

Análise exploratória e preditiva de dados de acidentes de trânsito com vítimas atendidas em um hospital universitário

Lino Marcos da Silva
UNIVASF – BA

Daniel Moreira Lopes
UNIVASF – BA

Marco Antonio de Jesus Saturnino
UNIVASF – BA

Márcio Ruan da Silva Carvalho
UNIVASF – BA

RESUMO

Acidentes envolvendo motocicletas chegam a representar mais de 75% do número de acidentes de trânsito no país causando grandes prejuízos à sociedade. Em Petrolina-PE, na região do Vale do São Francisco, de acordo com dados do Hospital Universitário da Universidade Federal do Vale do São Francisco (HU/Univasf), o número de entradas de pacientes em decorrência desse tipo de acidentes é superior ao número de outras modalidades de acidentes. Uma consequência imediata desse elevado índice é a superlotação dos leitos hospitalares. No entanto, as consequências extrapolam as limitações físicas dos hospitais e atingem outros setores. Por exemplo, o setor econômico é atingido diretamente pelos reflexos negativos desses resultados, devido ao afastamento ou supressão precoce de força de trabalho. Com efeito, estima-se que, em 2017, o impacto na economia devido aos acidentes com motocicletas foi de aproximadamente 200 bilhões de reais. Além disso, autoridades da área de saúde, ao analisarem os reflexos dos índices de acidentes de trânsito nos sistemas de saúde das grandes cidades brasileiras, são unânimes em afirmar que este é um problema sério para as redes hospitalares e consideram que há uma epidemia global de acidentes de trânsito, sendo a motocicleta a principal vilã.

Palavras-chave: Acidentes de trânsito, Hospital universitário, Saúde pública.

1 INTRODUÇÃO

Acidentes envolvendo motocicletas chegam a representar mais de 75% do número de acidentes de trânsito no país causando grandes prejuízos à sociedade [1, 2, 3]. Em Petrolina-PE, na região do Vale do São Francisco, de acordo com dados do Hospital Universitário da Universidade Federal do Vale do São Francisco (HU/Univasf), o número de entradas de pacientes em decorrência desse tipo de acidentes é superior ao número de outras modalidades de acidentes. Uma consequência imediata desse elevado índice é a superlotação dos leitos hospitalares. No entanto, as consequências extrapolam as limitações físicas dos hospitais e atingem outros setores. Por exemplo, o setor econômico é atingido diretamente pelos reflexos



negativos desses resultados, devido ao afastamento ou supressão precoce de força de trabalho. Com efeito, estima-se que, em 2017, o impacto na economia devido aos acidentes com motocicletas foi de aproximadamente 200 bilhões de reais [3]. Além disso, autoridades da área de saúde, ao analisarem os reflexos dos índices de acidentes de trânsito nos sistemas de saúde das grandes cidades brasileiras, são unânimes em afirmar que este é um problema sério para as redes hospitalares e consideram que há uma epidemia global de acidentes de trânsito, sendo a motocicleta a principal vilã [4].

Diante disso, é importante o desenvolvimento de estudos que possibilitem uma compreensão técnica e científica dessa problemática, que auxiliem o planejamento de atividades preventivas e assistenciais e, sobre tudo, promovam a educação para o trânsito com vistas a redução do número desse tipo de acidente. A problemática de acidentes de trânsito, particularmente aqueles envolvendo motocicletas tem sido estudados sob vários aspectos e em vários países [5, 6, 7, 8, 9, 10, 11].

Sendo assim, é relevante a realização de estudos, sobre acidentes de trânsito tanto no âmbito nacional, como no local, sob os mais variados enfoques e utilizando-se dos diversos recursos metodológicos disponíveis. Nesse sentido, propôs-se realizar neste trabalho o estudo dos dados relacionadas aos acidentes de trânsito envolvendo motocicletas cujos pacientes foram atendidos no HU-Univasf, utilizando-se como metodologia a coleta, tratamento e análise de dados, o desenvolvimento de modelos matemáticos exploratórios e preditivos, usando técnicas de ciências de dados.

2 OBJETIVO

Este estudo teve por objetivo geral investigar a existência de relações entre fatores sociodemográficos das vítimas de acidentes de trânsito, a gravidade dos acidentes e a evolução hospitalar. Os objetivos específicos foram: realizar análise exploratória dos dados dos acidentes envolvendo transportes terrestres e com vítimas atendidas no HU-Univasf; identificar relações entre variáveis associadas aos acidentes de trânsito e internação hospitalar; identificar fatores envolvidos nas causas dos acidentes envolvendo motociclistas; e desenvolver modelos matemáticos preditivos usando variáveis dos acidentes.

