

شناسایی عوامل موثر بر شدت تصادفات موتورسیکلت در راه‌های برون‌شهری با الگوریتم جنگل تصادفی (RF) و تحلیل SHAP

مقاله علمی-پژوهشی

*حامد سیفی (نویسنده مسئول)، گروه مهندسی عمران، دانشگاه پیام نور، تهران، ایران

شاهین شعبانی، گروه مهندسی عمران، دانشگاه پیام نور، تهران، ایران

محمد کوهی، دانش آموخته کارشناسی ارشد، گروه مهندسی عمران، دانشگاه پیام نور، تهران، ایران

*پست الکترونیکی نویسنده مسئول: hamedsaify@pnu.ac.ir

دریافت: ۱۴۰۴/۱۰/۲۰ - پذیرش: ۱۴۰۵/۰۳/۰۲

صفحه ۷۴-۵۵

چکیده

تصادفات موتورسیکلت به دلیل آسیب‌پذیری بالای این گروه از کاربران راه، سهم قابل توجهی از تلفات جاده‌ای را به خود اختصاص می‌دهند و شدت پیامدهای آن‌ها، به‌ویژه در راه‌های برون‌شهری، همواره یکی از چالش‌های اصلی ایمنی راه بوده است. هدف اصلی این پژوهش، شناسایی و تفسیر عوامل مؤثر بر شدت تصادفات موتورسیکلت در راه‌های برون‌شهری با استفاده از رویکردهای یادگیری ماشین تفسیرپذیر است. بدین منظور، از داده‌های تصادفات سه‌ساله (۱۳۹۶ تا ۱۳۹۸) موتورسیکلت در راه‌های برون‌شهری استان خراسان رضوی استفاده شد. متغیرهای مورد بررسی در قالب سیزده دسته شامل ویژگی‌های راه، وضعیت هندسی، نوع شانه، نقص مؤثر راه، عامل انسانی، نوع وسیله نقلیه مقصر، کاربری اراضی و سایر عوامل محیطی و ترافیکی تعریف شدند. در این پژوهش، سه الگوریتم یادگیری ماشین شامل جنگل تصادفی، XGBoost و LightGBM برای پیش‌بینی شدت تصادفات (خسارتی، جرحی و فوتی) به کار گرفته شدند. پس از پیش‌پردازش داده‌ها و متعادل‌سازی کلاس‌ها با استفاده از روش SMOTE عملکرد مدل‌ها با معیارهای recall precision و F1-score ارزیابی شد. نتایج نشان داد که مدل جنگل تصادفی با دقت کلی ۷۶ درصد، عملکرد بهتری نسبت به سایر مدل‌ها دارد و به‌عنوان مدل برتر انتخاب شد. به‌منظور تفسیر خروجی مدل منتخب، از روش SHAP استفاده گردید. نتایج تحلیل SHAP نشان داد که به ترتیب نوع شانه، نوع وسیله نقلیه مقصر، نقص مؤثر راه، عامل انسانی، کاربری اراضی، سرعت مجاز، نوع راه، وضعیت خط‌کشی و هندسه مسیر، مهم‌ترین عوامل مؤثر بر شدت تصادفات موتورسیکلت در راه‌های برون‌شهری هستند. یافته‌های این پژوهش نشان می‌دهد که شدت تصادفات موتورسیکلت حاصل برهم‌کنش پیچیده عوامل زیرساختی، رفتاری و وسیله نقلیه بوده و اتخاذ رویکردهای شدت‌محور در برنامه‌ریزی ایمنی راه می‌تواند نقش مؤثری در کاهش پیامدهای شدید تصادفات داشته باشد.

واژه‌های کلیدی: تصادفات موتورسیکلت، جنگل تصادفی (RF)، تحلیل SHAP

۱- مقدمه

موتورسیکلت سواران (Rasri et al., 2024; Fondzenyuy et al., 2024) جستجو کرد. این مسائل در کنار عدم تجهیزات حفاظتی کافی در موتورسیکلت (Kashifi, 2023) و عدم استفاده از کلاه ایمنی (Brockhus et al., 2024)، سبب شده که آمار مرگ و میر تصادفات موتورسیکلت به اندازه یک سوم (۲۸ درصد) مرگ و میر سالانه (۱,۲ میلیون

تصادفات موتورسیکلت در کشورهای در حال توسعه بخش قابل توجهی از تصادفات را به خود اختصاص داده است (Rasri et al., 2024). دلیل این امر را می‌توان در تعداد بسیار زیاد این نوع وسیله نقلیه در ناوگان حمل و نقل کشورهای در حال توسعه (به دلیل درآمد پایین مردم و هزینه خرید و نگهداری اندک موتورسیکلت) (Naqvi & Tiwari, 2017) و خطای انسانی

2020; Haruna et al., 2020; Fiorentini & Losa, 2020; Amiri et al., 2020) استفاده شده‌اند و بطور ویژه نیز در مدلسازی و شناسایی عوامل موثر بر شدت تصادفات موتورسیکلت به کار گرفته شده‌اند. برای مثال وهاب و جیانگ (Wahab & Jiang, 2019) از الگوریتم‌های طبقه‌بندی برای ایجاد مدل‌های پیشبینی شدت جراحات ناشی از تصادفات موتورسیکلت استفاده کردند. در این مطالعه قدرت مدل‌های پرسپترون چندلایه، استنتاج قانون و درخت طبقه‌بندی و رگرسیون با هم مقایسه شد. نتایج مطالعه نشان داد که مدل simpleCart با دقت متوسط ۷۳٫۸۱ درصد عملکرد بهتری دارد. همچنین مهمترین عوامل مرتبط با شدت تصادف موتورسیکلت شامل نوع موقعیت، نوع محل تصادف، زمان تصادف، نوع برخورد و طرف مقابل برخورد شناسایی شدند. حاجی دیمیتریو و همکاران (Hadjidimitriou et al., 2019) از روش‌های رگرسیون لجستیک، جنگل تصادفی، ماشین بردار پشتیبان و شبکه عصبی عمیق برای تحلیل شدت تصادفات موتورسیکلت و ویژگی‌هایی که می‌توان به طور منطقی در لحظه تصادف جمع‌آوری نمود استفاده کردند. متغیرهای به کار رفته در این مطالعه شامل سن، سرعت، نوع برخورد، هنگامه روز و هفته و ساعت تماس با مراکز خدمات اضطراری بود. نتایج نشان داد که می‌توان از یک سیستم هوشمند قابل اطمینان و مفید برای طبقه‌بندی شدت تصادفات موتورسیکلت با استفاده از ویژگی‌هایی که می‌توان از سنسورها و اتخاذ سیستم یادگیری ماشین غیرخطی به دست آورد استفاده نمود. رضاپور و همکاران (Rezapour et al., 2020) از درخت طبقه‌بندی و رگرسیون لجستیک برای پیشبینی شدت تصادف موتورسیکلت بر اساس متغیرهای سرعت مجاز، سن، رده عملکردی راه و سرعت تطبیقی استفاده کردند. نتایج نشان داد که عملکرد دو مدل با اختلاف بسیار جزئی به هم شبیه است. ایجاز و همکاران (Ijaz et al., 2021) از الگوریتم‌های جنگل تصمیم، جنگل تصادفی و درخت تصمیم برای تحلیل شدت تصادفات موتورسیکلت‌های سه‌چرخ در پاکستان استفاده کردند. داده‌های تصادف سه‌ساله (۲۰۱۷-۲۰۱۹) شامل ۲۷۴۳ تصادف موتورسیکلت سه‌چرخ برای تحلیل استفاده شد. دقت مدل‌ها مورد ارزیابی قرار گرفت و مدل جنگل تصمیم با دقت کلی ۸۳٫۷ درصد انتخاب شد. نتایج مطالعه نشان داد که عواملی مانند وضعیت روشنایی، رانندگان جوان ۲۰ تا ۳۰ ساله، معابر با سرعت مجاز زیاد، ساعات غیراوج و وضعیت تابش آفتاب

(نفر در سال) ناشی از تصادفات جاده‌ای در دنیا باشد (Brockhus et al., 2024). به منظور رسیدگی به این وضعیت، درک عوامل موثر بر تصادفات موتورسیکلت برای توسعه استراتژی‌های ایمنی به منظور کاهش مرگ و میر ناشی از تصادفات ضروری است (Hsu, 2024). در اکثر مطالعات ایمنی موتورسیکلت، به دلیل نبود حفاظت مناسب برای موتورسیکلت سواران در فرآیند وقوع تصادف، بر عوامل موثر بر شدت تصادف تمرکز شده است (Kashifi, 2023). این عوامل به چهار دسته مهم (۱) ویژگی‌های طرح هندسی راه شامل عرض میانه، تعداد خطوط، عرض خط، نوع راه، نوع شانه و موارد اینچنینی، (۲) خصوصیات و رفتار موتورسیکلت سوار شامل سن، جنسیت، تندرانی، استفاده از تجهیزات ایمنی و نوع مانور، (۳) ویژگی‌های موتورسیکلت شامل توان موتور و نوع یا بدنه موتورسیکلت و (۴) متغیرهای محیطی شامل وضعیت آب و هوا و روشنایی، محل تصادف و هنگامه روز تقسیم شده‌اند (Kashifi, 2023; Rifaat, 2012). با این حال برای شناسایی عوامل موثر بر شدت تصادفات، لازم است تا ابتدا مجموعه‌ای از عوامل مرتبط با هم مورد آزمون قرار گرفته و عوامل موثر شناسایی شوند. در این راستا رویکرد متداول، مدلسازی تصادفات است که در واقع پیش بینی شدت تصادفات بر اساس مجموعه‌ای از متغیرهای پیش بینی کننده است. در ادبیات مدلسازی شدت تصادفات، چهار دسته کلی مدلسازی وجود دارد (Chai et al., 2024): (۱) مدل‌های رگرسیونی آماری مانند لجوجیت و پروبیت، (۲) مدل‌های یادگیری ماشین مانند جنگل تصادفی (RF) و XGBoost، (۳) مدل‌های یادگیری عمیق مانند شبکه عصبی کانولوشن و حافظه طولانی کوتاه مدت و (۴) مدل‌های ترکیبی. چای و همکاران (Chai et al., 2024) با مرور مطالعات گذشته، کاربردهای چهار دسته فوق را مطابق جدول (۱) پیشنهاد داده‌اند. از میان چهار دسته فوق، امروزه روش‌های یادگیری ماشین بصورت چشمگیر برای مدلسازی فراوانی و شدت تصادفات استفاده می‌شوند زیرا قادر به در نظر گرفتن روابط پیچیده و غیرخطی بین متغیرها هستند و در نتیجه برای پیشبینی عوامل موثر بر شدت تصادفات از مدل‌های آماری بهترند (Kashifi et al., 2022 & 2023). این روش‌ها در مطالعات متعددی برای مدلسازی شدت تصادف (Khan et al., 2024; Hassan et al., 2023; Infanti et al., 2022; Panda et al., 2022; Siddik, 2021; Shetty & Liu, 2021; Donchenko et al., 2020; Ghasedi et al., 2021; Zhang et al., 2020; Tariq et al., 2020; Assi et al.,

