RBT

REVISTA BRASILEIRA DE TRANSPORTES

BRAZILIAN TRANSPORT JOURNAL

V.6 N.1 | JUL 2026



FGV TRANSPORTES

ARTIGO 6

SEGURANÇA VIÁRIA E GESTÃO URBANA: PADRÕES ESPACIAIS E PREDIÇÃO DE RISCO DE FATALIDADE ENTRE MOTOCICLISTAS NO SISTEMA RODOVIÁRIO DO DISTRITO FEDERAL (2014–2023)

Janduhy Pereira dos Santos (DER/DF), Valdir Adilson Steinke, Maurício Theodósio Mattos Marques, Ricardo Sameshima Taba, Andrea Amaziles A. A. de Carvalho Lousada, Bruno Maia Soriano Lousada, Jesus Mauro Vieira de Oliveira e Caroline Ribeiro Chahini

RESUMO

Este artigo investiga a dinâmica espacial e os fatores associados à gravidade de acidentes envolvendo motociclistas no Distrito Federal, com base em dados de acidentes rodoviários (2014–2023). Adota-se um fluxo analítico que combina análise espacial para detecção de áreas críticas (Getis-Ord Gi*), modelagem preditiva por gradient boosting (XGBoost) e balanceamento de classes (SMOTE), com interpretação da importância dos preditores via SHAP. A avaliação baseia-se em partição treino–teste estratificada, priorizando a comparação de desempenho e a leitura substantiva das variáveis. Os achados indicam padrões concentrados no espaço urbano e a relevância de características do evento, do contexto operacional e de aspectos temporais na explicação do risco. As evidências resultantes subsidiam estratégias de segurança viária voltadas à priorização territorial, à fiscalização direcionada e a ações de moderação de velocidade, contribuindo para o planejamento intersetorial e para a redução de desfechos graves entre motociclistas.

Palavras-chave: acidentes de trânsito; motocicletas; análise espacial

ABSTRACT

This study investigates the spatial dynamics and factors associated with the severity of motorcycle-involved road crashes in Brazil's Federal District, using crash records from 2014 to 2023. An analytical workflow is adopted that combines spatial statistics to detect critical areas (Getis-Ord Gi*), predictive modeling via gradient boosting (XGBoost), and class imbalance handling with SMOTE, with predictor importance interpreted using SHAP. Model evaluation relies on a stratified train–test split, prioritizing comparative performance assessment and the substantive interpretation of predictors. The results highlight the relevance of event-specific attributes, operational context, and temporal factors in explaining risk. The resulting evidence informs road safety strategies focused on geographic prioritization, targeted enforcement, and speed management measures, supporting cross-sector planning and the reduction of severe outcomes among motorcyclists.

Keywords: Road traffic accidents; Motorcycles; Spatial analysis.

1. INTRODUÇÃO

Nas últimas décadas, o crescimento urbano acelerado no Brasil tem provocado transformações significativas nas dinâmicas territoriais e nos padrões de mobilidade, especialmente nas grandes metrópoles. A expansão periférica, associada à insuficiência e à baixa atratividade dos sistemas de transporte coletivo, tem impulsionado o aumento da frota de veículos motorizados, com destaque para as motocicletas. Nesse contexto, esse modal passou a desempenhar papel estratégico na mobilidade urbana, seja como alternativa aos congestionamentos, seja como instrumento de inserção no mercado de trabalho, sobretudo em atividades mediadas por plataformas digitais.

No Distrito Federal, essa dinâmica se intensificou de forma expressiva, impulsionada por fatores econômicos e estruturais. A partir de 2015, a crise econômica e a reconfiguração do mercado de trabalho contribuíram para o crescimento de ocupações informais, muitas das quais dependentes do uso da motocicleta. Esse processo foi ainda mais acentuado durante a pandemia da COVID-19, com a ampliação dos serviços de entrega e transporte individual. Como resultado, observou-se aumento significativo da frota de motocicletas, acompanhado por elevação nos registros de acidentes de trânsito envolvendo esse grupo de usuários, que se destacam como um dos mais vulneráveis no sistema viário (SENATRAN, 2024; DETRAN-DF, 2024).

Do ponto de vista da segurança viária, a literatura recente tem avançado no sentido de superar abordagens centradas exclusivamente na identificação de pontos críticos, incorporando perspectivas sistêmicas baseadas na gestão do risco, como o paradigma do Safe System. Essa abordagem reconhece a vulnerabilidade dos usuários e a inevitabilidade do erro humano, atribuindo aos gestores do sistema a responsabilidade de estruturar vias, regular velocidades e organizar o tráfego de forma a minimizar a ocorrência de lesões graves e fatais (OECD/ITF, 2008; WHO, 2023).

Nesse contexto, a análise espacial e a modelagem preditiva têm se consolidado como ferramentas fundamentais para a compreensão da acidentalidade, permitindo não apenas identificar padrões territoriais, mas também estimar probabilidades de ocorrência de desfechos graves. Estudos recentes têm aplicado técnicas como estatística espacial e algoritmos de aprendizado de máquina para investigar fatores associados à severidade dos acidentes, evidenciando a relevância de variáveis relacionadas ao ambiente viário, ao comportamento dos usuários e às condições operacionais do sistema de transporte.

No entanto, apesar dos avanços metodológicos, ainda há lacunas importantes na literatura, especialmente no que se refere à integração entre análises espaciais tradicionais — voltadas à identificação de áreas de concentração de eventos — e abordagens preditivas capazes de estimar o risco condicional de fatalidade. Em muitos estudos, essas dimensões são tratadas de forma isolada, o que limita a compreensão integrada do fenômeno e reduz o potencial de aplicação dos resultados no planejamento da segurança viária.

Diante desse cenário, este estudo tem como objetivo (i) identificar padrões espaciais de acidentes envolvendo motociclistas no Sistema Rodoviário do Distrito Federal (2014–2023) e (ii) desenvolver um modelo preditivo de risco de fatalidade, com base em técnicas de aprendizado de máquina, com o objetivo de identificar fatores críticos associados à severidade dos eventos.

A principal contribuição deste artigo reside na integração entre análise espacial de concentração de acidentes (Getis-Ord Gi*) e modelagem preditiva de risco condicional de fatalidade (XGBoost), permitindo distinguir empiricamente duas dimensões frequentemente confundidas na literatura: (i) onde os acidentes são mais frequentes e (ii) onde a probabilidade de ocorrência de óbito é mais elevada. Essa abordagem combinada possibilita uma leitura mais precisa do risco viário, contribuindo para a priorização de intervenções diferenciadas no território.

Adicionalmente, o estudo avança ao aplicar essa integração metodológica em escala de rede rodoviária, incorporando variáveis operacionais e temporais e explorando técnicas de interpretação de modelos (SHAP), o que amplia a compreensão dos mecanismos associados à gravidade dos acidentes. Embora as técnicas empregadas sejam amplamente reconhecidas na literatura, sua aplicação conjunta, com foco na distinção entre risco absoluto e risco condicional, constitui um avanço analítico relevante no campo da segurança viária.

Os resultados obtidos oferecem subsídios para o planejamento e a gestão da mobilidade, contribuindo para o direcionamento de ações de engenharia, fiscalização e educação no trânsito, em consonância com abordagens orientadas por evidências e pela redução de mortes e lesões graves.

2. REVISÃO DE LITERATURA

Nos últimos trinta anos, o uso das motocicletas tem passado por mudanças significativas em sua finalidade. Antes, esses veículos eram usados principalmente para lazer e deslocamentos individuais, com pouca utilização para fins laborais. Contudo, o avanço do processo de metropolização nas grandes cidades, aliado ao aumento da demanda por serviços terceirizados e entregas rápidas, fez com que a frota de motocicletas crescesse de forma expressiva em diversos países.

Em transporte e segurança viária, o debate contemporâneo enfatiza que reduzir a mortalidade exige combinar gestão de risco, governança e monitoramento com intervenções integradas em infraestrutura, velocidades seguras, veículos, fiscalização e atendimento pós-acidente. Relatórios de síntese têm apontado que, embora haja leve redução recente, a carga global de mortes no trânsito permanece elevada, reforçando a necessidade de estratégias orientadas por evidências e por indicadores de desempenho e risco, e não apenas por contagens absolutas de eventos (WHO, 2023).

Apesar do consenso crescente na literatura quanto à necessidade de abordagens sistêmicas baseadas na gestão do risco, observa-se que muitos estudos ainda se concentram na identificação de padrões agregados de ocorrência de acidentes, sem incorporar explicitamente a distinção entre

frequência de eventos e probabilidade de desfechos graves. Essa limitação reduz o potencial de aplicação dos resultados no planejamento de intervenções mais direcionadas, especialmente em contextos onde a exposição e a severidade não coincidem espacialmente.

Para usuários vulneráveis — como motociclistas — essa agenda é particularmente relevante, pois a severidade tende a aumentar com a exposição a altas velocidades e com conflitos em ambientes viários que não foram desenhados para absorver impactos. Por isso, estudos baseados em localização e predição devem ser articulados a uma lógica de gestão do risco: onde a infraestrutura e a operação podem ser redesenhadas para reduzir energia de colisão e separar conflitos, em alinhamento com os princípios do Safe System (OECD/ITF, 2008; European Commission, 2023).

Assim, Santos Júnior *et al*. (2022) descrevem os acidentes de trânsito fatais como um problema complexo e desafiador para as autoridades públicas. Os autores salientam comportamentos de risco no trânsito urbano e os elevados custos sociais e econômicos dos acidentes de trânsito fatais, defendendo ações intersetoriais e o aperfeiçoamento da vigilância. Esses prejuízos incluem a sobrecarga dos serviços de saúde e assistência previdenciária, refletindo os elevados custos sociais e econômicos associados à acidentalidade envolvendo motocicletas.

A preocupação em garantir condições seguras para os motociclistas envolve todos os setores que compõem a mobilidade urbana. Nesse contexto, Brenac *et al*. (2006) destacam a importância da visibilidade dos motociclistas nas vias, enfatizando que a dificuldade de um objeto ou indivíduo ser percebido visualmente em meio ao ambiente é um fator recorrente nos acidentes ocorridos em áreas urbanas.

A mesma preocupação se estende às estradas rurais, onde o uso das motocicletas assume uma diversidade de funções além do simples deslocamento do condutor. Agyemang, Adanu e Jones (2021) analisam os acidentes com motocicletas nas zonas rurais de Gana e destacam que esses eventos são frequentemente negligenciados pelas políticas públicas. Os autores apontam que a precariedade das vias, o acesso limitado a serviços de emergência e o comportamento de risco por parte dos condutores afetam diretamente a gravidade das lesões registradas nesses contextos.

Na elaboração e execução de políticas públicas voltadas à segurança viária, instrumentos analíticos, como as ferramentas de geoinformação, tornam-se fundamentais para identificar áreas críticas e propor intervenções eficazes para a redução dos acidentes de trânsito. Estudos como os de Tippayanate *et al*. (2024) e Haque *et al*. (2010) evidenciam o potencial das técnicas de análise espacial na identificação de padrões territoriais de acidentes envolvendo motocicletas, contribuindo para o direcionamento de ações preventivas. De modo geral, esses estudos demonstram a eficácia das técnicas de análise espacial na identificação de áreas críticas. No entanto, a maior parte das abordagens permanece centrada na detecção de clusters de ocorrência, sem avançar na modelagem do risco condicional à ocorrência do evento. Essa limitação pode levar à priorização de áreas com alto volume de acidentes, mas não necessariamente

com maior probabilidade de fatalidade. Complementarmente, Miao *et al*. (2025) demonstram, por meio de modelos espaço-temporais aplicados à China, que a ausência do uso do capacete está fortemente associada à gravidade dos acidentes, reforçando a importância de dados georreferenciados para decisões baseadas em evidências.

A aplicação de técnicas de geoinformação tem se mostrado estratégica na formulação de políticas públicas de mobilidade e segurança viária. Estudos como o de Jiang, Yuen e Lee (2020) demonstram como a integração entre mineração de regras de associação e análise espacial, por meio de Sistemas de Informação Geográfica (SIG), possibilita não apenas identificar fatores críticos associados a acidentes envolvendo motociclistas, mas também mapear os territórios onde esses fatores convergem. Pesquisas internacionais que aplicam a técnica do XGBoost em predições de gravidade de acidentes viários, como Wisutwattanasak *et al*. (2024) em acidentes envolvendo motociclistas na Tailândia; Chen *et al*. (2025), com otimização MSCPO-XGBoost na China; Qian e Wang (2023), em acidentes britânicos; e Yang *et al*. (2022), em rodovias americanas, evidenciam a robustez da técnica para capturar fatores não lineares e interpretar variáveis contextuais em estudos de segurança viária. Apesar do avanço no uso de algoritmos de aprendizado de máquina para previsão da severidade dos acidentes, observa-se que essas abordagens frequentemente não incorporam explicitamente a dimensão espacial dos fenômenos, tratando a localização predominantemente como uma variável explicativa adicional, sem explorar plenamente sua estrutura espacial. Além disso, há pouca integração entre modelos preditivos e análises espaciais exploratórias, o que limita a compreensão conjunta dos padrões territoriais e dos mecanismos de risco.