3 METODOLOGIA

O estudo se caracterizou como uma pesquisa exploratória, de abordagem quantitativa e com o uso de técnicas e ferramentas das estatísticas descritiva e inferencial, da Ciência de Dados e Inteligência Artificial. O conjunto de dados fornecido

pelo HU contemplava o período entre 01/01/2018 e 31/12/2022 e continha 41 dados de um total de 40.345 vítimas.



Metodologicamente, adotou-se uma abordagem embasada nos princípios fundamentais da ciência de dados. Esses princípios compreendem um conjunto de etapas eficazes na manipulação e análise de grandes conjuntos de dados, com o propósito de extrair informações relevantes para fundamentar processos de tomada de decisão. As técnicas foram implementadas por meio das bibliotecas *sklearn*, *statsmodels* e *pycaret*, da linguagem de programação *Python* e também por meio do *software Orange*.

Inicialmente, procedeu-se a etapa do pré-processamento dos dados, onde estes foram submetidos a procedimentos de filtragem, limpeza e seleção, com o objetivo de obter um conjunto de dados íntegros e de alta qualidade. Essa etapa incluiu o tratamento de valores faltantes, a detecção de valores discrepantes ou anomalias, a conversão de variáveis categóricas em formatos numéricos, bem como a normalização dos dados, entre outros procedimentos pertinentes. Ao final desse processo, obteve-se um conjunto de dados com 40.213 instâncias.

A fim de obter uma compreensão preliminar do conjunto de dados, bem como a geração de *insights*, foi realizada uma análise exploratória. Em uma etapa subsequente, ocorreu uma seleção das variáveis que comporiam os modelos preditivos. Posteriormente, utilizou-se de técnicas de aprendizado de máquina para a geração de modelos. Como última etapa desse processo, as métricas de avaliação foram calculadas para quantificar o desempenho dos modelos gerados durante o processamento dos dados.

A técnica de aprendizado de máquina empregada neste trabalho foi a classificação, que consiste no uso de algoritmos chamados classificadores para identificar características das variáveis e rotular dados de saídas dentro de um contexto, em classes distintas.

O *Ensemble Learning* é uma técnica na área de aprendizado de máquina que envolve a combinação de vários modelos de aprendizado para melhorar a precisão e o desempenho geral da previsão ou classificação. Em vez de confiar em apenas um único modelo, o *Ensemble Learning* aproveita a diversidade de diferentes modelos para obter resultados mais confiáveis e robustos.

O conceito por trás do *Ensemble Learning* baseia-se na ideia de que diferentes modelos podem aprender padrões e relações nos dados de maneiras diferentes. Ao combinar esses modelos, é possível aproveitar as forças individuais e atenuar suas fraquezas, resultando em uma previsão geralmente mais precisa e estável.

Existem várias abordagens para o *Ensemble Learning*, sendo as mais comuns são *Bagging* e *Boosting*. No primeiro, vários modelos idênticos são treinados em conjuntos de dados diferentes, gerados por amostragem com reposição dos dados originais. Os resultados desses modelos individuais são então combinados, geralmente por meio de votação ou média, para formar a previsão final. No segundo, a técnica consiste em treinar os modelos sequencialmente. Cada modelo subsequente é treinado para corrigir os erros do modelo anterior, concentrando-se nos exemplos que foram classificados incorretamente. Isso permite



que os modelos se concentrem mais nos casos difíceis e melhorem gradualmente sua precisão. Neste trabalho, foi usado um algoritmo de cada tipo.