تصادفات دارند. کشیفی (Kashifi, 2023) از روش‌های XGBoost و تحلیل SHAP برای شناسایی عوامل موثر بر شدت تصادف استفاده کرد. نتایج این مطالعه که از مجموعه داده چهارساله تصادفات در فرانسه شامل سال‌های ۲۰۱۴ تا ۲۰۱۷ برای مدل‌سازی شدت تصادفات موتورسیکلت استفاده شد، نشان داد که گروه تصادف، رده راه، سطح شهرنشینی، نوع موتورسیکلت و سن کاربر به طور قابل توجهی بر شدت تصادفات موثر هستند، علاوه بر این احتمال وقوع آسیب‌های شدید در تصادفات موتورسیکلت در مناطق برونشهری، رانندگان مسن، عدم بستن کلاه ایمنی و خروج از راه بیشتر است.

(درخشندگی) در شدت تصادفات موتورسیکلت‌های سه‌چرخ موثر هستند. پانیکر و رامادورای (Panicker & Ramadurai, 2022) از الگوریتم‌های RF، جنگل استنتاج شرطی (Cforest) و مدل پروبیت تربیتی برای شناسایی عوامل موثر بر شدت تصادفات موتورسیکلت استفاده کردند. محققان داده‌های تصادفات سه ساله (۲۰۱۶-۲۰۱۸) شامل ۷۶۵۴ تصادف موتورسیکلت رخ داده در چنای هندوستان را به کار گرفتند. نتایج نشان داد که الگوریتم Cforest عملکرد بهتری نسبت به دو مدل دیگر دارد. همچنین نتایج نشان داد که نوع برخورد، سن راننده و قابلیت دید راه به ترتیب بیشترین تاثیر را بر شدت

جدول ۱. کاربرد روشهای مختلف مدل‌سازی تصادفات (Chai et al., 2024)

روش	کاربرد پیشنهادی
رگرسیون آماری	<ul style="list-style-type: none"> مجموعه داده با اندازه کوچک رسیدگی به روابط ساده‌تر بین ویژگی‌های تصادفات جاده‌ای
یادگیری ماشین	<ul style="list-style-type: none"> مجموعه داده با اندازه متوسط تا بزرگ تحلیل پیچیده و روابط غیرخطی بین ویژگی‌های تصادفات جاده‌ای رویکردهای متداول برای رسیدگی به کلان داده با تفسیرپذیری پیشبینی
یادگیری عمیق	<ul style="list-style-type: none"> مجموعه داده با اندازه بسیار بزرگ درک الگوهای پیچیده از مجموعه داده‌های با اندازه بسیار بزرگ از تصادفات جاده‌ای
ترکیبی	<ul style="list-style-type: none"> وظایف با نیازهای خاص ایده آل برای پیشبینی با چندین مرحله پردازش رسیدگی به نیاز برای الگوریتم‌های اضافی برای پشتیبانی الگوریتم اصلی

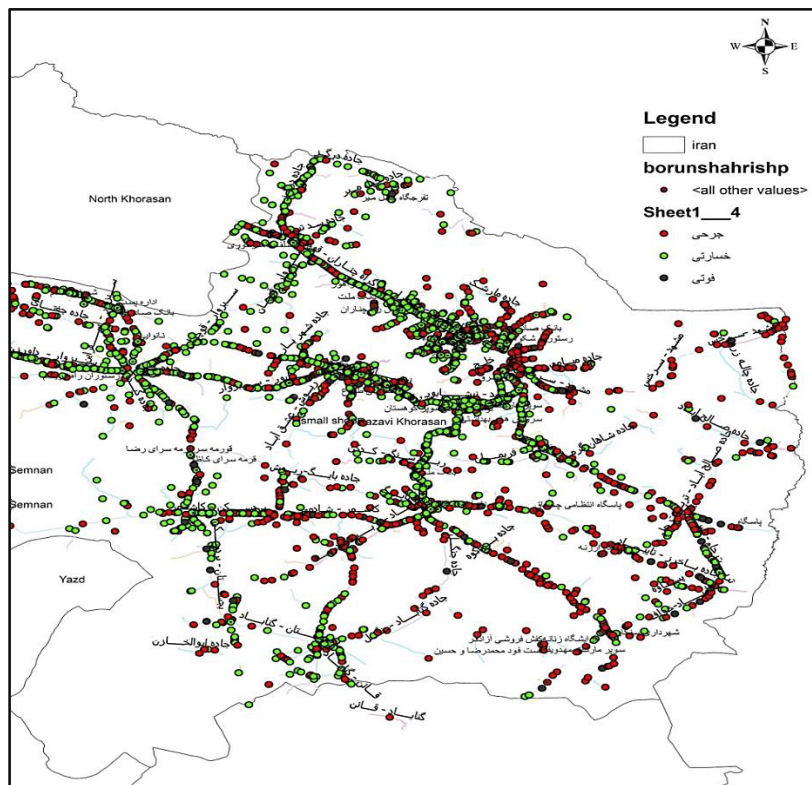
طبقه‌بندی (Chen & Guestrin, 2016) و LightGBM به دلیل افزایش سرعت آموزش و کارایی در کلان داده (Ke et al., 2017) برای پیشبینی شدت تصادفات موتورسیکلت استفاده شده است. در نهایت به منظور تفسیر پیشبینی مدل و شناسایی مهمترین عوامل موثر، از روش SHAP که یک روش پرکاربرد در این رابطه بوده و در مطالعات متعددی (Scarano et al., 2025; Yoon, 2025; Sunkpho et al., 2025; Akter et al., 2025; Hassan et al., 2024; Kashifi, 2023) به کار رفته، استفاده خواهد شد.

از آنجا که هدف اصلی این پژوهش، شناسایی عوامل موثر بر شدت تصادفات موتورسیکلت و تفسیر نتایج (اهمیت عوامل) است و مجموعه داده مورد استفاده نیز در اندازه متوسط تا بزرگ قرار می‌گیرد، بنابراین با توجه به توضیحات جدول (۱)، استفاده از روشهای یادگیری ماشین بهترین گزینه برای رسیدن به هدف پژوهش خواهد بود. همچنین با توجه به اینکه در مطالعات گذشته طیف متنوعی از مدل‌های یادگیری ماشین استفاده شده و همگی نیز کارایی لازم را داشته‌اند (Chai et al., 2024)، در این تحقیق از سه الگوریتم RF به دلیل پایداری زیاد و کاهش خطر بیش برآزش (Breiman, 2001)، XGboost به دلیل قدرت زیاد یادگیری گرادینتی و دقت چشمگیر در مسائل

۲- داده‌ها

همچنین برخی داده‌های اولیه مانند نوع راه بر اساس اطلاعات لایه‌های GIS تطبیق داده شد. در نهایت مجموعه داده نهایی شامل متغیرهای نشان داده شده در جدول (۲) است که توزیع تصادفات بر اساس ویژگی‌های هر متغیر را نشان می‌دهد. این داده‌ها در سیزده دسته شامل شدت، نوع راه، وضعیت هندسی، نوع شانه، نقص موثر راه، موقعیت تصادف، نوع کاربری، محدوده امکان، مانع دید، عارضه ترافیکی، عارضه حمل و نقلی، نوع وسیله نقلیه و نقص وسیله نقلیه دسته‌بندی شده‌اند. در جدول (۲) توزیع درصد عوامل هر دسته بر اساس بیشترین به کمترین مرتب شده است.

مجموعه داده مورد تحلیل در این تحقیق شامل تصادفات سه ساله (۱۳۹۶-۱۳۹۸) رخ داده موتورسیکلت در راه‌های برونشهری استان خراسان رضوی است که با مراجعه به پلیس راه استان خراسان رضوی جمع‌آوری شد. تعداد کل تصادفات شامل ۶۸۰۲ فقره و همگی دارای اطلاعات مکانی (نقاط GPS) بودند. با پیاده‌سازی نقاط تصادف بر روی شبکه راه استان خراسان رضوی با استفاده از نرم افزار GIS، پراکنش داده‌های تصادف مطابق شکل (۱) حاصل شد و اطلاعاتی مانند کاربری زمین و عوارض ترافیکی در محل‌های تصادف به‌دست آمد.



