

مدلسازی تعمیم و ساده‌سازی شبکه رودخانه‌ها با استفاده از شبکه‌های عصبی مبتنی بر گراف

محدثه حاجی ملا احمدی^۱، محمد کریمی^{۲*}، پرستو پيله فروش‌ها^۳

^۱ کارشناس ارشد سیستم اطلاعات مکانی - دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین

طوسی

mohadeseahmadi037@gmail.com

^۲ دانشیار دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی

mkarimi@kntu.ac.ir

^۳ استادیار دانشکده مهندسی نقشه‌برداری - دانشگاه صنعتی خواجه نصیرالدین طوسی

pilehforoosh@kntu.ac.ir

(دریافت: مهر ۱۴۰۴، تصویب: خرداد ۱۴۰۵)

چکیده

فرایند تعمیم نقشه، یک روش پیچیده به منظور تولید نقشه‌های کوچک مقیاس از نقشه‌های بزرگ مقیاس است که با حفظ ساختارها و ویژگی‌های عوارض، منجر به تبدیل عوارض نقشه به صورت خوانا در مقیاس هدف می‌شود. شبکه‌های رودخانه‌ای به دلیل ساختار شاخه‌ای، سلسله‌مراتبی و وابسته به توپولوژی، از چالش‌برانگیزترین عوارض در فرایند تعمیم به شمار می‌روند. اگرچه روش‌های داده‌محور و یادگیری عمیق در سال‌های اخیر برای تعمیم شبکه‌های رودخانه‌ای مورد استفاده قرار گرفته‌اند، بسیاری از آن‌ها ساختار سلسله‌مراتبی شبکه را به‌صورت صریح مدل نکرده یا بدون در نظر گرفتن قیود کارتوگرافیک، منجر به حذف یا حفظ نامناسب شاخه‌ها می‌شوند. در این پژوهش، یک چارچوب دومارحله‌ای برای تعمیم و ساده‌سازی شبکه رودخانه‌ها ارائه شده است. در گام نخست، ساختار سلسله‌مراتبی شبکه با ترکیب رتبه‌بندی هورتون-استرالر و تحلیل پیوستگی هندسی Stroke استخراج می‌شود. در گام دوم، فرایند انتخاب و حذف شاخه‌ها با استفاده از یک مدل یادگیری عمیق مبتنی بر گراف با معماری GraphSAGE و در چارچوب دستورالعمل‌های سازمان نقشه‌برداری انجام می‌گیرد. مدل پیشنهادی بر روی داده‌های شبکه رودخانه استان اصفهان در مقیاس ۱:۲۵,۰۰۰ اجرا و نتایج آن با نقشه مرجع ۱:۵۰,۰۰۰ مقایسه شد. برای ارزیابی دقت، از ضریب انطباق هندسی (CLC) استفاده گردید که میزان همپوشانی طولی شبکه تعمیم‌یافته با نقشه مرجع را نشان می‌دهد. نتایج نشان داد که مقدار CLC برابر با ۰/۹۶ است که بیانگر حفظ مناسب ساختار کلی شبکه و تطابق بالای خروجی مدل با تعمیم دستی کارتوگراف در مقیاس هدف می‌باشد.

واژه‌های کلیدی: تعمیم شبکه رودخانه، ساختار دهی شبکه رودخانه، فرایند حذف شاخه، یادگیری عمیق مبتنی بر گراف، پایگاه داده توپوگرافی.

* نویسنده رابط

۱- مقدمه

نقشه‌ها به‌عنوان ابزار اصلی برای نمایش و تحلیل اطلاعات مکانی شناخته می‌شوند و بدون وجود آن‌ها، انجام تحلیل‌های دقیق تقریباً غیرممکن خواهد بود [۱]. با گسترش استفاده از داده‌های مکانی در علوم مختلف و فناوری اطلاعات جغرافیایی، اهمیت نقشه‌ها فراتر از نمایش مکان‌ها رفته و در مقیاس‌های مختلف نیز برجسته شده است [۲، ۳]. بدین منظور، برای تولید و به‌روزرسانی نقشه‌ها در مقیاس‌های مختلف، استفاده از تعمیم نقشه ضروری است. این فرآیند که برای ایجاد نقشه‌های کوچک‌مقیاس از نقشه‌های بزرگ‌مقیاس به کار می‌رود، با ساده‌سازی نقشه و حذف جزئیات غیرضروری، ضمن کاهش زمان مورد نیاز، به حفظ اطلاعات اساسی و دقت تحلیل‌های جغرافیایی نیز کمک می‌کند [۱، ۴، ۵]. هدف از تعمیم تنها ساده‌سازی داده‌ها نیست، بلکه باید معنای مکانی پدیده‌ها نیز حفظ گردد [۶، ۲]. در این فرآیند، توجه ویژه‌ای به عوارض خطی، از جمله شبکه‌های رودخانه‌ای می‌شود [۵، ۷].

شبکه‌های رودخانه‌ای به عنوان یکی از عناصر اصلی نقشه‌های توپوگرافی، نقش مهمی در تحلیل‌های جغرافیایی ایفا می‌کنند [۸]. این شبکه‌ها معمولاً دارای جزئیات فراوان و ساختاری پیچیده هستند، که نمایش آن‌ها در مقیاس‌های کوچک چالشی اساسی به شمار می‌رود [۹]. در چنین شرایطی، تعمیم شبکه‌های رودخانه اهمیت ویژه‌ای پیدا می‌کند، زیرا امکان ساده‌سازی داده‌های پیچیده و درک بهتر روابط هیدرولوژیکی را فراهم می‌سازد.

در سال‌های اخیر، توجه پژوهشگران به ویژگی‌های مکانی شبکه‌های رودخانه‌ای در فرآیند تعمیم و ساده‌سازی افزایش یافته است [۱۰، ۱۱]. یکی از اهداف اصلی در تعمیم نقشه‌ها، کاهش تعداد عوارض در مقیاس کوچکتر است، در حالی که اطلاعات مهم و معنادار حفظ شود [۱۲]. در تعمیم شبکه‌های رودخانه‌ای، دو سوال اساسی مطرح است: اول، چند شاخه باید حذف یا انتخاب شوند و دوم، کدام شاخه‌ها در مقیاس هدف بی‌اهمیت هستند [۱۰، ۱۱]. برای پاسخ به این سوال، فرایند تعمیم شبکه رودخانه‌ای عمدتاً شامل دو مرحله انتخاب و ساده‌سازی است که در میان آن‌ها، مرحله انتخاب نقش کلیدی در حفظ بخش‌های مهم رودخانه و حذف بخش‌های غیرضروری دارد [۸].

روش‌های انتخاب در تعمیم و ساده‌سازی شبکه رودخانه به دو رویکرد اصلی تقسیم بندی می‌شوند: رویکردهای قاعده‌محور و داده‌محور [۲، ۱۳]. در روش‌های قاعده‌محور، ویژگی‌های شبکه رودخانه استخراج و بر اساس معیارهایی چون طول، اهمیت هیدرولوژیکی یا موقعیت جغرافیایی، تصمیم‌گیری برای حذف شاخه‌ها انجام می‌شود. این روش‌ها نیازمند تنظیمات دستی و توجه به ویژگی‌های ساده هستند. اما در روش‌های داده‌محور، به‌ویژه با استفاده از یادگیری عمیق، ویژگی‌ها به‌طور خودکار استخراج شده و مدل‌هایی برای پیش‌بینی شاخه‌های قابل حفظ یا حذف توسعه می‌یابند. این رویکرد به دلیل توانایی در پردازش داده‌های پیچیده و غیرخطی، تعمیم مؤثری در مقیاس‌های مختلف ارائه می‌دهد [۵].

تحقیقات زیادی به مدلسازی فرایند تعمیم شبکه رودخانه‌ها پرداخته‌اند. مطالعات موجود را می‌توان در دو بخش ساختار دهی شبکه رودخانه‌ها و مدلسازی فرایند عملگر هندسی طبقه بندی کرد. علاوه بر این، مطالعه توپولوژی شبکه‌های رودخانه‌ای یکی از مباحث مهم در ژئومورفولوژی است که به بررسی ساختار و ارتباطات داخلی این شبکه‌ها می‌پردازد [۱۴، ۱۵]. یکی از نخستین تلاش‌ها در این زمینه توسط هورتون صورت گرفت که سامانه کلاسیک رده‌بندی آبراهه‌ها را بر پایه جهت جریان آب ارائه کرد. این روش با تعیین مرتبه زیرساختارهای شبکه رودخانه‌ای، امکان تحلیل کمی ساختارهای شاخه‌ای را فراهم می‌کند [۱۶]. استرالر نیز با توسعه نظریات هورتون، سیستم کدگذاری جدیدی برای شناسایی ساختار درختی شبکه‌های رودخانه‌ای معرفی کرد که ویژگی‌های مورفولوژیکی و اندازه جریان‌ها را منعکس می‌سازد [۱۶]. Thomson برای سازماندهی شبکه‌های رودخانه‌ای بر اساس اصول گشتالت، مفهوم strokeها را مطرح کرده و جریان اصلی را با استفاده از ترتیب هورتون و طول strokeها شناسایی نمود. با این حال، استفاده از طولانی‌ترین مسیر برای تعیین جریان اصلی باعث خطاهای بزرگ می‌شود [۸]. Li (۲۰۱۸) روشی نوین برای ساختاردهی شبکه‌های رودخانه‌ای ارائه داده‌است که در آن از ترکیب ویژگی‌های هندسی و توپولوژیکی استفاده می‌شود. در این روش، ابتدا اتصالات stroke هر بخش رودخانه با توجه به ویژگی‌هایی مانند طول و زاویه ساخته شده و سپس ساختار سلسله‌مراتبی شبکه رودخانه شناسایی می‌شود. این رویکرد

باعث حفظ ساختار توپولوژیکی و بهره‌برداری از ویژگی‌های هندسی می‌شود، در حالی که بسیاری از روش‌های پیشین تنها به یکی از این جنبه‌ها توجه داشتند [۱۶].

در ادامه، مطالعات مختلف در زمینه تعمیم شبکه‌های رودخانه‌ای و به‌ویژه در حوزه انتخاب و حذف شاخه‌ها مورد بررسی قرار گرفته است. یکی از اهداف اصلی تعمیم شبکه‌های رودخانه‌ای، کاهش پیچیدگی شبکه در مقیاس‌های کوچکتر بدون از دست دادن ویژگی‌های اصلی و معنی‌دار است. مشکل اصلی در تعمیم و ساده سازی شبکه رودخانه‌ها، شناسایی شاخه‌های بی‌اهمیت و حذف آن‌هاست [۱۷]. بررسی‌ها نشان می‌دهد که روش‌های موجود در این زمینه را می‌توان به دو دسته اصلی روش‌های دانش‌مبنا و روش‌های داده‌مبنا طبقه بندی کرد. در روش‌های دانش‌مبنا تصمیم‌گیری‌ها بیشتر بر اساس دانش ساختاری، اصول و قوانین موجود در تحلیل‌ها انجام می‌شود [۲]. از جمله مهم‌ترین روش‌های پیشنهادی در این حوزه می‌توان به استفاده از مدل‌های تحلیل ریاضی فازی اشاره کرد که به‌طور ویژه به توزیع مکانی رودخانه‌ها توجه دارند. در این روش‌ها، شاخص‌هایی مانند طول رودخانه، تراکم، اهمیت نسبی و نوع ساختار شبکه رودخانه‌ای به‌عنوان معیارهای ارزیابی در نظر گرفته می‌شوند [۱۸]. روش‌های دیگری نیز مانند نظریه گراف برای انتخاب و حذف شاخه‌های رودخانه‌ای استفاده شده‌اند [۱۹]. در این روش‌ها، شاخه‌ها به‌عنوان گره‌هایی در یک درخت مدل‌سازی شده و بر اساس طول رودخانه، زاویه بین شاخه‌ها و درجه گره‌ها تعیین می‌شوند. از این طریق، شاخه‌های کم‌اهمیت شناسایی و حذف می‌شوند [۲۰]. یکی از مشکلات این روش‌ها عدم توانایی در تحلیل ویژگی‌های معنایی و ساختاری شبکه‌ها است، به‌ویژه در مواقعی که ویژگی‌های هندسی و توپولوژیکی در تضاد با هم قرار می‌گیرند [۲۰].

منظور از تضاد میان ویژگی‌های هندسی و توپولوژیکی، شرایطی است که اهمیت یک شاخه رودخانه بر اساس معیارهای هندسی با جایگاه آن در ساختار شبکه رودخانه هم‌خوانی ندارد. به‌عنوان مثال، ممکن است یک شاخه از نظر هندسی کوتاه و نسبتاً مستقیم باشد و بر اساس معیارهایی مانند طول یا میزان خمیدگی، کم‌اهمیت تلقی شود؛ در حالی که همان شاخه از نظر توپولوژیکی نقش مهمی در اتصال زیرشبکه‌ها، حفظ پیوستگی جریان یا تکمیل ساختار سلسله‌مراتبی شبکه ایفا کند. در مقابل، شاخه‌هایی وجود دارند که از نظر هندسی برجسته

و طولانی هستند، اما از نظر توپولوژیکی در مرتبه‌های پایین شبکه قرار داشته و حذف آن‌ها تأثیر قابل توجهی بر انسجام کلی شبکه ندارد. این ناهماهنگی میان معیارهای هندسی و توپولوژیکی موجب می‌شود روش‌های مبتنی بر قوانین ساده یا آستانه‌های تجربی، در تصمیم‌گیری‌های انتخاب یا حذف شاخه‌ها، به‌ویژه در شبکه‌های متراکم یا دارای الگوی زهکشی پیچیده، دچار خطا شوند.

از سوی دیگر، در مقایسه با رویکردهای دانش‌مبنا، رویکردهای داده‌مبنا از داده‌های گسترده‌تر و الگوریتم‌های یادگیری ماشین برای انتخاب و حذف شاخه‌ها استفاده می‌کنند. در این راستا، الگوریتم‌هایی مانند ماشین بردار پشتیبان، شبکه عصبی پس‌انتشار و درخت تصمیم مورد استفاده قرار گرفته‌اند [۸، ۲۱]. به عنوان مثال، Sen و Gokgoz (۲۰۱۵)، از الگوریتم یادگیری ماشین بردار پشتیبان برای تعمیم شبکه‌های رودخانه‌ای استفاده کردند. هدف این مطالعه، طبقه‌بندی آبراهه‌ها به دو دسته انتخاب و حذف بود. در این راستا، ویژگی‌های هندسی، توپولوژیکی و معنایی آبراهه‌ها استخراج و اهمیت هر ویژگی با استفاده از آزمون آماری کای-دو (χ^2) سنجیده شد. نتایج ارزیابی مدل نشان داد که دقت این روش معادل ۸۵ درصد است [۲۱]. Yan و همکاران (۲۰۲۲) از سه روش یادگیری ماشین شامل شبکه عصبی پس‌انتشار، ماشین بردار پشتیبان و درخت تصمیم برای انتخاب شبکه رودخانه استفاده کردند. آن‌ها هشت شاخص برای توصیف ویژگی‌های شبکه رودخانه در نظر گرفتند. نتایج نشان داد که شبکه عصبی پس‌انتشار و ماشین بردار پشتیبان با دقت ۰/۹۴ عملکرد خوبی داشتند، در حالی که روش درخت تصمیم با دقت ۰/۹۱ مزایای قابل توجهی در تفسیرپذیری و استخراج قوانین داشت [۸]. Wang و همکاران (۲۰۲۳) از مدل شبکه عصبی کانولوشنی گرافی (GraphSAGE) برای انتخاب هوشمند بخش‌های مهم از شبکه رودخانه‌ای استفاده کردند. این مدل با استخراج ویژگی‌هایی مانند مرتبه شاخه، طول، انحنا و قیدهای مکانی، روابط پیچیده میان اجزای شبکه را به‌صورت بردارهای تعبیه‌شده می‌آموزد. نتایج نشان داد که دقت مدل بیش از ۰/۹۵ است [۱۳].