4 RESULTADOS

4.1 ANÁLISE EXPLORATÓRIA

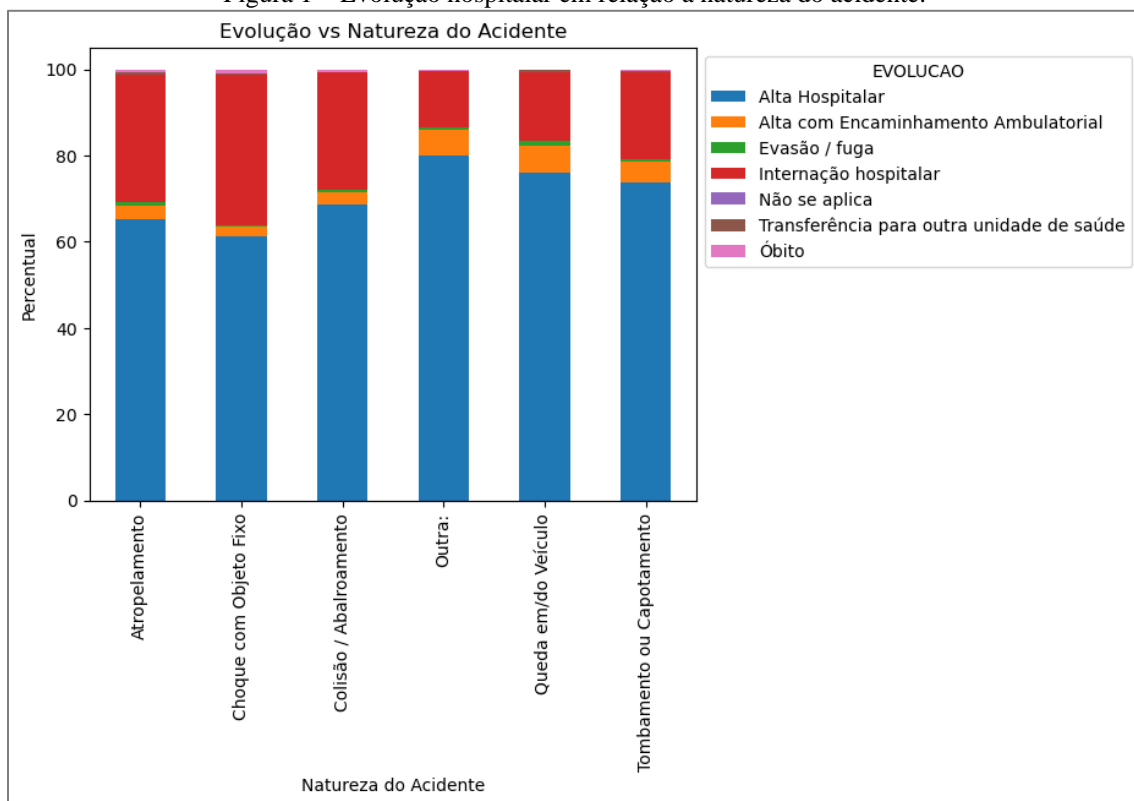
Os resultados da análise exploratória dos dados, dentre outros achados, corroboraram com a informação de que o atendimento às vítimas de acidentes envolvendo motocicletas são a principal demanda do hospital, chegando a representar cerca de 74% do total. O perfil sociodemográfico das vítimas revelou que 76,9% do total de pacientes atendidos eram do sexo masculino; 52,5% tinham entre 20 e 40 anos; e 94,8% das vítimas foram identificadas como pardas.

Em relação aos acidentes, observou-se que 73,7% envolvem motocicletas, 38,4% ocorrem aos sábados ou domingos; que cerca de 65,0% ocorrem a tarde ou a noite; que as lesões mais frequentes são fratura (21%), corte / laceração (11,8%) e traumatismo cranioencefálico (8,9%); que a parte do corpo mais atingida são os membros inferiores (33,9%), membros superiores (26, 9%) e cabeça (14, 1%); que os pacientes evoluíram para alta hospitalar em 71,5% dos casos, para internação hospitalar em 22, 9% e para alta com encaminhamento ambulatorial em 4,1 %. Os óbitos representaram cerca de 0, 5%. É importante destacar que em termos absolutos o número de internações e óbitos corresponde, respectivamente, a cerca de 9220 e 201, no período analisado.

A evolução hospitalar é uma variável de interesse, visto que é responsável por indicar a superlotação dos hospitais. Nesse sentido, nesta análise exploratória, buscou-se identificar variáveis que estivessem relacionadas com a evolução hospitalar da vítima. Os valores possíveis para essas variáveis são: alta hospitalar, alta com encaminhamento ambulatorial, evasão/fuga, internação hospitalar, transferência para outra unidade de saúde e óbito, além da opção 'não se aplica'.

A primeira variável analisada em relação a evolução hospitalar foi a **Natureza do Acidente** e os resultados estão apresentados na **Figura 1**.