شکل ۱. پراکنش تصادفات موتورسیکلت در شبکه راه استان خراسان رضوی به منظور دستیابی به اطلاعات مکانی و تطبیق داده‌های اولیه

جدول ۲. توصیف داده‌های نهایی و توزیع دسته عوامل و ویژگی‌های هر دسته

ردیف	دسته عامل	عوامل	درصد	عوامل	درصد
۱	شدت	خسارتی	۵۸,۰۹٪	فوتی	۲,۱۶٪
		جرحی	۳۹,۷۵٪		
۲	وضعیت هندسی	مستقیم، مسطح	۷۸,۹۹٪	مستقیم، سربالایی	۲,۵۳٪
		پیچ، مسطح	۵,۷۸٪	پیچ، سربالایی	۱,۹۳٪
		نامشخص	۴,۷۶٪	پل	۰,۲۵٪
		مستقیم، سربالایی	۳,۰۶٪	تونل	۰,۰۳٪
۳	وضعیت خطکشی	پیچ، سربالایی	۲,۶۸٪		
		منقطع	۶۳,۷۰٪	نامشخص	۵,۳۷٪

ردیف	دسته عامل	عوامل	درصد	عوامل	درصد
۴	نوع راه	ممتد	۱۹,۲۹٪	دوبل	۰,۲۴٪
		ندارد	۱۱,۴۱٪		
		بزرگراه بین شهری	۳۵,۸۴٪	راه خاکی	۰,۲۶٪
		راه اصلی	۲۵,۱۵٪	اتصال به راه اصلی	۰,۲۵٪
		راه فرعی درجه ۲	۱۵,۱۷٪	پیااده‌رو	۰,۳۶٪
		راه فرعی درجه ۱	۶,۷۸٪	راه خاکی کیفیت پایین	۰,۱۲٪
		مسکونی	۵,۷۸٪	اتصال به راه فرعی درجه ۱	۰,۱۰٪
		آزادراه	۴,۱۲٪	مسیرهای پیاده‌روی یا چندمنظوره	۰,۰۹٪
		نامشخص	۳,۳۲٪	اتصال به راه فرعی درجه ۲	۰,۰۴٪
		اتصال به بزرگراه بین شهری	۱,۴۰٪	رمپ‌های ورودی خروجی آزادراه	۰,۰۴٪
۵	نوع شانه	رود سرویس	۱,۱۶٪	مسیر ویژه دوچرخه	۰,۰۱٪
		شانه آسفالته	۴۰,۸۱٪	شانه خاکی	۲۴,۶۰٪
		شانه ندارد	۲۷,۰۵٪	نامشخص	۷,۵۴٪
		ندارد	۶۸,۳۸٪	وجود مانع دست انداز	۰,۲۶٪
		کم بودن عرض راه	۱۰,۲۸٪	نقص رویه آسفالت	۰,۲۴٪
		نامشخص	۹,۵۶٪	فقدان شانه خاکی و پارکینگ	۰,۲۴٪
		نقص علائم عمودی	۵,۴۵٪	فقدان حفاظ ایمنی کنار راه	۰,۱۳٪
		سایر...	۳,۰۱٪	اختلاف سطح بین آسفالت و شانه	۰,۱۳٪
		نقص علائم افقی	۰,۷۹٪	غیراستاندارد بودن حفاظ کنار راه	۰,۰۶٪
		نقص روشنایی راه معبر	۰,۶۳٪	شیب عرضی و طولی غیر استاندارد	۰,۰۶٪
۶	نقص موثر راه	نقص خط کشی راه معبر	۰,۴۴٪	نشست راه	۰,۰۳٪
		قوس با زاویه تند	۰,۳۱٪		
		باند سواره رو	۸۶,۲۰٪	کنار راه (حریم)	۱,۸۴٪
		نامشخص	۹,۱۲٪	سایر...	۰,۵۵٪
		شانه	۲,۰۲٪	خارج از حریم راه	۰,۲۸٪
		ندارد	۴۶,۹۷٪	جنگل	۲,۶۳٪
		مزرعه	۲۰,۲۰٪	منطقه حفاظت شده محیطزیست	۲,۴۴٪
		مسکونی	۱۶,۹۱٪	باغ میوه	۲,۳۵٪
		صنعتی	۵,۷۵٪	قبرستان	۰,۷۲٪
		سبزه زار	۰,۶۹٪	شالیزار	۰,۰۷٪
۷	موقعیت تصادف	پارک	۰,۴۰٪	معدن	۰,۰۷٪
		نظامی	۰,۳۱٪	تجاری	۰,۰۳٪
		باغ	۰,۲۸٪	مجتمع خدماتی	۰,۰۱٪
		چمنزار	۰,۱۵٪	باغ انگور	۰,۰۱٪
		عدم وجود ساختمان	۹۱,۰۰٪	درمانی	۰,۲۹٪
		نامشخص	۳,۵۹٪	فرهنگی	۰,۲۶٪
		خدمات خودرو	۱,۵۱٪	تفریحی	۰,۲۱٪
		انتظامی	۰,۷۹٪	اداری	۰,۲۱٪
		صنعتی	۰,۶۳٪	قبرستان	۰,۱۸٪
		مسکونی	۰,۳۵٪	آموزشی	۰,۱۵٪
۸	نوع کاربری	باند سواره رو	۸۶,۲۰٪	کنار راه (حریم)	۱,۸۴٪
		نامشخص	۹,۱۲٪	سایر...	۰,۵۵٪
		شانه	۲,۰۲٪	خارج از حریم راه	۰,۲۸٪
		ندارد	۴۶,۹۷٪	جنگل	۲,۶۳٪
		مزرعه	۲۰,۲۰٪	منطقه حفاظت شده محیطزیست	۲,۴۴٪
		مسکونی	۱۶,۹۱٪	باغ میوه	۲,۳۵٪
		صنعتی	۵,۷۵٪	قبرستان	۰,۷۲٪
		سبزه زار	۰,۶۹٪	شالیزار	۰,۰۷٪
		پارک	۰,۴۰٪	معدن	۰,۰۷٪
		نظامی	۰,۳۱٪	تجاری	۰,۰۳٪
۹	محدوده اماکن	باغ	۰,۲۸٪	مجتمع خدماتی	۰,۰۱٪
		چمنزار	۰,۱۵٪	باغ انگور	۰,۰۱٪
		عدم وجود ساختمان	۹۱,۰۰٪	درمانی	۰,۲۹٪
		نامشخص	۳,۵۹٪	فرهنگی	۰,۲۶٪
		خدمات خودرو	۱,۵۱٪	تفریحی	۰,۲۱٪
		انتظامی	۰,۷۹٪	اداری	۰,۲۱٪
		صنعتی	۰,۶۳٪	قبرستان	۰,۱۸٪
		مسکونی	۰,۳۵٪	آموزشی	۰,۱۵٪

ردیف	دسته عامل	عوامل	درصد	عوامل	درصد
۱۰	مانع دید	تجاری	۰,۲۹٪	توالث عمومی	۰,۱۳٪
		باستانی	۰,۲۹٪	جایگاه سوخت	۰,۱۰٪
		ندارد	۸۴,۶۷٪	تل خاک و مشابه آن	۰,۴۱٪
		نامشخص	۷,۵۶٪	شیب	۰,۴۰٪
		وسيله در حال حرکت	۲,۵۷٪	درخت یا بوته	۰,۳۸٪
		قوس قائم	۱,۱۳٪	نور چراغ وسیله نقلیه مقابل	۰,۱۳٪
		سایر...	۰,۸۴٪	کولاک	۰,۱۲٪
		مه یا دود	۰,۶۳٪	نور خورشید	۰,۰۹٪
		وسيله متوقف	۰,۵۱٪	طوفان شن	۰,۰۶٪
		ساختمان یا کیوسک	۰,۴۹٪	تابلو	۰,۰۱٪
۱۱	عارضه ترافیکی	ندارد	۹۰,۶۸٪	تقاطع	۰,۴۴٪
		ورودی خروجی بزرگراه	۳,۴۳٪	پارکینگ	۰,۳۲٪
		تیر روشنایی	۱,۶۶٪	آبرو	۰,۰۳٪
		دوربین کنترل سرعت	۱,۵۴٪	مجتمع	۰,۰۱٪
		جایگاه سوخت	۱,۳۴٪	مجتمع خدمات رفاهی	۰,۰۱٪
		دوربرگردان	۰,۵۱٪	چراغ راهنمایی	۰,۰۱٪
		ندارد	۹۷,۸۴٪	پد بالگرد اورژانس	۰,۲۲٪
		ایستگاه اتوبوس	۱,۹۱٪	ایستگاه اه آهن	۰,۰۳٪
		عجله و شتاب بی مورد	۵۳,۷۹٪	خستگی و خواب آلودگی	۳,۵۶٪
		نامشخص	۹,۵۰٪	عدم آشنایی با جاده	۰,۵۷٪
۱۲	عارضه حمل و نقلی	سایر عوامل ...	۹,۱۶٪	مهاری نکردن محموله به طرز صحیح	۰,۳۴٪
		ندارد	۹,۰۱٪	ضعف ناشی از کهولت سن	۰,۲۱٪
		عدم تشخیص سهم عبور سایرین	۷,۲۶٪	مصرف مشروبات الکلی	۰,۱۲٪
		نقص عضو موثر	۶,۴۷٪	استعمال مواد مخدر	۰,۰۱٪
		سواری	۶۱,۵۶٪	ادوات کشاورزی	۰,۳۵٪
		موتورسیکلت	۱۳,۸۶٪	مینی بوس	۰,۳۱٪
		وانت بار	۱۰,۱۴٪	نامشخص	۰,۱۵٪
		کامیون	۸,۴۵٪	ادوات راه سازی	۰,۰۶٪
		اتوبوس	۱,۵۷٪	تانکر حمل مواد خطرناک	۰,۰۳٪
		تریلی	۱,۵۱٪	آمبولانس	۰,۰۳٪
۱۳	عامل انسانی	کامیونت	۱,۲۲٪	سایر...	۰,۰۳٪
		دوچرخه	۰,۷۱٪	خودروپلیس	۰,۰۱٪
		ندارد	۸۷,۲۴٪	صاف بودن لاستیک	۰,۲۲٪
		نامشخص	۸,۳۲٪	فقدان زنجیر چرخ	۰,۰۶٪
		سایر نقایص ...	۳,۲۶٪	نقص دستگاه تعلیق	۰,۰۶٪
		نقص سیستم ترمز	۰,۴۱٪	نقص سیستم فرمان	۰,۰۶٪
		نقص سیستم روشنایی	۰,۳۷٪		
		نقص سیستم ترمز	۰,۴۱٪		
		نقص سیستم روشنایی	۰,۳۷٪		
		نقص سیستم ترمز	۰,۴۱٪		
۱۴	نوع وسیله نقلیه مقرر	سواری	۶۱,۵۶٪	ادوات کشاورزی	۰,۳۵٪
		موتورسیکلت	۱۳,۸۶٪	مینی بوس	۰,۳۱٪
		وانت بار	۱۰,۱۴٪	نامشخص	۰,۱۵٪
		کامیون	۸,۴۵٪	ادوات راه سازی	۰,۰۶٪
		اتوبوس	۱,۵۷٪	تانکر حمل مواد خطرناک	۰,۰۳٪
		تریلی	۱,۵۱٪	آمبولانس	۰,۰۳٪
		کامیونت	۱,۲۲٪	سایر...	۰,۰۳٪
		دوچرخه	۰,۷۱٪	خودروپلیس	۰,۰۱٪
		ندارد	۸۷,۲۴٪	صاف بودن لاستیک	۰,۲۲٪
		نامشخص	۸,۳۲٪	فقدان زنجیر چرخ	۰,۰۶٪
۱۵	نقص وسیله نقلیه	سایر نقایص ...	۳,۲۶٪	نقص دستگاه تعلیق	۰,۰۶٪
		نقص سیستم ترمز	۰,۴۱٪	نقص سیستم فرمان	۰,۰۶٪
		نقص سیستم روشنایی	۰,۳۷٪		
		نقص سیستم ترمز	۰,۴۱٪		
		نقص سیستم روشنایی	۰,۳۷٪		
		نقص سیستم ترمز	۰,۴۱٪		
		نقص سیستم روشنایی	۰,۳۷٪		
		نقص سیستم ترمز	۰,۴۱٪		
		نقص سیستم روشنایی	۰,۳۷٪		
		نقص سیستم ترمز	۰,۴۱٪		

۳- روش تحقیق

و i از ۱ تا N (تعداد نقاط داده) متغیر است. دسته پیش بینی شده نهایی C_i برای نقطه داده i ام مطابق رابطه (۱) تعیین می‌شود (Champahom et al., 2024).

$$C_i = \operatorname{argmax}_k p_{ik} \quad (1)$$

و بدین معنا است که دسته پیش‌بینی شده، دسته‌ای است که بالاترین احتمال پیش‌بینی شده را در بین تمام درختان جنگل تصادفی دارد.