با توجه به بررسی‌های صورت گرفته، در بخشی از مطالعات گذشته ساختاردهی شبکه‌های رودخانه‌ای عمدتاً بر اساس ویژگی‌های توپولوژیکی انجام شده و ویژگی‌های هندسی غالباً نادیده گرفته شده است. همچنین، فرآیندهای حذف و انتخاب معمولاً مبتنی بر روش‌های دانش‌محور با

قوانین از پیش تعیین‌شده بود که مستعد خطا و وابسته به تجربه انسانی بوده است. برای رفع این محدودیت‌ها، در این تحقیق، یک مدل مبتنی بر یادگیری عمیق به منظور تعمیم و ساده‌سازی شبکه رودخانه‌ها در دو گام ارائه شده است. در گام اول، ترکیبی از ویژگی‌های توپولوژیکی و هندسی برای ساختاردهی شبکه رودخانه به‌کار گرفته شده، و در گام دوم، فرآیند انتخاب و حذف به‌طور ترکیبی با استفاده از یادگیری عمیق و دستورالعمل‌های سازمان نقشه‌برداری انجام می‌شود. این رویکرد امکان تحلیل دقیق‌تر و کاهش محدودیت‌های روش‌های مبتنی بر دانش را فراهم می‌کند.

نوآوری این تحقیق در ارائه یک مدل دو مرحله‌ای برای تعمیم شبکه رودخانه است که در آن: (۱) ساختار سلسله‌مراتبی شبکه با ترکیب رتبه‌بندی هورتون-استرالر و پیوستگی هندسی Stroke استخراج می‌شود، (۲) انتخاب/حذف شاخه‌ها با مدل GraphSAGE بر مبنای نمایش گراف دوگان شاخه‌محور انجام می‌گیرد، و (۳) برای هم‌راستاسازی خروجی با قواعد رسمی کارتوگرافی، یک مرحله اصلاح مبتنی بر همسایگی مطابق دستورالعمل‌های سازمان نقشه‌برداری اعمال می‌شود. این ترکیب، تصمیم‌گیری داده‌محور را با قیود کارتوگرافیک تلفیق می‌کند و به تولید خروجی نزدیک‌تر به تعمیم مرجع کمک می‌نماید. همچنین، برخلاف مطالعات گذشته که عملگرهای هندسی را صرفاً به ساده‌سازی هندسی خطوط محدود می‌کنند، در این پژوهش عملگر هندسی در سطح تصمیم‌گیری کارتوگرافی تعریف شده است؛ به این معنا که تعیین می‌شود کدام شاخه‌ها در مقیاس هدف حفظ و کدام حذف شوند. این تصمیم‌ها با استفاده از یک مدل یادگیری عمیق مبتنی بر گراف و سپس اصلاح مبتنی بر همسایگی و قیود کارتوگرافیک اعمال می‌شوند.

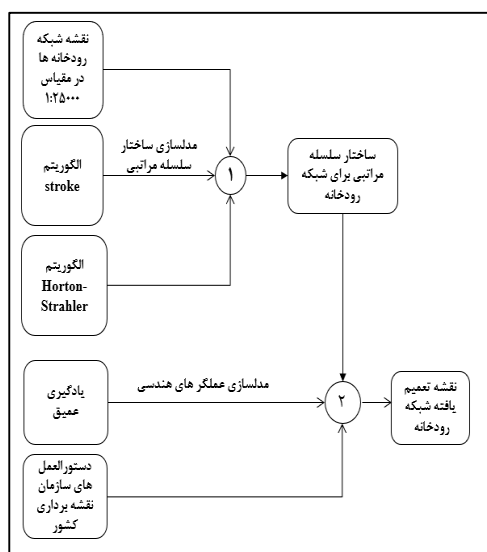
در ادامه، در بخش دوم مدل پیشنهادی این تحقیق ارائه گردیده است و نتایج حاصل از اجرای مدل پیشنهادی در یک منطقه مطالعه موردی در بخش سوم مورد بررسی و ارزیابی قرار می‌گیرد. در نهایت در بخش چهارم نتیجه‌گیری و در بخش پنجم پیشنهاد کارهای آینده ارائه می‌شود.

۲- مواد و روش‌ها

در این تحقیق، فرآیند مدل‌سازی تعمیم و ساده‌سازی شبکه‌های رودخانه‌ای شامل دو مرحله اصلی مدل‌سازی

ساختار سلسله‌مراتبی شبکه‌های رودخانه‌ای و مدل‌سازی فرآیند عملگر هندسی شبکه‌های رودخانه‌ای می‌باشد (شکل ۱). در گام اول، برای ایجاد ساختار سلسله‌مراتبی، شبکه رودخانه به‌عنوان یک ساختار درختی در نظر گرفته می‌شود که این ساختار می‌تواند از طریق الگوریتم Horton-Strahler تحلیل شود. این الگوریتم به‌طور بازگشتی برای رتبه‌بندی شاخه‌ها بر اساس روابط جریان و ساختار شبکه عمل می‌کند و به تعیین سلسله‌مراتب طبیعی شاخه‌ها کمک می‌کند. پس از تعیین سلسله‌مراتب توپولوژیکی، از الگوریتم Stroke برای تحلیل پیوستگی هندسی و تداوم جریان در مسیرهای رودخانه استفاده می‌شود. این الگوریتم با بررسی زاویه‌های بین شاخه‌های متصل به هر گره، پیوستگی و هم‌راستایی شاخه‌ها را ارزیابی کرده و ساختار هندسی شبکه را تقویت می‌کند. ترکیب نتایج حاصل از تحلیل‌های Horton-Strahler و Stroke یک مدل سلسله‌مراتبی از شبکه رودخانه ایجاد می‌کند که هم ساختار توپولوژیکی و هم ویژگی‌های هندسی و اتصال فضایی شاخه‌ها را پوشش می‌دهد. در گام دوم، به‌منظور اعمال عملگرهای هندسی، از رویکرد ترکیبی یادگیری عمیق و دستورالعمل‌های سازمان نقشه‌برداری برای انتخاب و حذف شاخه‌ها استفاده می‌شود. در نهایت، فرآیند تعمیم با استفاده از روش ارزیابی انطباق هندسی مورد ارزیابی قرار می‌گیرد.

لازم به ذکر است منظور از عملگرهای هندسی در این پژوهش، مجموعه تصمیم‌ها و تبدیلاتی است که مستقیماً بر نمایش هندسی شبکه رودخانه در مقیاس هدف اثر می‌گذارند و منجر به انتخاب یا حذف شاخه‌ها و در نتیجه کاهش پیچیدگی هندسی شبکه می‌شوند.



شکل ۱- مراحل انجام تحقیق

۲-۱- مدلسازی ساختار سلسله‌مراتبی شبکه رودخانه

ساختاردهی شبکه‌های رودخانه‌ای گامی اساسی در فرآیند تعمیم و ساده سازی به شمار می‌رود [۷، ۱۱]. برای این منظور، شبکه رودخانه معمولاً به صورت یک گراف $G = (V, E)$ مدل می‌شود، که در آن V مجموعه گره‌ها و نمایانگر نقاط تلاقی یا خروجی‌های رودخانه و E مجموعه یال‌ها و نمایانگر مسیرهای جریان آب هستند [۱۶، ۲۰]. این نحوه مدل‌سازی، زمینه را برای استفاده از الگوریتم‌ها و روش‌های مختلف به منظور سازمان‌دهی، رتبه‌بندی و استخراج ساختار سلسله‌مراتبی رودخانه‌ها فراهم می‌کند. در این راستا، رویکرد سلسله‌مراتبی و رویکرد مبتنی بر Stroke برای ساختاردهی شبکه‌های رودخانه‌ای پیشنهاد شده‌اند. در ادامه، هریک از این الگوریتم‌ها تشریح می‌شوند.

۲-۱-۱- رویکرد سلسله‌مراتبی

شبکه‌های رودخانه‌ای از پیچیده‌ترین و در عین حال منظم‌ترین ساختارهای طبیعی هستند و ویژگی سلسله‌مراتبی آن‌ها، در قالب ارتباط میان شاخه‌های اصلی و فرعی، نقش کلیدی در مدل‌سازی جریان و تعمیم نقشه‌های توپوگرافی ایفا می‌کند. الگوریتم‌های رتبه‌بندی با تعیین مرتبه هر جریان آبی، ساختار شبکه را به صورت درختی سازمان‌دهی و قابل تحلیل می‌کنند [۲۲]. روش‌هایی مانند درجه‌بندی هورتون، استرالر و شومر نمونه‌های شناخته‌شده این رویکرد هستند که هر یک بر اساس منطق خاص خود، جایگاه شاخه‌ها را در سلسله‌مراتب طبیعی شبکه مشخص می‌کنند.

یکی از روش‌های متداول برای تحلیل ساختار سلسله‌مراتبی شبکه‌های رودخانه‌ای، الگوریتم Horton-Strahler است. این الگوریتم، شاخه‌های رودخانه را بر اساس جایگاه آن‌ها در سلسله‌مراتب طبیعی شبکه رتبه‌بندی می‌کند و به این ترتیب، امکان تشخیص شاخه‌های اصلی و فرعی را فراهم می‌سازد. ورودی الگوریتم Horton-Strahler یک گراف جهت‌دار است که در آن، گره‌ها نمایانگر محل تلاقی یا انشعاب شاخه‌ها و یال‌ها نشان‌دهنده مسیر جریان آب هستند. علاوه بر ساختار اتصال گره‌ها، اطلاعاتی مانند نوع گره و جهت جریان نیز در گراف لحاظ می‌شود تا الگوریتم بتواند جریان طبیعی رودخانه را به‌درستی بازنمایی

کند. این الگوریتم تنها برای گراف‌های بدون چرخه معتبر است؛ به عبارت دیگر، هیچ مسیر بسته‌ای نباید وجود داشته باشد که از یک گره شروع شده و دوباره به همان گره بازگردد. پس از اطمینان از بدون حلقه بودن گراف، فرآیند رتبه‌بندی شاخه‌ها به صورت بازگشتی و بر پایه دو قانون اصلی انجام می‌شود:

۱. تخصیص رتبه به شاخه‌های آغازین: شاخه‌هایی که هیچ یال ورودی ندارند (گره‌های ابتدایی شبکه)، به عنوان سرشاخه‌های جریان شناخته شده و رتبه آن‌ها برابر با ۱ در نظر گرفته می‌شود.
۲. محاسبه رتبه شاخه‌های پایین‌دست: برای هر گره پایین‌دست، رتبه شاخه خروجی بر اساس رتبه شاخه‌های ورودی تعیین می‌شود. اگر تمام شاخه‌های ورودی دارای رتبه‌ای یکسان باشند، رتبه شاخه خروجی یک واحد بیشتر از آن رتبه خواهد بود. اگر شاخه‌های ورودی دارای رتبه‌های متفاوت باشند، رتبه شاخه خروجی برابر با بالاترین رتبه در میان شاخه‌های ورودی تعیین می‌شود.

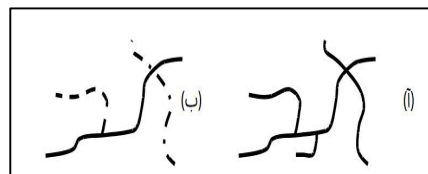
به این ترتیب، الگوریتم Horton-Strahler ساختار سلسله‌مراتبی مطابق با ساختار طبیعی شبکه رودخانه ایجاد می‌کند؛ به گونه‌ای که شاخه‌های اصلی دارای رتبه‌های بالاتر بوده و شاخه‌های فرعی با رتبه‌های پایین‌تر به آن‌ها متصل می‌شوند.

۲-۱-۲- رویکرد مبتنی بر stroke

Stroke بیانگر مسیرهای مستقیمی می‌باشد که با کمترین تغییر جهت و بیشترین تداوم هندسی امتداد می‌یابد. این مفهوم از اصول گشتالت، به ویژه تداوم خوب، الهام می‌گیرد [۲۳]. مسیرهای Stroke بر اساس سه قانون اصلی و به شرح زیر ساخته می‌شوند [۲۴]:

- (۱) قانون نام: تنها قوس‌هایی انتخاب می‌شوند که نام یکسانی (مانند نام رودخانه) دارند تا پیوستگی معنایی حفظ شود.
 - (۲) قانون طول: در میان قوس‌های متصل، مسیری انتخاب می‌شود که طول بیشتری دارد.
 - (۳) قانون زاویه: ادامه مسیر به قوسی داده می‌شود که کمترین تغییر جهت را نسبت به قوس جاری دارد.
- این برداشت ادراکی از پیوستگی و امتداد خطوط، در شکل ۲ نشان داده شده است. در این شکل، اصل تداوم

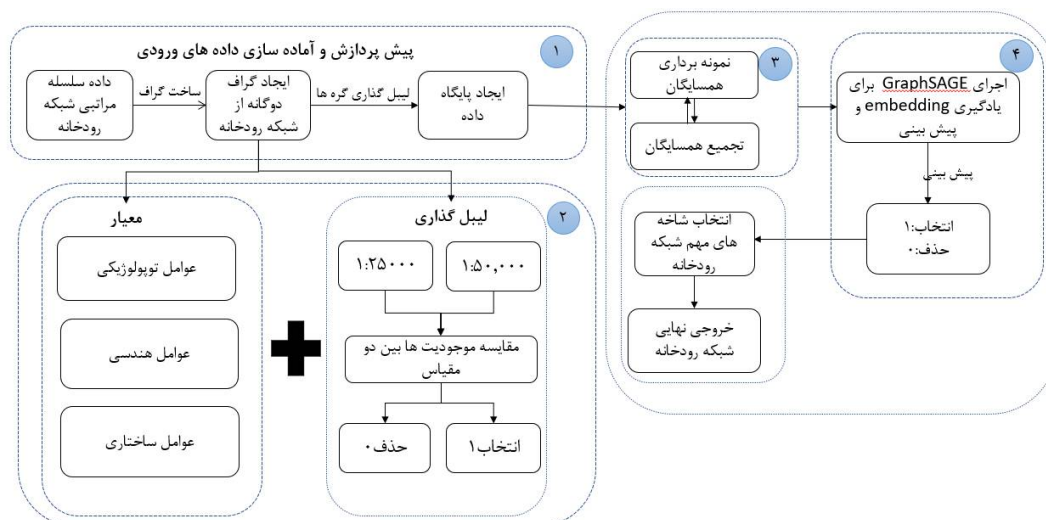
خوب موجب می‌شود که عناصر ترسیمی به‌طور طبیعی به‌صورت چهار منحنی پیوسته و صاف درک شوند [۲۳].