Figura 1 – Evolução hospitalar em relação a natureza do acidente.

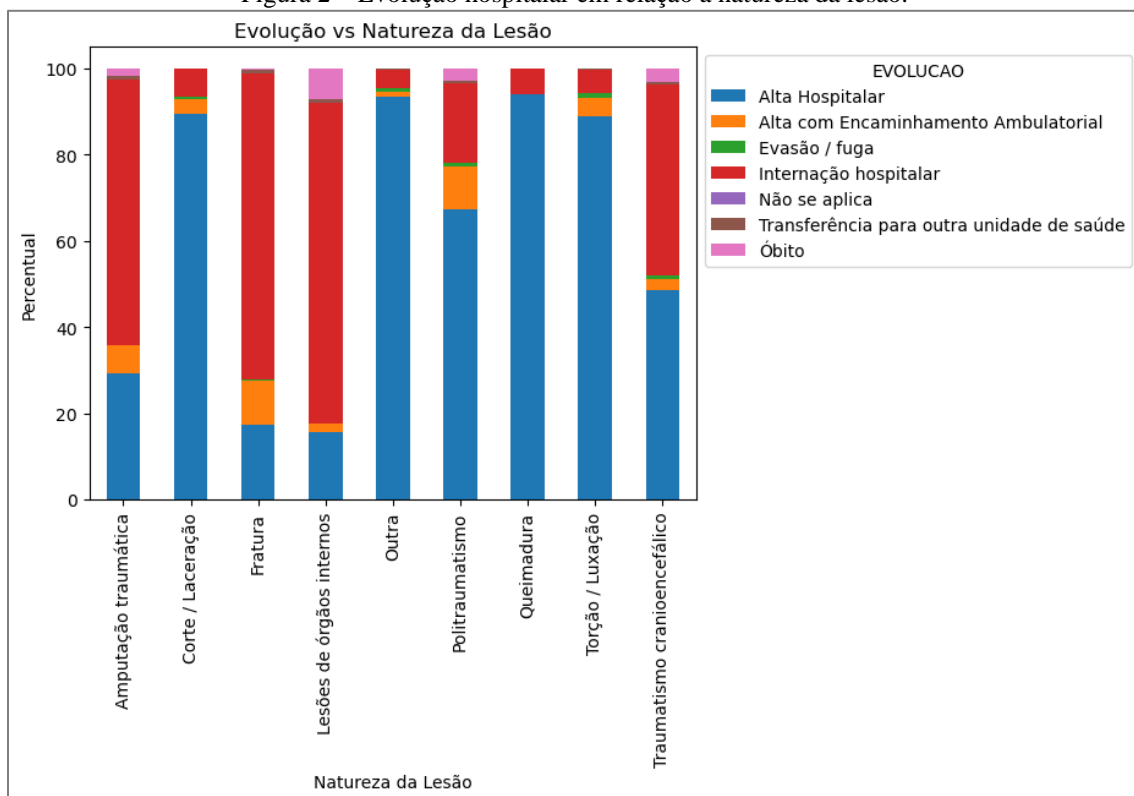


Fonte: dados da pesquisa

Nesta análise, observou-se que o quadro das vítimas de acidentes envolvendo choque com objeto físico, atropelamento e colisão/abalroamento foram os que mais evoluíram para internação hospitalar.

Quando se comparou a **Natureza da Lesão** sofrida pela vítima com a sua evolução hospitalar, observou-se que houve uma evolução maior para internação hospitalar nos casos de vítimas que tiveram amputação traumática, fratura, lesões de órgãos internos ou traumatismos cranioencefálico, conforme apresentado na **Figura 2**. Contudo, é importante destacar que as lesões do tipo amputação traumática e lesões de órgãos internos corresponderam a menos de 0,7% dos acidentes. Em relação aos pacientes que sofreram fraturas, observou-se ainda um número significativo de alta com encaminhamento ambulatorial. Ou seja, nesses casos, mesmo mediante alta, ainda há uma demanda por serviços hospitalares.