۳-۲- XGBoost

XGBoost در حوزه‌های مطالعاتی متعدد نتایج امیدوارکننده‌ای را به همراه داشته است (Hasan et al., 2024). این الگوریتم شامل مجموعه‌ای از درختان تصمیم است که هر کدام از درختان قبلی خود یاد می‌گیرند و بر درخت بعدی تأثیر می‌گذارند. فرمول بندی XGBoost با توابع درخت k به شرح رابطه (۲) است (Hasan et al., 2024).

$$\hat{q}_i^{(t)} = \sum_{k=1}^t f_k(p_i) = \hat{q}_i^{(t-1)} + f_t(p_i) \quad (2)$$

که $\hat{q}_i^{(t)}$ شدت تصادف تخمینی بعد از تکرارهای t ام، k تعداد درختان اضافی، t تعداد تکرارها، $f_k(p_i)$ تابع درخت k ام برای متغیرهای p_i مقدار $\hat{q}_i^{(t-1)}$ مقدار پاسخ پیش‌بینی شده برای تکرار نهایی و $f_t(p_i)$ تابع درخت تکرار t ام است. تابع هدف برای بهینه‌سازی $l(q_i, \hat{q}_i)$ از دست رفته را می‌توان بصورت زیر نوشت. که $\Omega(f_t)$ عبارت تنظیم برای جلوگیری از بیش‌برازش و کاهش پیچیدگی، T تعداد برگها، ω_j^2 نرم L_2 امتیازات گره t ام و n تعداد کل تصادفات در داده‌ها است.

$$obj = \sum_{k=1}^n l(q_i, \hat{q}_i) + \sum_{k=1}^t \Omega(f_k) \quad (3)$$

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (4)$$

در این پژوهش از سه الگوریتم RF، XGBoost و LightGBM برای مدلسازی شدت تصادفات موتورسیکلت استفاده می‌شود. روش RF به طور گسترده در ادبیات ایمنی راه برای تحلیل کلان داده با چندین متغیر مستقل بررسی شده است و محققان بیان داشته‌اند که برای تحقیقات ایمنی راه مناسب است (yan & shen, 2022; chen & chen, 2020). علاوه بر این، الگوریتم‌های تقویتی مانند XGBoost و LightGBM نیز برای پیش‌بینی تصادفات جاده‌ای و سطوح شدت آن‌ها مورد بررسی قرار گرفته‌اند و دقت پیش‌بینی بهتری با هزینه‌های محاسباتی کمتر به دست آمده است (Ma et al., 2021; Parsa et al., 2020; Qu et al., 2019).

۳-۱- جنگل تصادفی (RF)

RF یک طبقه‌بندی‌کننده مبتنی بر درخت است. RF از برچسب زدن (bagging) و انتخاب ویژگی تصادفی استفاده می‌کند که دو تکنیک مجزا برای دسته‌بندی هستند (Breiman, 2001). برچسب زدن، هر درخت را بصورت جداگانه ایجاد می‌کند، درحالی‌که مجموعه ویژگی تصادفی، درختان تصمیم را به سرعت ایجاد می‌کند. RF ویژگی‌های زیرگروه‌ها را بصورت تصادفی انتخاب می‌کند (برخلاف استفاده از هر ویژگی در درختان تصمیم) و خروجی مجموعه داده جدید را با میانگین‌گیری خروجی‌های داده‌های آموزشی bootstrap تصادفی جداگانه پیش‌بینی می‌کند (Hasan et al., 2024). از لحاظ ریاضی احتمالات دسته پیش‌بینی شده برای نقطه داده i ام با p_{ik} نشان می‌دهیم، که در آن k از ۱ تا k (تعداد دسته‌های جراثت)

LightGBM - ۳-۳

می کند که فرآیند یادگیری بر خطاهای بحرانی تر متمرکز است (Champahom et al., 2024). این رویکرد به LightGBM اجازه می دهد تا دقت یادگیری مبتنی بر گرادیان را حفظ کند و در عین حال پیچیدگی محاسباتی را به طور قابل توجهی کاهش دهد. بسته بندی ویژگی های انحصاری^۴ (EFB) : EFB با بسته بندی ویژگی های متقابلا انحصاری (آنهايي که بعيد است همزمان غیر صفر باشند)، به چالش فضاهای ویژگی با ابعاد بالا می پردازد و در نتیجه تعداد ویژگی ها را بدون از دست دادن قابل توجه اطلاعات کاهش دهد (Ke et al., 2017). این بر اساس مشاهده ای است که در بسیاری از مجموعه داده ها، به ویژه مجموعه های پراکنده، فضای ویژگی می تواند بدون تاثیر بر قدرت پیش بینی مدل فشرده شود (Champahom et al., 2024).

این الگوریتم با تمرکز بر سرعت و کارایی بدون به خطر انداختن دقت مدل، روش های سنتی تقویت گرادیان را بهینه می کند (Champahom et al., 2024). این الگوریتم از یک الگوریتم مبتنی بر هیستوگرام برای یادگیری درخت تصمیم استفاده می کند، استفاده از حافظه را کاهش می دهد و سرعت را بهبود می بخشد (Ke et al., 2017). نوآوری کلیدی در این الگوریتم شامل موارد زیر است: الف) نمونه برداری یک طرفه مبتنی بر گرادیان^۳ (GOSS) : GOSS یکی از ویژگی های کلیدی LightGBM است که کارایی فرآیند تقویت گرادیان را افزایش می دهد (Ke et al., 2017). این الگوریتم نمونه هایی با گرادیان های بزرگتر را در اولویت قرار می دهد، زیرا برای یادگیری آموزنده تر در نظر گرفته می شوند. این الگوریتم داده ها را به دو گروه تقسیم می کند: A (شامل نمونه هایی با بزرگترین گرادیان) و B (شامل نمونه های باقی مانده با گرادیان های کوچکتر). از یک ثابت α برای نگه داشتن همه نمونه ها در A و انتخاب تصادفی کسری (b) نمونه از B استفاده می شود و تضمین

$$A = \left\{ (x_i, y_i) \mid \text{بزرگترین گرادیان} \right\}, \left\{ (x_i, y_i) \mid \text{کوچکترین گرادیان} \right\} \quad (5)$$

۳-۴- ارزیابی عملکرد مدل

منتخب و دارای تعادل خوب بین Precision و Recall در نظر گرفته می شود و نشان دهنده اثربخشی آن در طبقه بندی شدت تصادفات است (Sunkpho et al., 2025).

کارایی و عملکرد مدل های یادگیری ماشین با استفاده از معیارهای کلیدی نشان داده شده در معادلات (۶) تا (۹) ارزیابی می شود (Sunkpho et al., 2025). مدلی که بیشترین امتیاز را در این معیارها داشته باشد، به عنوان مدل

$$Accuracy = \frac{\text{تعداد دسته صحیح}}{\text{تعداد کل دسته ها}} \quad (6)$$

$$Precision = \frac{\text{مثبت صحیح}}{\text{مثبت غلط} + \text{مثبت صحیح}} \quad (7)$$

$$Recall = \frac{\text{مثبت صحیح}}{\text{منفی غلط} + \text{مثبت صحیح}} \quad (8)$$

$$F1 \text{ score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

۳-۵- تحلیل SHAP

شود (Scarano et al., 2025) که در آن مقدار ϕ_i مقدار shapley ویژگی i است؛ S زیرمجموعه ویژگی به دست آمده از X است که مجموعه تمام ویژگی‌هاست؛ وزن احتمال S محاسبه شده بعد از جایگشت و ترکیب ویژگی است؛ $x_{S \cup \{i\}}$ و x_S مقادیر در ویژگی‌های ورودی در اجتماع مجموعه S با به ترتیب i و S است؛ $f_{S \cup \{i\}}(x_{S \cup \{i\}})$ پیش‌بینی مدل با در نظر گرفتن ویژگی i برای زیرمجموعه S است؛ $f_S(x_S)$ پیش‌بینی مدل بدون در نظر گرفتن ویژگی i برای زیرمجموعه S است؛ و $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$ سهم حاشیه‌ای ویژگی i برای پیش‌بینی را نشان می‌دهد.

پس از به دست آوردن مدل بهینه بر اساس معیارهای عملکرد، نتایج پیش‌بینی آن توضیح داده می‌شود. با الهام گرفتن از نظریه تئوری بازی، توضیحات اضافی Shapley (SHAP) تمام ویژگی‌ها را همانگونه که در خروجی نهایی $f(x)$ سهم هستند در نظر می‌گیرد (Scarano et al., 2025). تکنیک SHAP مبتنی بر مقادیر Shapley (Shapley, 1953) است به گونه‌ای که یک مقدار مهم را به هر ویژگی، مطابق با سهم حاشیه‌ای با در نظر گرفتن تمام سهم‌های ویژگی‌های ممکن، اختصاص می‌دهد. مقادیر SHAP پیش‌بینی مدل را به مجموع سهم‌ها از هر ویژگی تجزیه می‌کند. بنابراین پیش‌بینی می‌تواند بصورت مجموع متوسط پیش‌بینی و سهم‌های هر ویژگی توضیح داده شود. مقدار shapley ویژگی i بصورت معادله (۱۰) محاسبه می‌شود.

$$\phi_i = \sum_{S \subseteq X \setminus \{i\}} \frac{|S|!(|X|-|S|-1)!}{|X|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (10)$$

۴- تحلیل داده و نتایج

مراحل تحلیل داده و دستیابی به نتایج که در نرم افزار پایتون انجام شده به شرح زیر است.