شکل ۲- (الف) یک شبکه ساده شامل ۸ یال و ۹ گره، (ب) بازنمایی همان شبکه در قالب ۴ مسیر Stroke در این فرآیند، هیچ بخشی از شبکه حذف نشده و Stroke‌ها صرفاً نمایش مفهومی پیوستگی هندسی مسیرها هستند [۲۳].

۲-۲-۲ مدل‌سازی عملگرهای هندسی شبکه رودخانه‌ها

به منظور مدل‌سازی عملگرهای هندسی در این تحقیق یک مدل چهار مرحله‌ای ارائه شده است (شکل ۳). در



شکل ۳- مدل کلی ارائه‌شده به منظور اعمال عملگر هندسی شبکه رودخانه

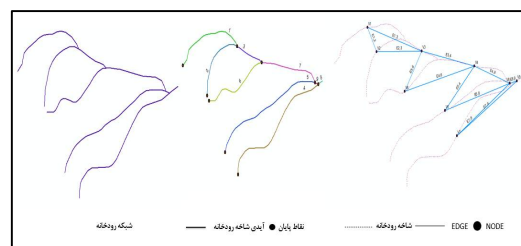
آن‌ها را بهتر نمایان می‌سازد. برای تبدیل شبکه رودخانه به ساختار گراف دوگان، این شبکه به‌صورت $G = (V, E)$ مدل می‌شود، که در آن V مجموعه گره‌هاست و هر گره نمایانگر یک بخش رودخانه است و E مجموعه یال‌ها است که ارتباط بین گره‌ها را مشخص می‌کند و یک ماتریس مجاورت $A \in \mathbb{R}^{N \times N}$ نشان می‌دهد کدام گره‌ها به هم متصل هستند؛ اگر گره i و j متصل باشد $A_{ij} = 1$ و در غیر این صورت مقدار صفر است [۱۳، ۲۵، ۲۶].

لازم به تأکید است که هدف از به‌کارگیری نمایش گراف دوگان در این پژوهش، تمرکز بر نمایش روابط توپولوژیکی

۲-۲-۱ تبدیل شبکه رودخانه به گراف دوگان

در گراف کلاسیک، گره‌ها معمولاً نقاط تقاطع یا محل اتصال شاخه‌ها را نشان می‌دهند و یال‌ها نمایانگر بخش‌های مختلف رودخانه هستند. اما در گراف دوگان، هر شاخه رودخانه به‌عنوان یک گره مستقل در نظر گرفته می‌شود و یال‌ها ارتباطات بین شاخه‌ها (مانند تلاقی‌ها یا انشعاب‌ها) را نشان می‌دهند. در مقایسه با گراف کلاسیک، گراف دوگان درک شهودی‌تری از ساختار و ارتباطات شبکه رودخانه ارائه می‌دهد و ماهیت بخش‌های رودخانه و روابط توپولوژیکی

و ساختاری ضروری برای تصمیم‌گیری کارتوگرافی در فرآیند تصمیم است. در این چارچوب، روابطی نظیر اتصال شاخه‌ها، روابط انشعابی بالادست-پایین‌دست، مشارکت شاخه‌ها در ساختار سلسله‌مراتبی و همسایگی توپولوژیکی به‌عنوان روابط ضروری در نظر گرفته شده‌اند، زیرا این روابط مستقیماً بر تصمیم حذف یا حفظ شاخه‌ها در مقیاس هدف اثرگذار هستند. در مقابل، برخی جنبه‌ها مانند شدت جریان، تغییرات زمانی دبی یا فرآیندهای دینامیکی هیدرولوژیکی که نقش مستقیمی در تصمیم‌های تصمیم‌گیری کارتوگرافی ندارند، در این سطح از مدل‌سازی لحاظ نشده‌اند.



شکل ۴- فرایند ساخت گراف دوگان

۲-۲-۲- الگوریتم ساخت گراف شبکه رودخانه

در این پژوهش، فرآیند ساخت گراف شامل چهار گام اصلی است. در گام نخست، هر بخش خطی رودخانه (بین دو گره انشعاب یا اتصال) به‌عنوان یک گره مستقل در گراف دوگان تعریف شد. در گام دوم، روابط توپولوژیکی بین بخش‌های مجاور رودخانه استخراج و به‌صورت یال‌های گراف مدل شدند؛ به‌گونه‌ای که اگر دو بخش رودخانه در یک گره مشترک به یکدیگر متصل باشند، یک یال بین گره‌های متناظر آن‌ها در گراف ایجاد می‌شود.

جهت‌دار بودن گراف بر اساس جهت جریان آب تعریف گردید. بدین منظور، جهت جریان هر بخش رودخانه با استفاده از اختلاف ارتفاع ابتدا و انتهای شاخه و با بهره‌گیری از داده مدل رقمی ارتفاع (DEM) تعیین شد. یال‌های گراف به‌صورت جهت‌دار از بخش بالادست به بخش پایین‌دست تعریف شدند تا روابط هیدرولوژیکی و انشعابی شبکه به‌درستی بازنمایی شود.

پیش از ساخت نهایی گراف، خطاهای رایج شبکه رودخانه شامل ناسازگاری در جهت جریان، گره‌های معلق، انشعاب‌های ناقص و خطاهای توپولوژیکی شناسایی و اصلاح شدند. به‌طور مشخص، شاخه‌هایی که جهت جریان آن‌ها با منطق ارتفاعی یا روابط همسایگی سازگار نبود، اصلاح

گردیدند و گره‌هایی که منجر به ایجاد چرخه یا بن‌بست غیرواقعی می‌شدند، بازبینی و تصحیح شدند. این مرحله تضمین می‌کند که گراف نهایی فاقد چرخه بوده و با فرض ساختار درختی شبکه رودخانه سازگار است. همچنین، تعیین جهت جریان مبتنی بر مقایسه ارتفاع ابتدا و انتهای هر شاخه و بررسی سازگاری آن با شاخه‌های مجاور در محل انشعاب‌ها انجام شده و در صورت مشاهده ناسازگاری، اصلاح جهت بر اساس قواعد توپولوژیکی جریان و پیوستگی شبکه اعمال گردیده است. در نهایت، گراف جهت‌دار حاصل به‌عنوان ورودی مدل GraphSAGE و مبنای استخراج ویژگی‌ها و فرآیند یادگیری قرار گرفت.

۲-۲-۳- استخراج ویژگی‌ها

برای تولید برجسب‌ها، داده‌های شبکه رودخانه در مقیاس‌های ۱:۲۵,۰۰۰ و ۱:۵۰,۰۰۰ با یکدیگر مقایسه شدند. در این فرآیند، شاخه‌هایی که همپوشانی فضایی قابل توجهی بین دو مقیاس داشتند، به‌عنوان شاخه‌های منتخب (برجسب ۱) تعیین شدند، و شاخه‌هایی که کمتر از این میزان همپوشانی داشتند، حذف یا به‌عنوان شاخه‌های منفی (برجسب ۰) در نظر گرفته شدند.

به‌منظور کمی‌سازی مفهوم همپوشانی فضایی قابل توجه، در این پژوهش از نسبت همپوشانی طولی استفاده شده است. بدین منظور، برای هر شاخه رودخانه در مقیاس ۱:۲۵,۰۰۰، نسبت طول بخش همپوشان با شبکه تعمیم‌یافته مرجع در مقیاس ۱:۵۰,۰۰۰ به طول کل شاخه محاسبه گردید. در صورتی که این نسبت برابر یا بزرگ‌تر از ۰/۵ باشد، شاخه به‌عنوان شاخه حفظ شده در مقیاس هدف تلقی شده و برجسب یک به آن اختصاص داده شد؛ در غیر این صورت، شاخه به‌عنوان شاخه حذف‌شده در نظر گرفته شده و برجسب صفر دریافت کرد.

ویژگی‌های داده‌های شبکه رودخانه نقش اساسی در آموزش مدل‌های یادگیری عمیق دارند و تأثیر تعیین‌کننده‌ای بر کیفیت نتایج تعمیم‌یافته دارند [۱۳]. در مدل گراف، هر گره با مجموعه‌ای از ویژگی‌ها توصیف می‌شود که نمایانگر خصوصیات هندسی، ساختاری و عملکردی آن در شبکه رودخانه است. این ویژگی‌ها شامل سه دسته اصلی عوامل ساختاری، توپولوژیکی و هندسی می‌باشند. عوامل مؤثر ساختاری رودخانه شامل رتبه رودخانه هستند. رتبه رودخانه به روابط سلسله‌مراتبی و

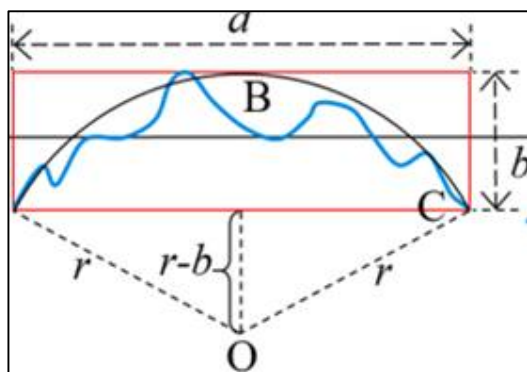
$$c = \frac{d}{b} \quad (2)$$

که در آن b پهنای مستطیل محاطی و d طول مستطیل محاطی است.

- شاخص انحناى خمیدگی شاخه‌ها: ماریچ بودن ویژگی معمول رودخانه‌هاست که نشان‌دهنده وجود منحنی‌های پیوسته و مکرر در مسیر رودخانه است. شاخص ماریچ بودن میزان انحراف پیوسته از مسیر مستقیم را اندازه‌گیری می‌کند و قادر است پیچیدگی خمیدگی‌های رودخانه را بهتر از معیارهای ساده‌تر نشان دهد (شکل ۶). این شاخص بر اساس طول مسیر، طول مستطیل کم‌حجم و شعاع طولانی‌ترین قوس در مسیر محاسبه می‌شود و عدد بزرگ‌تر نشان‌دهنده ماریچ بودن بیشتر رودخانه است (رابطه ۳) [۲۸].

$$m = \frac{L}{2r * \arcsin(\frac{d}{2r})} - 1 \quad (3)$$

که در آن L طول مسیر رودخانه، d طول مستطیل و r شعاع بزرگ‌ترین قوس می‌باشد.



شکل ۶- نمودار شماتیک محاسبه شاخص انحناى خمیدگی شاخه‌ها و شاخص انحراف مستطیل محاطی [۲۸]

- فاصله شعاعی: این معیار، موقعیت نسبی هر شاخه از رودخانه را نسبت به یک نقطه مرجع (معمولاً مرکز حوضه یا دهانه اصلی) به‌صورت کمی بیان می‌کند (شکل ۷). این ویژگی از اختلاف فاصله دو انتهای یک شاخه رودخانه تا نقطه مرجع محاسبه می‌شود و رابطه آن به‌صورت رابطه شماره (۴) است [۲۸].

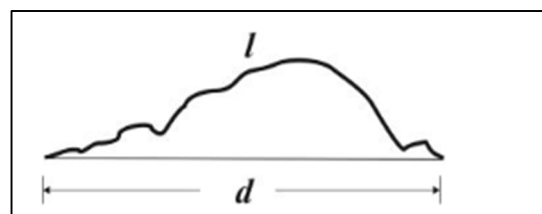
$$l_{OA} - l_{OB} = \Delta l \quad (4)$$

سیستم‌های رتبه‌بندی مختلفی مانند رتبه‌بندی‌های هورتون، استرالر و هک اشاره دارد که برای تحلیل و سازماندهی بخش‌های مختلف شبکه رودخانه‌ای مورد استفاده قرار می‌گیرند [۸]. عوامل مؤثر توپولوژیکی شامل تعداد رودخانه‌های فرعی و گره‌های مجاور هستند. تعداد رودخانه‌های فرعی به تعداد شاخه‌ها و سلسله‌مراتب اتصالات در شبکه رودخانه‌ای مربوط می‌شود. در نهایت، گره‌های مجاور نمایانگر تعداد گره‌هایی هستند که از نظر توپولوژیکی با یک گره خاص در ارتباط بوده و به آن متصل می‌شوند. [۸، ۱۳]. همچنین عوامل مؤثر هندسی شبکه رودخانه شامل طول رودخانه، طول stroke، انحنا، اختلاف فاصله نقطه مرجع، شاخص انحراف مستطیل محاطی و انحناى خمیدگی شاخه‌ها می‌باشد [۲۷، ۲۸] که در ادامه به اختصار تشریح شده‌اند.

- انحنا: انحناى رودخانه نشان‌دهنده میزان انحراف مسیر رودخانه از یک خط مستقیم بین نقاط شروع و پایان آن است (شکل ۵). ساده‌ترین شاخص کمی برای اندازه‌گیری آن، نسبت طول واقعی رودخانه به فاصله خطی میان نقاط شروع و پایان آن است (رابطه ۱) [۹]:

$$ca = \frac{l}{d} \quad (1)$$

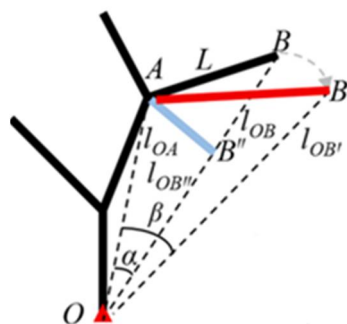
که در آن l طول واقعی رودخانه و d فاصله خطی بین نقطه شروع و پایان رودخانه است.



شکل ۵- انحناى شاخه رودخانه [۹]

- شاخص انحراف مستطیل محاطی: این شاخص به‌منظور اندازه‌گیری میزان انحراف یا خمیدگی مسیر یک بازه رودخانه از راستای خط مستقیم بین دو نقطه ابتدایی و انتهایی آن استفاده می‌شود (شکل ۶). برای این منظور، بازه مورد نظر در یک مستطیل محاطی قرار می‌گیرد و نسبت پهنای مستطیل به طول آن به عنوان شاخص انحراف محاسبه می‌شود. هر چه این نسبت بزرگ‌تر باشد، بیانگر خمیدگی یا انحناى بیشتر مسیر رودخانه نسبت به راستای مستقیم است (رابطه ۲) [۲۸].

که در آن $l_{OA} - l_{OB}$ به ترتیب فواصل انتهایی A و B شاخه مورد نظر تا نقطه مرجع O هستند؛ Δl شاخص فاصله شعاعی شاخه نسبت به مرکز شبکه است. هرچه مقدار Δl بیشتر باشد، نشان‌دهنده کشیدگی یا پراکندگی بیشتر شاخه در ساختار کلی شبکه رودخانه است.