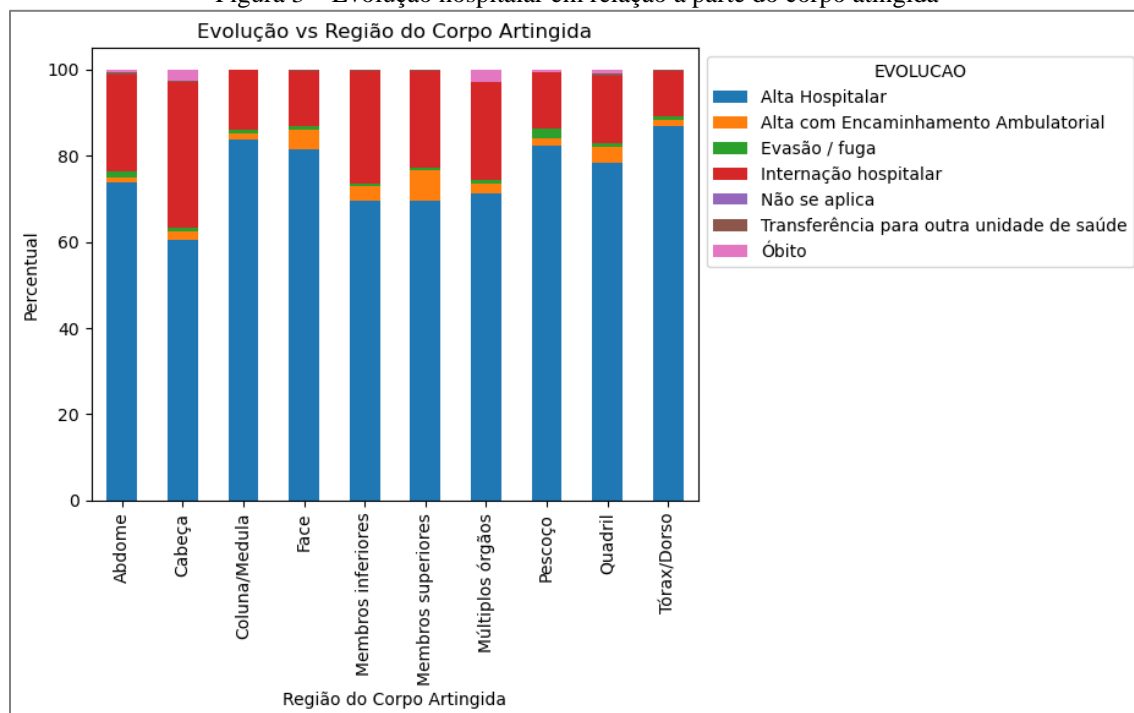
Figura 2 – Evolução hospitalar em relação a natureza da lesão.



Fonte: dados da pesquisa

Outra variável que foi analisada em relação a evolução hospitalar foi a **Região do Corpo Atingida**. Para essa variável, os resultados, que estão apresentados na **Figura 3**, indicam que os maiores índices de internação hospitalar se deram quando a cabeça e os membros inferiores foram atingidos no acidente. É importante destacar que essas duas regiões do corpo são as mais afetadas em cerca de 48% dos acidentes. Observou-se ainda que os maiores índices de alta hospitalar com acompanhamento ambulatorial se deram nos casos em que a face ou os membros superiores foram atingidos.

Figura 3 – Evolução hospitalar em relação a parte do corpo atingida



Fonte: dados da pesquisa

Outras variáveis como **Álcool**, **Condições de Habilitação** e **Zona de Ocorrência do Acidente** também podem indicar influência na hospitalização das vítimas.

Com efeito, foi observado uma taxa maior de internação hospitalar para as vítimas que fizeram uso de bebidas alcoólicas, não habilitadas e vítimas de acidentes ocorridos na zona rural.

Por sua vez, as variáveis **Sexo** e **Turno do Acidente** não apresentaram diferenças significativas em relação a evolução hospitalar, embora acidentes que ocorreram pela manhã ou com pessoas do sexo feminino tiveram uma taxa de internação levemente menor de internações.

Em relação a **Zona de Ocorrência** dos acidentes, observou-se que, embora a maioria dos acidentes tenham ocorridos na zona urbana (61,6%), os acidentes ocorridos na zona rural resultam em uma maior gravidade, visto que a taxa de internação hospitalar para esse tipo de acidente é aproximadamente o dobro das internações hospitalares ocorridas em decorrência de acidentes na zona urbana.

Considerando que a maioria dos acidentes envolveu motociclistas, ao voltar o olhar para o **Uso do Capacete** observou-se que, dentre as vítimas fatais desse tipo de acidente, 63,3% não estavam usando capacetes; e que em 65,6% dos casos em que houve traumatismo cranioencefálico a vítima não estava fazendo uso desse equipamento de segurança. Este dado reforça a percepção de que os capacetes desempenham um importante papel na conservação da vida.