۴-۲- مدل‌سازی

برای پیش‌بینی شدت تصادف، سه مدل یادگیری ماشین RF، XGBoost و LightGBM آموزش داده شدند. پس از آموزش، ماتریس سردرگمی مدل‌ها به دست آمد و عملکرد هر مدل با استفاده از معیار FI-score مورد ارزیابی قرار گرفت (جدول ۳).

ماتریس سردرگمی مدل RF (شکل ۲-الف) نشان می‌دهد که این مدل ۵۱۹ تصادف جرحی، ۵۱۶ تصادف خسارتی و ۷۴۸ تصادف فوتی را به درستی پیش‌بینی نموده است. همچنین دقت کل مدل ۷۶ درصد است.

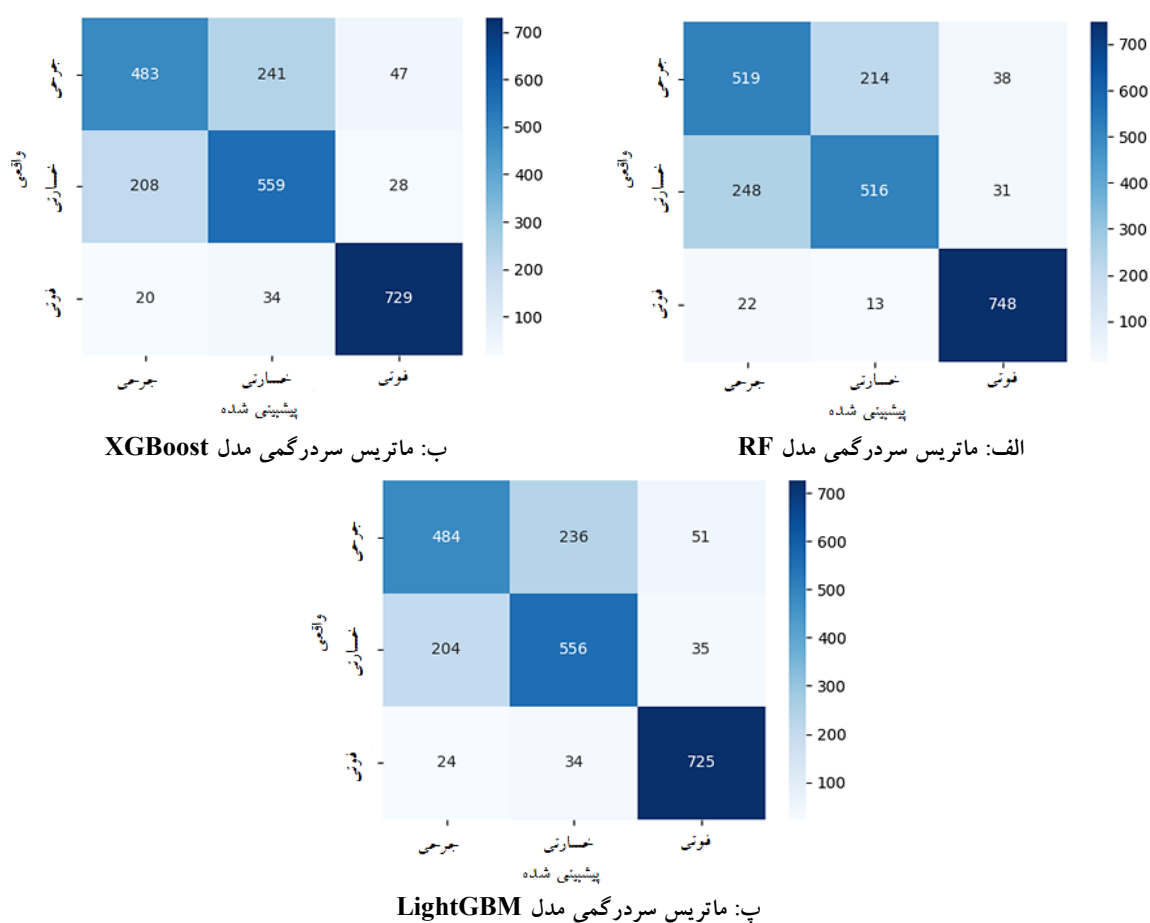
ماتریس سردرگمی مدل XGBoost (شکل ۲-ب) نشان می‌دهد که این مدل ۴۸۳ تصادف جرحی، ۵۵۹ تصادف خسارتی و ۷۲۹ تصادف فوتی را به درستی پیش‌بینی نموده است. همچنین دقت کل مدل ۷۵ درصد است.

۴-۱- پیش پردازش داده

برای آماده‌سازی داده‌ها ابتدا تمام متغیرهای متنی با استفاده از LabelEncoder کدگذاری و به مقادیر عددی تبدیل شدند. سپس با حذف ستون‌های غیرضروری مانند "کد منحصر به فرد تصادف" و تبدیل متغیر هدف "شدت تصادف" به اعداد، داده‌ها به دو بخش X (ویژگی‌ها) و Y (برچسب) تقسیم شدند. نهایتاً به منظور کاهش عدم توازن کلاس‌ها با استفاده از تکنیک SMOTE^۶ نمونه‌های کلاس‌های با فراوانی کمتر، با تولید نمونه‌های مصنوعی افزایش یافت تا کلاس‌ها متعادل شوند و مجموعه داده‌ای متوازن فراهم شود. داده‌های متوازن شده به نسبت ۸۰ به ۲۰ به ترتیب به دو مجموعه آموزش و آزمون تقسیم شدند و مدل‌ها با اعمال وزن‌دهی متوازن به کلاس‌ها آموزش دیدند.

مقایسه نتایج مدل‌سازی سه مدل یادگیری ماشین (جدول ۳) نشان می‌دهد که مدل RF بهترین دقت کلی و بهترین عملکرد را در پیشبینی تصادفات فوتی و جرحی داشته است. بنابراین برای ادامه کار و تحلیل SHAP انتخاب می‌شود.

ماتریس سردرگمی مدل LightGBM (شکل ۲-پ) نشان می‌دهد که این مدل ۴۸۴ تصادف جرحی، ۵۵۶ تصادف خسارتی و ۷۲۵ تصادف فوتی را به درستی پیش‌بینی نموده است. همچنین دقت کلی مدل ۷۵ درصد است.



شکل ۲. ماتریس سردرگمی مدل‌های RF، XGBoost و LightGBM

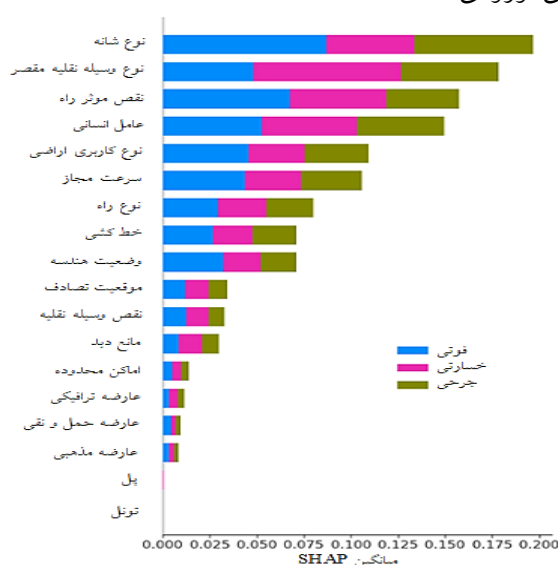
جدول ۳. نتایج ارزیابی عملکرد مدل‌های یادگیری ماشین

مدل	کلاس	precision	recall	f1-score
RF	جرحی	۰,۶۶	۰,۶۷	۰,۶۷
	خسارتی	۰,۶۹	۰,۶۵	۰,۶۷
	فوتی	۰,۹۲	۰,۹۶	۰,۹۴
	دقت کلی (accuracy)			۰,۷۶
XGBoost	جرحی	۰,۶۸	۰,۶۳	۰,۶۵
	خسارتی	۰,۶۷	۰,۷۰	۰,۶۹
	فوتی	۰,۹۱	۰,۹۳	۰,۹۲
	دقت کلی (accuracy)			۰,۷۵
LightGBM	جرحی	۰,۶۸	۰,۶۳	۰,۶۵
	خسارتی	۰,۶۷	۰,۷۰	۰,۶۹
	فوتی	۰,۸۹	۰,۹۳	۰,۹۱
	دقت کلی (accuracy)			۰,۷۵

۳-۴- تحلیل SHAP

در این تحلیل، از مدل RF به عنوان مدل برتر استفاده و مقادیر SHAP برای پیشبینی آن محاسبه گردید. شکل (۳) نمودار حاصل از تحلیل SHAP که تأثیر متوسط هر ویژگی بر خروجی مدل است را نمایش می‌دهد؛ به‌گونه‌ای که محور افقی میانگین قدر مطلق مقادیر SHAP (نشان‌دهنده شدت تأثیرگذاری ویژگی‌ها بر پیش‌بینی شدت تصادف مدل) و محور عمودی ویژگی‌های ورودی مدل است.

مطابق شکل (۳) از میان ویژگی‌های انتخاب‌شده برای تحلیل، متغیرهای نوع شانه، نوع وسیله نقلیه مقصر، راه، عامل انسانی مؤثر در تصادف، کاربری اراضی، سرعت مجاز، نوع راه، خط‌کشی و وضعیت هندسی به عنوان مهم ترین عوامل مؤثر بر شدت تصادفات موتورسیکلت در راه‌های برون‌شهری شناسایی شدند.



شکل ۳. میانگین اهمیت ویژگی‌ها بر اساس مقادیر SHAP

۴-۴- بحث

همچنین حاشیه راه عاری از مانع می‌تواند به کاهش شدت تصادفات موتورسیکلت کمک کند.