Assim sendo, conforme detalha Panumasvivat *et al*. (2024), a tomada de decisão em políticas públicas de mobilidade conta com as técnicas geoespaciais como suporte essencial para a identificação de fatores de risco, não apenas para os motociclistas, mas também para os demais usuários dos sistemas de transporte, tanto em áreas urbanas quanto rurais. Essas ferramentas possibilitam a implementação de medidas como intervenções de engenharia de tráfego, intensificação da fiscalização, melhorias na infraestrutura viária e campanhas de conscientização direcionadas a horários e locais específicos.

Ao gerar mapas preditivos e de aglomerados de alta incidência com embasamento estatístico, o estudo encaminha sugestões de intervenções em infraestrutura, fiscalização e campanhas educativas. Assim, as técnicas geoespaciais deixam de ser meros instrumentos técnicos para se consolidarem como ferramentas auxiliares de justiça territorial e prevenção de acidentes. Adicionalmente, embora diversos estudos apontem a relevância de variáveis relacionadas à geometria das vias, velocidades regulamentadas e volumes de tráfego na explicação da severidade dos acidentes, a disponibilidade e padronização desses dados ainda representam um desafio em muitos contextos institucionais. Essa limitação tem levado pesquisadores a utilizarem dados espaciais e operacionais, como a localização geográfica e características do evento, como forma de capturar parcialmente esses efeitos.

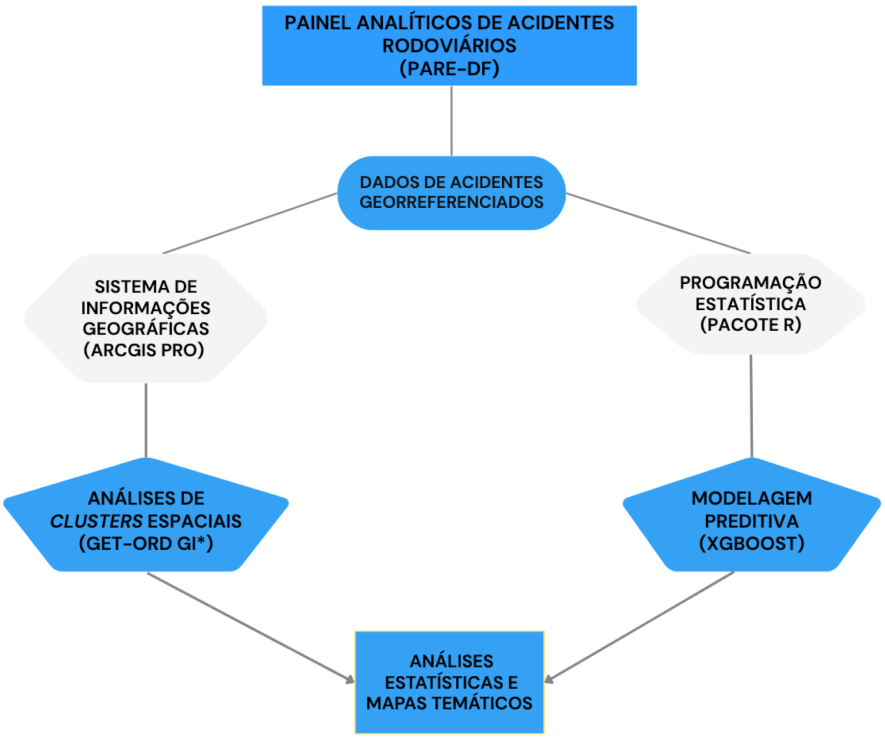
3. METODOLOGIA

A presente pesquisa adota uma abordagem quantitativa com ênfase em métodos de análise espacial, visando identificar padrões de distribuição e fatores associados aos acidentes de trânsito envolvendo motociclistas no Sistema Rodoviário do Distrito Federal (SRDF).

Assim, este estudo estrutura-se em três eixos interligados (Figura 1), conforme o organograma metodológico: (i) organização, limpeza e padronização da base de dados, com geocodificação dos registros e integração à malha rodoviária do SRDF; (ii) análise espacial por meio do método *Hot Spot Analysis* (Getis-Ord Gi*) para identificação de zonas críticas (*hot spots*) e de baixa incidência (*cold spots*) de acidentes envolvendo motociclistas; e (iii) modelagem preditiva utilizando o algoritmo XGBoost, visando identificar os fatores críticos que aumentam a probabilidade de fatalidades.

A integração dessas etapas resultou na produção de análises estatísticas e mapas temáticos, que permitem visualizar os padrões de ocorrência dos acidentes, localizar aglomerações significativas de acidentes envolvendo motociclistas e estimar probabilidades de risco em diferentes segmentos rodoviários.

Figura 1 — Organograma metodológico.



Fonte: Autores.

3.1. Organização e Tratamento dos Dados

Os registros utilizados na presente análise foram extraídos do geovisualizador DERGeo (Santos *et al.*, 2023), uma plataforma desenvolvida pelo Departamento de Estradas de Rodagem do Distrito Federal (DER-DF) que disponibiliza, por meio do módulo PARE-DF (Painel Analítico de Acidentes de Trânsito no Distrito Federal), informações georreferenciadas sobre acidentes ocorridos na malha rodoviária sob sua jurisdição. Essa ferramenta viabiliza o acesso a uma base espacial estruturada, o que possibilitou a aplicação de técnicas estatísticas e de geoinformação ao longo do estudo.

O conjunto de dados analisado compreende 9.055 registros de acidentes com motociclistas, georreferenciados em coordenadas UTM (fuso 23S, Datum SIRGAS 2000), ocorridos no SRDF entre os anos de 2014 e 2023. As ocorrências incluem informações detalhadas, como data da ocorrência, natureza do acidente e categorias de veículos envolvidos, além da geolocalização de cada evento.

Previamente às análises espaciais e estatísticas, a base de dados foi submetida a um processo de pré-processamento no software Microsoft Excel, com foco na preparação e estruturação dos registros. As etapas envolveram: identificação e eliminação de duplicidades e inconsistências, padronização das variáveis categóricas (como tipo e natureza do acidente, dia da semana e categoria do veículo), conversão de formatos temporais (data e hora) para estruturas reconhecíveis por sistemas de informações geográficas e validação das coordenadas geográficas. Esse procedimento assegurou a consistência semântica e a integridade sintática dos dados, viabilizando sua posterior importação e manipulação em ambiente geoespacial no ArcGIS Pro.

3.2. Análise espacial e preditiva

A presente etapa metodológica combina técnicas de análise espacial com métodos de aprendizado de máquina, visando aprofundar a compreensão dos padrões de risco associados a acidentes envolvendo motociclistas no Sistema Rodoviário do Distrito Federal.

Inicialmente, foi aplicada a técnica de *Hot Spot Analysis* (Getis-Ord Gi*), implementada no software ArcGIS Pro 3.5.1, com o objetivo de identificar agrupamentos estatisticamente significativos de alta e baixa concentração de eventos. Essa etapa caracteriza-se como uma análise exploratória espacial, voltada à identificação de padrões de ocorrência (frequência de acidentes) ao longo da rede viária.

Em uma segunda etapa, foi desenvolvido um modelo preditivo utilizando o algoritmo XGBoost, implementado no ambiente R (versão 4.3.2), com o objetivo de estimar a probabilidade de ocorrência de óbitos condicionada à ocorrência do acidente. Diferentemente da análise espacial exploratória, essa abordagem incorpora uma dimensão explicativa e prognóstica, permitindo identificar fatores associados à severidade dos eventos.

A utilização combinada dessas duas abordagens permite distinguir analiticamente duas dimensões complementares: (i) a concentração espacial de acidentes e (ii) o risco condicional de fatalidade, contribuindo para uma interpretação mais abrangente da segurança viária.

Neste estudo, a espacialidade foi tratada de forma complementar: de um lado, por análise espacial exploratória da concentração de acidentes (Getis-Ord Gi*); de outro, pela incorporação das coordenadas geográficas como proxies locacionais no modelo preditivo, ao lado de variáveis temporais, operacionais e relacionadas à natureza do acidente.

3.2.1. Algoritmo Getis-Ord Gi*

A identificação de áreas de concentração foi realizada pelo método *Hot Spot Analysis* (Getis-Ord Gi*), que estima a associação espacial local com base em matrizes de vizinhança e pesos espaciais, permitindo destacar *clusters* de valores elevados/baixos com significância estatística (Getis; Ord, 1992).

Um ponto quente é uma área ou local dentro de um limite identificável que apresenta uma concentração de incidentes. Os processos principais para estimar aglomerados de acidentes incluem a análise de acidentes de trânsito georreferenciados e o mapeamento de clusters utilizando as funções Getis-Ord Gi*.

A estatística local de associação espacial é dada por:

$$G_i^* = \frac{\sum_j w_{ij} x_j - \bar{X} \sum_j w_{ij}}{S \sqrt{\frac{n \sum_j w_{ij}^2 - \left(\sum_j w_{ij}\right)^2}{n-1}}} \quad (1)$$

Onde:

$Gi*$ é o valor do índice para a localização i.
xj é o valor do atributo na localização j.
wi, j é o peso espacial entre as localizações i e j.

X é a média dos valores do atributo.
S é o desvio padrão dos valores do atributo.
n é o número total de localizações.

O método avalia a intensidade da concentração espacial de eventos a partir dos valores atribuídos a cada feição e aos seus vizinhos imediatos, identificando agrupamentos estatisticamente significativos de altas ocorrências (*hot spots*) e baixas ocorrências (*cold spots*).

Para a definição da vizinhança espacial, adotou-se o critério de distância fixa de 1.500 metros. A escolha desse parâmetro baseou-se em testes preliminares de autocorrelação espacial, Índice de Moran Global (Global Moran's I), e na análise da conectividade da malha rodoviária do SRDF, considerando a escala média de interação entre segmentos viários urbanos e periurbanos. Distâncias inferiores resultaram em fragmentação excessiva dos clusters, enquanto distâncias superiores tenderam a diluir padrões locais relevantes.

Foi utilizada uma matriz de pesos espaciais do tipo distância fixa, garantindo comparabilidade entre feições com diferentes números de vizinhos. O nível de significância adotado foi de 95% ($p < 0,05$), com interpretação baseada em valores de *z-score* e *p-value*.

A operacionalização foi realizada integralmente no ArcGIS Pro, utilizando a ferramenta *Hot Spot Analysis* (Getis-Ord Gi*), permitindo a classificação dos resultados em cinco categorias (muito baixo, baixo, neutro, alto e muito alto).

3.2.2. Modelagem Preditiva Com Xgboost

Para a etapa preditiva do estudo, foi empregado o algoritmo XGBoost (*Extreme Gradient Boosting*), pertencente à classe dos métodos de *ensemble learning*, mais especificamente à técnica de *boosting*. Este algoritmo constrói, de forma sequencial, uma série de árvores de decisão (*weak learners*), onde cada nova árvore é ajustada para corrigir os erros das anteriores, otimizando uma função objetivo por meio de gradiente descendente (CHEN *et al.*, 2025).

A modelagem preditiva foi implementada no ambiente R (versão 4.3.2), utilizando o pacote *xgboost*, enquanto as etapas de análise espacial exploratória foram conduzidas no software ArcGIS Pro, versão

3.5.1. Essa separação reflete a natureza complementar das abordagens: o ambiente SIG para análise espacial e o ambiente estatístico para modelagem preditiva.

O XGBoost é particularmente eficaz em tarefas de classificação binária, como a adotada neste estudo — prever a ocorrência de acidentes com vítima fatal (1) ou não fatal (0) —, por sua capacidade de modelar relações não lineares, lidar com dados desbalanceados e por oferecer mecanismos integrados de regularização, que reduzem o risco de *overfitting*.

A função objetivo geral do XGBoost pode ser expressa como:

Onde:

ℓ é a função de perda (*loss functio*n);
$\hat{y}_i^{(t)}$ é a previsão no estágio t;
é o termo de regularização;
γ é o custo de complexidade por folha da árvore;
T é o número de folhas;
λ é o coeficiente de regularização L2;
w é o vetor de pesos da árvore.

$$\mathcal{L}^{(t)} = \Sigma_{i=1}^{n} \ell\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t), \Omega(f) = \gamma \cdot T + (\lambda/2) \cdot \Sigma_{j=1}^{T} w_j^2 \qquad (2)$$

A predição para a instância xi após t árvores é dada por:

$$\hat{y}_i = \Sigma_{k=1}^{K} f_k(x_i), f_k \in \mathfrak{F} \qquad (3)$$

Onde $\boldsymbol{F}$ representa o espaço das possíveis árvores de decisão.

Inicialmente, os dados passaram por uma etapa de tratamento e organização, que incluiu: (i) criação de variáveis categóricas; (ii) verificação de valores ausentes e eliminação de registros inconsistentes; (iii) transformação do campo de data para o formato adequado; e (iv) para lidar com desbalanceamento entre as classes, aplicou-se o SMOTE (*Synthetic Minority Over-Sampling Technique*), que gera instâncias sintéticas na vizinhança da minoria, reduzindo o viés do classificador sem mera duplicação de observações (Chawla *et al*., 2002).

O conjunto de dados foi dividido em 70% para treinamento e 30% para teste, utilizando amostragem estratificada. A opção por uma única divisão treino–teste teve como objetivo preservar a clareza do modelo e permitir a comparação direta com estudos similares da literatura. Contudo, reconhece-se que essa abordagem pode introduzir variabilidade nas estimativas, sendo recomendada, em estudos futuros, a utilização de validação cruzada, preferencialmente com bloqueio espacial ou temporal.