Quando a natureza da lesão corporal foi analisada a partir do uso ou não do capacete, observou-se um maior número de ocorrência de traumatismo cranioencefálico e corte/laceração quando o equipamento



não estava sendo utilizado. Além disso, ficou evidenciado que cabeça e face foram as partes do corpo mais atingidas nos casos de vítimas que não estavam fazendo uso do capacete.

4.2 MODELOS PREDITIVOS

Com base nos resultados da Análise Exploratória, buscou-se, por meio de um modelo de Classificação, categorizar a **Evolução Hospitalar** do Paciente com base na **Natureza da Lesão** sofrida e na **Região do Corpo Atingida**.

Dessa maneira, para a construção do modelo, a Evolução Hospitalar do paciente foi considerada como variável dependente, enquanto a Região do Corpo Afetada e a Natureza da lesão resultante do acidente foram considerados como variáveis independentes.

Para validar a escolha dessas variáveis, realizou-se uma análise de correlação empregando o coeficiente de correlação de Pearson entre as variáveis independentes. O coeficiente calculado foi de 0.5, indicando uma correlação de fraca magnitude conforme a teoria. Portanto, é justificável incorporar ambas as variáveis como independentes.

Para a construção do modelo classificador, inicialmente procedeu-se à divisão do conjunto de dados em um conjunto de treinamento e um conjunto de teste, na proporção 2/3 e 1/3, respectivamente.

Em seguida, o conjunto de treinamento foi submetido aos seguintes algoritmos de aprendizado de máquina: classificador *Random Forest*, que é uma técnica de *bagging*; e o classificador *Gradient Boosting*, que corresponde a uma abordagem de *boosting*.

Para avaliação dos classificadores foram utilizados as métricas Acurácia, Revocação, Precisão e *F1-Score*, que são comumente utilizadas. Os resultados obtidos estão apresentados na **Tabela 1** e indicam que o modelo de classificação do tipo *Gradient Boosting* apresentou um desempenho superior, com uma eficiência em torno de 85%.

Tabela 1 - Avaliação do modelos proposto

	Acurácia	Recall	Precisão	F1-score
Gradient Boosting	0,8512	0,8512	0,8227	0,8349
Random Forest	0,8403	0,8403	0,8031	0,8222

Fonte: dados da pesquisa

5 CONSIDERAÇÕES FINAIS

Neste estudo, baseado numa análise exploratória e preditiva realizada em um conjunto de dados sobre vítimas de acidentes de trânsito atendidas em um hospital universitário, foi observado que alguns fatores como a zona de ocorrência do acidente, a natureza do acidente, a condição de habilitação do



condutor, o consumo de bebidas alcoólicas, a natureza da lesão e a região do corpo atingida podem estar relacionados com a evolução hospitalar da vítima.

No entanto, foi observado que as variáveis indicadoras do tipo de lesão sofrida e da região do corpo afetada tiveram mais impacto na classificação da evolução hospitalar. Dessa maneira, essas duas variáveis foram utilizadas no desenvolvimento de um modelo preditivo para previsão da evolução hospitalar de cada vítima. A partir daí, criou-se um modelo de classificação, que obteve uma taxa de precisão de 85%.

Os resultados corroboraram estudos anteriores que indicaram que vítimas de acidentes envolvendo motocicletas são a maioria dos atendimentos hospitalares e que a maioria das vítimas são pessoas do sexo masculino com idade entre 20 e 40 anos [12]. Por outro lado, apontaram resultados diferentes em outros aspectos, como a zona de maior ocorrência [12].

Além disso, ficou evidenciado que, apesar da maioria dos acidentes ocorrerem na zona urbana, os que aconteceram na zona rural demandaram mais internação hospitalar, indicando a necessidade de ações preventivas direcionadas para este segmento da população. É importante destacar que Petrolina-PE e Juazeiro-BA são dois municípios com grande vocação agrícola e com grande parte da população morando ou trabalhando na zona rural.

Para uma maior compreensão sobre como os fatores sociodemográficos influenciam na demanda dos serviços hospitalares, cabe um estudo mais aprofundado incluindo mais dados relativos às vítimas e também aos acidentes.