عامل دوم- نوع وسیله نقلیه مقصر: این عامل، دومین عامل با اهمیت بالا در تحلیل SHAP است. اگرچه خودروهای سواری بیشترین سهم را در داده‌های توصیفی دارند، اما اهمیت این متغیر بیش از آنکه ناشی از فراوانی باشد، به تفاوت جرم، سختی بدنه و سرعت برخورد بازمی‌گردد. برخورد نامتقارن میان موتورسیکلت و وسایل نقلیه سبک و سنگین (به دلیل تفاوت حجم و سختی بدنه)، منجر به انتقال انرژی بالاتر به بدن موتورسیکلت‌سوار و شدت بیشتر تصادفات می‌شود (Quddus et al., 2002; Pai & Saleh, 2008; Pai & Saleh, 2007; Lapparent, 2006). مدل RF این الگوی فیزیکی را به خوبی درک کرده و آن را به‌عنوان یک عامل تشدیدکننده شدت شناسایی نموده است. این نتیجه تأکیدی بر آسیب‌پذیری ذاتی موتورسیکلت‌سواران در سیستم حمل‌ونقل ترکیبی است. در این راستا اتخاذ تصمیمات سیاسی و ملی برای تجهیز نمودن موتورسیکلت به تجهیزات حفاظتی بیشتر و همچنین اعمال قانون و فرهنگ سازی در خصوص استفاده از تجهیزات حفاظت فردی مانند کلاه ایمنی می‌تواند به کاهش شدت تصادفات موتورسیکلت کمک نماید.

عامل سوم- نقص موثر راه: این عامل علی‌رغم سهم نسبتاً کم در جدول توصیفی، اهمیت بالایی در SHAP دارد. این یافته نشان می‌دهد که نقص راه یک متغیر کم‌وقوع اما پریسک است. وجود نقص‌هایی نظیر کمبود عرض راه، نقص علائم، خرابی رویه یا فقدان تجهیزات ایمنی، در صورت وقوع تصادف، احتمال بروز پیامدهای شدید را به‌طور قابل‌توجهی افزایش می‌دهد (Savolainen & Mannering, 2008; Schneider & Savolainen, 2011; Geedipally et al., 2011). این نتیجه نشان‌دهنده برتری رویکرد یادگیری ماشین در شناسایی اثرات غیرخطی و شرطی متغیرهاست و بیان می‌کند که اولویت‌بندی مداخلات ایمنی نباید صرفاً براساس فراوانی نقص‌ها انجام شود. در کل اهمیت نقص راه در شدت تصادفات موتورسیکلت نشان‌دهنده اهمیت طراحی و نگهداری

همانگونه که اشاره شد، تحلیل SHAP مهمترین عوامل موثر بر شدت تصادفات موتورسیکلت در راه‌های برونشهری را نشان داد (شکل ۵). این بخش با هدف تفسیر نتایج و ایجاد پیوند منسجم میان آمار توصیفی داده‌ها، عملکرد مدل منتخب (RF) و خروجی تحلیل SHAP ارائه شده است. رویکرد اتخاذشده در این بحث، عبور از تفسیر صرف فراوانی متغیرها و تمرکز بر نقش آن‌ها در تشدید پیامد تصادف است. بدین معنا که نتایج SHAP نشان می‌دهند کدام عوامل، مستقل از میزان وقوع، در انتقال تصادف از سطح خسارتی به جرحی و فوتی نقش کلیدی ایفا می‌کنند. یافته‌های این پژوهش نشان می‌دهد که شدت تصادفات موتورسیکلت در راه‌های برون‌شهری حاصل برهم‌کنش پیچیده‌ای از ویژگی‌های زیرساختی، نوع وسیله نقلیه درگیر و رفتار انسانی است. مدل جنگل تصادفی، به‌واسطه توانایی در شناسایی روابط غیرخطی و اثرات شرطی، موفق شده است این برهم‌کنش‌ها را به‌صورت معنادار استخراج کند.

عامل اول- نوع شانه: نوع شانه راه به‌عنوان مهم‌ترین عامل موثر بر شدت تصادفات شناسایی شد. اگرچه در نگاه توصیفی، شانه آسفالتی بیشترین سهم وقوع را دارد، اما اهمیت بالایی آن در SHAP بیانگر نقش رفتار ساز این عنصر زیرساختی است. شانه آسفالتی، به‌ویژه در راه‌های برون‌شهری، عملاً به‌عنوان یک باند غیررسمی برای تردد موتورسیکلت یا سبقت‌گیری تلقی می‌شود. این ویژگی موجب افزایش سرعت، کاهش فاصله عرضی با جریان ترافیک و افزایش احتمال برخوردهای شدید می‌گردد. از سوی دیگر، شانه‌های خاکی یا فاقد شانه، با ایجاد ناپایداری حرکتی، احتمال خروج از راه را افزایش داده و این نوع تصادفات غالباً با شدت بالاتری همراه هستند. بنابراین، نتایج نشان می‌دهد که صرف وجود شانه، بدون توجه به نوع، عرض و کیفیت آن، تضمین‌کننده ایمنی نیست. در نتیجه وجود شانه‌های عریض و آسفالتی و

اصولی راه و زیرساخت موجود است. کنترل طرح هندسی و نگهداری به موقع و دوره‌ای علائم و تجهیزات و خرابی‌های سطح راه می‌تواند در کاهش شدت تصادفات موتورسیکلت موثر باشد.

عامل چهارم- عامل انسانی: این عامل به‌عنوان یکی از متغیرهای کلیدی، نقش تعیین‌کننده‌ای در شدت تصادفات دارد. داده‌های توصیفی نشان می‌دهند که عجله و شتاب بی‌مورد سهم غالب را در میان عوامل انسانی دارند و SHAP این عامل را به‌عنوان تشدیدکننده شدت معرفی می‌کند. این یافته که با نتایج مطالعات پیشین (Wang et al., 2014; Xin et al., 2017) نیز منطبق است نشان می‌دهد که رفتارهای پرخطر، حتی در شرایط زیرساختی نسبتاً ایمن، می‌توانند شدت تصادف را به سطح جرحی یا فوتی سوق دهند. در نتیجه سیاست‌های ایمنی مبتنی صرف بر بهبود زیرساخت، بدون مداخلات رفتاری و کنترلی، اثربخشی محدودی خواهند داشت. فرهنگ سازی مناسب و آموزش‌های دوره‌ای از طریق مراکز گواهینامه و رسانه‌های ملی و محلی می‌تواند سبب رعایت قوانین ترافیکی توسط موتورسیکلت‌سواران و کاهش شدت تصادفات آنها شود.

عامل پنجم- کاربری اراضی: این عامل به‌عنوان پنجمین عامل مهم، نشان‌دهنده نقش محیط در شدت تصادفات است. تمرکز تصادفات در اراضی کشاورزی و نواحی فاقد کاربری مشخص، بیانگر تردد بالای موتورسیکلت برای فعالیت‌های شغلی محلی و دسترسی‌های متعدد و غیرایمن به راه‌های برون‌شهری است. در چنین محیط‌هایی، ترکیب سرعت بالای وسایل نقلیه عبوری و حضور کاربران آسیب‌پذیر، احتمال بروز تصادفات شدید را افزایش می‌دهد. این نتیجه نشان می‌دهد که ایمنی موتورسیکلت در راه‌های برون‌شهری نیازمند رویکردهای مبتنی بر مکان و کاربری زمین است. تعریض راه‌های برون‌شهری و هموارسازی و ایمن سازی حاشیه راه می‌تواند در کاهش وقوع و شدت تصادفات موتورسیکلت به‌ویژه در کاربری‌های کشاورزی موثر باشد.

عامل ششم- سرعت مجاز: سرعت اولین عامل در تقریباً یک چهارم تلفات جاده‌ای است (Montella et al., 2024). از این رو سرعت مجاز می‌بایست تعادل مناسبی بین ایمنی و جریان ترافیک ایجاد کند (Wang et al., 2018). در این تحقیق، سرعت مجاز، نوع راه و وضعیت هندسی مسیر، یک خوشه اثرگذار را در نتایج SHAP تشکیل می‌دهند. اهمیت هم‌زمان این متغیرها نشان می‌دهد که شدت تصادفات موتورسیکلت بیش از آنکه تابع یک عامل منفرد باشد، حاصل سیستم سرعت حاکم بر راه است. راه‌های مستقیم و مسطح با سرعت مجاز بالا، از طریق ایجاد احساس ایمنی کاذب، رفتارهای پرخطر را تشدید کرده و در صورت وقوع تصادف، پیامدهای شدیدتری رقم می‌زنند. این نتیجه با نظریه جبران ریسک همخوانی دارد و نشان می‌دهد که یکنواختی هندسی می‌تواند به‌طور غیرمستقیم شدت تصادف را افزایش دهد. اگرچه فرهنگ سازی مناسب و آموزش دقیق می‌تواند موثر باشد، اما رویکرد مهندسی در این رابطه، تدقیق سرعت است (مطالعه مجدد سرعت در مسیر و تنظیم مجدد سرعت مجاز یا سرعت منطقه‌ای). مطالعات نشان داده‌اند که کاهش سرعت مجاز می‌تواند سبب کاهش شدت تصادفات شود (Lanzaro & Andrade, 2023). بنابراین لازم است که متولیان ایمنی، راه‌های برون‌شهری موجود را مورد مطالعات سرعت قرار داده و در صورت نیاز سرعت مجاز را کاهش دهند تا ایمنی لازم فراهم شود.

