مدل‌سازی داده‌محور عوامل مؤثر بر واریانس سرعت در راه‌های برون‌شهری با بهره‌گیری از رویکرد یادگیری ماشین (مطالعه موردی: استان همدان)

مقاله علمی-پژوهشی

*علی توکلی کاشانی (نویسنده مسئول)، دانشیار، دانشکده مهندسی عمران و مرکز تحقیقات ایمنی کاربردی حمل‌ونقل جاده‌ای، دانشگاه علم و صنعت ایران، تهران، ایران
محمد نجف پور درو، دانشجوی دکتری، دانشکده مهندسی عمران و مرکز تحقیقات ایمنی کاربردی حمل‌ونقل جاده‌ای، دانشگاه علم و صنعت ایران، تهران، ایران
امیررضا اسکندری، دانشجوی کارشناسی ارشد، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران
مجید نجفی بقرآباد، دانشجوی کارشناسی ارشد، دانشکده مهندسی عمران، دانشگاه علم و صنعت ایران، تهران، ایران
*پست الکترونیکی نویسنده مسئول: alitavakoli@iust.ac.ir

دریافت: ۱۴۰۴/۰۵/۲۸- پذیرش: ۱۴۰۴/۱۱/۰۲
صفحه ۳۲۸-۳۱۵

چکیده

واریانس سرعت به‌عنوان شاخصی کلیدی در تحلیل پایداری جریان ترافیک و ارزیابی ایمنی جاده‌ای شناخته می‌شود؛ چراکه نوسانات سرعت میان وسایل نقلیه، یکی از زمینه‌های اصلی افزایش احتمال بروز تصادف است. این پژوهش با هدف شناسایی و تبیین عوامل مؤثر بر واریانس سرعت در معابر برون شهری ایران و با تکیه بر روش‌های نوین یادگیری ماشین، به‌ویژه مدل جنگل تصادفی انجام شده است. دلیل انتخاب این مدل، توانایی بالای آن در مدیریت داده‌های غیرخطی، تشخیص الگوهای پیچیده و کاهش خطر بیش‌برازش است. داده‌های موردنیاز از ۲۰۰ نقطه منتخب از شبکه راه‌های برون‌شهری استان همدان از طریق پیمایش میدانی گردآوری شد و مجموعه‌ای از متغیرهای ترافیکی، زیرساختی و محیطی به‌عنوان ورودی مدل مورد استفاده قرار گرفت. یافته‌ها نشان دادند که در صد تخطی از سرعت مجاز مهم‌ترین عامل افزایش واریانس سرعت است و رابطه‌ای مستقیم با نوسانات جریان ترافیک دارد. همچنین، میانگین سرعت وسایل نقلیه تأثیری غیرخطی بر واریانس داشته و در محدوده‌های سرعت بالا، تغییرات سرعت به صورت نمایی افزایش می‌یابد. از سوی دیگر، افزایش سهم وسایل نقلیه سنگین تا حدود ۳۰ تا ۴۰ در صد موجب تشدید نوسانات سرعت شده و پس از این حد، با تثبیت جریان، این نوسانات کاهش می‌یابد. در بین عوامل زیرساختی، افزایش تعداد خطوط عبوری اثر مثبت و وجود سرعت‌کاه‌ها اثر منفی معناداری بر واریانس سرعت داشته‌اند. در حوزه عوامل محیطی نیز، تراکم جمعیتی پیرامون راه و کاربری کشاورزی موجب کاهش واریانس سرعت شده‌اند، در حالی‌که افزایش حجم ترافیک ساعتی با یکنواخت‌تر شدن جریان حرکت، نوسانات سرعت را کاهش می‌دهد. بر اساس نتایج، بازنگری در محدودیت‌های سرعت، کنترل رفتار رانندگان در مقاطع پرتردد و استفاده هدفمند از سرعت‌کاه‌ها می‌تواند به‌طور مؤثری به کاهش واریانس سرعت و ارتقای سطح ایمنی جاده‌ها منجر شود. یافته‌های این پژوهش می‌تواند به‌عنوان مبنایی برای سیاست‌گذاری‌های مبتنی بر داده در حوزه ایمنی راه‌های برون‌شهری ایران مورد استفاده قرار گیرد.

واژه‌های کلیدی: واریانس سرعت، ایمنی جاده‌ای، جنگل تصادفی، یادگیری ماشین، تحلیل ترافیک، تصادفات جاده‌ای

۱-مقدمه

تصادفات جاده‌ای یکی از معضلات اساسی جوامع معاصر به‌ویژه در کشورهای درحال‌توسعه است و همچنان به‌عنوان یکی از مهم‌ترین چالش‌های سلامت عمومی و توسعه پایدار شناخته می شود. بر ا ساس گزارش سازمان بهدا شت جهانی، سوانح ترافیکی دوازدهمین عامل مرگ‌ومیر در جهان و اصلی‌ترین علت مرگ جوانان زیر ۲۹ سال هستند (Word

Health Organization(WHO), 2023). ایران نیز از این قاعده مستثنا نبوده و طی سال‌های اخیر با روند افزایشی تصادفات و تلفات مواجه بوده است. مطابق آمار رسمی، در سال ۱۴۰۲ حدود ۲۰،۰۴۵ نفر در اثر تصادفات جاده‌ای جان خود را از دست داده‌اند که نسبت به سال پیش از آن حدود شش در صد افزایش را نشان می‌دهد (سازمان پزشکی قانونی کشور، ۱۴۰۲). نرخ مرگ‌ومیر ناشی از تصادفات در ایران به‌طور میانگین ۲۴,۳۸ نفر به ازای هر صد هزار نفر جمعیت برآورد شده است که فاصله قابل‌توجهی با میانگین جهانی (۱۵ نفر) دارد. این اختلاف بیانگر ضرورت بازنگری در سیاست‌های ایمنی ترافیک کشور و اجرای مداخلات مبتنی بر داده است.