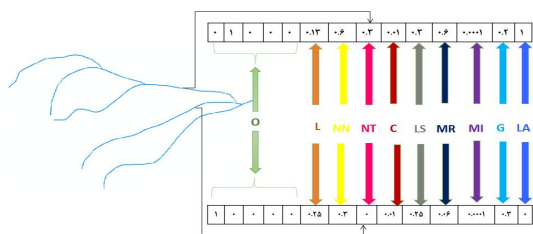


شکل ۷- نمودار شماتیک محاسبه اختلاف فاصله نسبت به نقطه مرجع [۲۸]

لازم به ذکر است که الگوریتم GraphSAGE تنها قادر به مدیریت داده‌های عددی است. ویژگی‌های معنایی معمولاً زمان آموزش طولانی‌تری برای شبکه‌های عصبی ایجاد می‌کنند و باعث عدم همگرایی مدل می‌شوند، به همین منظور، در این تحقیق دو روش استانداردسازی به کار گرفته شد: ویژگی‌های ساختاری مانند رتبه شاخه‌ها با استفاده از روش One-Hot Encoding به داده‌های عددی تبدیل شدند و ویژگی‌های عددی مانند طول شاخه، میزان انحنای و تعداد گره‌های مجاور با روش Min-Max به بازه $[0,1]$ نگاشت می‌شوند. به این ترتیب، داده‌های ورودی مدل به صورت مجموعه‌ای از بردارهای ویژگی استاندارد شده برای تمامی گره‌ها آماده شدند که امکان مقایسه‌پذیری، هم‌مقیاس‌سازی و یادگیری مؤثر در مدل گرافی را فراهم می‌سازد.

لازم به توضیح است که ویژگی‌های ساختاری مورد استفاده در این پژوهش، از جمله رتبه‌های سلسله‌مراتبی شاخه‌های رودخانه، ماهیتی گسسته و طبقه‌ای دارند و بیانگر سطوح متمایز ساختاری در شبکه رودخانه‌ای هستند، نه کمیت‌های پیوسته با فاصله عددی معنادار. از این‌رو، نرمال‌سازی این ویژگی‌ها به بازه $[0,1]$ می‌تواند به القای روابط عددی مصنوعی و نادرست میان سطوح سلسله‌مراتبی منجر شود. به‌منظور حفظ ماهیت طبقه‌ای این ویژگی‌ها و جلوگیری از تحمیل ترتیب یا فاصله عددی غیرواقعی به مدل یادگیری، از روش One-Hot Encoding برای بازنمایی ویژگی‌های ساختاری استفاده شد. اگرچه این روش منجر به

افزایش بعد بردار ورودی می‌شود، اما با توجه به معماری GraphSAGE و توانایی آن در پردازش مؤثر بردارهای ویژگی با ابعاد بالاتر، این افزایش بعد تأثیر منفی معناداری بر فرآیند آموزش و عملکرد مدل نداشته است.



شکل ۸- استانداردسازی ویژگی‌های کیفی و عددی

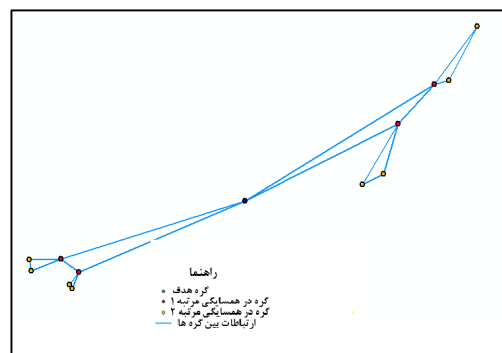
همچنین ذکر این نکته قابل توجه است که مرجع برچسب‌گذاری در این پژوهش، نقشه تعمیم‌یافته ۱:۵۰,۰۰۰ تولیدشده توسط کارتوگراف‌های سازمان نقشه‌برداری کشور است که فرآیند انتخاب/حذف و ساده‌سازی شاخه‌ها را مطابق دستورالعمل‌ها و استانداردهای رسمی تعمیم عوارض هیدروگرافی انجام داده‌اند.

۲-۴- نمونه‌برداری و تجمیع

الگوریتم GraphSAGE به‌منظور بهبود مقیاس‌پذیری شبکه‌های عصبی گرافی معرفی شد. برخلاف مدل‌های اولیه مانند GCN که در آن‌ها تمام همسایگان گره برای به‌روزرسانی ویژگی‌ها در نظر گرفته می‌شوند، نوآوری اصلی GraphSAGE در این است که به‌جای استفاده از تمامی همسایگان، یک زیرمجموعه تصادفی از آن‌ها را انتخاب کرده و سپس با بهره‌گیری از توابع تجمیع مختلف مانند Mean، Max یا Sum، ویژگی‌های همسایگان را ترکیب می‌کند. این راهکار علاوه بر کاهش پیچیدگی محاسباتی، امکان یادگیری بازنمایی‌های تعمیم‌پذیر و کارآمد را در گراف‌های بسیار بزرگ فراهم می‌سازد. در نتیجه GraphSAGE نه تنها مسئله مقیاس‌پذیری را برطرف می‌کند، بلکه انعطاف‌پذیری بیشتری در استخراج ساختارهای پنهان داده‌ها نسبت به مدل‌های کلاسیک GNN ارائه می‌دهد. این ویژگی آن را به گزینه‌ای مناسب برای مدل‌سازی ساختارهای پویا مانند شبکه‌های رودخانه‌ای تبدیل کرده است [۱۳، ۲۷، ۲۸، ۲۹].

فرآیند اصلی در GraphSAGE شامل دو مرحله‌ی کلیدی نمونه‌گیری و تجمیع است: در روش نمونه‌گیری،

برای هر گره مجموعه ای تصادفی از همسایگان آن انتخاب می‌شود.



شکل ۹- نمونه‌ای از فرایند نمونه‌برداری

این فرآیند برای هر لایه از مدل به صورت تکراری انجام می‌شود [۱۳، ۲۷، ۲۸، ۲۹]. تا گره‌ها بتوانند اطلاعات بیشتری را از همسایگان دورتر خود جمع‌آوری کنند فرض کنید $N(v)$ نمایانگر مجموعه همسایگان گره v باشد، در این صورت فرآیند نمونه‌گیری به صورت رابطه (۵) تعریف می‌شود [۱۳]:

$$sample_k(N(v)) = N_k(v) \quad (5)$$

که در آن $N_k(v)$ مجموعه ای از همسایگان گره v در عمق k است و $sample_k$ فرآیند نمونه‌گیری برای انتخاب k همسایه است. در این فرآیند، برای هر گره هدف به‌طور تصادفی ابتدا ۴ همسایه از مرتبه اول و سپس ۸ همسایه از مرتبه دوم نمونه‌برداری شده و یک زیرگراف محلی که شامل گره هدف و همسایگان آن باشد، تشکیل گردید. این رویکرد باعث افزایش مقیاس‌پذیری مدل و کاهش نیاز محاسباتی در مراحل آموزش و ارزیابی می‌شود. پس از نمونه‌برداری و انتخاب زیرگراف محلی، مرحله‌ی بعدی ترکیب اطلاعات گره هدف با همسایگان آن است. این فرآیند در هر لایه‌ی مدل با استفاده از عملگرهای خاصی انجام می‌شود که در GraphSAGE به آن‌ها تجمیع‌کننده گفته می‌شود. هدف از این مرحله، انتقال اطلاعات خصوصیات همسایگان به گره هدف است تا نمایشی به‌روز شده از آن گره به دست آید. این کار با استفاده از توابع تجمیع مختلف انجام می‌شود که در GraphSAGE شامل روش‌های ماکزیمم، مجموع و میانگین هستند [۱۳، ۲۷]. در این تحقیق، با توجه به ماهیت درختی و پراکندگی اتصالات در شبکه رودخانه، از عملگر Max استفاده شد (رابطه ۶). در

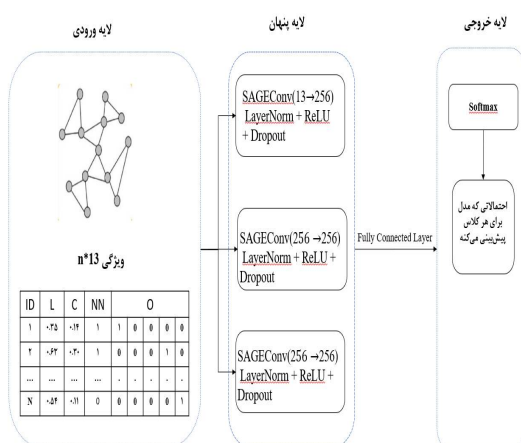
این عملگر برای هر بعد از ویژگی‌ها، بیشترین مقدار در بین همسایگان انتخاب می‌شود [۲۷].

$$\sigma\left(W^k \cdot \text{MAX}\left(\{h_u^{(k-1)}, \forall u \in N(v)\}\right)\right) = h_v^k \quad (6)$$

که در آن h_v^k نمایش گره v در لایه k است، $N(v)$ مجموعه همسایگان گره v ، MAX عملگر max روی بردار های ویژگی همسایگان، W^k ماتریس وزن قابل آموزش در لایه k و σ تابع فعالسازی است.

۲-۲-۵- ساخت مدل انتخاب

در این پژوهش، مدل انتخاب و طبقه‌بندی شاخه‌های رودخانه بر پایه‌ی چارچوب GraphSAGE طراحی گردید و شامل سه بخش اصلی است: لایه ورودی، لایه‌های پنهان و لایه خروجی.



شکل ۱۰- معماری مدل طبقه‌بندی گراف بر پایه GraphSAGE

لایه ورودی از دو بخش تشکیل می‌شود. بخش نخست ماتریس ویژگی‌ها است که در مرحله‌ی استخراج معیارها محاسبه شده‌اند. هر گره در شبکه رودخانه‌ای با یک بردار شامل ۱۳ ویژگی هندسی و هیدرولوژیکی توصیف می‌شود. این بردارها در کنار یکدیگر ماتریس ویژگی‌ها را تشکیل می‌دهند که به صورت $X \in \mathbb{R}^{N \times F}$ نمایش داده می‌شود؛ در این رابطه N تعداد گره‌ها و F تعداد ویژگی‌ها (۱۳ ویژگی) است. بخش دوم مربوط به ساختار توپولوژیکی شبکه رودخانه است که ارتباط بین گره‌ها را نشان می‌دهد. این ساختار با استفاده از ماتریس مجاورت بازنمایی شده و بیانگر روابط انشعابی و اتصال شاخه‌ها است. ترکیب ماتریس ویژگی‌ها و ماتریس مجاورت، ورودی اصلی مدل را تشکیل

می‌دهد و به عنوان پایه‌ی یادگیری مدل در نظر گرفته می‌شود. در شکل ۱۰ معماری مدل طبقه‌بندی گراف بر پایه GraphSAGE نشان داده می‌شود.

در این پژوهش از سه لایه‌ی متوالی SAGEConv با ۲۵۶ کانال پنهان در هر لایه و تابع فعال‌سازی ReLU استفاده شده است. تابع ReLU مقادیر منفی را صفر و مقادیر غیرمنفی را بدون تغییر عبور می‌دهد (رابطه ۷). ReLU باعث ساده‌تر شدن محاسبات و کاهش مشکل ناپدید شدن گرادیان می‌شود [۳۰].

$$\max(0, x) = \text{ReLU}(X) \quad (7)$$

لایه خروجی مدل یک لایه کاملاً متصل است که خروجی لایه پنهان دوم را به عنوان ورودی دریافت کرده و توزیع احتمال طبقه‌بندی گره‌ها را به عنوان خروجی برمی‌گرداند. این توزیع احتمال توسط یک لایه طبقه‌بندی خطی به برداری با اندازه ۲۵۶ کانال تبدیل می‌شود. رابطه ریاضی برای طبقه‌بندی به صورت رابطه (۸) می‌باشد.

$$F(x, w) = wx + b \quad (8)$$

که در آن $F(x, w)$ بردار خروجی است، x برداری وردی است، w ماتریس پارامتر یا وزن‌ها است و b بردار انحراف است. همچنین، تابع فعال‌سازی Softmax برای نرمال‌سازی خروجی شبکه عصبی به یک توزیع احتمالی استفاده می‌شود، جایی که هر گره خروجی نمایانگر یک دسته است و مقدار خروجی هر گره نمایانگر احتمال آن دسته می‌باشد. معادله ریاضی آن به صورت رابطه (۹) است.

$$\text{softmax}(x_i) = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}} \quad (9)$$

این تابع هر مقدار z_i را به یک احتمال برای تعلق به کلاس i تبدیل می‌کند، مجموع تمام خروجی‌ها برای تابع Softmax برابر با ۱ است.

توابع هزینه در GNN نقش اساسی در آموزش مدل‌ها برای یادگیری بازنمایی‌های معنادار و انجام پیش‌بینی‌های دقیق روی داده‌های گرافی دارند. انتخاب تابع هزینه بسته به نوع وظیفه (کلاس‌بندی رأس، پیش‌بینی یال، کلاس‌بندی گراف) و ویژگی‌های داده‌ها می‌تواند تأثیر قابل‌توجهی بر عملکرد مدل داشته باشد. تابع Cross-Entropy یکی از رایج‌ترین و پرکاربردترین توابع هزینه در

مسائل دسته‌بندی است و به ویژه در شبکه‌های عصبی گرافی برای کلاس‌بندی رأس‌ها و پیش‌بینی یال‌ها استفاده می‌شود. این تابع اختلاف بین توزیع پیش‌بینی‌شده مدل و توزیع برچسب‌های واقعی را اندازه‌گیری می‌کند و سعی دارد این اختلاف را کمینه کند [۳۰، ۳۱].

$$\mathcal{L}_{CE}(y, p) = - \sum_{i=1}^C y_i \log p_i \quad (10)$$

که در آن C تعداد کلاس‌ها است و p_i احتمال پیش‌بینی‌شده توسط مدل برای کلاس i است. مزیت اصلی CE این است که با خروجی Softmax شبکه‌های عصبی هم‌خوانی دارد و می‌توان آن را از منظر احتمالاتی نیز تفسیر کرد، زیرا CE در واقع لگاریتم منفی احتمال دسته‌بندی است. به این ترتیب، کمینه‌سازی CE باعث می‌شود مدل، توزیع پیش‌بینی خود را به توزیع واقعی برچسب‌ها نزدیک کند و اعتماد مدل به پیش‌بینی‌های درست افزایش یابد [۳۱]. لازم به ذکر است با توجه به عدم تعادل کلاس‌ها در داده‌های شاخه‌های رودخانه، وزن کلاس‌ها برای مقابله با سوگیری مدل در پیش‌بینی‌ها محاسبه شد. در این پژوهش، کلاس صفر نسبت به کلاس یک تعداد نمونه کمتری داشت، بنابراین وزن بیشتری به آن اختصاص یافت. وزن هر کلاس به صورت معکوس تعداد نمونه‌های آن کلاس تعیین و سپس نرمال‌سازی شد تا مجموع وزن‌ها برابر با یک شود. این وزن‌ها به تابع CrossEntropyLoss منتقل شدند تا مدل هنگام آموزش به هر دو کلاس توجه متوازن داشته باشد و از سوگیری به سمت کلاس غالب جلوگیری شود. استفاده از این روش موجب افزایش حساسیت مدل نسبت به کلاس کمتر مشاهده‌شده و بهبود عملکرد کلی مدل در شرایط عدم تعادل داده‌ها شد.

