاپای نقشه طبقه‌بندی تراکم در مقایسه با قطعات نمونه زمینی به ترتیب ۸۵/۱۹ درصد و ۰/۵۶۱ به دست آمد.

خطا از  $k=1$  به  $k=6$  در حدود ۲۱/۱۸ درصد بود و بعد از این مقدار، تنها ۲/۴۵ درصد کاهش نشان داده شد. براساس این نتایج، جست‌وجو برای یافتن مقدار بهینه  $k$  به‌عنوان ابزاری مناسب برای دستیابی به صحت‌های بیشتر و خطای کمتر ضروری است.

پس از تدوین پایگاه اطلاعات مرجع، تعیین  $k$  مناسب ( $k=6$ ) و معیار بهینه فاصله (فاصله اقلیدسی)



شکل ۳. رابطه همبستگی معیارهای بهینه فاصله و تعداد  $k$



شکل ۴. نقشه تراکم (تعداد درختان در هکتار) به روش kNN

## نتیجه گیری

هدف این پژوهش بهینه‌سازی طبقه‌بندی مشخصه تراکم (تعداد درختان در هکتار) در تصاویر ماهواره ای در توده‌های جنگلی با استفاده از روش ناپارامتریک kNN بود. هرچند که صحت کلی و ضریب کاپای حاصل در این تحقیق (به ترتیب ۸۵/۱۹ درصد و ۰/۵۶۱) مقادیر قابل توجهی در این زمینه هستند، ولی مقدار ضریب کاپا به طور محسوسی از صحت کلی کمتر است. لذا

نمی‌توان با اطمینان از قابلیت زیاد این رویکرد در جنگلهای هیرکانی صحبت کرد. از آنجا که این روش در کشور ما جدید است و هنوز جایگاه خود را در بررسی‌های علوم جنگل باز نکرده است، پیشنهاد می‌شود بررسی‌های بیشتری در زمینه قابلیت‌های دیگر داده‌های سنجش از دور و برآورد و طبقه‌بندی دیگر مشخصه‌های جنگل با روش kNN انجام گیرد.

## References

- [1]. Torabzadeh, H., Morsdorf, F., and Schaepman, M.E. (2014). Fusion of imaging spectroscopy and airborne laser scanning data for characterization of forest ecosystems –A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 97: 25–35.
- [2]. Hernandez, C., Nunes, L., Lopes, D., and Graña, M. (2014). Data fusion for high spatial resolution LAI estimation. *Information Fusion*, 16: 59–67.
- [3]. Lemay, V., Maedel, J., and Coops, N.C. (2008). Estimating stand structural details using nearest neighbor analyses to link ground data, forest cover maps, and Landsat imagery. *Remote sensing of environment*, 112: 2578-2591.
- [4]. Nguyen, T.T.H. (2010). Applying different methods for prediction of stand volume using SPOT 5 data. In: proceeding of The 31th Asian conference remote sensing ACRS, TS02-4. Nov. 1-5 Ha Noi. Vietnam.
- [5]. Samaniego, L., Bardossy, A., and Schulz, K. (2007). Supervised Classification of Remotely Sensed Imagery using a Modified k-NN Technique. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7): 2112 – 2125.
- [6]. Mohammadi, J., and Shataee, Sh. (2007). Forest stand density mapping using Landsat ETM+ data, Loveh forests, North of Iran. In: proceedings of the 28th Asian association on remote sensing ACRS, Nov. 12-16, Kuala Lumpur, Malaysia. 7 pp.
- [7]. Lemay, V., and Temesgen, H. (2005). Comparison of Nearest Neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables. *Forest Science*, 51: 109-119.
- [8]. Finley, A.O., Ek, A. R., Bai, Y., and Bauer, M.E. (2003). K-Nearest Neighbor Estimation of Forest Attributes: Improving Mapping Efficiency. 5th Annual Forest Inventory and Analysis Symposium, New Orleans, LA. 62-68.
- [9]. Franco-Lopez, H., Ek, A.R., and Bauer, M.E. (2001). Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sensing of Environment*, 77: 251– 274.
- [10]. Lorenzo, B., Chirici, Gh., Corona, P., Marchetti, M., Maselli, F., and Ricotta, C. (2005). Stand basal area extension over forestland by different classification algorithms applied to Landsat ETM+ imagery. In: proceedings of Forest SAT Symposium Heriot Watt University, August 5-9, Edinburgh.



- [11]. Reese, H., Nilsson, M., Pahlén, T.G., Hagner, O., Joyce, S., Tingelöf, U., Egberth, M., and Olsson H. (2003). Countrywide Estimates of Forest Variables Using Satellite Data and Field Data from the National Forest Inventory. *Ambio* 32(8): 542-548.
- [12]. Makela, H., and Pekkarinen, A. (2004). Estimation of forest stand volumes by Landsat TM imagery and stand-level field-inventory data. *Forest Ecology and Management*, 196: 245-255.
- [13]. Huiyan, G., Limin, D., Gang, W., Dong, X., Shunzhong, W., and Hui, W. (2006). Estimation of forest volume by integrating Landsat TM imagery and forest inventory data. *Science in China: series e technological science*, 49: 54-62.
- [14]. Labrecque, S., Fournier, R.A., Luther, J.E., and Piercey, D. (2006). A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. *Forest Ecology and Management*, 226: 129-144.
- [15]. Mohammadi, J., Shataee, Sh., Habashi, H., and Yaghmaee, F. (2008). Comparison of Remote Sensing and Geostatistics Techniques in forest tree density estimation, Case Study Loveh Forests, Gorgan. *Journal of Agriculture Sciences and Natural Resources*, 15(1): 10-21.
- [16]. Pir Bavaghar, M. (2011). Evaluation of capability of IRS-P6 satellite data for predicting quantitative attributes of forests (case study: Northern Zagros forests). *Iranian Journal of Forest*, 3(4): 277-289.
- [17]. Pir Bavaghar, M., Ghahramani, L., and Fatehi P. (2011). Evaluation of the capability of SPOT5-HRG data for predicting tree density in the northern Zagros forests. *Iranian Journal of Forest and Poplar Research*, 19(2): 242-253.
- [18]. Souza, R., Rittner, L., and Lotufo, R. (2013). A comparison between k-optimum path forest and k-nearest neighbors supervised classifiers, *Pattern recognition letters*, 39: 2-10.
- [19]. Ghaemi, M., and Feizi-Derakhshi, M.R. (2014). Forest Optimization Algorithm. *Expert Systems with Applications*, 41(15): 6676-6687.
- [20]. Fassnacht, F.E., Hartig, F., Latifi, H., Berger, C., Hernandez, J., Corvalan, P., and Koch, B. (2014) Importance of sample size, data type and prediction method for remote sensing-based estimation of aboveground forest biomass. *Remote sensing of environment*, 154: 102-114.
- [21]. Latifi, H., Fassnacht, F.E., Hartig, F., Berger, C., Hernandez, J., Corvalan, P., and Koch, B. (2015) Stratified aboveground forest biomass estimation by remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 38: 229-241.
- [22]. Scheuber, M. (2010) Potentials and limits of the k-nearest-neighbor method for regionalizing sample-based data in forestry. *European Journal of Forest Research*, 129: 825-832.
- [23]. Rahman M.M. (2006). Tropical forest biomass estimation and mapping using k-nearest neighbour (KNN) method, In: S. Nayak, S. K. Pathan and J. K.Garg (Eds.): proceedings of Symposium of ISPRS Commission IV, Sept. 25.-30. Goa, India: 860-865.