تلفات ناشی از تصادفات جاده‌ای تنها پیامد انسانی ندارند، بلکه بار اقتصادی سنگینی را نیز به دولت‌ها و جوامع تحمیل می‌کنند. برآوردها نشان می‌دهد در کشورهای توسعه‌یافته، هزینه‌های ناشی از تصادفات حدود سه درصد از تولید ناخالص داخلی(GDP) را شامل می شود، در حالی‌که در کشورهای با درآمد متوسط یا پایین، از جمله ایران، این سهم ممکن است تا پنج درصد افزایش یابد (Lord et al., 2021). بنابراین، کاهش تصادفات جاده‌ای نه‌تنها یک هدف ایمنی، بلکه ضرورتی اقتصادی محسوب می‌شود که تحقق آن مستلزم شناسایی علمی عوامل مؤثر بر رفتار و پویایی جریان ترافیک است. در میان شاخص‌های مرتبط با ایمنی ترافیک، واریانس سرعت به‌عنوان یکی از پارامترهای کلیدی شناخته می‌شود. مطالعات متعدد نشان داده‌اند که افزایش پراکندگی سرعت در میان وسایل نقلیه، احتمال وقوع تصادف را به‌طور معناداری افزایش می‌دهد (Quddus, 2013; Gargoum & El-Basyouny, 2016). در حالی‌که میانگین سرعت بیش‌تر با شدت تصادفات مرتبط است، واریانس سرعت رابطه مستقیمی با احتمال وقوع آن دارد؛ بدین معنا که ناهمگنی در رفتار حرکتی وسایل نقلیه، احتمال برخورد را افزایش می‌دهد. از این رو، کاهش واریانس سرعت یکی از اهداف مهم در ارتقای سطح ایمنی راه‌ها به‌شمار می‌آید. با وجود پژوهش های متعدد در این زمینه، اغلب مطالعات پیشین مبتنی بر شبیه سازی‌های نرم‌افزاری یا داده‌های خارج از بسترهای بومی بوده‌اند که تعمیم‌پذیری نتایج آن‌ها برای شرایط خاص جاده‌های ایران را محدود می‌کند. از سوی دیگر، پیچیدگی روابط میان متغیرهای ترافیکی، زیرساختی و محیطی نیازمند استفاده از روش‌های تحلیلی پیشرفته‌ای است که بتوانند رفتارهای غیرخطی و تعاملات چندمتغیره را به‌درستی مدل کنند. در این راستا، پژوهش حاضر با هدف شناسایی و تحلیل عوامل مؤثر بر واریانس سرعت در معابر برون‌شهری ایران و با بهره‌گیری از مدل جنگل تصادفی به‌عنوان یکی از الگوریتم‌های قدرتمند یادگیری ماشین، طراحی شده است. در این تحقیق از داده‌های واقعی برداشت شده از سطح راه‌های کشور استفاده شده تا نتایج حاصل از آن دارای اعتبار میدانی و قابلیت کاربرد در سیاست‌گذاری‌های ملی ایمنی جاده‌ای باشد. یافته‌های پژوهش می‌توانند مبنایی برای بازنگری در محدودیت های سرعت، کنترل رفتار رانندگان و طراحی زیرساخت‌های ایمن‌تر در شبکه راه‌های برون‌شهری ایران فراهم آورند.

۲-پیشینه تحقیق

بررسی ادبیات پژوهش نشان می‌دهد که واریانس سرعت یکی از شاخص‌های کلیدی در تحلیل ایمنی ترافیک و پویایی جریان حرکتی است. مطالعات اولیه در این حوزه تأکید کرده‌اند که کاهش نوسانات سرعت میان وسایل نقلیه، یکی از مؤثرترین راهکار ها برای بهبود ایمنی و کاهش احتمال برخورد های ترافیکی است (Graves et al., 1993). به‌عنوان نمونه، گریوز و همکاران در توسعه مدل‌های تحلیلی خود نشان دادند که کاهش واریانس سرعت، علاوه بر بهینه‌سازی شرایط عملکرد شبکه، می‌تواند امکان افزایش محدودیت سرعت به‌صورت ایمن را فراهم آورد و اثرات خارجی تصادفات را کاهش دهد (Graves et al., 1993). در ادامه، وانگ و همکاران نیز بر اهمیت شناسایی متغیرهای مؤثر بر واریانس سرعت تأکید کردند و استدلال نمودند که مدیریت این متغیر نقش اساسی در سیاست‌گذاری‌های کلان ایمنی ترافیک دارد (Wang et al., 2018).

از نخستین مطالعات میدانی مرتبط، گاربر و گادیرائو بودند که نشان دادند واریانس سرعت با افزایش سرعت متوسط تا حدودی کاهش یافته و در بازه‌های خاصی مجدداً افزایش می یابد. آنان همچنین اختلاف زیاد بین سرعت طراحی و سرعت مجاز را از دلایل اصلی افزایش نوسانات سرعت معرفی

کرد ند (Garber & Gadirau, 1988). به‌طور مشـــا به، هو گزارش کرد که در مقاطع دارای سرعت متو سط بالا و نرخ تخطی زیاد از محدودیت ســـرعت، پراکندگی ســـرعت میان وسایل نقلیه بیشتر است (Hu, 2017).

در حوزه ترکیب ترافیکی، شـــن و یانگ با بررســـی اثر حضـــور وســـایل نقلیه ســـنگین دریافتند که واریانس ســـرعت در جاده با افزایش درصـــد خودروهای ســـنگین ابتدا افزایش یافته و در محدوده ۳۰ تا ۴۰ درصـــد به حداکثر می‌رســـد، اما پس از آن، به دلیل تثبیت جر یان حرکتی کاهش می یا بد (Shen & Yang, 2020). در مقابل، روه و همکاران نتیجه گرفتند که حضـــور وســـایل نقلیه ســـنگین می‌تواند به یکنواخت‌تر شـــدن جریان و کاهش تفاوت ســـرعت‌ها منجر شـــود، زیرا ســـایر رانندگان ناچار به تبعیت از ســـرعت گروهی مشـــابه می‌شـــوند (Roh et al., 2021). این نتایج نشـــان می‌دهد که اثر ترکیب ترافیک بر واریانس سرعت ممکن ا ست به شرایط محلی، نوع معبر و رفتار رانندگان وابسته باشد.

از منظر حجم ترافـ یک، هوا نگ رابـ طه‌ای معکوس بین حجم عبوری و وار یانس ســـر عت گزارش کرد، به‌گو نه‌ای که در ترافیک‌های ســـنگین‌تر، جریان حرکت یکنواخت‌تر و نوســـانات سرعت کمتر ا ست (Huang et al., 2013). در بـستر ایران، علیزاده و همکاران نـ شان دادند که وجود بافت‌های مـ سکونی و تراکم جمعیتی در حا شیه راه‌ها موجب کاهش واریانس سرعت می‌شـــود، زیرا حضـــور عابرین و دســـترســـی‌های متعدد باعث کاهش ســـرعت و احتیاط بیشـــتر رانندگان می‌گردد (Alizade et al., 2023). همچنین، کاربری‌های کشـــاورزی پیرامون راه نیز از طریق افزایش تعداد دســـترســـی‌ها و تردد ماشـــین‌آلات کشـــاورزی، موجب کاهش ســـرعت و کاهش واریانس ســـرعت می‌شوند (Panschin & Vitikainen, 2010).

در بین عوامل زیرساختی، تعداد خطوط عبوری به‌عنوان متغیری تأثیرگذار شـــناخته شـــده اســـت. افزایش تعداد خطوط معمولاً آزادی مانور بیشـــتری ایجاد کرده و در نتیجه باعث افزایش تغییرات ســـر عت و تغییر خط در بین رانـ ند گان می‌شـــود (Huang et al., 2013; Xiaobao, 2008). همچنین، وجود میانه راه ممکن اســـت به دلیل ایجاد مانورهای تغییر جهت، کاهش یا افزایش ســـرعت، منجر به افزایش واریانس سرعت گردد (Mohanty & Dey, 2019). از سوی دیگر، وجود سرعت‌کاه‌ها به‌صورت معناداری موجب کاهش واریانس سرعت در مقاطع نـصب آن‌ها شده ا ست، چراکه این تجهیزات موجب همگن شدن رفتار حرکتی رانندگان می‌شوند (Zhong & Hou, 2010).