همچنین در آموزش شبکه‌های عصبی گرافی، بهینه‌سازی مدل نقش مهمی در بهبود دقت و توان تعمیم‌دهی آن دارد. هدف اصلی بهینه‌سازی، کمینه کردن تابع هزینه و تنظیم وزن‌های شبکه به گونه‌ای است که پیش‌بینی‌های مدل بیشترین تطابق را با برچسب‌های واقعی داشته باشند. Adam الگوریتمی پیشرفته‌تر است که ویژگی‌های گرادیان نزولی و مومنتوم را ترکیب می‌کند. در Adam علاوه بر گرادیان فعلی، اطلاعاتی از لحظه اول (میانگین گرادیان‌ها) و لحظه دوم (میانگین مربعات گرادیان‌ها) ذخیره شده و نرخ یادگیری برای هر پارامتر

به صورت خودکار تنظیم می‌شود. به روزرسانی وزن‌ها در Adam به صورت رابطه (۱۱) انجام می‌شود [۳۳]:

$$\begin{aligned} m_t &= \mu * m_{t-1} + (1 - \mu) * g_t \\ n_t &= \vartheta * n_{t-1} + (1 - \vartheta) * g_t^2 \\ \hat{m}_t &= \frac{m_t}{1 - \mu^t} \\ \hat{v}_t &= \frac{n_t}{1 - \vartheta^t} \end{aligned} \quad (11)$$

که در آن g_t گرادیان در زمان t ، m_t و $\hat{v}_t$ میانگین‌های اصلاح‌شده لحظه اول و دوم، μ و ϑ ضرایب ضربه مومنوم و ε یک مقدار کوچک برای جلوگیری از تقسیم بر صفر است. برای افزایش پایداری آموزش و جلوگیری از بیش‌برازش، در هر لایه از نرمال‌سازی ویژگی‌ها و حذف تصادفی واحدها استفاده می‌شود. Dropout در هر لایه به طور تصادفی برخی از واحدهای شبکه را غیرفعال می‌کند تا مدل به جای یادگیری ویژگی‌های خاص داده‌های آموزشی، ویژگی‌های عمومی‌تری را یاد بگیرد.

Layer Normalization ویژگی‌های هر لایه را نرمال‌سازی می‌کند به طوری که میانگین صفر و واریانس یک داشته باشند، که باعث پایداری گرادیان‌ها و کاهش مشکلاتی مانند انفجار یا فروپاشی گرادیان‌ها می‌شود. این تکنیک همچنین اجازه می‌دهد که مدل بدون حساسیت زیاد به مقیاس ویژگی‌های ورودی، روابط بین گره‌ها را بهتر یاد بگیرد. شرط توقف آموزش مدل در این پژوهش بر اساس مکانیزم Early Stopping تعریف شده است. به طور مشخص، پس از هر دوره، عملکرد مدل روی مجموعه اعتبارسنجی ارزیابی می‌شود و معیارهای بررسی شامل دقت و تابع زیان هستند. اگر دقت اعتبارسنجی نسبت به بهترین مقدار قبلی بهبود قابل توجهی داشته باشد یا تابع زیان کاهش معناداری پیدا کند، بهبود در نظر گرفته می‌شود و شمارنده توقف صفر می‌شود و وزن‌های مدل ذخیره می‌شوند. در صورتی که هیچ بهبودی رخ ندهد، شمارنده توقف افزایش پیدا می‌کند. هنگامی که شمارنده توقف به مقدار تعیین شده برسد و هیچ بهبودی مشاهده نشود، آموزش متوقف شده و مدل به مرحله ارزیابی نهایی روی داده‌های آزمون منتقل می‌شود. به بیان ساده، مدل آموزش می‌بیند تا زمانی که بهبود معناداری در عملکرد روی داده‌های اعتبارسنجی حاصل شود و اگر برای تعداد

مشخصی از دوره‌ها هیچ بهبودی رخ ندهد، آموزش متوقف شده و از overfitting جلوگیری می‌شود.

در این پژوهش، پس از تولید نقشه تعمیم شده توسط مدل GraphSAGE، برای بهبود دقت و هم‌راستاسازی با نقشه‌های دستی کارتوگرافی، یک فرآیند اصلاح مبتنی بر همسایگی تعریف شد. مطابق دستورالعمل‌های سازمان نقشه‌برداری، در اطراف هر شاخه فرعی شبکه رودخانه‌ای یک شعاع ۲۰۰ متر در نظر گرفته شد و تمامی شاخه‌هایی که در این محدوده قرار داشتند، بررسی شدند. سپس با استفاده از معیار طول شاخه‌ها، شاخه‌ای که طول بیشتری داشت به عنوان شاخه مهم‌تر در نظر گرفته شد و مقدار پیش‌بینی مدل برای آن شاخه اصلاح شد. به این ترتیب، مدل علاوه بر توانایی پیش‌بینی خودکار، با رعایت قواعد فاصله و اهمیت هندسی شاخه‌ها (اعم از شاخه اصلی و فرعی) و مطابق با استانداردهای سازمان نقشه‌برداری، نتایج خود را با نقشه‌های کارتوگرافی هم‌راستا کرده و دقت نهایی در بازنمایی شبکه رودخانه‌ای افزایش یافت.

۲-۳- ارزیابی در فرایند تعمیم شبکه رودخانه

ارزیابی نتایج تعمیم فرآیندی است که طی آن نتایج حاصل بررسی و کنترل می‌شوند تا میزان مطابقت و رضایت‌بخش بودن آن‌ها با اهداف و مقیاس نقشه سنجیده شود. مدل‌سازی فرایند تعمیم و ساده‌سازی باید به گونه‌ای انجام شود که تا حد امکان با فرایند دستی کارتوگراف مطابقت داشته باشد [۳۰]. برای این منظور، فرایند تعمیم و ساده‌سازی در مقیاس ۱:۵۰۰۰۰ با مجموعه داده تعمیم شده موجود در همان مقیاس، که توسط کارتوگراف و بر اساس دستورالعمل‌های سازمان نقشه‌برداری از نقشه‌های ۱:۲۵۰۰۰ تولید شده است، مقایسه می‌شود. به طور کلی، دو روش ارزیابی برای تعمیم و ساده‌سازی وجود دارد:

- روش ارزیابی انطباق هندسی: در این روش، هدف آن است که فرایند تعمیم و ساده‌سازی به گونه‌ای مدل‌سازی شود که نقشه تعمیم شده با نقشه پایه متناسب و هماهنگ باشد، به طوری که ویژگی‌های اصلی نقشه پایه حفظ شوند. به این منظور، در این روش می‌توان از ضریب تطابق CLC مطابق رابطه (۱۲) برای ارزیابی میزان تطابق و حفظ ویژگی‌های اصلی استفاده کرد [۳۱]. CLC نشان می‌دهد که چقدر دو مجموعه خطوط که معیارهای مشابهی

را در زمین نمایش می‌دهند، با یکدیگر همپوشانی دارند [۳۱]:

$$clc = \frac{M}{M + O + C} \quad (12)$$

که در آن M مجموع طول خطوط مطابقتی، O مجموع طول خطوطی که از مجموعه داده تعمیم شده حذف شده‌اند و C مجموع طول خطوطی در مجموعه داده تعمیم شده که مطابقتی در مجموعه داده مرجع ندارند. هرچه مقدار CLC بالاتر باشد، شبکه استخراج‌شده با شبکه اصلی همسان‌تر است و ویژگی‌های اصلی شبکه بهتر حفظ شده‌اند. اگر CLC پایین باشد، جریان‌های اضافه یا کم‌نمایش زیادی وجود دارد و تطابق شبکه با مرجع ضعیف‌تر است [۳۱].

• روش ارزیابی تطبیقی: این روش، هدف آن است که مدل‌سازی فرایند تعمیم و ساده سازی به‌گونه‌ای باشد که نقشه تعمیم شده تا حد ممکن با فرایند دستی کارتوگرافی هماهنگ و نزدیک باشد. شاخه‌های تعمیم شده صحیح، شاخه‌هایی هستند که با شاخه‌های تعمیم شده توسط کارتوگراف برابرند و با TP مشخص می‌شوند. شاخه‌های تعمیم شده نادرست، شاخه‌هایی هستند که با شاخه‌های کارتوگراف برابر نیستند و با FP مشخص می‌شوند. شاخه‌هایی که توسط مدل در نظر گرفته نشده‌اند، با FN مشخص می‌شوند و شاخه‌هایی که هم توسط مدل و هم توسط کارتوگراف در نظر گرفته نشده‌اند، با TN مشخص می‌شوند. بنابراین، میزان صحت، کامل بودن و دقت به ترتیب با استفاده از روابط (۱۳) الی (۱۵) محاسبه می‌شوند.

$$precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} \quad (15)$$

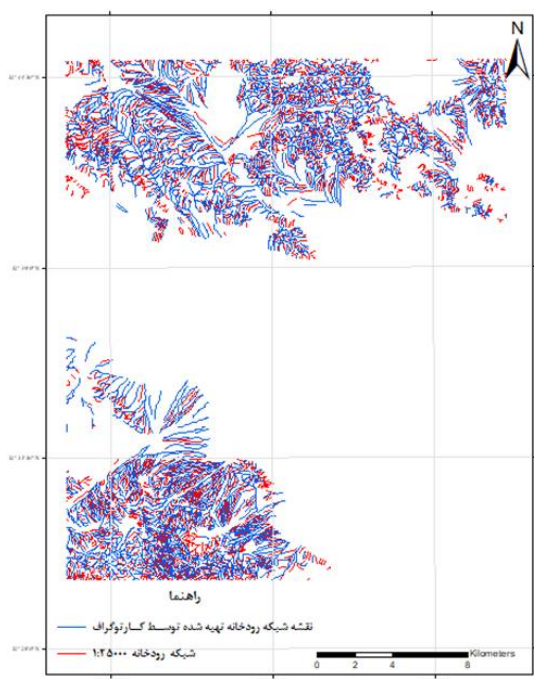
۳- پیاده‌سازی

در این بخش ابتدا منطقه مورد مطالعه در بخش ۱-۳ ارائه شده است. سپس، ساختار سلسله مراتبی شبکه رودخانه با استفاده از اطلاعات هندسی و توپولوژیکی عوارض در بخش ۲-۳، اعمال عملگر هندسی شبکه رودخانه در

بخش ۳-۳ و ارزیابی تعمیم و ساده سازی شبکه رودخانه در بخش ۳-۴ ارائه می‌گردد.

۳-۱- منطقه مورد مطالعه

در این تحقیق، یک مجموعه داده‌ی ورودی در مقیاس ۱:۲۵۰۰۰ و یک مجموعه داده‌ی تست در مقیاس ۱:۵۰۰۰۰ از پایگاه داده‌ی توپوگرافی استان اصفهان به‌عنوان منطقه‌ی مورد مطالعه انتخاب شده است (شکل ۱۱). پیاده‌سازی مدل پیشنهادی در محیط‌های نرم‌افزاری مختلف و متناسب با ماهیت داده‌ها و الگوریتم‌ها انجام شده است. پیش‌پردازش داده‌های مکانی، استخراج ویژگی‌ها، ساخت گراف‌ها و اعمال تبدیلات هندسی شبکه راه‌ها در محیط نرم‌افزار ArcGIS Pro انجام گرفت. مدل‌های یادگیری ماشین و یادگیری عمیق با استفاده از زبان برنامه‌نویسی Python پیاده‌سازی شدند. الگوریتم Random Forest با بهره‌گیری از کتابخانه scikit-learn اجرا شد و پیاده‌سازی شبکه یادگیری عمیق مبتنی بر گراف با استفاده از کتابخانه PyTorch انجام گرفت.



شکل ۱۱- منطقه مورد مطالعه

لازم به تأکید است که در چارچوب پیشنهادی این پژوهش، تمامی ویژگی‌های گره‌ها، روابط همسایگی و ساختار گراف صرفاً از شبکه رودخانه در مقیاس ۱:۲۵,۰۰۰ استخراج

تعیین شد و ناسازگاری‌های احتمالی در محل تلاقی‌ها اصلاح شد. بررسی داده‌ها نشان داد که از مجموع ۶۸۶۵ شاخه، تعداد ۱۴۰۳ شاخه (حدود ۱۵ درصد) دارای خطای ارتفاع بودند؛ به‌گونه‌ای که ارتفاع انتهایی شاخه بیشتر از ارتفاع ابتدای آن بود. این خطاها اصلاح شدند تا جهت جریان طبیعی آبراهه‌ها حفظ گردد.

همچنین، در محل تلاقی شاخه‌های رودخانه‌ای، ممکن است ناسازگاری در جهت جریان رخ دهد؛ به این معنا که جهت جریان برخی شاخه‌ها با شاخه‌ها مجاور هم‌راستا نباشد و در نتیجه، هندسه شبکه رودخانه‌ای دچار اختلال گردد. برای رفع این ناسازگاری‌ها، لازم است هندسه و توپولوژی شاخه‌های متصل به گره به صورت همزمان تحلیل شوند؛ در این پژوهش حدود ۱۰ درصد از شاخه‌ها دارای خطا در همسایگی بودند که با استفاده از قاعده طول و زاویه در محل تلاقی‌ها اصلاح شدند. شکل‌های ۱۲ و ۱۳ به ترتیب نشان دهنده وجود خطای جهت جریان و نتیجه اصلاح آن می‌باشد.

پس از اصلاح جهت جریان شاخه‌ها و اطمینان از صحت آن، الگوریتم‌های Horton-Strahler روی شبکه رودخانه پیاده‌سازی می‌شوند. با توجه به ویژگی‌ها و داده‌های موجود، نتیجه اجرای این الگوریتم‌ها در این مطالعه، شبکه‌ای با ۵ مرتبه شاخه‌ها را نشان می‌دهد (شکل ۱۴).