عامل هفتم- نوع راه: این عامل رابطه مستقیمی با نوع کاربری، حجم تردد موتورسیکلت و سرعت تردد دارد. در بزرگراه‌های برون‌شهری به دلیل سرعت بیشتر و احتمال خطای انسانی، احتمال وقوع و شدت تصادفات موتورسیکلت بیشتر است. همچنین در راه‌های اصلی و فرعی برون‌شهری به دلیل وجود کاربری‌های کشاورزی بیشتر و تردد بیشتر موتورسیکلت، احتمال وقوع و شدت بیشتر تصادفات موتورسیکلت محتمل است. در کل شدت تصادفات موتورسیکلت در راه‌های برون‌شهری به دلیل سرعت بیشتر ترافیک و همچنین احتمال رخ دادن تصادفات

احتمال وقوع و شدت تصادف را بیشتر کند. مطالعات پیشین این مورد را تایید نموده‌اند (Savolainen & Mannering, 2008; Pai & Saleh, 2008; Shaheed et al., 2013). همچنین عدم کنترل موتورسیکلت در قوس‌ها که می‌تواند به دلیل سرعت زیاد باشد، سبب خواهد شد که موتورسیکلت از مسیر خارج شده که معمولاً با شدت بیشتر تصادف همراه خواهد بود. مطالعات پیشین (Chang et al., 2016; Cunto & Ferreira, 2017) نیز تایید کرده‌اند که تصادفات خروج از راه موتورسیکلت با شدت بیشتری همراه هستند. آموزش مناسب و فرهنگ‌سازی به همراه استفاده از تجهیزات مناسب مانند حفاظ کناری در قوس‌ها می‌تواند در کاهش شدت تصادفات موتورسیکلت در این خصوص موثر باشد. در مجموع، تحلیل SHAP نشان می‌دهد که عوامل زیرساختی، در تعامل با رفتار انسانی و نوع وسیله نقلیه، نقش محوری در شدت تصادفات موتورسیکلت دارند.

رخ به رخ، بیشتر است. این مورد در مطالعات پیشین (Kashifi, 2023) ذکر شده است. همانگونه که اشاره شد بازطراحی و تعریض راه، همچنین نگهداری دوره‌ای و مناسب راه‌های برونشهری می‌تواند در کاهش شدت تصادفات موتورسیکلت موثر باشد.

عامل هشتم - خط‌کشی: رابطه وضعیت خط‌کشی با شدت تصادف در رعایت قوانین ترافیکی است. اکثر تصادفات موتورسیکلت در راه‌های برونشهری با خطوط منقطع و ممتد رخ داده است. سبقت غیرمجاز اصلی‌ترین دلیل در این مورد است. آموزش و فرهنگ‌سازی موثرترین روش در کاهش شدت تصادف برای این مورد است.

عامل نهم - وضعیت هندسه مسیر: اکثر تصادفات موتورسیکلت در مسیرهای مستقیم مسطح و پیچ مسطح رخ داده‌اند. در مسیرهای مستقیم و به دلیل هموار بودن، موتورسیکلت سوار اقدام به تندرانی کرده که این عامل در کنار عدم تجهیزات حفاظت کافی موتورسیکلت می‌تواند

۵- نتیجه‌گیری

هدف اصلی این پژوهش، شناسایی عوامل کلیدی مؤثر بر شدت تصادفات موتورسیکلت در راه‌های برونشهری با استفاده از رویکردهای یادگیری ماشین تفسیرپذیر بود. بدین منظور، از مجموعه داده تصادفات سه‌ساله (۱۳۹۶-۱۳۹۸) استان خراسان رضوی شامل ۶۸۰۲ فقره تصادف موتورسیکلت استفاده شد که پس از تکمیل اطلاعات مکانی و توصیفی، در قالب سیزده دسته عامل مورد تحلیل قرار گرفت. در این تحقیق، سه الگوریتم یادگیری ماشین شامل جنگل تصادفی (RF)، XGBoost و LightGBM برای پیش‌بینی شدت تصادفات به کار گرفته شدند. نتایج ارزیابی عملکرد نشان داد که مدل RF با دقت کلی ۷۶ درصد و عملکرد بهتر در تفکیک تصادفات فوتی و جرحی، نسبت به دو مدل دیگر برتری دارد. این نتیجه بیانگر توانایی بالاتر RF در مدل‌سازی روابط غیرخطی و اثرات تعاملی میان متغیرهای مؤثر بر شدت تصادف است. برای تفسیر خروجی مدل منتخب و شناسایی

عوامل تعیین‌کننده شدت، از تحلیل SHAP استفاده شد. نتایج این تحلیل نشان داد که مهم‌ترین عوامل مؤثر بر شدت تصادفات موتورسیکلت در راه‌های برونشهری به ترتیب شامل نوع شانه، نوع وسیله نقلیه مقصر، نقص مؤثر راه، عامل انسانی، کاربری اراضی، سرعت مجاز، نوع راه، وضعیت خط‌کشی و هندسه مسیر هستند. نکته مهم آن است که برخی از این عوامل، علی‌رغم فراوانی کمتر در داده‌های توصیفی، در صورت وقوع نقش تشدیدکننده‌ای در انتقال تصادف به سطوح جرحی و فوتی ایفا می‌کنند. یافته‌های این پژوهش نشان می‌دهد که شدت تصادفات موتورسیکلت بیش از آنکه تابع یک عامل منفرد باشد، حاصل برهم‌کنش هم‌زمان ویژگی‌های زیرساختی راه، رفتار کاربران و ناهمگنی وسایل نقلیه است. به‌طورویژه، عناصر زیرساختی مانند نوع شانه و نقص راه، علاوه بر نقش فیزیکی، به‌عنوان عوامل رفتارساز عمل کرده و می‌توانند به‌طور غیرمستقیم رفتارهای پرخطر را تشدید کنند. این

پیامدهای شدید تصادفات داشته باشند. در نهایت، استفاده هم‌زمان از مدل‌های یادگیری ماشین و روش‌های تفسیرپذیر مانند SHAP، چارچوبی قدرتمند برای تحلیل ایمنی راه فراهم می‌کند که می‌تواند شکاف میان دقت پیش‌بینی و قابلیت تفسیر نتایج را پر کند. این چارچوب قابلیت تعمیم به سایر استان‌ها و انواع کاربران آسیب‌پذیر راه را دارد و می‌تواند به‌عنوان مبنایی علمی برای تصمیم‌گیری‌های سیاسی و مهندسی در حوزه ایمنی راه مورد استفاده قرار گیرد.

نتیجه بر ضرورت بازنگری در نگاه سنتی به ایمنی راه، که صرفاً بر کاهش تعداد تصادفات متمرکز است، تأکید دارد. از منظر کاربردی، نتایج این تحقیق نشان می‌دهد که راهبردهای ایمنی موتورسیکلت در راه‌های برون‌شهری باید به‌صورت شدت‌محور طراحی شوند. بهبود کیفیت و عرض شانه‌ها، حذف یا اصلاح نقص‌های مؤثر راه، بازنگری در سرعت‌های مجاز متناسب با کاربری اراضی و هندسه مسیر و همچنین اجرای برنامه‌های هدفمند آموزش و کنترل رفتار موتورسیکلت‌سواران، می‌توانند نقش مؤثری در کاهش

۶- پی‌نوشت‌ها

1. Random Forest
2. Conditional Inference Forest
3. Gradient-based One-Side Sampling
4. Exclusive Feature Bundling
5. Shapley Additive exPlanations
6. Synthetic Minority Over-Sampling Technique

۷-مراجع

-Aker, R., Susilawati, S., Zubair, H., Chor, W, T., (2025). Analyzing feature importance for older pedestrian crash severity: A comparative study of DNN models, emphasizing road and vehicle types with SHAP interpretation, *Multimodal Transportation*, 4, 100203.

-Amiri, A.M., Sadri, A., Nadimi, N., Shams, M., (2020). A comparison between Artificial Neural Network and Hybrid Intelligent Genetic Algorithm in predicting the severity of fixed object crashes among elderly drivers. *Accid. Anal. Prev.* 138, 105468.

-Assi, K., Rahman, S.M., Mansoor, U., Ratrou, N., (2020), Predicting crash injury severity with machine learning algorithm synergized with clustering technique: a promising protocol. *Int. J. Environ. Res. Publ. Health* 17, 5497.

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

-Brockhusa, L. A., Liasidisb, P., Lewisb, M., Jakoba, D. A., Demetriadesb, D., (2024). Injury patterns and outcomes in motorcycle driver crashes in the United States: The effect of helmet use, *Injury*, 55, 111196.

-Chai, A. B. Zh., Lau, B. T., Tee, M. K. T., McCarthyb, C., (2024). Enhancing road safety with machine learning: Current advances and future directions in accident prediction using non-visual data, *Engineering Applications of Artificial Intelligence*, 137, 109086.

-Champahom, T., Se, CH., Watcharamaisakul, F., Jomnonkwao, S., Karoonsoontawong, A., Ratanavaraha, V., (2024). Tree-based approaches to understanding factors influencing crash severity across roadway classes: A Thailand case study, *IATSS Research* 48, 464-476.

- Chang, F., Li, M., Xu, P., Zhou, H., Haque, M., Huang, H., (2016). Injury severity of motorcycle riders involved in traffic crashes in Hunan, China: a mixed ordered logit approach, *Int. J. Environ. Res. Public Health*, 13, 714.
- Chen, M.M. Chen, M.C., (2020), Modeling road accident severity with comparisons of logistic regression, decision tree and random forest, *Inf. 11*.
- Chen, T., Guestrin, C., (2016). XGboost: A scalable tree boosting system, in proceeding of the 22nd ACM SIGKDD Conference, 785-794.
- Cunto, F.J.C., Ferreira, S., (2017). An analysis of the injury severity of motorcycle crashes in Brazil using mixed ordered response models, *J. Transp. Saf. Secur*, 9, 33-46.
- Donchenko, D., Sadovnikova, N., Parygin, D., (2020). Prediction of road accidents' severity on Russian roads using machine learning techniques. In: Proceedings of the 5th International Conference on Industrial Engineering (ICIE 2019). *Springer*, 1493-1501.
- Fiorentini, N., Losa, M., (2020). Handling imbalanced data in road crash severity prediction by machine learning algorithms. *Infrastructure 5*, 61.
- Fondzenyuya, S. K., Usamia, D. SH., Gonz'alez-Hern'andezb, B., Brownc, L., Morrisc, A., Persia, L., (2024). Developing improved crash prevention approaches through in-depth investigation of motorcycle crash causation patterns, *Heliyon10*, e32866.
- Ghasedi, M., Sarfjoo, M., Bargegol, I., (2021). Prediction and analysis of the severity and number of suburban accidents using Logit model, factor analysis and machine learning: a case study in a developing country. *SN Appl. Sci.* 3, 13.
- Hadjidimitriou, N.S., Lippi, M., Dell'Amico, M., Skiera, A., (2019). Machine learning for severity classification of accidents involving powered two wheelers, *IEEE Trans. Intell. Transp. Syst.* 21, 4308-4317.
- Haruna, L., Sallehuddin, R., Radzi, H.M., (2020). Discrete particle swarm optimization based filter feature selection technique for the severity of road traffic accident prediction. In: *Advances in Intelligent Systems and Computing, Springer*.
- Hasan, A. S., Jalayer, M., Das, S., Bin Kabir, A., (2024). Application of machine learning models and SHAP to examine crashes involving young drivers in New Jersey, *International Journal of Transportation Science and Technology*, 14, 156-170.
- Hsu, C. K., (2024). Heat-induced risks of road crashes among older motorcyclists: Evidence from three motorcycle-dominant cities in Taiwan, *Journal of Transport & Health*, 35, 101754.
- Ijaz, M., Ian, L., Zahid, M., Jamal, A., (2021). A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw. *Accid. Anal. Prev.* 154, 106094.
- Infante, P., Jacinto, G., Afonso, A., Rego, L., Nogueira, V., Quaresma, P., Saias, J., Santos, D., Nogueira, P., Silva, M., Costa, R.P., Gois, P., Manuel, P.R., (2022). Comparison of statistical and machine-learning models on road traffic accident severity classification. *Computers 11*, 80.
- Kashifi, M. T., (2023), Investigating two-wheelers risk factors for severe crashes using an interpretable machine learning approach and SHAP analysis, *IATSS Research 47*, 357-371.
- Kashifi, M. T., Ahmad, I., (2022). Efficient histogram-based gradient boosting approach for accident severity prediction with multisource data, *Transp. Res. Rec.*, 036119812210743.
- Ke, G., Meng, Q., Finely, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., (2017). LightGBM: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, 30.