مرور کلی ادبیات فوق نشـــان می‌دهد که واریانس ســـرعت تحت‌تأثیر هم‌زمان عوامل ترافیکی، زیرســـاختی و محیطی قرار دارد و روابط بین این متغیرها غالباً غیرخطی و تعاملی هســـتند. بیشتر پژوهش‌های موجود یا بر مبنای شبیه‌سازی انجام شده‌اند یا داده‌های آن‌ها متعلق به کشورهای توسعه‌یافته است. ازاین‌رو، یافته‌های آن‌ها به‌صورت مستقیم قابل تعمیم به شرایط جاده‌های ایران نیســـتند؛ به‌ویژه با توجه به تفاوت‌های قابل‌توجه در رفتار رانندگان، وضـــعیت زیرســـاخت‌ها و ترکیب ناوگان. بنابراین، شکاف اصلی پژوهش حاضر در بررسی اثرات چندبعدی عوامل مؤثر بر وار یانس ســـر عت با اســـتـ فاده از داده های واقعی برداشت‌شده از شبکه راه‌های برون‌شهری ایران و بهره‌گیری از مدل‌های یادگیری ما شین با توانایی مدل سازی روابط غیرخطی نهفته اســـت. چنین رویکردی می‌تواند درک دقیق‌تری از پویایی جریان ترافیک و عوامل مولد ناهمگنی در ســـرعت‌ها ارائه دهد و مبنای علمی برای تدوین ســـیاســـت‌های ایمنی داده‌محور در کشور فراهم سازد.

۳-روش تحقیق

در این بخش، رویکرد پژوهش، داده‌های مورد استفاده و ساختار مدل تحلیلی تشـــریح می‌گرد ند. همچنین منطق به کارگیری الگوریتم انتخابی و نحوه‌ی ارز یابی عملکرد آن توضـــیح داده شده ا ست تا چارچوبی شفاف از مراحل مدل سازی و تحلیل فراهم شود.

۳-۱-داده‌های مورد استفاده

داده‌های این پژوهش حاصـــل برداشـــت میدانی از ۲۰۰ نقطه منتخب در شـــبکه‌ی راه‌های برون‌شـــهری اســـتان همدان اســـت. انتخاب این نقاط بر مبنای تنوع شرایط محیطی و زیر ساختی و با تأیید کارشناسان حمل‌ونقل و ایمنی ترافیک انجام شده است. موقعیت مکانی بخشی از مقاطع برداشت‌شده در شکل ۱ نمایش داده شده است.

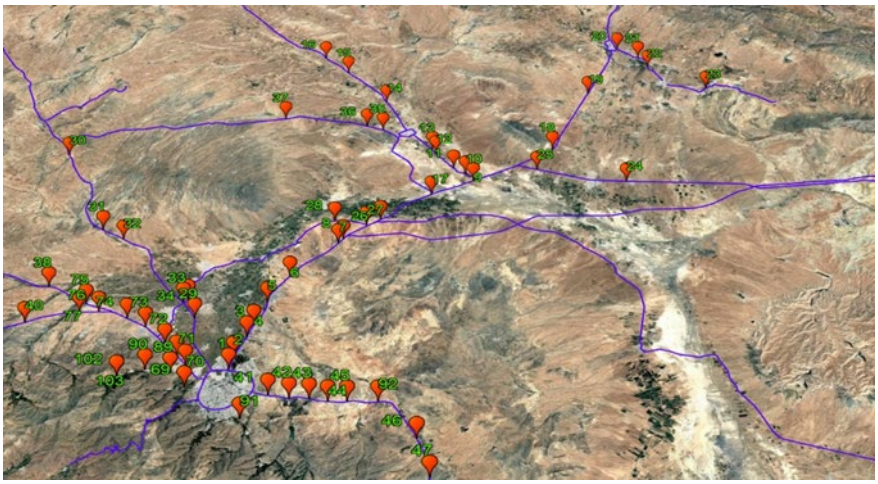
داده‌ها در سه گروه اصلی طبقه‌بندی شده‌اند.

ویژگی‌های ترافیکی: شامل حجم عبوری و سایل نقلیه، سرعت متوسط جریان، درصد تخطی از سرعت مجاز و نسبت وسایل نقلیه سنگین؛ متغیرهایی که مستقیماً بر تغییرپذیری سرعت تأثیر می‌گذارند.

ویژگی های محیطی: شامل تراکم جمعیت اطراف راه، نوع کاربری‌های مجاور (مسکونی، کشاورزی و صنعتی) و الگوی توسعه شهری در حریم مسیر، که می‌توانند بر رفتار رانندگی و ادراک خطر اثرگذار باشند.

ویژگی‌های زیرساختی: مانند وجود میانه فیزیکی، سرعت‌کاه، تعداد خطوط عبور و وضعیت هندسی مسیر (شیب و قوس)، که از عوامل تعیین‌کننده‌ی پایداری سرعت محسوب می‌شوند.

تمامی داده‌ها پس از کنترل کیفی و حذف موارد ناقص، در قالب یک پایگاه داده‌ی ساخت‌یافته سازمان‌دهی شدند تا امکان تحلیل کمی و مدل‌سازی آماری فراهم گردد. جزئیات هر متغیر و نحوه‌ی پردازش داده‌ها در جدول ۱ ارائه شده است.



شکل ۱. نمونه‌ای از شبکه راه‌های پیمایش شده استان همدان

جدول ۱. جزییات داده‌های مورداستفادهبرای مدل‌سازی پیش‌بینی واریانس سرعت

متغیر	نوع متغیر		متغیر	نوع متغیر
درصد تخطی از سرعت مجاز	کمی		جمعیت	کمی
درصد خودرو سنگین	کمی		کاربری کشاورزی	کیفی
حجم تردد ساعتی	کمی		شهرسازی دوطرفه اطراف راه	کیفی
وجود سرعت‌کاه	کیفی		میانه	کیفی
تعداد خطوط	کمی		واریانس سرعت	کمی

۳-۲-مدل جنگل تصادفی

در سال‌های اخیر، مدل‌های یادگیری ماشین به دلیل توانایی در شناسایی روابط غیرخطی و پیچیده میان متغیرها، جایگزین مناسبی برای روش‌های سنتی خطی محسوب می‌شوند. این مدل‌ها با حذف پیش‌فرض‌های سخت‌گیرانه درباره‌ی توزیع داده‌ها، در مواجهه با داده‌های پرحجم، نویزی یا ناهمگن کارایی بالاتری نشان می‌دهند. یکی از الگوریتم‌های قدرتمند در این حوزه، مدل جنگل تصادفی است که نخستین بار توسط بریمن معرفی شد (Breiman, 2001). مدل جنگل تصادفی با ترکیب خروجی چندین درخت تصمیم مستقل، مدلی پایدار و مقاوم ایجاد می‌کند که در برابر نوسانات داده‌ای حساسیت کمتری داشته و کمتر دچار بیش برازش می شود. این الگوریتم بر دو مفهوم کلیدی استوار است.

نمونه‌گیری بوت‌استرپ (Bagging): در هر مرحله، زیرمجموعه‌ای تصادفی از داده‌های آموزشی با جایگزینی انتخاب می‌شود تا برای آموزش یک درخت تصمیم مستقل به‌کار رود. انتخاب تصادفی ویژگی‌ها: در زمان تقسیم هر گره، تنها بخشی از متغیرها به‌صورت تصادفی برای تصمیم‌گیری مورد استفاده قرار می‌گیرند تا از هم‌بستگی زیاد میان درخت‌ها جلوگیری شود. در نهایت، تصمیم نهایی مدل بر اساس رأی‌گیری اکثریت میان خروجی درخت‌ها تعیین می‌شود. اگر خروجی درخت‌ها برای نمونه‌ی ورودی x به ترتیب $h_1(x), h_1(x), \dots, h_k(x)$ باشند، خروجی نهایی مدل از رابطه‌ی زیر به‌دست می‌آید (Yan & Shen, 2022):

$$H(x) = mode\{h_1(x), h_2(x), \dots, h_k(x)\}$$

در این رابطه، تابع mode مقداری را بازمی‌گرداند که بیشترین فراوانی را میان پیش‌بینی‌ها دارد.