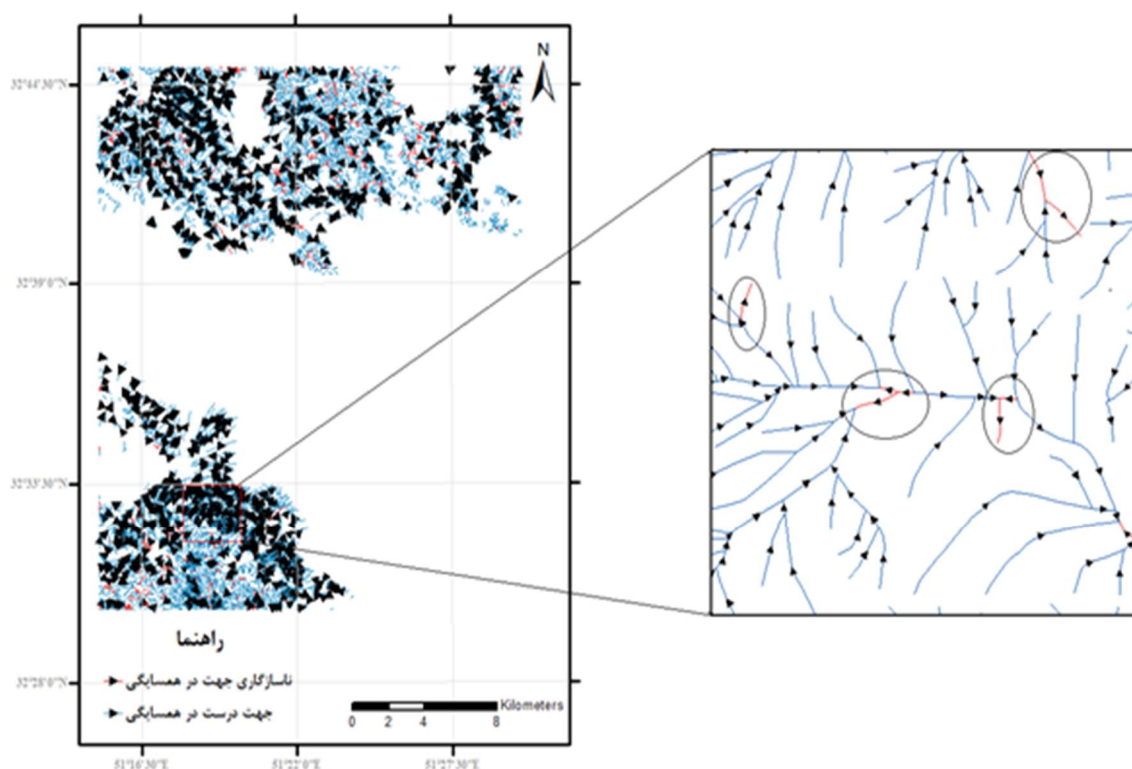
شده‌اند و نقشه مقیاس ۱:۵۰,۰۰۰ هیچ نقشی در ساخت ورودی مدل نداشته است. نقشه ۱:۵۰,۰۰۰ صرفاً به‌عنوان مرجع تعمیم‌یافته برای تولید برچسب‌های آموزشی و ارزیابی عملکرد مدل به‌کار رفته است. بدین ترتیب، مدل پیشنهادی در حال یادگیری نگاشت تصمیم‌های تعمیم کارتوگرافی از فضای ویژگی‌های شبکه بزرگ‌مقیاس به تصمیم‌های انتخاب یا حذف شاخه‌ها در مقیاس هدف است.

۳-۲- نتایج پیاده‌سازی ساختار سلسله‌مراتبی شبکه رودخانه

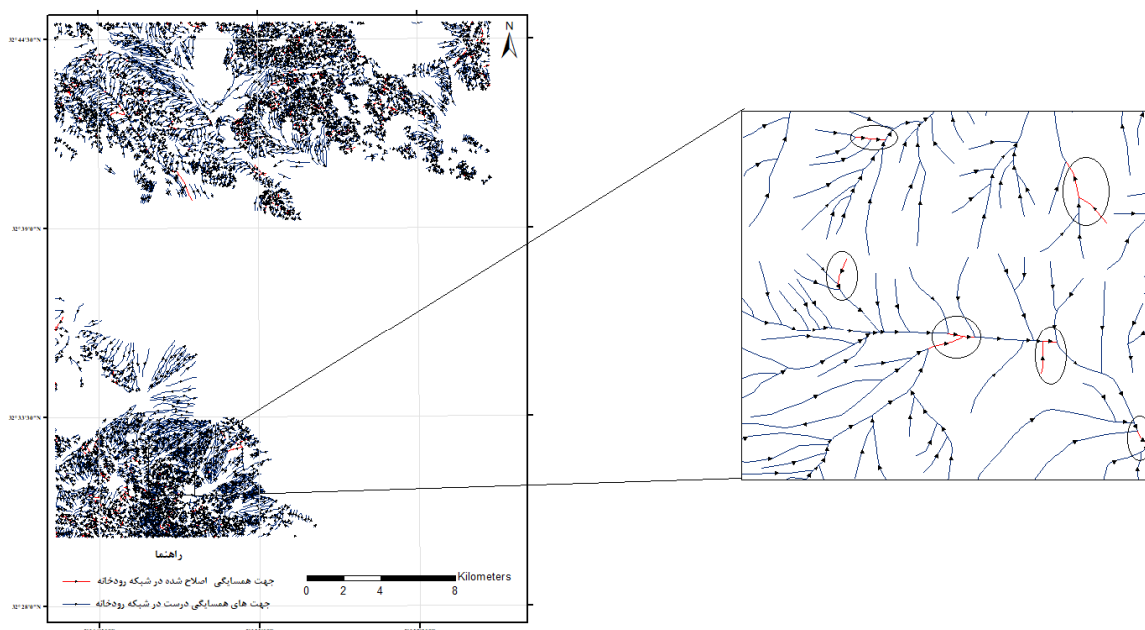
در این بخش، به بررسی خروجی‌های الگوریتم Horton-Strahler و الگوریتم Stroke به ترتیب در بخش‌های ۳-۲ و ۱-۲ پرداخته شده است.

٣-٢-١- نتائج الگوریتم Horton-Strahler

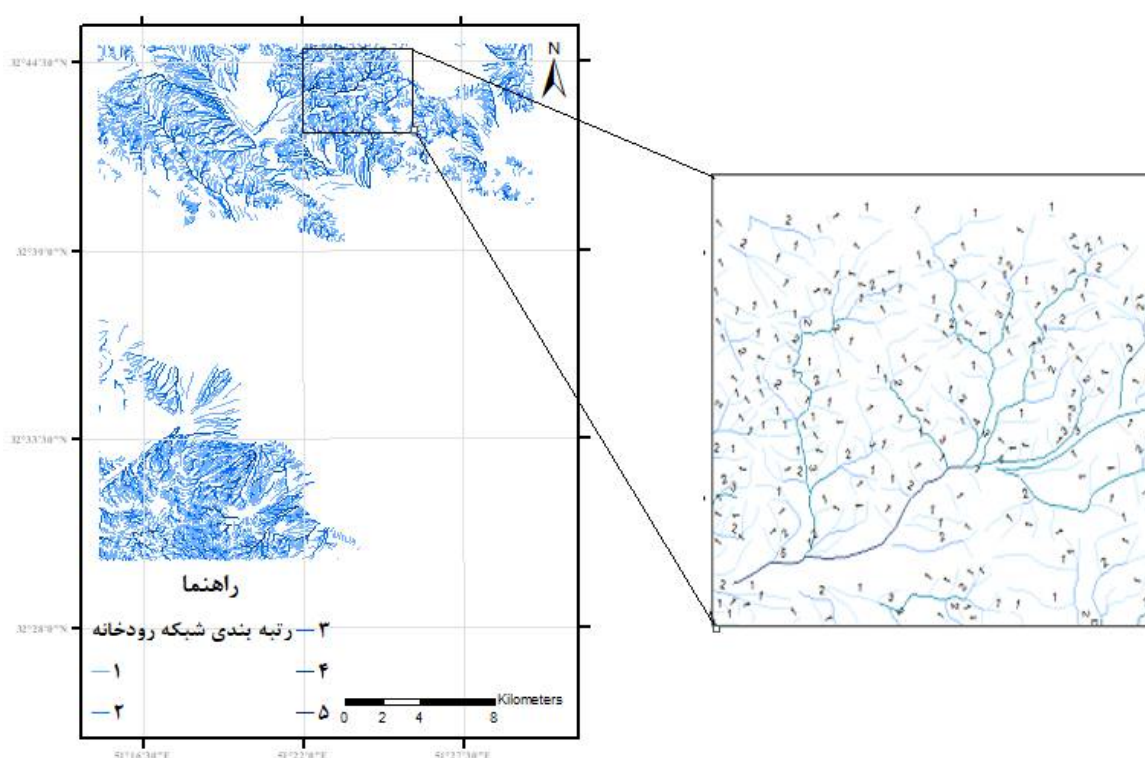
پیش از اجرای مدل ساختار سلسله‌مراتبی، لازم است خطاهای توپولوژیکی مانند رشدگی و نرسیدگی شاخه‌ها شناسایی و اصلاح شوند. پس از رفع این خطاها، جهت جریان بررسی و اصلاح گردید. بدین منظور، جهت جریان شاخه‌ها بر اساس اختلاف ارتفاع و ویژگی‌های هندسی آن‌ها



شکل ۱۲- خطا جهت جریان در همسایگی شاخه‌ها



شکل ۱۳- شبکه رودخانه پس از اصلاح خطا همسایگی



شکل ۱۴- نتیجه الگوریتم Horton-strahler

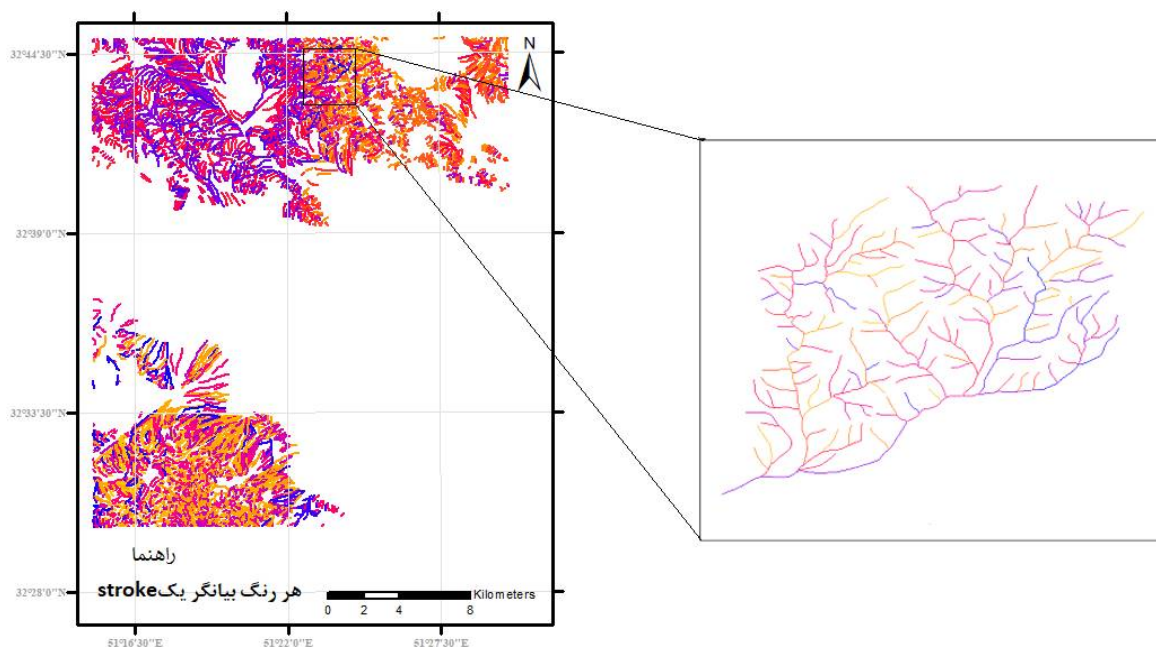
۳-۲-۲- نتایج الگوریتم stroke

در این مرحله، بخش‌های خطی شبکه رودخانه که تغییر جهت کمی دارند و زاویه بین آن‌ها تقریباً ۱۸۰ درجه

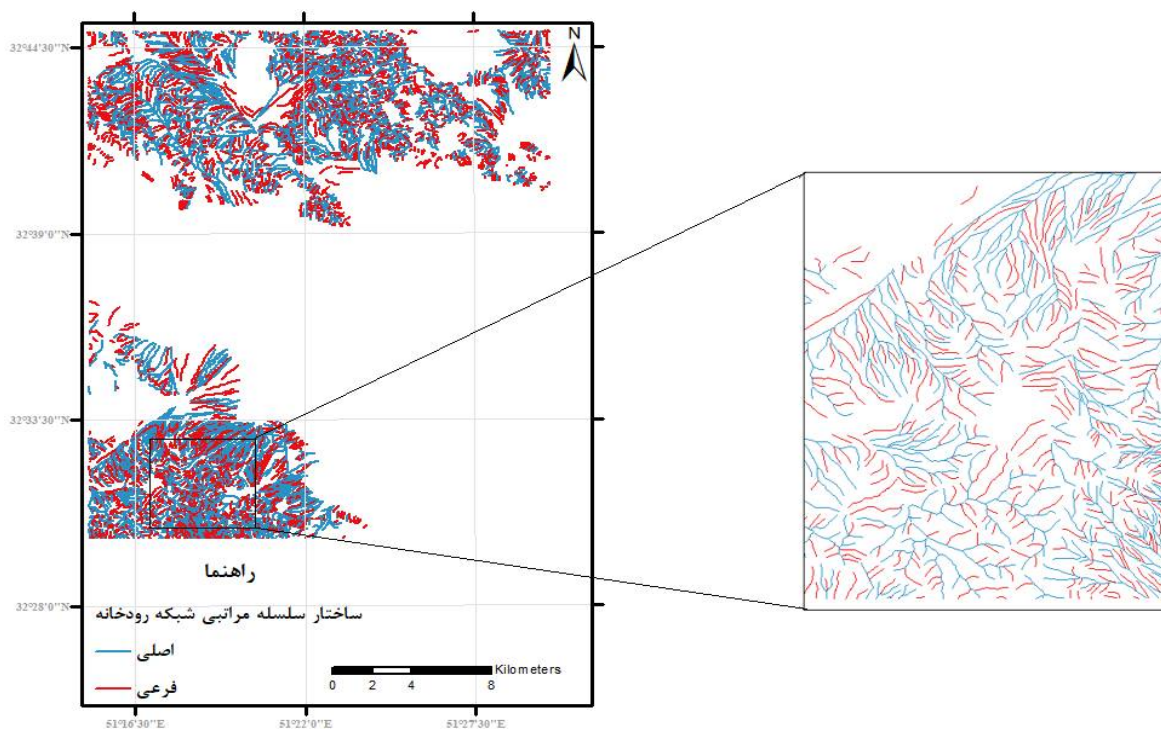
است، شناسایی شدند. این شاخه‌ها به صورت تقریباً مستقیم جریان آب را دنبال می‌کنند و با حفظ جهت پیوسته خود، در یک Stroke ترکیب می‌شوند. به عبارت دیگر، یک Stroke شامل مجموعه‌ای از شاخه‌های رودخانه

است که تغییر جهت اندکی دارند و پیوستگی طولی آن‌ها حفظ شده است. در نتیجه، ۴۰۳۹ Stroke شناسایی و ایجاد شد (شکل ۱۵). هر استروک شامل یک یا چند بخش خطی متوالی از شبکه است. پس از شناسایی Stroke، هر Stroke بر اساس روش Horton-Strahler و تعداد اتصالات ساختار

سلسله‌مراتبی ایجاد شد که هم ارتباطات توپولوژیکی و هم ویژگی‌های هندسی شبکه رودخانه را در نظر بگیرد. از مجموع استروک‌ها، ۱۶۷۹ استروک شاخه اصلی و ۲۳۶۰ استروک شاخه فرعی تشکیل شد (شکل ۱۶).