- Khan, M.N., Das, A., Ahmed, M.M., (2024). Prediction of truck-involved crash severity on a rural mountainous freeway using transfer learning with ResNet-50 deep neural network. *J Transp Eng a Syst* 150, 04023131.
- Lanzaro, G., Andrade, M., (2023). Afuzzy expert system for setting Brazilian highway speed limits, *International Journal of Transportation Science and Technology*, 12, 505-524.
- Lapparent, M. de, (2006). Empirical Bayesian analysis of accident severity for motorcyclists in large French urban areas, *Accid. Anal. Prev.* 38, 260–268.
- Ma, Z., Mei, G., Cuomo, S., (2021), An analytic framework using deep learning for prediction of traffic accident injury severity based on contributing factors, *Accid. Anal. Prev.* 160.
- Montellaa, A., Calvib, A., D’Amicob, F., Ferranteb, C., Galantea, F., Maurielloa, F., Riccardia, M, R., Scaranoa, A., (2024). A methodology for setting credible speed limits based on numerical analyses and driving simulator experiments, *Transportation Research Part F: Psychology and Behaviour*, 100, 289–307.
- Naqvi, H. M., Tiwari, G., (2017), Factors Contributing to Motorcycle Fatal Crashes on National Highways in India, *Transportation Research Procedia*, 25, 2084–2097.
- Pai, C.W., Saleh, W., (2007). An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK, *Saf. Sci.*, 45, 832–847.
- Pai, C.-W., Saleh, W., (2008). Modelling motorcyclist injury severity by various crash types at T-junctions in the UK, *Saf. Sci.*, 46, 1234–1247.
- Panda, C., Mishra, A.K., Dash, A.K., Nawab, H., (2022). Predicting and explaining severity of road accident using artificial intelligence techniques, SHAP and feature analysis. *Int. J. Crashworthiness* 28, 186–201.
- Panicker, A.K., Ramadurai, G., (2022). Injury severity prediction model for two-wheeler crashes at mid-block road sections, *Int. J. Crashworthiness* 27, 328–336.
- Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., Kouros Mohammadian, A., (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis, *Accid. Anal. Prev.* 136.
- Qu, Y., Lin, Z., Li, H., Zhang, X., (2019). Feature recognition of urban road traffic accidents based on GA-XGBoost in the context of big data, *IEEE Access* 7, 170106–170115.
- Rasria, N., Satiennama, T., Satiennama, W., Sukkhoa, P., Jaensirisakb, S., Kronprasertc, N., Champahomd, T., (2024). Factors influencing wrong way driving behavior of motorcycle riders in Thailand, *Transportation Research Interdisciplinary Perspectives*, 25, 101122.
- Rezapour, M., Mehrara Molan, A., Ksaibati, K., (2020). Analyzing injury severity of motorcycle at-fault crashes using machine learning techniques, decision tree and logistic regression models. *International Journal of Transportation Science and Technology* 9, 89–99.
- Rifaat, S.M., Tay, R., De Barros, A., (2012). Severity of motorcycle crashes in Calgary, *Accid. Anal. Prev.* 49, 44–49.
- Sajid Hasan, A, B., Jalayer, M., Das, S., Kabir, A, B., (2024). Application of machine learning models and SHAP to examine crashes involving young drivers in New Jersey, *International Journal of Transportation Science and Technology* 14, 156–170.
- Savolainen, P., Mannering, F., (2007). Probabilistic models of motorcyclists’ injury severities in single-and multi-vehicle crashes, *Accid. Anal. Prev.*, 39, 955–963.

- Scarano, A., Sadeghi, M., Mauriello, F., Riccardi, M. R., Aghabayk, K., Montella, A., (2025). Cyclist crash severity modeling: A hybrid approach of XGBoost-SHAP and random parameters logit with heterogeneity in means and variances, *Journal of Safety Research*, 93, 373–398.
- Shaheed, M.S.B., Gkritza, K., Zhang, W., Hans, Z., (2013). A mixed logit analysis of two- vehicle crash severities involving a motorcycle, *Accid. Anal. Prev.*, 61, 119–128.
- Shapley, L. S. (1953). A value for n-person games. *Contrib. Theory Games*, 2(28), 307–317.
- Siddik, MdA.B., Arman, MdS., Hasan, A., Jahan, M.R., Islam, M., Biplob, K.B.B., (2021). Predicting the death of road accidents in Bangladesh using machine learning algorithms. In: *Communications in Computer and Information Science. Springer Science and Business Media Deutschland GmbH*, 160–171.
- Sunkpho, J., Se, C., Wipulanusat, W., Ratanavaraha, V., (2025). SHAP-based convolutional neural network modeling for intersection crash severity on Thailand's highways, *IATSS Research* 49, 27–41.
- Tariq, A.-M., Xiao, H., Jean, P.N., Eshrak, N.A., Ibrahim, A., (2020). Exploiting machine learning algorithms for predicting crash injury severity in Yemen: hospital case study. *Appl. Comput. Math.* 9, 164.
- Wahab, L., Jiang, H., (2019). A comparative study on machine learning based algorithms for prediction of motorcycle crash severity, *PLoS One* 14, e0214966.
- Wahab, L., Jiang, H., (2020). Severity prediction of motorcycle crashes with machine learning methods. *Int. J. Crashworthiness* 25, 485–492.
- Wang, X., Zhou, Q., Quddaus, M., Fang, S., (2018). Speed, speed variation and crash relationship for urban arterials, *Accid. Anal. Prev.*, 113, 236-243.
- Wang, Z., Lee, C., Lin, P.-S., (2014). Modeling injury severity of single-motorcycle crashes on curved roadway segments, in: *93rd Annual Meeting of the Transportation Research Board*, Washington, DC.
- Xin, C., Wang, Z., Lee, C., Lin, P.-S., (2017). Modeling safety effects of horizontal curve design on injury severity of single-motorcycle crashes with mixed-effects logistic model, *Transp. Res. Rec.* 2637 38–46.
- Yan, M., Shen, Y., (2022). Traffic accident severity prediction based on random forest, *Sustain* 14.
- Yoon, J., (2025). Prediction of high-risk areas using the interpretable machine learning: Based on each determinant for the severity of pedestrian crashes, *Journal of Transport Geography*, 126, 104216.
- Zhang, C., He, J., Wang, Y., Yan, X., Zhang, Changjian, Chen, Y., Liu, Z., Zhou, B., (2020). A crash severity prediction method based on improved neural network and factor analysis. *Discrete Dynam Nat. Soc.* 1–13.

Identification of Factors Affecting Motorcycle Crash Severity on Intercity Roads Using the Random Forest (RF) Algorithm and SHAP Analysis

Hamed Sefi, Department of Civil Engineering, Payame Noor University (PNU), Tehran, Iran.

Shahin Shabani, Department of Civil Engineering, Payame Noor University (PNU), Tehran, Iran.

Mohammad Koochi, M.Sc., Grad., Department of Civil Engineering, Payame Noor University (PNU), Tehran, Iran.

E-mail: hamedsaify@pnu.ac.ir

Received: February 2026- Accepted: May 2026

ABSTRACT

Motorcycle crashes account for a substantial share of road traffic fatalities due to the high vulnerability of this group of road users, and the severity of their consequences—particularly on intercity roads—has consistently remained one of the major challenges in road safety. The primary objective of this study is to identify and interpret the factors influencing motorcycle crash severity on intercity roads using interpretable machine learning approaches. To this end, three-year motorcycle crash data (2017–2019) from intercity roads in Razavi Khorasan Province were utilized. The explanatory variables were classified into thirteen categories, including road characteristics, geometric conditions, shoulder type, contributing road defects, human factors, at-fault vehicle type, land use, and other environmental and traffic-related factors. In this study, three machine learning algorithms—Random Forest, XGBoost, and LightGBM—were employed to predict crash severity levels (property damage only, injury, and fatal). After data preprocessing and class balancing using the SMOTE method, model performance was evaluated using precision, recall, and F1-score metrics. The results indicated that the Random Forest model achieved superior performance compared to the other models, with an overall accuracy of 76%, and was therefore selected as the optimal model. To interpret the output of the selected model, the SHAP method was applied. The SHAP analysis revealed that shoulder type, at-fault vehicle type, contributing road defects, human factors, land use, posted speed limit, road type, pavement marking condition, and roadway geometry were, respectively, the most influential factors affecting motorcycle crash severity on intercity roads. The findings of this study demonstrate that motorcycle crash severity results from a complex interaction of infrastructural, behavioral, and vehicle-related factors, and that adopting severity-oriented approaches in road safety planning can play an effective role in reducing severe crash outcomes.

Keywords: Motorcycle Crashes, Random Forest (RF), SHAP Analysis