برای ارزیابی عملکرد مدل، از نمونه‌های خارج از کیسه استفاده شده است؛ به این ترتیب که حدود یک‌سوم داده‌ها در هر مرحله از آموزش درخت حضور ندارند و برای آزمون همان درخت مورد استفاده قرار می‌گیرند. این روش، برآوردی دقیق از خطای مدل بدون نیاز به مجموعه‌ی اعتبارسنجی جداگانه فراهم می‌کند. علاوه بر این، دقت نهایی مدل با شاخص میانگین مربعات خطا (MSE) سنجیده شده است.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2$$

علاوه بر شاخص MSE، برای ارزیابی جامع‌تر دقت پیش‌بینی مدل از معیار جذر میانگین مربعات خطا (RMSE) نیز استفاده شد. این شاخص، جذر مقدار MSE است و به‌صورت زیر تعریف می‌شود:

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

که در آن‌ها y_i مقدار واقعی و $\hat{y}_i$ مقدار پیش‌بینی‌شده برای نمونه‌ی i است.

مقدار RMSE همواره در واحد متغیر وابسته بیان می شود و نشان‌دهنده میزان انحراف میانگین پیش‌بینی‌ها از مقادیر واقعی است. هرچه مقدارRMSE و MSE کمتر باشد، مدل از دقت پیش‌بینی بالاتری برخوردار است.

در این پژوهش، از RMSE به‌عنوان شاخص مکمل MSE جهت ارزیابی عملکرد مدل جنگل تصادفی استفاده شد تا میزان پراکندگی خطای پیش‌بینی به‌صورت قابل تفسیرتر بیان گردد.

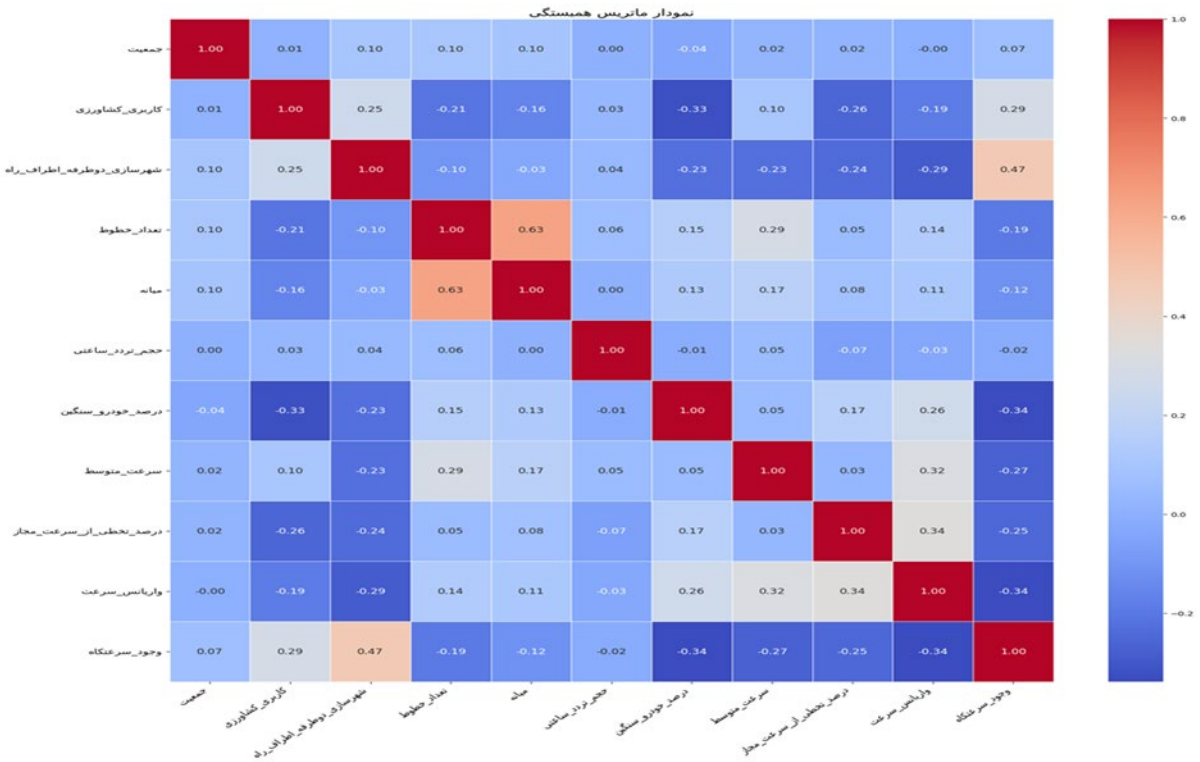
در نهایت، خروجی مدل به‌صورت احتمال تعلق هر مقطع به طبقه‌ی «پرخطر» محاسبه شد. این احتمال‌ها، همراه با مقادیر OOB و MSE، مبنای اولویت‌بندی نقاط بحرانی و برنامه‌ریزی اقدامات ایمن‌سازی در راه‌های برون‌شهری قرار گرفتند.

٤-نتایج

این بخش به تبیین یافته‌های حاصل از مدل جنگل تصادفی در پیش‌بینی واریانس سرعت وسایل نقلیه اختصاص دارد. نتایج ارائه‌شده شامل ارزیابی عملکرد مدل، بهینه‌سازی پارامترهای آن و تحلیل متغیرهای اثرگذار بر تغییرپذیری سرعت است.

٤-١-تحلیل همبستگی میان متغیرها

پیش از اجرای مدل، ماتریس همبستگی متغیرها به‌منظور بررسی روابط درونی و جلوگیری از بروز هم‌خطی چندگانه توسعه داده شد. همان‌گونه که در شکل ۲ مشاهده می‌شود، ضرایب همبستگی میان تمامی متغیرها کمتر از ۰٫۷ است؛ بنابراین، استقلال نسبی میان متغیرها برقرار بوده و داده‌ها از منظر چندهمخطی برای ورود به مدل جنگل تصادفی مناسب ارزیابی شدند. این موضوع از بروز سوگیری در نتایج و کاهش دقت مدل جلوگیری می‌کند.



شکل ۲. نمودار ماتریس همبستگی بین متغیرها

٤-۲-بهینه‌سازی مدل جنگل تصادفی

فرایند پیاده‌سازی مدل با استفاده از زبان برنامه‌نویسی Python و کتابخانه‌ی Scikit-learn انجام گرفت. به‌منظور افزایش دقت و پایداری مدل، بهینه‌سازی ابرپارامترها (Hyperparameters) با استفاده از الگوریتم جست‌وجوی احتمالی Hyperopt صورت گرفت. انتخاب مقادیر مناسب برای پارامترهایی نظیر تعداد درخت‌ها، عمق بیشینه، و حداقل نمونه‌های مورد نیاز در هر برگ، تأثیر مستقیم بر عملکرد مدل و کنترل بیش‌برازش دارد.

در این پژوهش، واریانس سرعت به‌صورت عددی در واحد مجذورکیلومتر بر مجذورساعت $\left(\frac{Km}{hr}\right)^2$ محاسبه شده است و میانگین مقدار واریانس در داده‌های مورد استفاده برابر با ۱۹۱,۳۰ $\left(\frac{Km}{hr}\right)^2$ است که بیانگر پراکندگی نسبتاً بالای سرعت در مقاطع موردمطالعه است.

نتایج تنظیم پارامترها در جدول ۲ ارائه شده است. بر اساس ارزیابی معیار میانگین خطای مطلق (MAE)، مقدار خطا پس از تنظیم پارامترها کاهش معناداری یافته است (جدول ۳)، که بیانگر بهبود چشمگیر توان پیش‌بینی مدل است.