A avaliação do desempenho considerou a acurácia e a curva AUC-ROC (*Receiver Operating Characteristic*), garantindo um balanço entre precisão e sensibilidade. Para interpretação do modelo e identificação dos fatores mais relevantes, foram gerados gráficos de importância de variáveis com base nos valores de ganho e aplicados os valores da SHAP (*Shapley Additive Explanations*) para a compreensão dos efeitos marginais de cada variável.

Essa abordagem permitiu estimar a probabilidade de ocorrência de acidentes fatais em diferentes segmentos rodoviários. Além disso, possibilitou identificar os principais preditores do risco de fatalidade, destacando-se a natureza do acidente, o envolvimento entre motocicleta e caminhão, a localização

geográfica (LAT/LONG) e fatores temporais, como o sábado.

Os principais hiperparâmetros utilizados no modelo foram: *eta* = 0,1; *max_depth* = 6; *nrounds* = 100; *subsample* = 0,8; *colsample_bytree* = 0,8; *objective = binary:logistic*; *eval_metric = auc*. A escolha desses parâmetros baseou-se em testes empíricos e recomendações da literatura, buscando o equilíbrio entre desempenho preditivo e controle de *overfitting*.

A interpretação utilizou SHAP, que decompõe a predição em contribuições aditivas por variável (valores de Shapley), permitindo analisar a importância global e os efeitos locais (Lundberg; Lee, 2017).

Cabe destacar que a inclusão das coordenadas geográficas (latitude e longitude) como variáveis explicativas pode introduzir uma sobreposição de informação espacial, uma vez que essas variáveis podem refletir padrões espaciais já presentes na variável resposta. Neste estudo, optou-se por manter essas variáveis com o objetivo de captar efeitos espaciais não observados diretamente, funcionando como proxies de características do ambiente viário. No entanto, os resultados associados a essas variáveis devem ser interpretados com cautela, não sendo considerados como relações causais, mas sim como indicativos de dependência espacial residual.

4. RESULTADOS

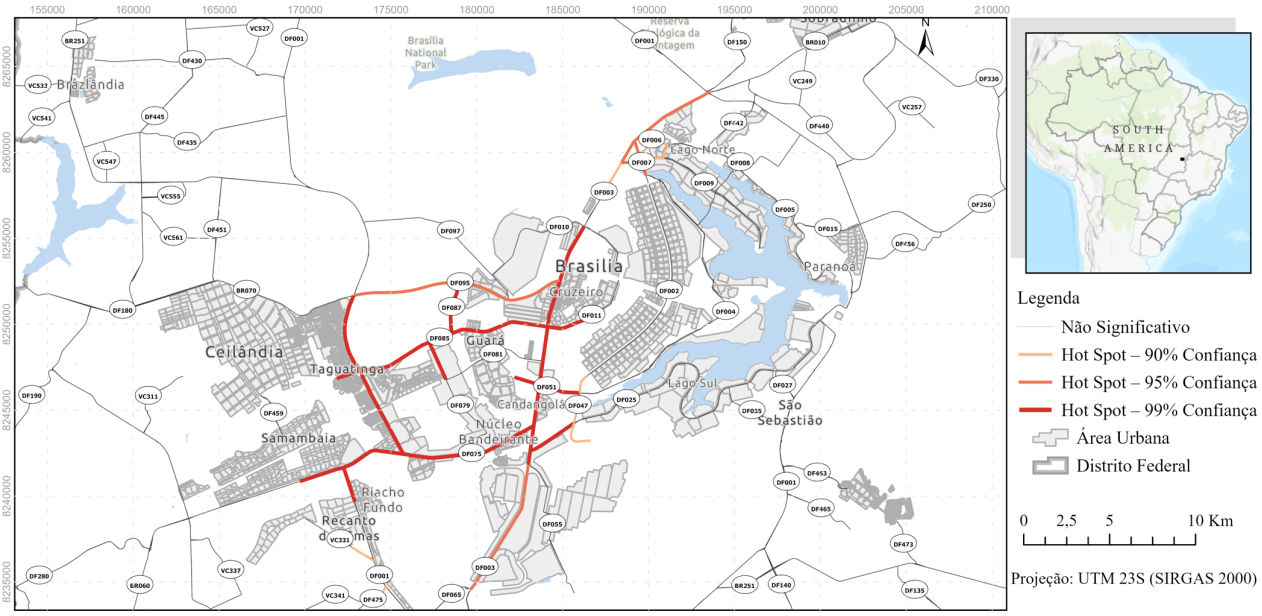
Entre 2014 e 2023, foram registrados 9.055 acidentes envolvendo motociclistas no SRDF, dos quais 439 resultaram em óbito e 8.616 em vítimas não fatais. Esses totais referem-se aos registros geocodificados e validados utilizados nas análises subsequentes.

A aplicação da técnica de Hot Spot Analysis (Getis-Ord Gi*) permitiu identificar padrões estatisticamente significativos de concentração espacial de acidentes envolvendo motociclistas no SRDF (Figura 2). Essa abordagem possibilitou diferenciar trechos com incidência elevada (hot spots) daqueles com ocorrência reduzida (cold spots), considerando o contexto geográfico e a distribuição espacial dos eventos. O método, ao incorporar a autocorrelação espacial, evidencia áreas onde os valores observados não se distribuem aleatoriamente, mas se agrupam de forma consistente, refletindo a presença de fatores locais que potencializam ou atenuam o risco de acidentes.

A análise revelou hot spots com alta significância estatística, distribuídos principalmente em rodovias estruturais que conectam o centro de Brasília às regiões administrativas e cidades do Entorno. Entre os trechos destacados, em ordem numérica, encontram-se: DF-001, DF-003, DF-075, DF-085, e DF-095. Esses segmentos apresentam agrupamentos contínuos de registros, formando corredores críticos dentro da malha rodoviária. As áreas não destacadas no mapa correspondem a trechos sem significância estatística, enquanto algumas regiões periféricas apresentam valores reduzidos de ocorrência, caracterizando cold spots. A delimitação espacial obtida evidencia que os agrupamentos mais intensos se concentram na porção centro-sudeste do território, abrangendo vias de alta conectividade no fluxo diário do Distrito Federal.

Para detalhar os corredores críticos identificados na Figura 2, a Tabela 1 apresenta, por rodovia, os trechos classificados como hot spots, o maior nível de significância observado, a extensão total (km) e a taxa média de acidentes por quilômetro. Essa síntese complementa o mapa e facilita a priorização das vias com maior intensidade e extensão de risco.

Figura 2 — Hot spots (Gi*) de acidentes envolvendo motociclistas no SRDF, 2014–2023.



Fonte: Autores.

Tabela 1 - Trechos de concentração significativa de acidentes com motociclistas (Getis-Ord Gi*) no SRDF, 2014–2023.

Rodovia	Extensão total (km)	Taxa média (acidentes/km)	Maior nível de significância	Níveis presentes
DF-001	16,05	68,68	99%	99%
DF-003	31,16	45,37	99%	99%, 95%, 90%
DF-005	0,79	15,18	90%	90%
DF-006	2,25	2,66	90%	90%
DF-007	2,12	23,62	95%	95%
DF-009	2,53	15,01	95%	95%, 90%
DF-011	2,24	50,96	99%	99%
DF-025	2,98	26,81	99%	99%
DF-047	5,1	17,25	90%	90%
DF-051	3,94	21,59	99%	99%
DF-075	8,02	85,71	99%	99%
DF-079	2,37	18,17	99%	99%
DF-085	13,01	53,49	99%	99%
DF-087	3,03	18,16	99%	99%
DF-095	12,54	35,97	95%	95%
DF-475	1,51	9,96	90%	90%

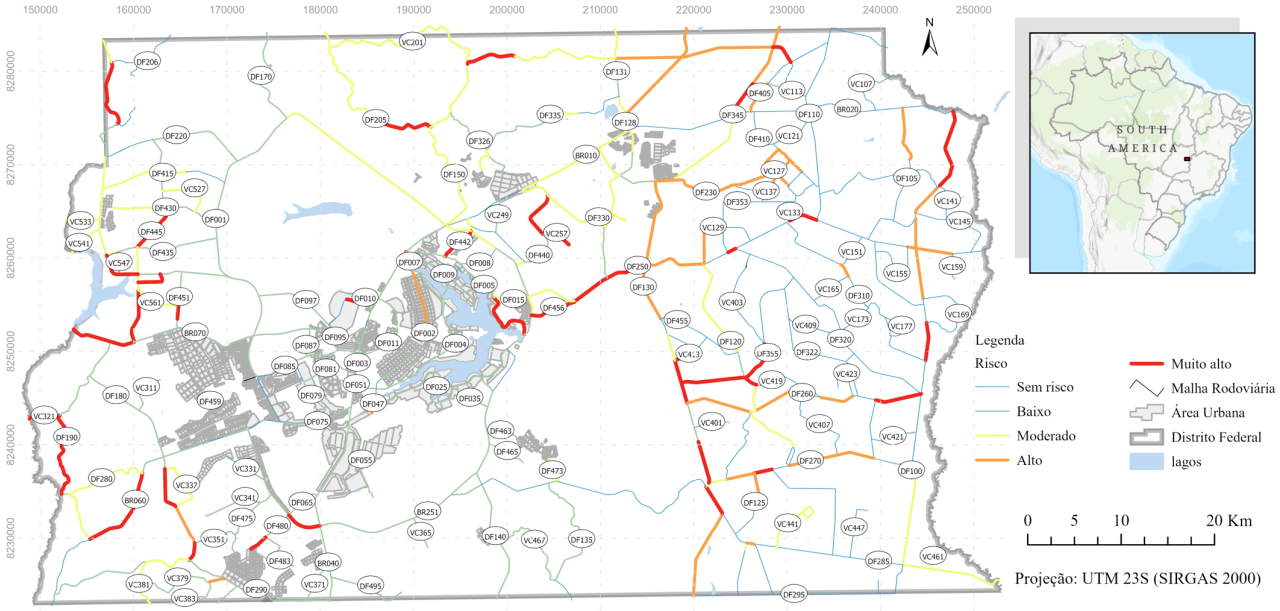
Fonte: Autores.

No conjunto da rede, os trechos classificados como hot spots somam 113,78 km (46 segmentos). Pela significância estatística, prevalecem áreas de 99% (68,69 km; 60,4% do total), seguidas de 95% (30,69 km; 27,0%) e 90% (14,40 km; 12,7%). Em termos de cobertura por via, a Tabela 1 indica 18 rodovias com trechos críticos; em 10 delas, o maior nível identificado foi 99%; em 3, o maior nível foi 95%; e, em 5, foi 90%. Esses padrões espaciais estão representados na Figura 2, na forma de corredores contínuos de maior intensidade.

Por rodovia, os maiores comprimentos de hot spots concentram-se na DF-003 (31,16 km; 27,4% do total; 99% predominante, com níveis de 99%, 95% e 90% presentes), DF-001 (16,05 km; 14,1%; 99%), DF-085 (13,01 km; 11,4%; 99%), DF-095 (12,54 km; 11,0%; 95%) e DF-075 (8,02 km; 7,0%; 99%). Considerando a taxa média de acidentes por km (ponderada pelo comprimento), destacam-se a DF-075 (85,71 acidentes/km), a DF-001 (68,68 acidentes/km) e a DF-085 (53,49 acidentes/km), enquanto a DF-003 apresenta 45,37 acidentes/km. No cruzamento atual, a DF-002 não registrou trechos classificados como hot spot. Esses resultados convergem com a Figura 2, na qual os eixos DF-001, DF-003, DF-075 e DF-085 aparecem como os principais corredores críticos em extensão e intensidade de significância.

Como complemento ao diagnóstico espacial, a Figura 3 apresenta o Mapa de Risco derivado do XGBoost para acidentes com motociclistas no SRDF, classificado em cinco níveis (muito baixo → muito alto). A performance do classificador está sintetizada na Tabela 3 (matriz de confusão) e na Tabela 4 (métricas globais; AUC-ROC = 0,8843), enquanto os principais preditores, segundo o ganho, são listados na Tabela 2, destacando natureza do acidente, interação motocicleta × caminhão, a localização e o dia da semana.

Figura 3 — Mapa de risco de fatalidade em acidentes com motociclistas no SRDF.



Fonte: Autores.

O Mapa de Risco representa a propensão à fatalidade, condicional à ocorrência do acidente, e não o volume de eventos. Assim, trechos rurais e de menor fluxo podem figurar como alto/muito alto: velocidades operacionais elevadas, menor iluminação, acostamentos estreitos e maior tempo de resposta de atendimento aumentam a chance de óbito caso o evento ocorra. Em contraste, nos corredores estruturantes (DF-085/EPTG, DF-075/EPNB, DF-095/EPCL, DF-003/EPIA), predomina risco baixo/moderado, com picos localizados em entroncamentos, retornos e acessos à rodovia DF-001, bem como às rodovias federais que cortam o Distrito Federal. Esse padrão é pontual e descontínuo, concentrado onde a complexidade operacional é maior, e é coerente com os preditores mais influentes (Tabela 2), notadamente a natureza do acidente, a interação motocicleta × caminhão, as coordenadas (LAT/LONG) e o componente semanal (sábado).