شکل ۱۵- الگوریتم stroke



شکل ۱۶- تحلیل ترکیبی توپولوژیکی و هندسی شبکه رودخانه

۳-۳- نتایج عملکرد هندسی شبکه رودخانه

در این بخش، فرآیند انتخاب و حذف شاخه‌های رودخانه با بهره‌گیری از الگوریتم GraphSAGE و تحلیل همسایگان هر شاخه مورد بررسی قرار گرفته است. برای استفاده از روش‌های یادگیری عمیق، شبکه رودخانه ابتدا به ساختار گراف دوگان تبدیل می‌شود. در این گراف، هر شاخه رودخانه دارای یک شناسه‌ی منحصر به فرد است و این شناسه‌ها به گره‌های متناظر در گراف نگاشت می‌شوند. در مجموعه داده مورد استفاده، تعداد کل گره‌ها ۶۸۶۵ گره و تعداد یال‌ها ۹۰۰۲ یال است که این مقادیر نشان دهنده تعداد شاخه‌ها و ارتباطات بین آنها در شبکه رودخانه می‌باشد.

۳-۳-۱- نتایج استخراج ویژگی‌ها و ساخت پایگاه داده نمونه

در این مرحله، ویژگی‌های هندسی، هیدرولوژیکی و توپولوژیکی شبکه رودخانه به گره‌های گراف اختصاص داده شد. سپس داده‌ها در دو مقیاس ورودی و هدف مقایسه شده و به هر نمونه بر اساس حضور در هر دو مقیاس، برچسب ۱ یا ۰ اختصاص یافت. در نهایت، توزیع برچسب به گونه‌ای است که حدود ۶۵ درصد داده‌ها (۴۴۶۲ نمونه) دارای برچسب ۱ و حدود ۳۵ درصد داده‌ها (۲۴۰۳ نمونه) دارای برچسب ۰ هستند.

برای آموزش مدل، داده‌ها به سه مجموعه اصلی تقسیم شدند: ۷۰ درصد از داده‌ها (۴۸۰۵ نمونه) به عنوان داده‌های آموزشی، ۲۰ درصد (۱۳۷۳ نمونه) به عنوان داده‌های تست و ۱۰ درصد (۶۸۶ نمونه) به عنوان داده‌های اعتبارسنجی تخصیص یافتند.

۳-۳-۲- نتایج مدل انتخاب

در این پژوهش، از یک مدل ۴ لایه GraphSAGE شامل سه لایه کانولوشنی، یک لایه خطی طبقه‌بندی استفاده شد. لایه ورودی مدل شامل ۱۳ ویژگی هندسی و هیدرولوژیکی، لایه تعبیه دارای ۲۵۶ بُعد و لایه کاملاً متصل دارای ۲۵۶ ورودی و ۲ خروجی می‌باشد. برای افزایش پایداری و جلوگیری از بیش‌برازش، در هر لایه از Dropout با نرخ ۰/۳ و Layer Normalization استفاده شد. توابع فعال‌سازی به کار رفته ReLU بوده و نرخ یادگیری برابر با ۰/۰۰۳ تنظیم شده است.

شرط توقف آموزش مدل در این پژوهش بر اساس مکانیزم Early Stopping با مقدار patience برابر ۳۵ تعریف شد. مدل حداکثر برای ۵۰۰ دوره آموزش داده شد. در طول آموزش، روند یادگیری با بررسی تغییرات تابع زیان و دقت در هر دوره مورد ارزیابی قرار گرفت. لازم به ذکر است، با توجه به نامتوازن بودن داده‌ها و تعداد بیشتر نمونه‌های کلاس ۱ نسبت به کلاس ۰، در فرآیند آموزش از وزن‌دهی به کلاس‌ها استفاده شده است.

برای ارزیابی عملکرد مدل GraphSAGE، از چندین معیار کلیدی استفاده شد که هرکدام جنبه‌های مختلف دقت و کارایی مدل را مورد بررسی قرار می‌دهند. این معیارها شامل Accuracy، Precision، Recall و F1-Score هستند. Accuracy نشان‌دهنده درصد پیش‌بینی‌های صحیح مدل از مجموع پیش‌بینی‌ها است و به طور کلی میزان دقت مدل را ارزیابی می‌کند. Precision درصد پیش‌بینی‌های صحیح برای کلاس مثبت (برچسب ۱) از تمام پیش‌بینی‌های مثبت مدل است که به ما می‌گوید مدل چقدر در پیش‌بینی درست کلاس‌های مثبت دقیق بوده است. Recall نیز درصد نمونه‌های واقعی مثبت که به درستی شناسایی شده‌اند را نشان می‌دهد و نشان‌دهنده توانایی مدل در شناسایی تمام نمونه‌های مثبت است. در نهایت، F1-Score میانگین هندسی Precision و Recall است که به طور متوازن به هر دو معیار توجه دارد و به‌ویژه در مواقعی که داده‌ها نامتعادل هستند، می‌تواند معیار بهتری برای ارزیابی باشد. این معیارها به منظور ارزیابی کیفیت پیش‌بینی‌های مدل در داده‌های آموزش، ارزیابی و تست محاسبه شدند و نتایج دقیق ارزیابی مدل بر اساس این معیارها در جدول ۱ نمایش داده می‌شود.

جدول ۱- آمار نتایج مدل روی مجموعه داده‌ها

شاخص ارزیابی	آموزش	ارزیابی	تست
Accuracy	۸۳/۵۵	۷۷/۹۵	۸۲/۵۸
Precision (class1)	۸۸/۲۷	۸۶/۲۸	۸۷/۱۲
Precision (class0)	۷۵/۴۶	۶۳/۲۶	۷۳/۷۸
Recall (class1)	۸۸/۲۷	۸۶/۲۸	۸۷/۱۲
Recall (class0)	۷۸/۹۳	۷۲/۳۴	۷۴/۷۳
F1-Score (class1)	۸۷/۱۵	۸۳/۳۱	۸۶/۸۳
F1-Score (class0)	۷۷/۱۶	۶۷/۴۹	۷۴/۲۶

پس از آموزش مدل GraphSAGE، پیش‌بینی‌ها روی تمام گره‌های گراف انجام شد (شکل ۱۷).

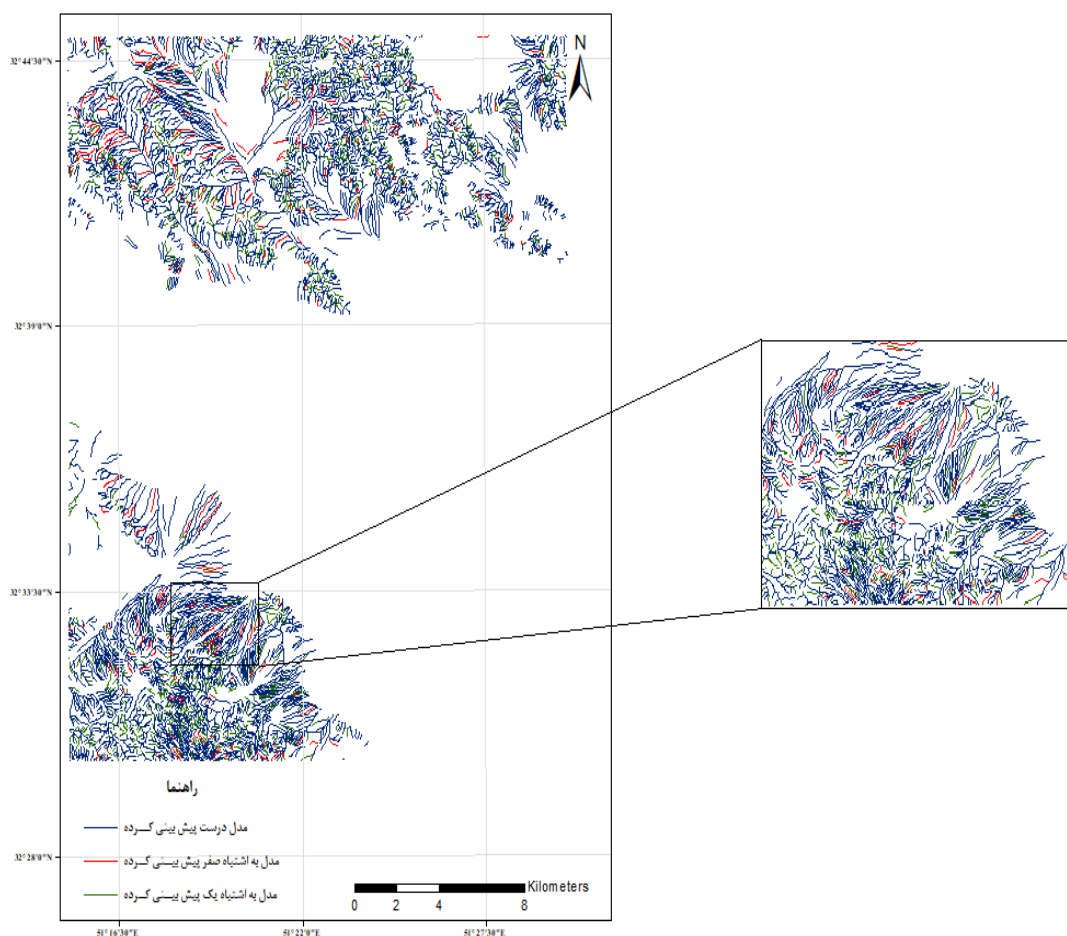
جدول ۲- نتایج مدل یادگیری عمیق

شاخص ارزیابی	آموزش	ارزیابی	تست
Accuracy	۹۷/۷۲	۹۴/۹۲	۹۶/۸۲
Precision	۹۸/۶۱	۹۵/۲۹	۹۷/۲۵
Recall	۹۷/۸۳	۹۷/۲۲	۹۷/۹۰
F1-Score	۹۸/۲۲	۹۶/۲۵	۹۷/۵۷

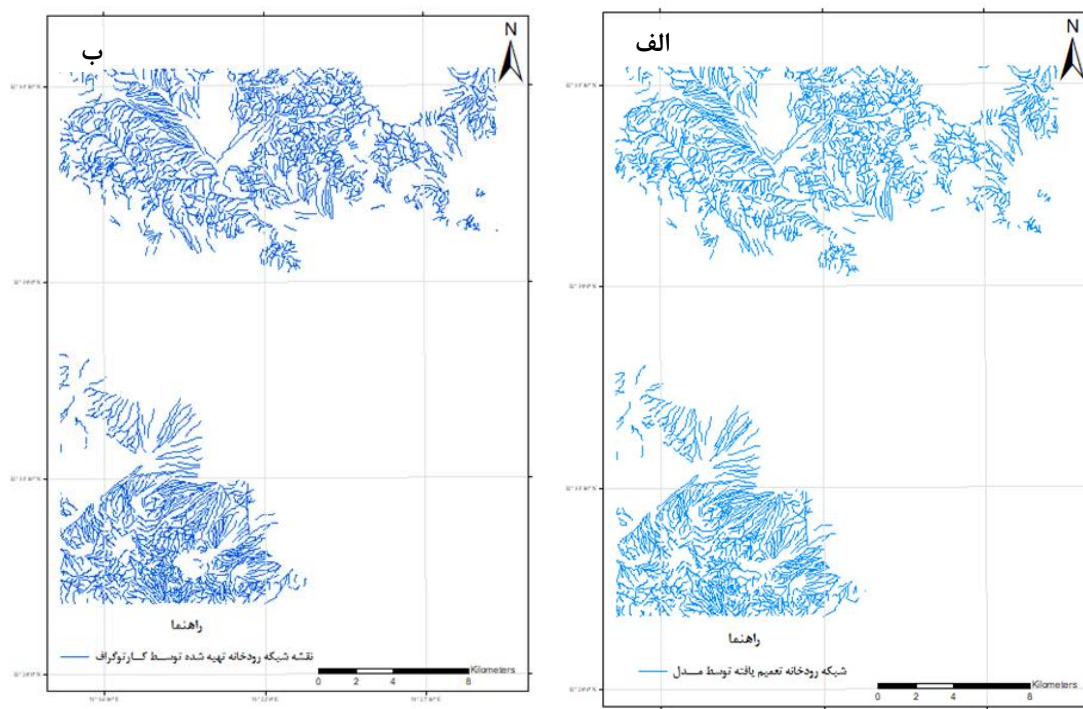
براساس نتایج حاصل شده، مدل ارائه شده در این پژوهش کلاس ۵۵۸۲ گره را به درستی پیش‌بینی کرده (با رنگ آبی نمایش داده شده‌اند)، در حالی که گره False Positives (گره‌هایی که در واقعیت برچسب صفر داشتند ولی مدل برای آن‌ها برچسب ۱ پیش‌بینی کرده) با رنگ سبز و گره False Negatives (گره‌هایی که در واقعیت برچسب یک

داشتند ولی مدل برای آن‌ها برچسب صفر پیش‌بینی کرده) با رنگ قرمز مشخص شده‌اند. نتایج خروجی مدل در جدول ۲ ارائه شده است. نقشه نهایی تولید شده توسط مدل ارائه شده در این تحقیق در مقیاس ۱:۵۰۰۰۰ در شکل ۱۸ (الف) و نقشه موجود تهیه شده توسط کارتوگراف در مقیاس ۱:۵۰۰۰۰ در شکل ۱۸ (ب) ارائه شده است.

لازم به تأکید است که عملکرد مدل پیشنهادی به کیفیت توپولوژی و صحت اطلاعات هیدرولوژیکی شبکه رودخانه وابسته است. در این پژوهش، پیش از اجرای مدل، خطاهای رایج داده‌های مکانی از جمله جهت جریان نادرست، ناسازگاری‌های همسایگی در محل انشعابات و خطاهای ارتفاعی شاخه‌ها شناسایی و اصلاح شده‌اند. بررسی داده‌ها نشان داد که حدود ۱۵ درصد از شاخه‌های شبکه دارای خطای جهت جریان بوده‌اند که پیش از فرآیند یادگیری اصلاح گردیدند.



شکل ۱۷- خروجی نتایج مدل GraphSAGE



شکل ۱۸- خروجی نقشه تعمیم یافته توسط مدل (الف) خروجی نقشه تهیه شده توسط کارتوگراف (ب)

۳-۴- نتایج ارزیابی تعمیم و ساده سازی شبکه رودخانه

ارزیابی تعمیم و ساده سازی شبکه رودخانه فرآیندی است که طی آن میزان انطباق نتایج حاصل از مدل پیشنهادی با اهداف کارتوگرافی و نقشه مرجع سنجیده می شود. در این پژوهش، ارزیابی با هدف بررسی میزان حفظ ساختار شبکه رودخانه، تطابق با تعمیم دستی کارتوگراف و تحلیل رفتار مدل در شرایط مختلف انجام شده است. بر این اساس، ارزیابی در سه محور مکمل شامل ارزیابی انطباق هندسی و تحلیل کیفی خطا و موارد شکست مدل صورت گرفت.