علاوه بر این، مقدار ضریب تعیین (R^2) برابر با ۰,٦۱ به‌دست آمد که با توجه به ماهیت داده‌های میدانی و پراکندگی آن‌ها، نشان‌دهنده‌ی برازش قابل‌قبول مدل بر داده‌های آزمون است. در شکل ۳، یکی از درخت‌های تصمیم حاصل از میان ۳۰۰ درخت تشکیل‌دهنده‌ی جنگل تصادفی نمایش داده شده است تا ساختار درونی مدل و فرآیند تصمیم‌گیری آن به‌صورت بصری قابل درک باشد.

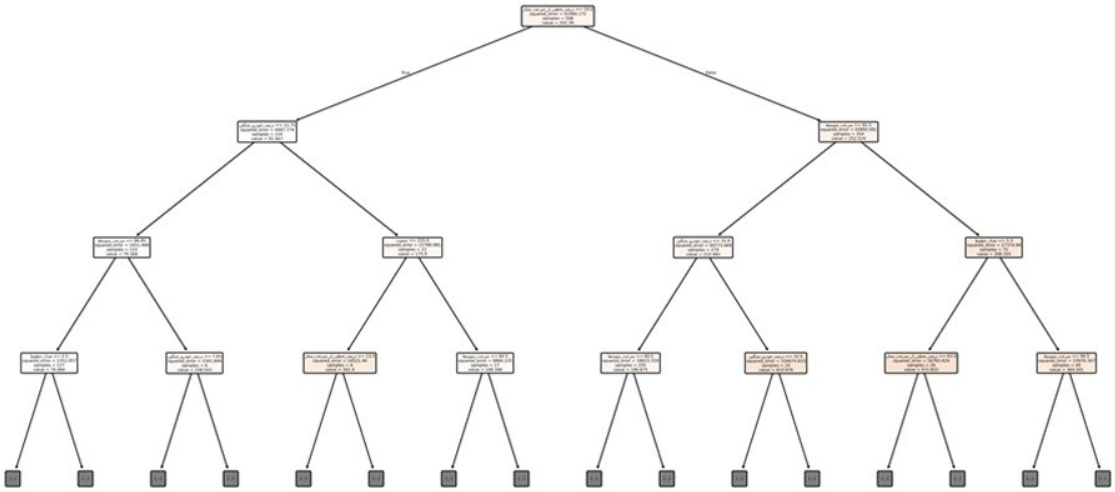
به‌منظور ارزیابی دقیق‌تر عملکرد مدل، شاخص جذر میانگین مربعات خطا (RMSE) نیز محاسبه شد که مقدار آن برابر با ۱۰۵,۳ است. مقدار پایین RMSE در مقایسه با دامنه تغییرات داده‌ها نشان‌دهنده دقت مناسب مدل در پیش‌بینی واریانس سرعت است.

جدول ۲. تنظیمات هایپرپارامترهای مدل جنگل تصادفی

هایپرپارامتر	توضیح	مقدار
max_depth	حداکثر عمق هر درخت در جنگل.	۱۰
min_samples_leaf	حداقل تعداد نمونه‌های موردنیاز برای تشکیل یک برگ.	۲
min_samples_split	حداقل تعداد نمونه‌های موردنیاز برای تقسیم یک گره داخلی.	۲
n_estimators	تعداد درخت‌های موجود در جنگل.	۳۰۰

جدول ۳. تنظیمات هایپرپارامترهای مدل جنگل تصادفی

شاخص	شماره بخش (Fold)	مقدار (قبل از تعیین هایپرپارامتر)	مقدار (بعد از تعیین هایپرپارامتر)
MAE	۱	۷۵,۹۶۶۹۹۹	۷۴,۸۸۱۵۳۱
MAE	۲	۸۸,۴۴۰۱۰۷	۸۶,۶۵۲۰۶۵
MAE	۳	۷۸,۲۶۶۳۲۶	۷۷,۱۷۰۷۸
MAE	۴	۷۳,۴۰۵۴۱۷۶	۷۱,۰۶۷۷۲۵
MAE	۵	۸۸,۵۴۱۲۴۷	۸۹,۳۶۰۶۹۹
MAE	کل	۸۰,۹۲۴۰۱۹	۷۵,۸۲۶۵۶۱



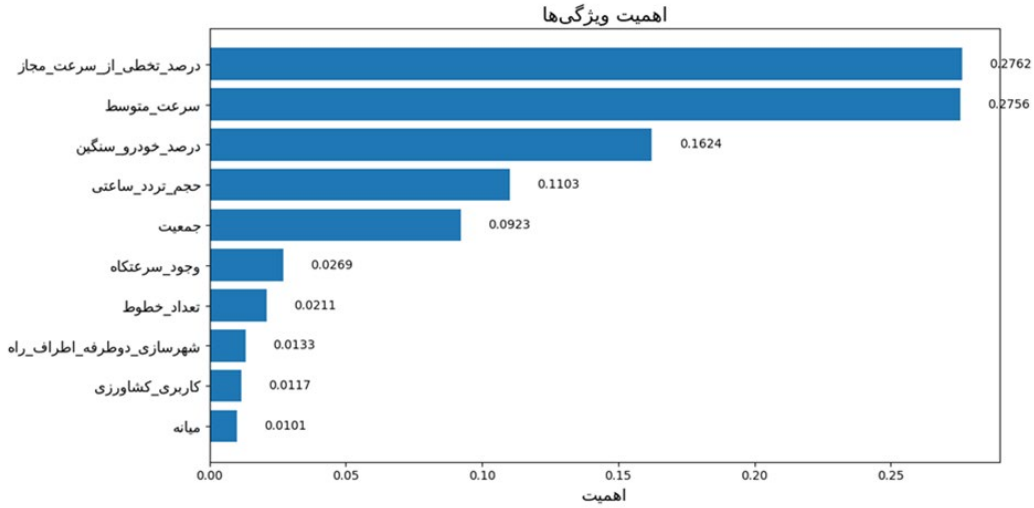
شکل ۳. نمونه‌ای از درخت‌های توسعه داده شده طی فرایند مدلسازی