Tabela 2 - Importância das variáveis

Posição	Variável	Ganho
1	Ano do registro	0,309
2	Natureza – Colisão frontal	0,093
3	Natureza – Choque com objeto fixo	0,083
4	Localização – Coordenadas (LAT/LONG)	0,076
5	Envolvimento – Motocicleta × Caminhão	0,040
6	Natureza – Colisão traseira	0,036
7	Natureza – Colisão lateral	0,032
8	Dia da semana – Sábado	0,031
9	Natureza – Queda de veículo	0,028

Fonte: Os autores.

As importâncias foram estimadas pelo ganho do XGBoost (redução média da função-objetivo nos nós em que cada variável participa, acumulada ao longo das árvores). Os nove preditores listados concentram 72,8% do ganho total. O ano do registro apresentou a maior contribuição (ganho = 0,309; ≈42%), seguido por categorias de natureza do acidente — colisão frontal (0,093), choque com objeto fixo (0,083), colisão traseira (0,036), colisão lateral (0,032) e queda de veículo (0,028) —, pelo envolvimento motocicleta × caminhão (0,040), pela localização – coordenadas (LAT/LONG) (agregado = 0,076; ≈10%) e por dia da semana – sábado (0,031). A proeminência de ANO é compatível com tendência temporal da série; as coordenadas capturam autocorrelação espacial remanescente; e o bloco de natureza aponta mecanismos lesivos de maior severidade. Ressalte-se que "ganho" é uma métrica específica do algoritmo e não implica causalidade; além disso, preditores temporais/espaciais podem introduzir vazamento temporal/espacial e inflar a discriminação sob partições aleatórias.

Do ponto de vista classificatório, o modelo obteve bom poder discriminante, com Área sob a Curva da Característica de Operação do Receptor (AUC-ROC) = 0,8843. Na amostra de teste, a matriz

de confusão (Tabela 3) apresentou Verdadeiros Negativos (TN) = 1.724, Falsos Positivos (FP) = 9, Falsos Negativos (FN) = 43 e Verdadeiros Positivos (TP) = 35, resultando em acurácia = 0,9713 (IC95%: 0,9625–0,9785), Coeficiente Kappa (κ) = 0,560 e Acurácia Balanceada = 0,722; observou-se Especificidade elevada para a classe Não Fatal (0,995) e Sensibilidade moderada para a classe Fatal (0,449).

As métricas globais detalhadas na Tabela 4 consolidam o desempenho — Área sob a Curva (AUC), acurácia, Coeficiente Kappa (κ), Sensibilidade, Especificidade, Valor Preditivo Positivo (PPV), Valor Preditivo Negativo (NPV) e Acurácia Balanceada —, oferecendo um panorama sintético para comparação com ajustes futuros de limiar e balanceamento.

Tabela 3 – Dados das manobras registradas

	Referência: Não Fatal	Referência: Fatal
Previsto: Não Fatal	TN = 1.724	FN = 43
Previsto: Fatal	FP = 9	TP = 35

Fonte: Os autores.

Tabela 4 - Métricas de desempenho no conjunto de teste.

Métrica	Valor
AUC-ROC	0,8843
Acurácia	0,9713
IC95% (Acurácia)	0,9625 – 0,9785
Kappa	0,560
Sensibilidade (Fatal)	0,449
Especificidade (Não Fatal)	0,995
PPV (Fatal)	0,795
NPV (Não Fatal)	0,976
Acurácia balanceada	0,722
Total (N)	1.811

Fonte: Os autores.

As métricas reportadas avaliam o desempenho sob diferentes ângulos: AUC-ROC mede o poder discriminante global do modelo (probabilidade de pontuar um caso “fatal” acima de um “não fatal”, independentemente do limiar); a acurácia é a proporção total de acertos; a sensibilidade indica a fração de casos fatais corretamente identificados; a especificidade, a fração de não fatais corretamente descartados; o PPV é a probabilidade de um previsto “fatal” ser, de fato, fatal; o NPV, de um previsto “não fatal” ser realmente não fatal; o coeficiente Kappa expressa o acordo além do acaso (útil em bases desbalanceadas); e a acurácia balanceada corresponde à média entre sensibilidade e especificidade,

mitigando o efeito do desbalanceamento de classes. Como a avaliação utilizou uma única partição treino–teste, as estimativas estão sujeitas a maior variabilidade; recomenda-se empregar validação cruzada e/ou *early stopping* em trabalhos futuros para reforçar a robustez dos resultados.

5. DISCUSSÃO

A análise espacial por meio do método Getis-Ord Gi* evidenciou a formação de corredores críticos de acidentes envolvendo motociclistas no Sistema Rodoviário do Distrito Federal, com destaque para rodovias estruturais como DF-001, DF-003, DF-075, DF-085 e DF-095. Esse padrão confirma que a distribuição dos eventos não ocorre de forma aleatória, mas acompanha eixos de elevada conectividade, forte circulação diária e intensa interação entre fluxos urbanos e periurbanos. Em termos aplicados, isso indica que a frequência dos acidentes está fortemente associada à estrutura funcional da rede viária, especialmente em corredores de grande demanda e elevada complexidade operacional.

Esses achados convergem com estudos que demonstram a utilidade das técnicas de análise espacial para identificar áreas críticas e subsidiar intervenções territorialmente orientadas. Trabalhos como os de Tippayanate *et al*. (2024), Haque *et al*. (2010) e Mohammed *et al*. (2023) mostram que a distribuição dos acidentes tende a refletir características locacionais, padrões de conectividade e condições de operação do sistema viário. No caso do Distrito Federal, a concentração dos hot spots em rodovias estruturais sugere que a combinação entre alto fluxo, conflitos de tráfego e articulação metropolitana constitui elemento central para compreender a concentração dos eventos envolvendo motociclistas.

Tomados em conjunto, os resultados não indicam uma sobreposição ampla entre “onde mais ocorrem acidentes” e “onde é maior a chance de óbito caso o evento ocorra”. A análise de hot spot revela corredores urbanos de alta frequência, ao passo que o mapa preditivo derivado do XGBoost destaca, sobretudo, segmentos periurbanos e rurais, nos quais a probabilidade condicional de fatalidade tende a ser maior. Esse contraste é relevante porque evidencia duas dimensões distintas do risco viário: a frequência dos eventos e a severidade esperada. Assim, áreas com grande número de ocorrências não coincidem, necessariamente, com aquelas onde o desfecho fatal é mais provável.

Essa distinção é coerente com a literatura recente sobre segurança viária orientada por risco. Estudos alinhados ao paradigma do Safe System indicam que ambientes com velocidades operacionais mais elevadas, menor iluminação, menor proteção lateral e tempos de resposta mais longos tendem a aumentar a severidade dos acidentes, mesmo quando o volume de eventos é relativamente menor. Nesse sentido, os resultados do presente estudo sugerem que, no SRDF, corredores urbanos demandam ações voltadas à redução da frequência dos acidentes, enquanto trechos periurbanos e rurais requerem intervenções prioritariamente direcionadas à mitigação da severidade.

No âmbito da modelagem preditiva, os resultados do XGBoost indicaram que colisões frontais, choques com objeto fixo, colisões laterais e traseiras, além da interação entre motocicletas e caminhões, estão entre os principais fatores associados ao aumento da probabilidade de fatalidade. Esse resultado é compatível com evidências internacionais que apontam maior severidade em eventos com maior transferência de energia, maior exposição corporal do motociclista e

interação com veículos de maior porte. Estudos como os de Agyemang, Adanu e Jones (2021), Wisutwattanasak *et al*. (2024) e Yang *et al*. (2022) reforçam que a severidade dos acidentes com motociclistas depende não apenas da ocorrência em si, mas do contexto operacional, do tipo de conflito e das condições do ambiente viário.

A relevância de variáveis temporais e operacionais, como sábado e natureza do acidente, também sugere que o risco não é apenas espacial, mas resulta da combinação entre localização, dinâmica de circulação e contexto de uso da via. Esse ponto é importante porque reforça a necessidade de políticas intersetoriais que articulem engenharia, fiscalização e gestão da velocidade com recortes territoriais e temporais específicos. Em outras palavras, a utilidade do modelo não está apenas em predizer eventos severos, mas em oferecer evidências para intervenções mais seletivas e territorialmente calibradas.

Do ponto de vista do desempenho preditivo, o modelo apresentou AUC-ROC de 0,8843 e acurácia elevada, indicando boa capacidade discriminatória. Ainda assim, a sensibilidade moderada revela que parte dos casos fatais não foi capturada pelo classificador, o que recomenda cautela na interpretação operacional do modelo. Em contextos de segurança viária, essa limitação é particularmente importante, pois a utilidade prática de um modelo não depende apenas da boa performance global, mas também da capacidade de identificar eventos raros e de maior gravidade.

Algumas limitações devem ser destacadas. Em primeiro lugar, a utilização de uma única partição treino–teste pode aumentar a variabilidade das estimativas, de modo que aplicações futuras podem se beneficiar de validação cruzada e, idealmente, de estratégias com bloqueio espacial ou temporal. Em segundo lugar, a inclusão de latitude e longitude como variáveis explicativas, embora útil para capturar dependência espacial residual, pode introduzir sobreposição de informação espacial, devendo seus efeitos ser interpretados como proxies locacionais e não como relações causais. Em terceiro lugar, a ausência de variáveis diretamente relacionadas à geometria viária, às velocidades regulamentadas e aos volumes de tráfego limita a capacidade explicativa do modelo, ainda que parte desses efeitos possa ter sido parcialmente capturada por variáveis locacionais e operacionais.

Mesmo com essas restrições, os resultados demonstram que a combinação entre análise espacial e modelagem preditiva oferece ganhos analíticos importantes. Enquanto o Getis-Ord Gi* permite localizar corredores com maior concentração de acidentes, o XGBoost acrescenta uma leitura sobre a severidade esperada dos eventos. Essa abordagem integrada contribui para uma interpretação mais consistente do risco viário e amplia o potencial de aplicação dos resultados no planejamento da segurança viária no Distrito Federal.

6. CONCLUSÃO

Este estudo demonstrou que a combinação entre análise espacial (Getis-Ord Gi*) e modelagem preditiva (XGBoost) constitui uma estratégia analítica útil para compreender a acidentalidade envolvendo motociclistas no Sistema Rodoviário do Distrito Federal. Os resultados confirmaram a existência de corredores com forte concentração de acidentes, sobretudo em rodovias estruturais de elevada conectividade, ao mesmo tempo em que revelaram que a maior probabilidade condicional de fatalidade tende a se distribuir de forma distinta, com destaque para segmentos periurbanos e rurais.

Assim, o principal achado da pesquisa foi evidenciar que frequência de acidentes e risco de fatalidade não correspondem à mesma dimensão analítica. Enquanto a análise espacial orienta a identificação de áreas com maior volume de ocorrências, o modelo preditivo permite reconhecer trechos onde a severidade esperada é mais elevada, ainda que a quantidade de eventos seja menor. Essa distinção reforça a necessidade de estratégias diferenciadas de intervenção no território.

Do ponto de vista aplicado, os resultados sugerem que corredores urbanos mais congestionados e com maior concentração de acidentes demandam medidas voltadas à redução da frequência dos eventos, como reorganização de fluxos, fiscalização e melhorias operacionais. Já os segmentos com maior probabilidade de desfecho fatal requerem ações direcionadas à mitigação da severidade, especialmente por meio da gestão da velocidade, do reforço da segurança lateral, de melhorias de visibilidade e de atenção a conflitos envolvendo veículos de grande porte.

Em termos metodológicos, o estudo contribui ao integrar duas abordagens frequentemente tratadas de forma isolada na literatura: a identificação de clusters espaciais e a predição da severidade dos acidentes. Embora as técnicas utilizadas sejam conhecidas, sua aplicação combinada, com ênfase na distinção entre concentração de eventos e risco condicional de fatalidade, amplia o potencial analítico para o planejamento da segurança viária.

Por fim, recomenda-se que estudos futuros incorporem variáveis mais diretamente relacionadas às características da infraestrutura viária, às velocidades regulamentadas, aos volumes de tráfego e a dados complementares de fiscalização e atendimento. Também se recomenda o uso de estratégias de validação cruzada com bloqueio espacial ou temporal, de modo a reforçar a robustez dos modelos e reduzir possíveis efeitos de dependência espacial residual.

BIBLIOGRAFIA

AGYEMANG, W.; ADANU, E. K.; JONES, S. (2021). Understanding the factors that are associated with motorcycle crash severity in rural and urban areas of Ghana. Journal of Advanced Transportation, **2021**, e6336517. DOI: https://doi.org/10.1155/2021/6336517.

BRASIL. Secretaria Nacional de Trânsito (SENATRAN). (2024). **Frota de Veículos – 2024 (Registro Nacional de Veículos Automotores – RENAVAM)**. Ministério dos Transportes, Brasília, DF.

BRENAC, T.; CLABAUX, N.; PERRIN, C.; VAN ELSLANDE, P. (2006). Motorcyclist conspicuity-related accidents in urban areas: a speed problem? Advances in Transportation Studies – Section A, n. **8**, p. 23–29.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, v. **16**, p. 321–357. DOI: https://doi.org/10.1613/jair.953.