Por fim, cabe destacar que este estudo serve como ponto de partida para o entendimento do perfil das vítimas atendidas no HU-Univasf, dos acidentes de trânsito dos quais foram vítimas, bem como para mapear condições de risco que contribuem com a gravidade do acidente fornecendo assim subsídio para a tomada de decisão mais assertiva no âmbito das políticas públicas de saúde no contexto do trauma envolvendo acidentes com transportes terrestres. Além disso, outros modelos preditivos podem ser analisados com vistas a obtenção de melhores resultados.

Os pesquisadores agradecem ao HU-Univasf, a Facepe e a Fapesb pelo apoio concedido para o desenvolvimento deste trabalho.



REFERÊNCIAS

- PREVIDÊNCIA SOCIAL. O impacto dos acidentes de trânsito para a previdência social. Boletim Epidemiológico 2. Disponível em: <http://sa.previdencia.gov.br/site/2017/03/3%C2%B0-Quadrimestre-Boletim-2-Impacto-Acidentes-de-Tr%C3%A2nsito.pdf>. Acesso em: 29 de mar. de 2023.
- TV SENADO. Trânsito brasileiro: 45 mil mortes e r\$ 50 bilhões de prejuízo econômico. Disponível em: <https://www12.senado.leg.br/tv/programas/em-discussao/2022/09/transito-brasileiro-45-mil-mortes-e-r-50-bilhoes-de-prejuizo-economico>. Acesso em: 29 de mar. De 2023.
- Acidentes no trânsito têm impacto de r\$ 199 bi na economia. Correio Braziliense, 14, agosto, 2018. Brasil. Disponível em: <https://www.correiobraziliense.com.br/app/noticia/brasil/2018/05/14/interna-brasil,680658/acidentes-no-transito-tem-impacto-de-r-199-bi-na-economia.shtml>. Acesso em: 09, maio 2022.
- RAMOS, M. Acidentes de trânsito têm impacto direto na rede de saúde pública do rio, diz secretário. O GLOBO, 2017. Disponível em: <https://oglobo.globo.com/rio/acidentes-de-transito-tem-impacto-direto-na-rede-de-saude-publica-do-rio-diz-secretario-21914148>. Acesso em: 09, maio 2022.
- RODRIGUES, C. L.; ARMOND, J. E.; GORIOS, C.; SOUZA, P. C. Acidentes que envolvem motociclistas e ciclistas no município de são paulo: caracterização e tendências. Revista Brasileira de Ortopedia. 49(6):602–606, 2014.
- MENDONÇA, M. F. S. SILVA, A. P. S. C.; CASTRO, C. C. L. Análise espacial dos acidentes de trânsito urbano atendidos pelo serviço de atendimento móvel de urgência: um recorte no espaço e no tempo. Revista Brasileira de Epidemiologia. 20(4): 727-741; OUT-DEZ 2017.
- NYAKYI, V.; KUZNETSOV, D.; NKANSAH-GYEKYE, Y. Mathematical model to assess motorcycle accidents in tanzania. Mathematical Theory and Modeling. Vol.4, No.9, 2014.
- MIŠKINIS, P.; VALUNTAITE, V. Mathematical simulation of the correlation between the frequency of road traffic accidents and driving experience. Transport, 25:3, 237-243, 2010. DOI:10.3846/transport.2010.29.
- LHUEZE, C. C. ONWRAH; UCHENDU, O. Road traffic accidents prediction modelling: an analysis of anambra state, nigeria. Accident analysis & prevention, Elsevier. V 112, p.21-29, 2018.
- HAYNES, S.; ESTIN, P. C. LAZAREVSKI, S.; SOOSAY, M.; KOR, AH-LIAN. Data analytics: factors of traffic accidents in the UK. The 10h IEEE International Conference on Dependable Systems, Services and Technologies, DESSERT'2019 5-7 June, 2019, Leeds, United Kingdom.
- NOUR, M.; NASEER, A.; ALKAZEMI, B.; JAMIL, M. A. Road traffic accidents injury data analytics. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 12, 2020.
- FERNANDES, F.E.C.V.; MELO, R.A.; ARAÚJO, F.S.A.; BORGES, F.K.B.; HOLANDA, O.Q.; CAMPOS, M.E.A. Acidentes por motocicleta e fatores associados à condição de habilitação dos condutores. Archives of Health Sciences, 26(2) 130-135, 2019.