۴-۳-تفسیر مدل و تحلیل متغیرهای اثرگذار

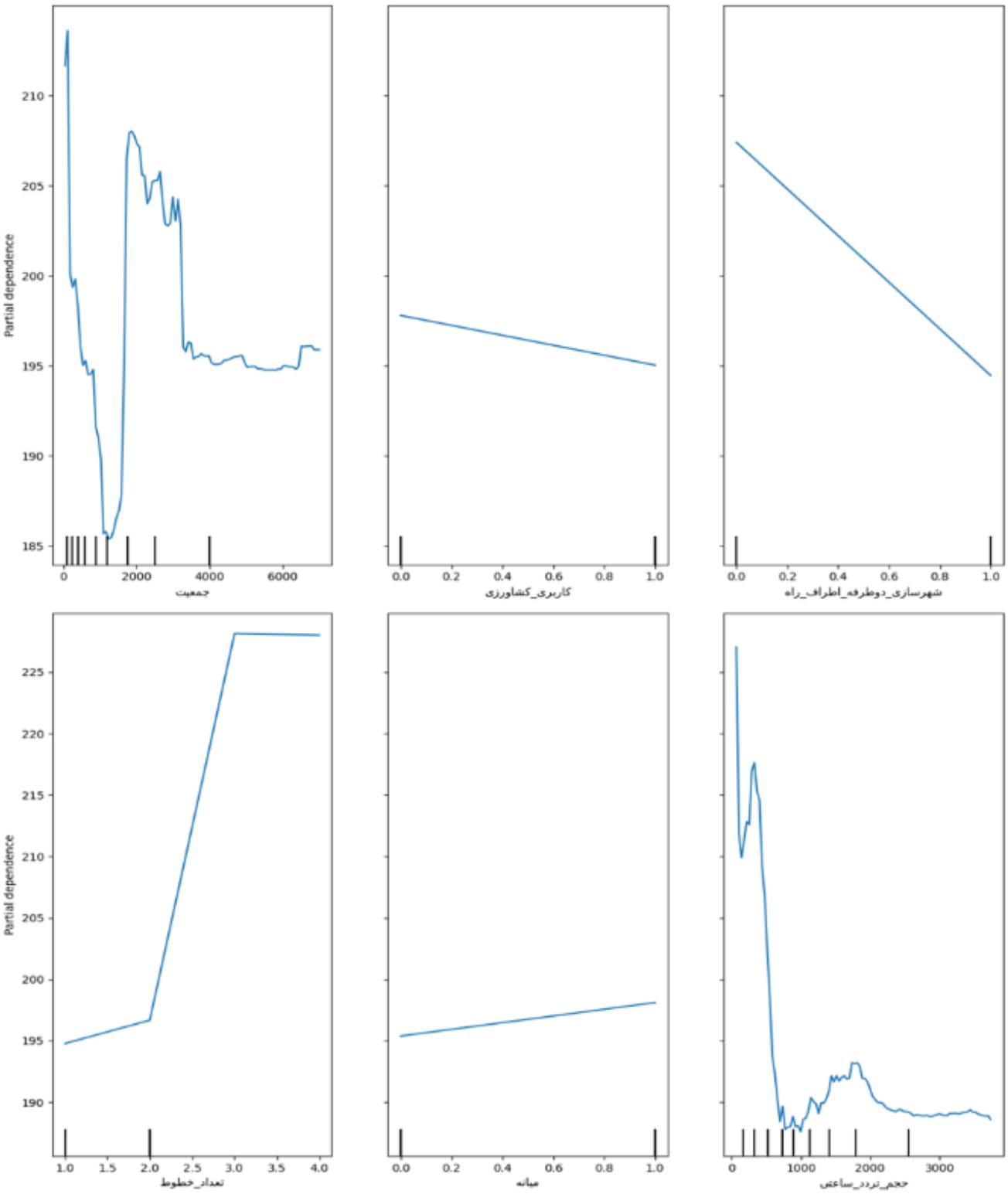
در شکل ٤ میزان اهمیت نسبی متغیرها در پیش‌بینی واریانس سرعت نشان داده شده است. نتایج حاکی از آن است که ویژگی‌های ترافیکی سهم غالب را در توضیح تغییرپذیری سرعت دارند. در میان آن‌ها، درصد تخطی از سرعت مجاز با اهمیت نسبی ۰,۲۷۶۲ بیشترین تأثیر را دارد، که بیانگر نقش مستقیم رفتار رانندگان و میزان تبعیت آن‌ها از قوانین محدودیت سرعت در نوسانات سرعت است. پس از آن، سرعت متوسط جریان (۰,۲۷۵٦) و درصد وسایل نقلیه سنگین (۰,۱٦۲٤) به‌عنوان عوامل کلیدی شناسایی شدند که به ترتیب بیانگر تأثیر ویژگی‌های جریان و ترکیب ترافیکی بر پایداری سرعت هستند.

در میان متغیرهای محیطی، حجم تردد ساعتی (۰,۱۱۰۳) و جمعیت پیرامونی (۰,۰۹۲۳) اهمیت بالاتری دارند. این نتایج نشان می‌دهد افزایش تراکم جمعیتی و حجم عبور، موجب افزایش تعاملات رانندگان و در نتیجه رشد نوسانات سرعت می‌شود. از سوی دیگر، متغیرهای زیرساختی مانند وجود سرعت‌کاه (۰,۰۲٦۹)، تعداد خطوط عبوری (۰,۰۲۱۱)، نوع توسعه شهری در حاشیه راه (۰,۰۱۳۳)، کاربری کشاورزی (۰,۰۱۱۷) و وجود میانه (۰,۰۱۰۱) نیز اثرگذارند، هرچند شدت تأثیر آن‌ها در مقایسه با عوامل ترافیکی کمتر است.

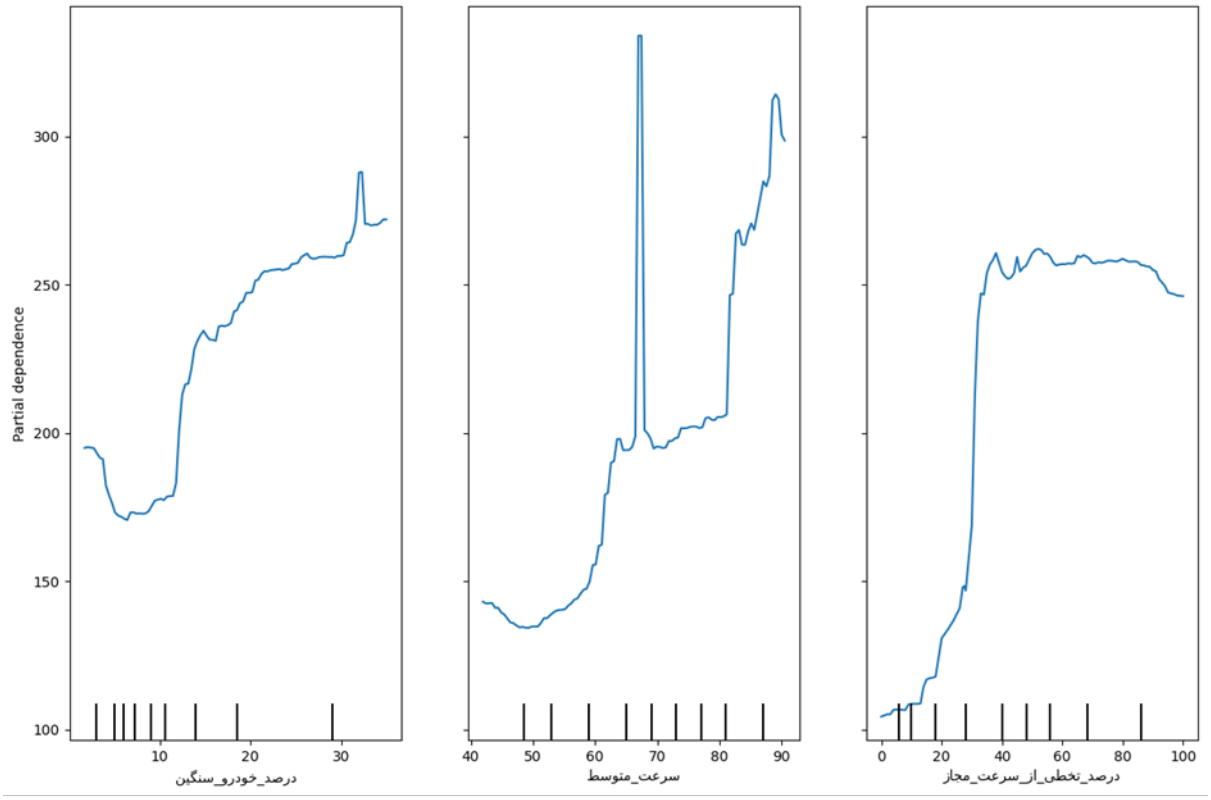
به‌طور کلی، یافته‌ها بیانگر آن است که سه دسته عامل شامل ویژگی‌های جریان ترافیک، رفتار رانندگان و شرایط هندسی مسیر نقش اساسی در شکل‌گیری واریانس سرعت دارند.