CHEN, F.; ZHANG, Q.; WANG, Z.; *et al*. (2025). Traffic accident severity prediction based on an enhanced MSCPO-XGBoost hybrid model. Scientific Reports, v. **15**, art. 25729. DOI: https://doi.org/10.1038/s41598-025-00797-7.

DISTRITO FEDERAL. Departamento de Trânsito do Distrito Federal (DETRAN-DF). (2024). **Frota de veículos registrados no DF em circulação — 2º trimestre de 2024**. Detran-DF, Brasília, DF.

EUROPEAN COMMISSION. (2023). **Road Safety Thematic Report: Safe System Approach**. European Commission, Brussels.

GETIS, A.; ORD, J. K. (1992). The analysis of spatial association by use of distance statistics. Geographical Analysis, v. **24**, n. **3**, p. 189–206. DOI: https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

HAQUE, M. M.; CHIN, H. C.; HUANG, H. (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Analysis & Prevention, v. **42**, n. **1**, p. 203–212. DOI: https://doi.org/10.1016/j.aap.2009.07.022.

JIANG, F.; YUEN, K. K. R.; LEE, E. W. M. (2020). Analysis of motorcycle accidents using association rule mining-based framework with parameter optimization and GIS technology. Journal of Safety Research, v. **75**, p. 292–309. DOI: https://doi.org/10.1016/j.jsr.2020.09.004.

LUNDBERG, S. M.; LEE, S.-I. (2017). A unified approach to interpreting model predictions. In: **Advances in Neural Information Processing Systems (NeurIPS 2017)**. DOI: https://doi.org/10.48550/arxiv.1705.07874.

MIAO, J.; PAN, Y.; ZHAO, K. (2025). Analysis of the spatiotemporal effects on the severity of motorcycle accidents without helmets and strategies for building sustainable traffic safety. Sustainability, v. **17**, n. **8**, art. 3280. DOI: https://doi.org/10.3390/su17083280.

MOHAMMED, S.; ALKHEREIBI, A. H.; ABULIBDEH, A.; JAWARNEH, R. N.; BALAKRISHNAN, P. (2023). GIS-based spatiotemporal analysis for road traffic crashes: in support of sustainable transportation planning. Transportation Research Interdisciplinary Perspectives, v. **20**, 100836. DOI: https://doi.org/10.1016/j.trip.2023.100836.

OECD/ITF. (2008). **Towards Zero: Ambitious Road Safety Targets and the Safe System Approach**. OECD/ITF, Paris.

PANUMASVIVAT, J.; *et al*. (2024). Unveiling the road to safety: understanding the factors influencing motorcycle accidents among riders in rural Chiang Mai, Thailand. Heliyon, v. **10**, n. **3**, e25698. DOI: https://doi.org/10.1016/j.heliyon.2024.e25698.

QIAN, R. Y.; WANG, X. (2023). Prediction of road traffic accident severity based on XGBoost-BP neural network. Advances in Transportation Studies, v. **61**, p. 19–36. DOI: https://doi.org/10.53136/97912218091902.

SANTOS JÚNIOR, C. J.; SILVA, J. P.; ROCHA, T. J. M. (2022). Análise da ocorrência de acidentes de trânsito fatais nas capitais do Nordeste do Brasil. Hygeia – Revista Brasileira de Geografia Médica e da Saúde, v. **18**, p. 44–53. DOI: https://doi.org/10.14393/hygeia1858395.

SANTOS, J. P.; GUIMARÃES, P. H.; SILVA, B. A. L.; CAMPANI, A. R.; OLIVEIRA, J. M. V. (2023). DERGeo: geovisualizador de informações geográficas para gestão do Sistema Rodoviário do Distrito Federal. Revista Brasileira de Transportes, v. **3**, n. **1**, p. 62–85. DOI: https://doi.org/10.12660/rbt.v3n1.2023.89030.

SIQUEIRA, J. S.; WERNECK, G. L.; PENA, P. G. L.; FERNANDES, R. C. P. (2025). Trabalho subordinado de entregadores em empresas-plataforma de entrega e acidentes nas ruas: magnitude e fatores associados. Cadernos de Saúde Pública, v. **41**, n. **3**, e00204723. DOI: https://doi.org/10.1590/0102-311xpt204723.

TIPPAYANATE, N.; IMPOOL, T.; SUJAYANONT, P.; MUTTITANON, W.; CHEMIN, Y.; SOM-ARD, J. (2024). A geospatial analysis of motorcycle accident risk factors in Khon Kaen Municipality, Thailand: examining the chain of survival and potential strategies. International Journal of Geoinformatics, v. **20**, n. **9**, p. 71–82. DOI: https://doi.org/10.52939/ijg.v20i9.3547.

WORLD HEALTH ORGANIZATION (WHO). (2023). **Global status report on road safety 2023**. WHO, Geneva.

WISUTWATTANASAK, P.; SE, C.; CHAMPAHOM, T.; *et al*. (2024). Factors affecting single and multivehicle motorcycle crashes: insights from day and night analysis using XGBoost-SHAP algorithm. Big Data and Cognitive Computing, v. **8**, n. **10**, 128. DOI: https://doi.org/10.3390/bdcc8100128.

YANG, Y.; LI, Y.; WANG, J.; *et al*. (2022). Predicting freeway traffic crash severity using XGBoost-Bayesian network model with consideration of features interaction. Journal of Advanced Transportation, v. **2022**, 4257865. DOI: https://doi.org/10.1155/2022/4257865.

ARTICLE 6

URBAN ROAD SAFETY AND MANAGEMENT: SPATIAL PATTERNS AND PREDICTED FATALITY RISKS AMONG MOTORCYCLISTS IN BRAZIL'S FEDERAL DISTRICT ROAD SYSTEM (2014–2023)

Janduhy Pereira dos Santos (DER/DF), Valdir Adilson Steinke, Maurício Theodósio Mattos Marques, Ricardo Sameshima Taba, Andrea Amaziles A. A. de Carvalho Lousada, Bruno Maia Soriano Lousada, Jesus Mauro Vieira de Oliveira e Caroline Ribeiro Chahini

RESUMO

Este artigo investiga a dinâmica espacial e os fatores associados à gravidade de acidentes envolvendo motociclistas no Distrito Federal, com base em dados de acidentes rodoviários (2014–2023). Adota-se um fluxo analítico que combina análise espacial para detecção de áreas críticas (Getis-Ord Gi*), modelagem preditiva por gradient boosting (XGBoost) e balanceamento de classes (SMOTE), com interpretação da importância dos preditores via SHAP. A avaliação baseia-se em partição treino–teste estratificada, priorizando a comparação de desempenho e a leitura substantiva das variáveis. Os achados indicam padrões concentrados no espaço urbano e a relevância de características do evento, do contexto operacional e de aspectos temporais na explicação do risco. As evidências resultantes subsidiam estratégias de segurança viária voltadas à priorização territorial, à fiscalização direcionada e a ações de moderação de velocidade, contribuindo para o planejamento intersetorial e para a redução de desfechos graves entre motociclistas.

Palavras-chave: acidentes de trânsito; motocicletas; análise espacial

ABSTRACT

This study investigates the spatial dynamics and factors associated with the severity of motorcycle-involved road crashes in Brazil's Federal District, using crash records from 2014 to 2023. An analytical workflow is adopted that combines spatial statistics to detect critical areas (Getis-Ord Gi*), predictive modeling via gradient boosting (XGBoost), and class imbalance handling with SMOTE, with predictor importance interpreted using SHAP. Model evaluation relies on a stratified train–test split, prioritizing comparative performance assessment and the substantive interpretation of predictors. The results highlight the relevance of event-specific attributes, operational context, and temporal factors in explaining risk. The resulting evidence informs road safety strategies focused on geographic prioritization, targeted enforcement, and speed management measures, supporting cross-sector planning and the reduction of severe outcomes among motorcyclists.

Keywords: Road traffic accidents; Motorcycles; Spatial analysis.

1. INTRODUCTION

During the past few decades, rapid urban growth in Brazil has triggered significant transformations in territorial dynamics and mobility patterns, especially in major metropolises. Associated with inadequate and unappealing public transport systems, spreading urban sprawl has fuelled the expansion of motor vehicle fleets, especially motorcycles. This mode now plays a strategic role in urban mobility within this context, as either an effective way of avoiding traffic jams or as a tool for entering the labour market, especially through activities running on digital platforms.

In Brazil's Federal District, this dynamic has intensified significantly, driven by economic and structural factors. From 2015 onwards, an economic crisis and the reconfiguration of the labour market fuelled the growth of casual work and unregistered jobs, many dependent on motorcycles. This process was intensified by the COVID-19 pandemic, with soaring demands for individual transport and delivery services. The outcomes were a significant expansion of the motorcycle fleet, accompanied by an upsurge in traffic accidents involving this group, ranked among the most vulnerable road system users (SENATRAN, 2024; DETRAN-DF, 2024).

From the road safety standpoint, recent literature is moving away from approaches focused only on identifying critical points, instead including systemic viewpoints based on risk management, such as the Safe System paradigm. This approach acknowledges user vulnerability and the inevitability of human error, assigning system managers responsibilities for structuring roads, regulating speeds, and organising traffic, in order to minimise the occurrence of serious accidents and fatal injuries (OECD/ITF, 2008; WHO, 2023).

In this context, spatial analyses and predictive modelling have firmed up their status as vital tools for understanding accidentality, not only identifying territorial patterns, but also estimating occurrence probabilities for serious outcomes. Recent studies have deployed techniques such as spatial statistics and machine learning algorithms to investigate factors associated with the accident severity, highlighting the relevance of variables related to road environments, user behaviours, and the transport system operating conditions.

However, despite this methodological progress, there are still important gaps in the literature, especially for integrating traditional spatial analyses that identify areas where events are concentrated, and predictive approaches able to estimate the conditional fatality risks. In many studies, these aspects are addressed individually, which hampers an integrated understanding of these phenomena, with less potential for applying these findings to road safety planning.

This is the context within which this study strives to (i) identify spatial patterns for motorcycle accidents in the Federal District Highway System (2014–2023); and (ii) develop a predictive fatality risk model based on machine learning techniques, for the purpose of identifying critical factors associated with event severity.

The main contribution of this paper is the integration between spatial analyses of accident concentration (Getis-Ord Gi*) and predictive modelling of conditional fatality risks (XGBoost). This allows two aspects to be distinguished empirically, which are frequently confounded in the literature: (i) where accidents are more frequent; and (ii) where the probabilities of death are higher. This combined approach ensures more accurate road risk estimates, thus helping prioritise customised interventions in this field.

Moreover, this study takes a further step forward by applying this methodological integration at the road network scale, including operating and time related variables while exploring model interpretation techniques through the SHapley Additive exPlanations (SHAP) game-theory-based framework, for a better understanding of mechanisms associated with accident severity. Although these techniques are widely acknowledged in the literature, their joint application constitutes a relevant analytical advance in the road safety field, focused on the distinction between absolute risk and conditional risk.

Providing input for mobility planning and management, these findings help steer traffic engineering, oversight, and education actions, aligned with evidence-driven approaches pursuing reductions in deaths and serious injuries.

2. LITERATURE REVIEW

Over the past thirty years, motorcycle uses and purposes have changed significantly. Once preferred for recreation and individual commuting, these vehicles were of little use for work purposes. However, spreading metropolitanisation processes in large towns and cities, together with rising demands for outsourced services and fast deliveries, have fuelled significant growth for motorcycle fleets in many countries.

In the transport and road safety fields, contemporary discussions stress that lowering mortality requires a combination of risk management, governance, and oversight with integrated interventions in infrastructure, safe speeds, vehicles, enforcement, and post-accident care. Summarised reports note that global traffic death burdens remain high, despite recent dips, underscoring the need for evidence-based strategies with performance and risk indicators, looking not only at absolute event figures (WHO, 2023).

Despite growing consensus in the literature on the need for systemic on risk-based management approaches, it is nevertheless clear that many studies still focus on identifying aggregate accident occurrence patterns, without explicitly addressing the distinction between event frequency and the probabilities of serious outcomes. This limitation weakens their potential for applying these findings to planning better targeted interventions, especially in contexts where exposure and severity are not spatially coincident.

For vulnerable users such as motorcyclists, this agenda is particularly important, as severity tends to rise in parallel with exposure to high speeds and conflicts in road environments not designed to absorb impacts. Location-based predictive studies must be grounded in risk management logic: where infrastructures and operations

may be redesigned to reduce collision energy and segregate conflicts, aligned with Safe System principles (OECD/ITF, 2008; European Commission, 2023).

Similarly, Santos Júnior et al. (2022) describe fatal traffic accidents as a complex and challenging problem for public authorities. The authors highlight risky behaviours in urban traffic and the high social and economic costs of fatal traffic accidents, advocating intersectoral actions and better oversight. These losses include overburdening healthcare and social security services, reflecting high social and economic costs associated with motorcycle accidents.

This concern for ensuring safe conditions for motorcyclists encompasses all other mobility sectors. In this context, Brenac et al. (2006) highlight the importance of motorcyclist visibility on roads, emphasising that difficulties in visually perceiving nearby objects or people is a recurrent factor in urban accidents.