۳-۴-۱- ارزیابی انطباق هندسی شبکه رودخانه

به منظور سنجش میزان انطباق شبکه رودخانه تعمیم یافته با نقشه مرجع، از شاخص انطباق هندسی CLC استفاده شد. این شاخص میزان همپوشانی طولی بین شبکه تولید شده توسط مدل و شبکه تعمیم یافته مرجع را نشان می دهد. نتایج ارزیابی نشان داد که مقدار CLC برابر با ۰/۹۶ است. این مقدار بالا بیانگر آن است که مدل GraphSAGE توانسته است بیش از ۹۶ درصد طول شبکه رودخانه ای را با نقشه مرجع همپوشانی دهد و ویژگی های اصلی شبکه

رودخانه ای، از جمله ساختار کلی، پیوستگی شاخه ها و الگوی توزیع فضایی، به خوبی حفظ شده اند. این نتیجه نشان می دهد که مدل پیشنهادی در بازتولید تصمیم های تعمیم دستی کارتوگراف، به ویژه در حذف شاخه های کم اهمیت و حفظ شاخه های اصلی، عملکرد مناسبی داشته و توانسته است شبکه ای ساده تر اما منسجم و خوانا در مقیاس هدف تولید کند.

۳-۴-۲- تحلیل کیفی خطا و موارد شکست مدل

اگرچه معیارهای کلی ارزیابی بیانگر عملکرد مناسب مدل پیشنهادی هستند، بررسی نمونه های دارای خطا نشان می دهد که برخی الگوهای خاص منجر به بروز خطا در پیش بینی مدل می شوند. تحلیل شاخه های دارای خطای False Positive و False Negative نشان می دهد که بخش عمده ای از خطاها مربوط به شاخه های کوتاه و کم طول واقع در نواحی با تراکم بالای انشعاب است؛ به ویژه شاخه هایی که از نظر هندسی کم اهمیت اند اما در ساختار توپولوژیکی شبکه نقش اتصال دهنده دارند.

در برخی موارد، مدل شاخه هایی را که در شبکه های متراکم بالادست قرار دارند و از نظر ساختاری اهمیت نسبی دارند، به اشتباه حذف کرده است (False Negative). از

سوی دیگر، برخی خطاهای False Positive مربوط به شاخه‌هایی است که به دلیل پیوستگی هندسی بالا یا طول نسبی بیشتر توسط مدل حفظ شده‌اند، در حالی که در تعمیم دستی کار توگرافی حذف شده‌اند. این موارد اغلب در شبکه‌هایی با پیچیدگی هندسی بالا و انحناهای پیوسته مشاهده می‌شوند.

لازم به تأکید است که ارزیابی انجام‌شده در این پژوهش به‌صورت هدفمند به یک شبکه رودخانه‌ای و یک حوضه آبریز مشخص محدود شده است. شبکه‌های رودخانه‌ای از نظر الگوی زهکشی، میزان انشعاب، نظم سلسله‌مراتبی و تأثیر عوامل ژئومورفولوژیکی می‌توانند در حوضه‌های مختلف تفاوت‌های اساسی داشته باشند. از این‌رو، نتایج ارائه‌شده در این پژوهش نباید به‌عنوان شواهدی برای تعمیم‌پذیری آماری مدل به تمامی حوضه‌های آبریز تلقی شود. با این حال، چارچوب پیشنهادی به‌گونه‌ای طراحی شده است که از نظر مفهومی قابلیت به‌کارگیری در حوضه‌های مختلف را دارد، مشروط بر آنکه مدل با داده مرجع متناظر با هر حوضه و با در نظر گرفتن ویژگی‌های ژئومورفولوژیکی خاص آن آموزش داده شود.

اگرچه هدف کلی پژوهش، تعمیم شبکه رودخانه‌ها برای تولید نقشه در مقیاس‌های کوچک‌تر عنوان شده است، ارزیابی مدل پیشنهادی در این مطالعه به‌صورت هدفمند به یک محدوده مقیاسی مشخص (۱:۲۵,۰۰۰ به ۱:۵۰,۰۰۰) محدود شده است. فرآیند تعمیم شبکه رودخانه‌ها ذاتاً وابسته به مقیاس هدف است و تغییر مقیاس منجر به تغییر قواعد حذف-انتخاب، آستانه‌های هندسی، میزان تجمع شاخه‌ها و حتی اهمیت نسبی ویژگی‌های توپولوژیکی و هندسی می‌شود. از این‌رو، به منظور حفظ پایدار ساختار سلسله‌مراتبی و ادراکی شبکه رودخانه توسط یک مدل آموزش‌دیده در یک مقیاس مشخص، در مقیاس‌های کوچک‌تر نیاز به بازآموزی یا بازتعریف ویژگی‌ها دارد.

۴- نتیجه‌گیری و پیشنهادات

هدف از انجام این تحقیق مدل‌سازی تعمیم شبکه رودخانه‌ها بود. روش توسعه داده شده در این تحقیق شامل ساختار سلسله‌مراتبی شبکه رودخانه و فرآیند اعمال عملگر هندسی می‌باشد. در مدل اول به منظور رفع چالش تشخیص دقیق جریان اصلی و شاخه‌های فرعی، از ترکیب

دو الگوریتم Horton-Strahler و Stroke استفاده گردید. الگوریتم Horton-Strahler با تمرکز بر روابط توپولوژیکی شبکه، شاخه‌ها را بر اساس جریان رتبه‌بندی کرده و امکان تعیین سلسله‌مراتب طبیعی و شناسایی شاخه‌های اصلی و فرعی را فراهم ساخت. در کنار این رویکرد توپولوژیکی، الگوریتم Stroke برای تحلیل هندسی شبکه به کار گرفته شد. این الگوریتم با محاسبه طول شاخه‌ها و زاویه‌های بین آن‌ها در محل تلاقی‌ها، پیوستگی و هم‌راستایی شاخه‌ها را بررسی کرد و شاخه‌های پایدارتر و هم‌راستا را به عنوان مسیر اصلی انتخاب نمود. ترکیب این دو الگوریتم باعث شد که ساختار سلسله‌مراتبی ایجاد شود که هم ارتباطات توپولوژیکی و هم ویژگی‌های هندسی شبکه رودخانه را در نظر بگیرد و دقت در تشخیص شاخه‌های اصلی و فرعی را به شکل قابل توجهی افزایش دهد.

در مدل دوم، به منظور حل چالش اعمال عملگر هندسی برای انتخاب و حذف خودکار شاخه‌ها، رویکردی ترکیبی مبتنی بر شبکه‌های عصبی گرافی و به ویژه الگوریتم GraphSAGE همراه با فرآیند اصلاح مبتنی بر همسایگی ارائه شد. در این رویکرد در مرحله اول، شاخه‌های رودخانه به‌عنوان واحد پایه در نظر گرفته شده و ویژگی‌های ساختاری، هندسی، مورفولوژیک و توپولوژیک هر شاخه به‌عنوان ویژگی‌های گراف استخراج می‌شوند. روابط توپولوژیکی بین شاخه‌ها نیز در یک ساختار گراف نمایش داده می‌شود. گراف به همراه برجسب‌ها و ویژگی‌های گره‌ها به‌عنوان داده ورودی مدل طبقه‌بندی استفاده می‌شود. سپس از طریق آموزش نظارت‌شده، روش‌های نمونه‌برداری و تجمع به کار گرفته شدند تا ویژگی‌های پیچیده‌تری از شبکه استخراج شود. مدل GraphSAGE برای طبقه‌بندی شاخه‌های رودخانه و انجام فرایند انتخاب و حذف در نقشه‌های بزرگ‌مقیاس شبکه رودخانه به کار گرفته شد. در مرحله دوم، برای افزایش دقت مدل، فرآیند اصلاح مبتنی بر همسایگی اعمال گردید. نتایج نشان داد که با استفاده از نقشه‌های مقیاس ۱:۲۵,۰۰۰، دقت مدل به ۹۶ درصد رسید.

در این پژوهش تلاش شد تا با ترکیب تحلیل توپولوژیکی و هندسی شاخه‌ها و استفاده از الگوریتم‌های پیشرفته یادگیری عمیق، فرآیند انتخاب و حذف شاخه‌ها مدل‌سازی شود. با این حال، برخی کاستی‌ها و زمینه‌های بهبود در این مدل‌ها شناسایی شد که می‌تواند موضوع بررسی‌های آتی قرار گیرد.

بررسی عملکرد مدل پیشنهادی در حوضه‌های آبریز با الگوهای زهکشی متفاوت و تحلیل تأثیر ویژگی‌های ژئومورفولوژیکی بر دقت مدل می‌تواند به‌عنوان مسیر تحقیقاتی آتی مدنظر قرار گیرد. همچنین، توسعه مدل پیشنهادی در یک چارچوب چندمقیاسی و ارزیابی عملکرد آن در محدوده‌های مقیاسی مختلف (مانند ۱:۵۰,۰۰۰ به ۱:۱۰۰,۰۰۰ و ۱:۲۵۰,۰۰۰) با بازتعریف ویژگی‌ها و داده مرجع متناظر، می‌تواند به‌عنوان مسیر پژوهشی آتی مورد توجه قرار گیرد. بعلاوه، با توجه به اینکه معماری‌های مختلف شبکه‌های عصبی گرافی، از جمله GCN و GraphSAGE، هر یک دارای مزایا و محدودیت‌های خاص خود هستند، مقایسه

مراجع

تجربی عملکرد این معماری‌ها در مسئله تعمیم شبکه‌های رودخانه‌ای می‌تواند به‌عنوان یک مسیر پژوهشی آتی مدنظر قرار گیرد. به‌ویژه بررسی تفاوت عملکرد مدل‌ها از نظر دقت، پایداری و هزینه محاسباتی در شبکه‌های بزرگ می‌تواند به انتخاب بهینه معماری در کاربردهای عملی کمک نماید.

سپاسگزاری

این تحقیق با حمایت دانشگاه صنعتی خواجه نصیرالدین طوسی (بنیاد علم ایران (INSF) با کد گرنت ۴۰۴۰۹۵۰۱ انجام شده است.

- [۱] B. "Constraint-Based Spatial Data Management for Cartographic Representation at Different Scales," *Surveying and Geoinformatics Engineering*, 2022.
- [۲] Yan, "Deep learning in automatic map generalization: achievements and challenges," *Geo-spatial Information Science*, 2025.
- [۳] D. "Scale matters: a survey of the concepts of scale used in spatial disciplines," *European Journal of Remote Sensing*, 2019.
- [۴] Renard, "Urban structure generalization in multi-agent process by use of reactional agents," *Transactions in GIS*, 2014.
- [۵] Wang, "An intelligent simplification method for river networks with an unsupervised variational autoencoder," *International Journal of Digital Earth*, 2025.
- [۶] yan, "Description and Generalization of River Networks," *In Description Approaches and Automated Generalization Algorithms for Groups of Map Objects*, 2019.
- [۷] Wu, "Automatic recognition method of primary path for an anastomosing river based on its typical features," *Transactions in GIS*, 2021.
- [۸] Yan, "A Comparison of Machine Learning Methods Applied to the Automated Selection of River Networks," *The Cartographic Journal*, 2022.
- [۹] Xu, "Drainage pattern recognition of river network based on graph convolutional neural network," *ISPRS International Journal of Geo-Information*, 2023.
- [۱۰] Ai, "The hierarchical watershed partitioning and data simplification of river network," in *In Progress in Spatial Data Handling: 12th International Symposium on Spatial Data Handling*, 2006.
- [۱۱] Zhang, "Evaluation of River Network Generalization Methods for preserving the drainage pattern," *ISPRS International Journal of Geo-Information*, 2016.
- [۱۲] Touya, "River Network Selection based on Structure and Pattern," *International Cartographic Conference*, 2007.
- [۱۳] Wang, "Graph neural network method for the intelligent selection of river system," *Geocarto International*, 2023.
- [۱۴] Dodds, "Geometry of river networks," *Physical Review*, 2000.
- [۱۵] Mantilla, "Extending generalized Horton laws to test embedding algorithms for topologic," *Geomorphology*, 2012.
- [۱۶] L. "Hierarchical elimination selection method of dendritic river network generalization," *Plos one*, 2018.

- [۱۷] Cheng, "An improved integrated generalization method for contours and rivers using importance sequence of all points from these two elements," *Geocarto International*, 2022.
- [۱۸] He, Principle and method of map data processing model, Wuhan University Press Wuhan, China, 2004.
- [۱۹] A. "GAP-tree extensions based on skeletons," in *10th International Symposium on Spatial Data Handling*, 2002.
- [۲۰] Qiu, "River network dynamic selection with spatial data and attribute data based on rough set," in *International Conference on Geoinformatics*, 2010.
- [۲۱] Sen, "An experimental approach for selection/elimination in stream network generalization using support vector machines," *Geocarto International*, 2015.
- [۲۲] Jiang, "THE GENERALIZATION METHOD RESEARCH OF RIVER NETWORK BASED ON MORPH STRUCTURE AND CATCHMENTS' CHARACTER KNOWLEDGE," in *ISPRS Archives*, 2008.
- [۲۳] Thomson, "The Stroke Concept in Geographic Network Generalization and Analysis," in *In Progress in Spatial Data Handling: 12th International Symposium on Spatial Data Handling*, 2006.
- [۲۴] Liu, "Road selection based on Voronoi diagrams and "strokes" in map generalization," *International Journal of Applied Earth Observation and Geoinformation*, 2010.
- [۲۵] Franceschi, "Learning Discrete Structures for Graph Neural Networks," 2020.
- [۲۶] Zhu, "DII-GCN: Dropedge Based Deep Graph Convolutional Networks," *Symmetry*, 2022.
- [۲۷] Z. "A Genetic Algorithm for Tributary Selection with Consideration of Multiple Factors," *Transactions in GIS*, 2016.
- [۲۸] Yu, "A recognition method for drainage patterns using a graph convolutional network," *International Journal of Applied Earth Observation and Geoinformation*, 2022.
- [۲۹] Hamilton, "Inductive Representation Learning on Large Graphs," in *Conference on Neural Information Processing Systems*, 2017.
- [۳۰] Yu, "Integrating domain knowledge and graph convolutional neural networks to support river network selection," *Transactions in GIS*, 2023.
- [۳۱] Duarte, "How do loss functions impact the performance of graph neural networks?," in *Brazilian Conference on Intelligent Systems*, 2021.
- [۳۲] Abbas, "Evaluating Loss Functions for Graph Neural Networks: Towards Pretraining and Generalization," *ACM Computing Surveys*, 2025.
- [۳۳] Izadi, "Optimization of Graph Neural Networks with Natural Gradient Descent," in *Electrical Engineering, Applied and Computational Mathematics and Statistics University of Notre Dame*, 2020.
- [۳۴] Battenfield, "Generalization of Hydrographic Features and Automated Metric Assessment Through Bootstrapping," in *13th Workshop of the International Cartographic Association commission on Generalisation and Multiple Representation.*, 2010.
- [۳۵] Li, "Selection Method of Dendritic River Networks Based on Hybrid Coding for Topographic Map Generalization," *SPRS International Journal of Geo-Information*, 2020.
- [۳۶] D. "Optimization of river network representation data models for web-based systems," *Earth and Space Science*, 2017.