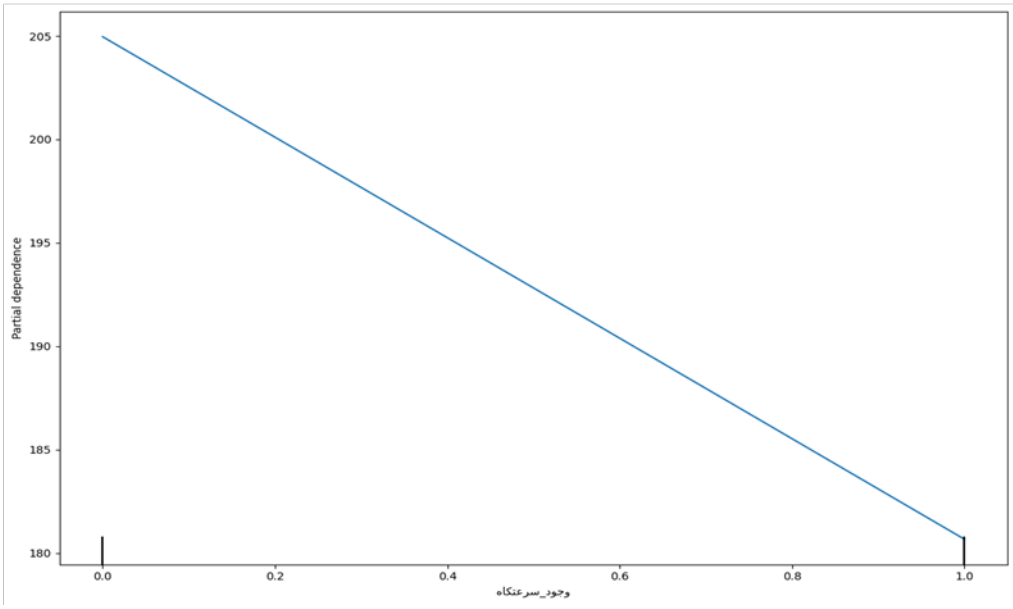


شکل ٤. اهمیت ویژگی‌های مختلف در مدل پیش‌بینی واریانس سرعت با استفاده از مدل جنگل تصادفی



شکل ۵. نمودار وابستگی جزئی عوامل کلیدی تأثیرگذار بر واریانس سرعت





ادامه شکل ۵. نمودار وابستگی جزئی عوامل کلیدی تأثیرگذار بر واریانس سرعت

۵- نتیجه‌گیری

به‌منظور درک جامع از نحوه‌ی اثرگذاری متغیرهای کلیدی بر واریانس سرعت، از مدل جنگل تصادفی و نمودارهای وابستگی جزئی (PDP) استفاده شد (شکل ۵). نتایج نشان داد که روابط میان واریانس سرعت و متغیرهای ورودی عمدتاً غیرخطی و تعاملی هستند و متغیرهای ترافیکی، زیرساختی و محیطی به‌صورت توأمان بر تغییرپذیری سرعت اثر می‌گذارند.

در بین متغیرهای ترافیکی، درصد تخطی از سرعت مجاز بیشترین تأثیر را در افزایش واریانس سرعت داشت. افزایش نرخ تخطی موجب ناهمگونی در رفتار رانندگان و بروز تغییرات ناگهانی در جریان ترافیک می‌شود، که با یافته‌های هو (Hu, 2017) مطابقت دارد. سرعت متوسط نیز رابطه‌ای غیرخطی با واریانس دارد؛ در سرعت‌های پایین تغییرپذیری محدود است اما با افزایش سرعت، نوسانات به‌صورت تصاعدی افزایش می‌یابد، که با نتایج گاربر و گادیرائو (Garber & Gadirau, 1988) هم‌سو است.

درصد وسایل نقلیه سنگین نیز اثر دوگانه‌ای بر واریانس سرعت دارد: در ابتدا با افزایش سهم این وسایل تا حدود ۳۰٪، واریانس رشد می‌کند، اما در سطوح بالاتر به دلیل یکنواختی جریان حرکتی، کاهش می‌یابد (Shen & Yang, 2020; Roh et al., 2021). این نتیجه نشان می‌دهد که در مسیرهایی با ترکیب ترافیکی متنوع، تفکیک خطوط ویژه برای وسایل نقلیه سنگین می‌تواند به کاهش نوسانات کمک کند.

در میان متغیرهای زیرساختی، تعداد خطوط عبوری تأثیر مثبت و وجود سرعت‌کاه تأثیر منفی بر واریانس داشت. افزایش تعداد خطوط به رانندگان آزادی عمل بیشتری در تغییر مسیر می‌دهد که باعث افزایش نوسانات سرعت می‌شود، در حالی که وجود سرعت‌کاه‌ها جریان ترافیک را یکنواخت‌تر می‌سازد و واریانس را کاهش می‌دهد. این یافته با نتایج پژوهش‌های هوانگ، ژانگ و شیائوبائو (Huang et al., 2013; Xiaobao, 2016) و ژانگ و هو (Zhong & Hou, 2010) هم‌خوان است. همچنین، وجود میانه در مسیر باعث افزایش جزئی واریانس گردید که احتمالاً ناشی از توقف‌های لحظه‌ای و مانورهای تغییر جهت است (Mohanty & Dey, 2019).

تحلیل تعاملی مدل جنگل تصادفی نشان داد که اثر برخی از متغیرهای زیرساختی مانند تعداد خطوط عبوری صرفاً مستقیم نیست و از طریق تعامل با عوامل ترافیکی بروز می‌کند. افزایش تعداد خطوط در شرایط ترافیک سبک باعث افزایش آزادی مانور و رشد نوسانات می‌شود، اما در شرایط ازدحام بالا، تأثیر آن کاهشی است، زیرا تراکم زیاد، مانورپذیری را محدود می‌کند و جریان حرکت را یکنواخت‌تر می‌سازد. به‌عبارت دیگر، اثر تعداد خطوط بر واریانس سرعت وابسته به سطح سرویس مسیر و حجم ترافیک است. این تعامل غیرخطی نشان می‌دهد که طراحی هندسی مسیرها باید همراه با مدیریت جریان ترافیک ارزیابی شود. همچنین، وجود میانه در مقاطع چندخطه ممکن است به‌صورت غیرمستقیم با تغییر رفتار سبقت‌گیری و کاهش ناگهانی سرعت در حین مانور، موجب افزایش نسبی واریانس گردد.در حوزه متغیرهای محیطی، تراکم جمعیتی پیرامون مسیر و کاربری کشاورزی اثر کاهشی بر واریانس داشتند. در مناطق با تراکم جمعیت بالا، احتیاط رانندگان بیشتر و رفتار حرکتی منظم‌تر است. در نواحی با کاربری کشاورزی نیز به‌دلیل وجود دسترسی‌های متعدد و حضور ماشین‌آلات کندرو، جریان حرکت یکنواخت‌تر می‌شود (Alizade et al., 2023; Panschin & Vitikainen, 2010).