This concern extends to rural roads, where motorcycle use absorbs a variety of functions beyond merely transporting riders. When examining motorcycle accidents in rural Ghana, Agyemang, Adanu, and Jones (2021) note that these events are often overlooked by public policies. These authors point out that the poor road conditions, limited access to emergency services, and risky radar behaviours directly affect the severity of injuries recorded in these contexts.

The use of analytical instruments, such as geoinformation tools, has become essential for the preparation and implementation of public policies pursuing road safety, identifying critical areas and proposing effective interventions designed to reduce traffic accidents. Studies such as those by Tippayanate et al. (2024) and Haque et al. (2010) demonstrate the potential of spatial analysis techniques for identifying territorial motorcycle accident patterns, thus helping steer preventive actions. Overall, these studies confirm the effectiveness of spatial analysis techniques in identifying critical areas. However, most approaches remain focused on detecting occurrence clusters, with little progress in event-conditional risk modelling. This constraint may lead to prioritising areas with high accident rates, but not necessarily higher fatality probabilities. Furthermore, Miao et al. (2025) use spatio-temporal models applied to China to demonstrate that helmet absence is closely linked to accident severity, underscoring the importance of georeferenced data for evidence-based decisions.

The application of geoinformation techniques has proven strategic for drawing up public policies focussed on mobility and road safety. Studies such as that by Jiang, Yuen and Lee (2020) demonstrate how the integration between association rule mining and spatial analysis through Geographic Information Systems (GIS), allows the identification of critical factors associated with motorcycle accidents, while also mapping areas where these factors converge. International studies deploying the XGBoost technique for road accident severity predictions, such as analyses by Wisutwattanasak et al. (2024) on motorcycle accidents in Thailand; Chen et al. (2025) on MSCPO-XGBoost optimisation in China; Qian and Wang (2023) on British accidents; and Yang et al. (2022) on US highways, confirm the robustness of this

technique for capturing nonlinear factors and interpreting contextual variables in road safety studies. Despite progress in the use of machine learning algorithms to predict accident severity, it is noted that these approaches often fail to explicitly include the spatial aspects of these phenomena, treating location mainly as an additional explanatory variable, without fully exploring its spatial structure. Moreover, there is little integration between predictive models and exploratory spatial analyses, limiting joint understandings of territorial patterns and risk mechanisms.

Consequently, as shown by Panumasvivat et al. (2024), decision-making in public mobility policies relies on geospatial techniques as essential input for identifying risk factors, not only for motorcyclists, but also for other transport system users, in both urban and rural areas. These tools pave the way for steps such as traffic engineering interventions, more intensive oversight and inspections, road infrastructure upgrades, and awareness campaigns run at specific times and places.

By generating predictive maps and high-incidence clusters with statistical bases, this study presents suggestions for interventions in infrastructure, oversight, and educational campaigns. No longer merely high-tech toys, geospatial techniques firm up their status as ancillary tools for pursuing territorial justice and preventing accidents. Although several studies underline the relevance of variables related to road geometry, speed limits, and traffic volumes for explaining accident severity, the availability and standardisation of these data are still challenging in many institutional contexts. This constraint has prompted researchers to use spatial and operational data, such as geographic location and event characteristics, as a way to partially capture these effects.

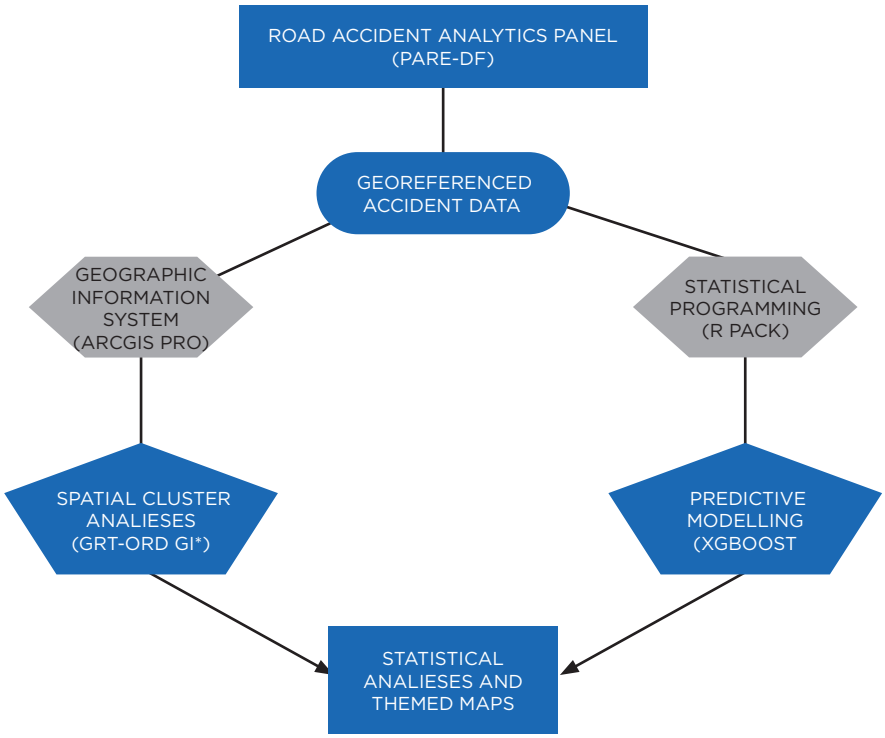
3. METHODOLOGY

This study adopts a quantitative approach, using spatial analysis methods to identify distribution patterns and factors associated with traffic motorcycle accidents in the Federal District Highway System (FDHS).

This study is consequently structured into three interconnected sections (Figure 1), as shown in the methodology flowchart: (i) database organisation, cleaning and standardisation, with record geocoding and integration into the FDHS road network; (ii) spatial analysis using the Hot Spot Analysis method (Getis-Ord Gi*) to identify critical zones (hot spots) and low incidence (cold spots) for motorcycle accidents; and (iii) predictive modelling using the XGBoost algorithm to identify critical factors that ramp up fatality probabilities.

The integration of these steps produced statistical analyses and thematic maps showing accident occurrence patterns, pinpointing significant motorcycle accident clusters, and estimating risk probabilities for different road segments.

Figure 1 – Methodological flowchart.



Source: Authors.

3.1. Data Organisation and Processing

The records used in this analysis were extracted from the DERGeo geoviewer (Santos et al., 2023), a platform developed by the Federal District Highway Department (DER-DF). This provides georeferenced information on accidents in the road network under its jurisdiction, through the Federal District Road Accident Analytics Panel (PARE-DF) module. This tool offers access to a structured spatial base that handled the statistical and geoinformation techniques throughout this study.

The analysed dataset encompasses 9,055 records of motorcycle accidents in the FDHS between 2014 and 2023, georeferenced in UTM coordinates (spindle 23S, Datum SIRGAS 2000). These entries include detailed information, such as occurrence dates, accident types, and vehicle categories involved, in addition to the geolocation of each event.

Prior to these spatial and statistical analyses, this database was run through a pre-processing procedure, using Microsoft Excel software to prepare and structure these records. These steps involved: identifying and eliminating duplicates and inconsistencies, standardising categorical variables (such as accident types, days of the week, and vehicle categories), converting date and time entries into formats readable by geographic information systems, and validating geographic coordinates. This procedure ensured the semantic consistency and syntactic integrity of the data, for subsequent import and processing in a geospatial environment through the ArcGIS Pro software.

3.2. Spatial and predictive analysis

This methodological stage combines spatial analysis techniques with machine learning methods, designed for a deeper understanding of motorcycle accident risk patterns in the Federal District Highway System.

Initially, the Hot Spot Analysis (Getis-Ord Gi*) technique was applied through the ArcGIS Pro 3.5.1 software, identifying statistically significant clusters with high and low event concentrations. Classified as a spatial exploratory analysis, this stage identifies accident frequency occurrence patterns in the road network.

Next, a predictive model was developed using the XGBoost algorithm, implemented in the R environment (version 4.3.2) to estimate the probabilities of deaths resulting from these accidents. Unlike exploratory spatial analyses, this approach includes an explanatory and prognostic aspect that identifies factors associated with event severity.

The combined use of these two approaches analytically distinguishes two complementary aspects: (i) spatial accident clusters; and (ii) conditional fatality risks, leading to more comprehensive road safety interpretations.

In this study, spatiality was treated in a complementary manner: through exploratory spatial analysis of accident clusters (Getis-Ord Gi*) on the one hand, and on the other by including geographic coordinates as locational proxies in the predictive model, together with time, operating and other related variables.

3.2.1. Getis-Ord Gi Algorithm*

Clusters were identified through the Hot Spot Analysis method (Getis-Ord Gi*), which estimates local spatial associations based on neighbourhood matrixes and spatial weights, pinpointing high and low value clusters with statistical significance (Getis; Ord, 1992).

A hot spot is an area or location with a cluster of accidents within an identifiable boundary. The main ways of estimating accident clusters include georeferenced traffic accident analyses and cluster mapping using the Getis-Ord Gi* functions.

Local spatial association statistics are given by:

$$G_i^* = \frac{\sum_j w_{ij} x_j - \bar{X} \sum_j w_{ij}}{S \sqrt{\frac{n \sum_j w_{ij}^2 - \left(\sum_j w_{ij}\right)^2}{n-1}}} \quad (1)$$

Where:

$Gi*$ is the index value for location i.
xj is the value of the attribute in location j.
wi, j is the spatial weight between the i and j locations.

X is the average of the attribute values.
S is the standard deviation of the attribute values.
n is the total number of locations.

This method assesses spatial event cluster intensities through values attributed to each feature and its immediate neighbours, identifying statistically significant groupings of frequent occurrences (hot spots) and few occurrences (cold spots).

Spatial neighbourhoods were defined by a distance criterion of 1,500 metres. The selection of this parameter was based on preliminary spatial autocorrelation tests, Global Moran's Index (Global Moran's I),

and a connectivity analysis of the Federal District Highway System, using the average scale of interactions between urban and peri-urban road segments. Shorter distances resulted in excessive cluster fragmentation, while longer distances tended to dilute significant local patterns.

A fixed-distance spatial weight matrix was used to ensure comparability among features with different neighbour headcounts. The adopted significance level was 95% ($p < 0.05$) with interpretations based on z-scores and p-values.

Calculations were handled entirely through the ArcGIS Pro software, using a Hot Spot Analysis tool (*Getis-Ord Gi**) to classify the findings into five categories (Very Low, Low, Neutral, High, and Very High).

3.2.2. Predictive modelling with xgboost

For the predictive stage of this study, the Extreme Gradient Boosting (XGBoost) algorithm was used. Belonging to the ensemble learning methods class, specifically the boosting technique, this algorithm sequentially constructs decision tree sets (weak learners), where each new tree is adjusted to correct the errors of its predecessors, optimising objective functions through descending gradients (CHEN et al., 2025).

Predictive modelling was implemented in the R environment (version 4.3.2) through the XGBoost package, while the exploratory spatial analysis steps were run on ArcGIS Pro software, version 3.5.1. This separation reflects the complementary nature of these approaches: the GIS environment for spatial analyses and the statistical environment for predictive modelling.

XGBoost is particularly effective for binary classification tasks, such as that adopted in this study – predicting the occurrence of fatal (1) or non-fatal (0) accidents – due to its ability to model nonlinear relationships, handle unbalanced data, and provide integrated regularisation mechanisms that lower overfitting risks.

The general objective function of XGBoost may be expressed as follows:

Where:

ℓ is the loss function;
ŷi^(t) is the t stage prediction;
Ω(ft) I is the regularisation term;
γ is the complexity cost per tree leaf;
T is the number of leaves;
λ is the L2 regularisation coefficient;
w is the tree weight vector.

$$\mathcal{L}^{(t)} = \Sigma_{i=1}^{n} \ell \left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i) \right) + \Omega(f_t), \Omega(f) = \gamma \cdot T + (\lambda/2) \cdot \Sigma_{j=1}^{T} w_j^2 \quad (2)$$

The prediction for the xi instance after t trees is given by:

$$\hat{y}_i = \Sigma_{k=1}^{K} f_k(x_i), f_k \in \mathfrak{F} \quad (3)$$

Where $\boldsymbol{F}$ represents a possible decision tree space.

Initially, these data were run through a treatment and organisation stage that included: (i) developing categorical variables; (ii) checking for missing values and eliminating inconsistent records; (iii) changing date fields to the appropriate format; and (iv) addressing imbalances among classes. The Synthetic Minority Over-Sampling Technique (SMOTE) was used here, generating synthetic instances in minority neighbourhoods and lessening classifier bias with no mere duplication of observations (Chawla et al., 2002).

The dataset was divided into 70% for training and 30% for testing, using stratified sampling. The option for a single training-test division was intended to preserve the clarity of the model and allow direct comparisons with similar studies in the literature. However, it is acknowledged that this approach may lead to estimate variability, and the use of cross-validation is recommended in future studies, preferably with spatial or temporal blocks.

The performance assessment evaluated accuracy and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), ensuring an even balance between precision and sensitivity. To interpret the model and identify the most relevant factors, variable significance graphs were generated from gain values; Shapley Additive Explanations (SHAP) values were used to understand the marginal effects of each variable.

This approach helped estimate the probabilities of fatal accidents along different road segments. It also allowed identification of major fatality risk predictors, highlighting accident types, motorcycle and truck involvement, geographic location (LAT/LONG) and temporal factors, such as days of the week (particularly Saturday).

The main hyperparameters used in the model were: *eta* = 0.1; *max_depth* = 6; *nrounds* = 100; *subsample* = 0.8; *colsample_bytree* = 0.8; objective = *binary:logistic; eval_metric = auc.* The choice of these parameters was steered by empirical tests and recommendations drawn from the literature, seeking an even balance between predictive performance and overfitting control.

The interpretation stage used SHAP, breaking predictions down into additive contributions by variables (Shapley values), thus underpinning global importance and local effects analyses (Lundberg; Lee, 2017).

It should be noted that the inclusion of geographic coordinates (latitude and longitude) as explanatory variables may result in overlapping spatial information, as they may reflect spatial patterns already present in response variables. This study retains these variables, in order to capture spatial effects not directly observed, serving as proxies for road environment characteristics. However, the findings associated with these variables should be interpreted with caution, considered not as causal relationships, but rather as residual spatial dependence indicators.

4. FINDINGS

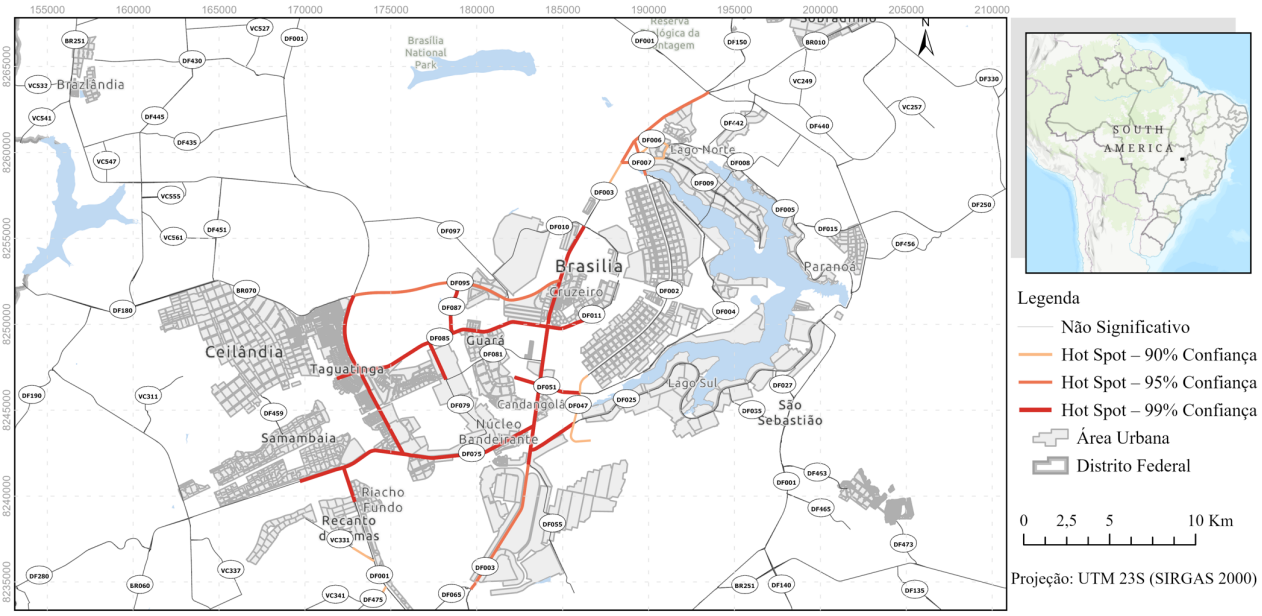
Between 2014 and 2023, 9,055 motorcycle accidents were recorded in the FDHS, of which 439 resulted in deaths and 8,616 had non-fatal victims. These totals refer to the geocoded and validated records used in subsequent analyses.

Application of the Hot Spot Analysis technique (Getis-Ord Gi*) identified statistically significant patterns for spatial accident clusters involving motorcyclists in the FDHS (Figure 2). This approach distinguishes segments with high accident rates (hot spots) from others with few accidents (cold spots), by examining their geographic contexts and the spatial distribution of these events. By including spatial autocorrelation, this method pinpoints areas where observed values are not distributed randomly, but are rather grouped consistently, reflecting the presence of local factors that increase or decrease accident risks.

This analysis highlighted hot spots with high statistical significance, distributed mainly along structural thoroughfares connecting downtown Brasília to outlying administrative regions and nearby towns. Highlighted segments include the DF-001, DF-003, DF-075, DF-085, and DF-095 highways, in numerical order. With continuous record clusters, these segments form critical corridors within this road network. Areas not highlighted on the map show stretches without statistical significance, while some outlying regions present low occurrence values, indicating cold spots. This spatial demarcation confirms that the heaviest clusters are grouped in the central-southeastern portion of the Federal District, along high connectivity routes with heavy daily traffic flows.

Presenting details of critical corridors identified in Figure 2, Table 1 lists road segments rated as hot spots, together with the highest observed significance levels, total length (km), and average accident rates per kilometre. This overview complements the map, for easier ranking of roads with heavier traffic and higher risks.

Figure 2 – Motorcycle accident hot spots (Gi*) in FDHS, 2014–2023.



Source: Authors.

Table 1 – Segments with significant motorcycle accident clusters (Getis-Ord Gi*) in the FDHS, 2014–2023.

Highway	Total length (km)	Average rate (accidents/km)	Higher significance level	Present levels
DF-001	16,05	68,68	99%	99%
DF-003	31,16	45,37	99%	99%, 95%, 90%
DF-005	0,79	15,18	90%	90%
DF-006	2,25	2,66	90%	90%
DF-007	2,12	23,62	95%	95%
DF-009	2,53	15,01	95%	95%, 90%
DF-011	2,24	50,96	99%	99%
DF-025	2,98	26,81	99%	99%
DF-047	5,1	17,25	90%	90%
DF-051	3,94	21,59	99%	99%
DF-075	8,02	85,71	99%	99%
DF-079	2,37	18,17	99%	99%
DF-085	13,01	53,49	99%	99%
DF-087	3,03	18,16	99%	99%
DF-095	12,54	35,97	95%	95%
DF-475	1,51	9,96	90%	90%

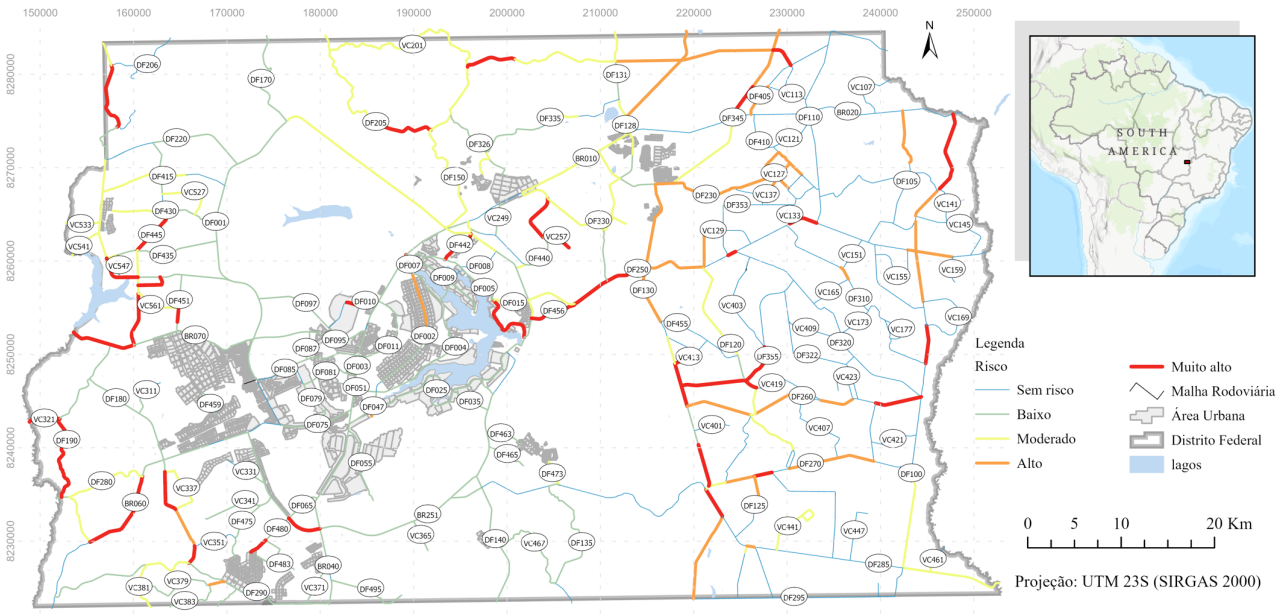
Source: Authors.

In the road network as a whole, 46 segments ranked as hot spots, with a total length of 113.78 km. Based on statistical significance, areas at 99% (68.69 km; 60.4% of the total) prevail, followed by 95% (30.69 km; 27.0%) and 90% (14.40 km; 12.7%). In terms of road coverage, Table 1 lists 18 highways with critical segments; along ten of them, the highest identified level was 99%; for three, the highest level was 95%; reaching 90% for the remaining five. These spatial patterns are presented in Figure 2, shown as continuous corridors with heavier traffic flows.

By highway, the longest hot spots are clustered in DF-003 (31.16 km; 27.4% of the total; 99% predominant, with levels of 99%, 95% and 90%), DF-001 (16.05 km; 14.1%; 99%), DF-085 (13.01 km; 11.4%; 99%), DF-095 (12.54 km; 11.0%; 95%) and DF-075 (8.02 km; 7.0%; 99%). Considering the average accident rate per km (weighted by length), DF-075 (85.71 accidents/km), DF-001 (68.68 accidents/km) and DF-085 (53.49 accidents/km) are notable, while DF-003 has 45.37 accidents/km. At the current intersection, DF-002 had no segments ranked as hot spots. These findings are convergent with Figure 2, showing the DF-001, DF-003, DF-075 and DF-085 highways as the main critical corridors by length and significance.

As a complement to the spatial diagnosis, Figure 3 presents the XGBoost Risk Map for motorcycle accidents in the FDHS, divided into five levels that range from Very Low to Very High. The classifier performance is summarised in Table 3 (confusion matrix) and Table 4 (global metrics; AUC-ROC = 0.8843), while the main predictors are listed by gain in Table 2, listing accident types, motorcycle ×truck interactions, locations, and days of the week.

Figure 3 – Fatal motorcycle accident risk map for the FDHS.



Source: Authors.

This Risk Map presents the likelihood of fatalities resulting from accidents, rather than event volumes. Consequently, rural areas and segments with lighter traffic flows may appear as high or very high risk, as speeding traffic, poor streetlighting, narrow shoulders, and longer service response times increase the chances of death should an event occur. In contrast, low to moderate risks predominate in the structuring corridors (DF-085/EPTG, DF-075/EPNB, DF-095/EPCL, DF-003/EPIA), peaking at junctions, roundabouts, and access roads leading to the DF-001 highway, as well as federal highways cutting through the Federal District. This patchy pattern is location-specific, clustered where traffic operations are most complex. It is consistent with the most influential predictors (Table 2), notably accident types, motorcycle x truck interactions, geographic coordinates (LAT/LONG), and days of the week component (Saturday).

Table 2 - Variables Ranking

Rank	Variable	Gain
1	Registration Year	0.309
2	Type – Head-on collision	0.093
3	Type – Collision with fixed object	0.083
4	Location – Coordinates (LAT/LONG)	0.076
5	Motorcycle × truck involvement	0.040
6	Type – Rear-end collision	0.036
7	Type – Side collision	0.032
8	Day of the week – Saturday	0.031
9	Type – Vehicle fall	0.028

Source: Authors.

These figures were estimated by XGBoost gains (mean reduction in the objective function of the nodes, where each variable participates, accumulated along the trees). These nine predictors cluster 72.8% of total gains. The highest contribution was from registration year (gain = 0.309; ≈ 42%), followed by accident type – head-on collision (0.093), collision with a fixed object (0.083), rear-end collision (0.036), side collision (0.032) and vehicle fall (0.028); motorcycle × truck involvement (0.040); location (LAT/LONG) coordinates (aggregate = 0.076; ≈ 10%); and by day of the week – Saturday (0.031). The prominence of the registration year is compatible with the series time trend; coordinates capture remaining spatial autocorrelations, while the type group indicates more severe harmful mechanisms. It should be noted that "gain" is a specific metric of the algorithm and does not imply causality; moreover, temporal/spatial predictors can introduce temporal/spatial leakage and inflate discrimination under random partitions.

From the classificatory standpoint, the model achieved good discriminant power, with the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) = 0.8843. In the test sample, the confusion matrix (Table 3) showed True Negatives (TN) = 1,724; False Positives

(FP) = 9; False Negatives (FN) = 43; and True Positives (TP) = 35; with accuracy = 0.9713 (95%CI: 0.9625–0.9785); Kappa Coefficient (κ) = 0.560; and Balanced Accuracy = 0.722. High specificity was noted for the Non-Fatal class (0.995), with moderate sensitivity for the Fatal class (0.449).

The global metrics presented in Table 4 consolidate Performance as the Area Under the Curve (AUC), Accuracy, Kappa Coefficient (κ), Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and Balanced Accuracy, providing an overview for comparison with future threshold and balancing adjustments.

Table 3 - Confusion Matrix

	Referência: Não Fatal	Referência: Fatal
Predicted: Non-Fatal	TN = 1.724	FN = 43
Predicted: Fatal	FP = 9	TP = 35

Source: Authors.