حجم ترافیک ساعتی نیز رابطه‌ای معکوس با واریانس داشت؛ در شرایط ترافیک سبک، نوسانات بیشتر و در ترافیک سنگین، حرکت منظم‌تر است (Hung et al., 2013).

در جمع‌بندی نهایی، نتایج نشان داد که کنترل و مدیریت واریانس سرعت مستلزم رویکردی چندبعدی است که هم‌زمان ابعاد ترافیکی، هندسی و محیطی را در برگیرد. بازنگری در محدوده‌های سرعت مجاز با استفاده از داده‌های واقعی، توسعه سامانه‌های هوشمند نظارت بر تخطی از سرعت، تفکیک جریان و سایل نقلیه سنگین در مسیرهای پرتردد، استفاده هدفمند از سرعت‌کاه در مقاطع با نوسانات بالا، و بازطراحی هندسی مسیرهای چندخطه بر اساس رفتار واقعی رانندگان می‌تواند به‌طور مؤثری موجب کاهش تغییرپذیری سرعت و ارتقای ایمنی جاده‌ها شود. از منظر پژوهشی، ترکیب مدل‌های یادگیری ماشین با داده‌های لحظه‌ای (Real-Time) و متغیرهای رفتاری رانندگان می‌تواند مسیر آینده تحقیقات در حوزه پیش‌بینی رفتار پرریسک رانندگی را هموار سازد.

۶-سپاسگزاری

نویسندگان بدین‌وسیله از اداره‌کل راهداری و حمل‌ونقل جاده‌ای استان همدان بابت همکاری در گردآوری داده‌های میدانی و همچنین از جناب آقای مهندس مصطفی وصالی به‌دلیل همراهی و پشتیبانی فنی در مراحل اجرای پژوهش، صمیمانه قدردانی می‌نمایند.

۷-مراجع

-سازمان پزشکی قانونی کشور. (۱۴۰۲). آمار متوفیات و مصدومین حوادث رانندگی در سال ۱۴۰۲. تهران، *سازمان پزشکی قانونی کشور.*

-Alizadeh, M., Davoodi, S. R., & Shaaban, K. (2023). Drivers' speeding behavior in residential streets: A structural equation modeling approach. *Infrastructures,* 8(1), 11-12.

-Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

-Hu, W. (2017). Raising the speed limit from 75 to 80 mph on Utah rural interstates: Effects on vehicle speeds and speed variance. *Journal of Safety Research,* 61, 83–92.

-Mohanty, M., & Dey, P. P. (2019). Major stream delay under limited priority conditions. Journal of Transportation Engineering, Part A: *Systems,* 145(3), 1–10.

-Yan, M., & Shen, Y. (2022). Traffic accident severity prediction based on random forest. Sustainability, 14(3), 1729.

-Zhong, L., & Hou, D. (2010, August). Evaluation of work zone speed reduction measures. In ICCTP 2010: Integrated Transportation Systems – Green, Intelligent, Reliable, *Beijing: ASCE.* 267–275.

-Alomari, A. H., Al-Salem, M., Al-Basyouny, K., & El-Basyouny, S. (2023). Modeling speed mean and variance for different enforcement conditions on multilane highways. *Journal of Transportation Engineering, Part A: Systems,* 149(8), 04023067.

-Garber, N. J., & Gadirau, R. (1988). Speed variance and its influence on accidents. Washington, *DC: Transportation Research Record.*

-Gargoum, S. A., & El-Basyouny, K. (2016). Exploring the association between speed and safety: A path analysis approach. *Accident Analysis & Prevention,* 93, 32–40.

-Graves, P. E., Lee, D. R., & Sexton, R. L. (1993). Speed variance, enforcement, and the optimal speed limit. *Economics Letters,* 42(2–3), 237–243.

-Huang, B., Zhao, L., Wang, W., & Liu, H. (2013). A simulation study for minimizing operating speed variation of multilane highways by controlling access. *Procedia – Social and Behavioral Sciences,* 96, 2767–2781.

-Lord, D., Qin, X., & Geedipally, S. R. (2021). Highway safety analytics and modeling. *Elsevier.*

-Panschin, S., & Vitikainen, A. (2010, April). Land consolidation and agricultural traffic: Instrument for calculating the traffic volume. In XXIV FIG International Congress, Sydney, Australia (pp. 11–16). *FIG – International Federation of Surveyors.*

-Quddus, M. (2013). Exploring the relationship between average speed, speed variation, and accident rates using spatial statistical models and GIS. *Journal of Transportation Safety & Security,* 5(1), 27–45.

-Roh, C. G., Jeon, H., & Son, B. (2021). Do heavy vehicles always have a negative effect on traffic flow? *Applied Sciences,* 11(12), 1–15.

-Shen, J., & Yang, G. (2020). Crash risk assessment for heterogeneous traffic and different vehicle-following patterns using microscopic traffic flow data. *Sustainability,* 12(23), 1–18.

-Wang, X., Abdel-Aty, M., & Lee, J. (2018). Speed, speed variation, and crash relationships for urban arterials. Accident Analysis & Prevention, 113, 236–243.

-World Health Organization. (2023). Global status report on road safety 2023. Geneva: *World Health Organization.*

-Xiaobao, Y. Y. (2008). Changes in traffic characteristics affected by number of lanes on freeways. Beijing: *China Academy of Transportation Research.*

Data-Driven Modeling of Factors Affecting Speed Variance on Rural Roads Using a Machine Learning Approach (Case Study: Hamadan Province)

Ali Tavakoli Kashani, *Associate Professor, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran and Road Safety Research Center, Iran University of Science and Technology, Tehran, Iran.*
Mohammad Najafpour Dero, Ph.D., Stud., School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran and Road Safety Research Center, Iran University of Science and Technology, Tehran, Iran.
Amirreza Eskandari, M.Sc., Student, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.
Majid Najafi, M.Sc., Student, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.

E-mail: alitavakoli@iust.ac.ir

Received: September 2025- Accepted: February 2026

ABSTRACT
Speed variance is widely acknowledged as a critical indicator of traffic flow stability and a major determinant of road safety, as excessive fluctuations in vehicle speed significantly elevate the likelihood and severity of crashes. This study aims to identify and model the key factors influencing speed variance on rural highways in Iran using the Random Forest algorithm—a robust machine learning method capable of capturing complex nonlinear relationships and reducing overfitting risk. A total of 200 observation points were collected from suburban road network of Hamadan province through detailed field surveys. The dataset included traffic, infrastructural, and environmental variables that potentially affect speed dispersion. The Random Forest results revealed that the percentage of speed limit violations was the most influential predictor, showing a strong positive association with the magnitude of speed variability. Average speed exhibited a nonlinear effect, remaining relatively stable at lower levels but sharply amplifying variance at higher speeds. Moreover, a moderate share of heavy vehicles (30–40%) increased speed fluctuations, whereas higher proportions contributed to more uniform traffic flow. Among infrastructural attributes, the number of lanes had a positive impact on speed variance, reflecting the greater freedom for lane changes and maneuvering, while the presence of speed bumps significantly reduced speed dispersion. Environmentally, higher roadside population density and agricultural land use were associated with decreased variance, suggesting more cautious driving behavior in such areas. Overall, the findings highlight the necessity of integrating traffic enforcement, infrastructure design, and environmental planning into comprehensive speed management strategies. The proposed machine learning framework provides a data-driven foundation for identifying high-risk segments and informing targeted interventions to enhance rural road safety.

Keywords: Speed Variance; Road Safety; Random Forest; Machine Learning; Traffic Behavior; Rural Highways