Table 4 - Test suite performance metrics.

Métrica	Valor
AUC-ROC	0.8843
Accuracy	0.9713
95%CI (Accuracy)	0.9625 – 0.9785
Kappa	0.560
Sensitivity (Fatal)	0.449
Specificity (Non-Fatal)	0.995
PPV (Fatal)	0.795
NPV (Non-Fatal)	0.976
Balanced Accuracy	0.722
Total (N)	1.811

Source: Authors.

These metrics assess performances from different angles: AUC-ROC measures overall model discriminant power other (probability of scoring a "fatal" case above a "non-fatal" one, regardless of the threshold); Accuracy is the total proportion of correct answers; Sensitivity indicates the fraction of correctly identified fatal cases; Specificity is the fraction of non-fatal cases correctly discarded; PPV is the probability of a predicted "fatal" being an actual death; NPV is the probability of a predicted "non-fatal" event being actually non-fatal; the Kappa coefficient expresses over-the-odds agreement (useful for unbalanced bases); and balanced accuracy corresponds to the mean between sensitivity and specificity,

mitigating class imbalance effects. As this assessment used a single training-test partition, these estimates are subject to greater variability; it is recommended that cross-validation and/or early stopping should be used in future studies to ensure the robustness of the findings.

5. DISCUSSION

The spatial analysis using the Getis-Ord Gi* method confirmed the formation of critical corridors for motorcycle accidents in the Federal District Highway System, particularly along structural thoroughfares such as the DF-001, DF-003, DF-075, DF-085 and DF-095 highways. This pattern confirms that the event distribution is not random, but is instead aligned with high connectivity highways carrying heavy daily traffic, and with intensive interactions between urban and peri-urban flows. In applied terms, this indicates that the accident frequency is closely associated with the functional structure of the road network, especially along high-demand corridors with very complex operations.

These findings converge with studies demonstrating the usefulness of spatial analysis techniques for identifying critical areas, with support for territorially oriented interventions. Studies such as those by Tippayanate et al. (2024), Haque et al. (2010) and Mohammed et al. (2023) show that accident distribution tends to reflect location-specific characteristics, connectivity patterns, and road system operating conditions. In Brazil's Federal District, hot spots clustered along structural thoroughfares suggest that the combination of heavy traffic flows and driver x rider conflicts along metropolitan road networks is a core element for understanding event clusters involving motorcyclists.

Taken together, the findings do not indicate a broad overlap between "where accidents occur most" and "where the odds of death are higher if an event occurs". This Hot Spot analysis pinpoints high-frequency urban corridors, while the XGBoost predictive map indicates mainly peri-urban and rural segments, where conditional probabilities of fatalities tend to be higher. This contrast is significant, as it underscores two distinct road risk aspects: event frequency and expected severity. Consequently, areas with large numbers of occurrences do not necessarily coincide with locations where fatal outcomes are more likely.

This distinction is consistent with recent literature on risk-steered road safety. Studies aligned with the Safe System paradigm indicate that areas with higher driving speeds, poorer streetlighting, narrower protective shoulders, and longer response times tend to ramp up accident severity, even when event volumes are relatively smaller. Along these lines, the findings of this study suggest that urban corridors in the FDHS require actions designed to reduce accident frequency, while peri-urban and rural segments need interventions focused primarily on mitigating their severity.

For predictive modelling, the XGBoost findings indicated that head-on collisions, collisions with fixed objects, side and rear collisions, and motorcycle x truck interactions are among the main factors associated with higher fatality probabilities. These findings are compatible with international evidence pointing to greater severity in events with higher energy transfers, more body exposure for motorcyclists, and interactions with larger vehicles. Studies such as those by Agyemang, Adanu and Jones (2021), Wisutwattanasak et al. (2024) and Yang et al. (2022) stress that motorcycle accident severity depends not only on the actual occurrences, but also on operating contexts, conflict type, and surrounding road conditions.

The relevance of temporal and operational variables, such as Saturdays and accident types, also suggests that risks are not merely spatial. They also result from combinations of location, traffic dynamics, and road use contexts. This point is important because it underlines the need for intersectoral policies that blend engineering, oversight, and speed management with specific territorial and temporal approaches. In other words, the model's usefulness lies not only in predicting severe events; it is also a source of evidence for more selective and territorially calibrated interventions.

From the predictive performance standpoint, the model presented a highly accurate AUC-ROC of 0.8843, indicating good discriminatory ability. Nevertheless, its moderate sensitivity shows that some of the deaths were not captured by the classifier, with caution recommended for operational interpretations of this model. In road safety contexts, this constraint is particularly important, as the practical usefulness of a model depends not only on good overall performance, but also on its ability to identify rare and more serious events.

Some other limitations should also be mentioned. First, the use of a single training-test partition may increase estimate variability, so future applications could benefit from cross-validation and, ideally, from strategies with spatial or temporal lockouts. Next, although the inclusion of latitude and longitude as explanatory variables is useful for capturing residual spatial dependence, this may lead to overlapping spatial information, with its effects being interpreted as locational proxies, rather than causal relationships. Finally, the absence of variables related directly to road geometry, speed limits, and traffic volumes limits the explanatory capacity of the model, although some of these effects may have been partially captured by locational and operational variables.

Despite these constraints, the findings demonstrate that a combination of spatial analysis and predictive modelling offers important analytical benefits. While Getis-Ord Gi* helps pinpoint corridors with more accident clusters, XGBoost adds expected event severity data. This integrated approach contributes to a more consistent interpretation of road risks, extending the potential for using these findings for road safety planning in Brazil’s Federal District.

6. CONCLUSION

This study showed that a combination of spatial analysis (Getis-Ord Gi*) and predictive modelling (XGBoost) is a useful analytical strategy for understanding motorcycle accidents in the Federal District Highway System. The findings confirmed the existence of corridors with more accident clusters, especially along structural thoroughfares with high connectivity, while disclosing that the highest conditional fatality probability distribution tends to be different, especially along peri-urban and rural segments.

The main findings of this study thus show that accident frequency and fatality risks do not slot into the same analytical dimension. While spatial analysis helps pinpoint areas with higher event volumes, the predictive model identifies segments with higher expected severity, despite fewer events. This distinction underscores the need for specific intervention strategies in Brazil's Federal District.

From an applied standpoint, the findings suggest that busier urban corridors with more accident clusters require steps designed to reduce event frequency, such as restructuring traffic flows, oversight activities, and operating upgrades. On the other hand, segments with higher probabilities of fatal outcomes require actions designed to mitigate event severity, especially through speed management, wider shoulders, better visibility, and more attention to conflicts involving large vehicles.

In methodological terms, the study contributes by integrating two approaches that are often addressed individually in the literature: identifying spatial clusters and predicting accident severity. Although these techniques are already known, their combined deployment, with emphasis on the distinction between event clusters and conditional fatality risks, extends the analytical potential for road safety planning.

Finally, it is recommended that future studies include variables related more directly to characteristics such as road infrastructure, speed limits, traffic volumes, and complementary oversight and service data. The use of cross-validation strategies is also recommended, with spatial or temporal blocks, thus underscoring model robustness, while lessening the possible effects of residual spatial dependence.

BIBLIOGRAPHY

AGYEMANG, W.; ADANU, E. K.; JONES, S. (2021). Understanding the factors that are associated with motorcycle crash severity in rural and urban areas of Ghana. Journal of Advanced Transportation, 2021, e6336517. DOI: https://doi.org/10.1155/2021/6336517.

BRASIL. Secretaria Nacional de Trânsito (SENATRAN). (2024). Frota de Veículos – 2024 (Registro Nacional de Veículos Automotores – RENAVAM). Ministério dos Transportes, Brasília, DF.

BRENAC, T.; CLABAUX, N.; PERRIN, C.; VAN ELSLANDE, P. (2006). Motorcyclist conspicuity-related accidents in urban areas: a speed problem? Advances in Transportation Studies – Section A, n. 8, p. 23–29.

CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, v. 16, p. 321–357. DOI: https://doi.org/10.1613/jair.953.

CHEN, F.; ZHANG, Q.; WANG, Z.; et al. (2025). Traffic accident severity prediction based on an enhanced MSCPO-XGBoost hybrid model. Scientific Reports, v. 15, art. 25729. DOI: https://doi.org/10.1038/s41598-025-00797-7.

DISTRITO FEDERAL. Departamento de Trânsito do Distrito Federal (DETRAN-DF). (2024). Frota de veículos registrados no DF em circulação — 2º trimestre de 2024. Detran-DF, Brasília, DF.

EUROPEAN COMMISSION. (2023). Road Safety Thematic Report: Safe System Approach. European Commission, Brussels.

GETIS, A.; ORD, J. K. (1992). The analysis of spatial association by use of distance statistics. Geographical Analysis, v. 24, n. 3, p. 189–206. DOI: https://doi.org/10.1111/j.1538-4632.1992.tb00261.x.

HAQUE, M. M.; CHIN, H. C.; HUANG, H. (2010). Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Analysis & Prevention, v. 42, n. 1, p. 203–212. DOI: https://doi.org/10.1016/j.aap.2009.07.022.

JIANG, F.; YUEN, K. K. R.; LEE, E. W. M. (2020). Analysis of motorcycle accidents using association rule mining-based framework with parameter optimization and GIS technology. Journal of Safety Research, v. 75, p. 292–309. DOI: https://doi.org/10.1016/j.jsr.2020.09.004.

LUNDBERG, S. M.; LEE, S.-I. (2017). A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS 2017). DOI: https://doi.org/10.48550/arxiv.1705.07874.

MIAO, J.; PAN, Y.; ZHAO, K. (2025). Analysis of the spatiotemporal effects on the severity of motorcycle accidents without helmets and strategies for building sustainable traffic safety. Sustainability, v. 17, n. 8, art. 3280. DOI: https://doi.org/10.3390/su17083280.

MOHAMMED, S.; ALKHEREIBI, A. H.; ABULIBDEH, A.; JAWARNEH, R. N.; BALAKRISHNAN, P. (2023). GIS-based spatiotemporal analysis for road traffic crashes: in support of sustainable transportation planning. Transportation Research Interdisciplinary Perspectives, v. 20, 100836. DOI: https://doi.org/10.1016/j.trip.2023.100836.

OECD/ITF. (2008). Towards Zero: Ambitious Road Safety Targets and the Safe System Approach. OECD/ITF, Paris.

PANUMASVIVAT, J.; et al. (2024). Unveiling the road to safety: understanding the factors influencing motorcycle accidents among riders in rural Chiang Mai, Thailand. Heliyon, v. 10, n. 3, e25698. DOI: https://doi.org/10.1016/j.heliyon.2024.e25698.

QIAN, R. Y.; WANG, X. (2023). Prediction of road traffic accident severity based on XGBoost-BP neural network. Advances in Transportation Studies, v. 61, p. 19–36. DOI: https://doi.org/10.53136/97912218091902.

SANTOS JÚNIOR, C. J.; SILVA, J. P.; ROCHA, T. J. M. (2022). Análise da ocorrência de acidentes de trânsito fatais nas capitais do Nordeste do Brasil. Hygeia – Revista Brasileira de Geografia Médica e da Saúde, v. 18, p. 44–53. DOI: https://doi.org/10.14393/hygeia1858395.

SANTOS, J. P.; GUIMARÃES, P. H.; SILVA, B. A. L.; CAMPANI, A. R.; OLIVEIRA, J. M. V. (2023). DERGeo: geovisualizador de informações geográficas para gestão do Sistema Rodoviário do Distrito Federal. Revista Brasileira de Transportes, v. 3, n. 1, p. 62–85. DOI: https://doi.org/10.12660/rbt.v3n1.2023.89030.

SIQUEIRA, J. S.; WERNECK, G. L.; PENA, P. G. L.; FERNANDES, R. C. P. (2025). Trabalho subordinado de entregadores em empresas-plataforma de entrega e acidentes nas ruas: magnitude e fatores associados. Cadernos de Saúde Pública, v. 41, n. 3, e00204723. DOI: https://doi.org/10.1590/0102-311xpt204723.

TIPPAYANATE, N.; IMPOOL, T.; SUJAYANONT, P.; MUTTITANON, W.; CHEMIN, Y.; SOM-ARD, J. (2024). A geospatial analysis of motorcycle accident risk factors in Khon Kaen Municipality, Thailand: examining the chain of survival and potential strategies. International Journal of Geoinformatics, v. 20, n. 9, p. 71–82. DOI: https://doi.org/10.52939/ijg.v20i9.3547.

WORLD HEALTH ORGANIZATION (WHO). (2023). Global status report on road safety 2023. WHO, Geneva.

WISUTWATTANASAK, P.; SE, C.; CHAMPAHOM, T.; et al. (2024). Factors affecting single and multivehicle motorcycle crashes: insights from day and night analysis using XGBoost-SHAP algorithm. Big Data and Cognitive Computing, v. 8, n. 10, 128. DOI: https://doi.org/10.3390/bdcc8100128.

YANG, Y.; LI, Y.; WANG, J.; et al. (2022). Predicting freeway traffic crash severity using XGBoost-Bayesian network model with consideration of features interaction. Journal of Advanced Transportation, v. 2022, 4257865. DOI: https://doi.org/10.1155/2022/4257865.