

## Uso de técnicas de aprendizado de máquina na modelagem da segurança viária: mapeamento sistemático

### Use of machine learning techniques in road safety modeling: mapping study

<sup>1</sup>Philippe Barbosa Silva, <sup>2</sup>Michelle Andrade, <sup>3</sup>Sara Ferreira

<sup>1</sup>Doutor em Transportes, Professor do Curso de Engenharia Civil – Instituto Federal Goiano – Campus Rio Verde (philippe.silva@ifgoiano.edu.br)

<sup>2</sup>Doutora em Transportes, Professora Adjunta do Departamento de Engenharia Civil e Ambiental, Programa de Pós-Graduação em Transportes – Universidade de Brasília (michelle.andrade@unb.br)

<sup>3</sup>Doutora em Engenharia Civil, Professora Auxiliar do Departamento de Engenharia Civil, Centro de Investigação do Território, Transportes e Ambiente – Universidade do Porto (asra@fe.up.pt)

---

**RESUMO:** A modelagem da segurança viária se coloca como importante alternativa na otimização de recursos e esforços para promoção de mobilidade segura. Este trabalho apresenta o mapeamento sistemático de artigos que tratam do desenvolvimento de modelos de previsão de acidentes, especialmente em rodovias, com uso de técnicas de aprendizado de máquina (AM). Para tanto, foi aplicado um protocolo de condução da revisão, utilizando como bases de dados o Portal de Periódicos Capes e Google Acadêmico. Inicialmente alguns aspectos bibliométricos foram apresentados, seguido de uma análise qualitativa. Como resultados fez-se a identificação das principais abordagens metodológicas e suas características, desempenho dos modelos e variáveis explicativas. Desta forma, o mapeamento foi importante para traçar o panorama da área de pesquisa, apontar limitações e oportunidades de investigação e ainda, evidenciar o potencial de utilização de AM para análise de acidentes de trânsito.

**Palavras Chave:** Acidentes de trânsito. Segurança viária. Previsão de acidentes. Aprendizado de máquina. Mapeamento sistemático.

**ABSTRACT:** The road safety modeling is an important alternative in the optimization of resources and efforts to promote safe mobility. This paper presents the mapping study of papers about the development of accident prediction models, especially on highways, using machine learning (ML) techniques. For this purpose, a revision management protocol was applied, using the Portal of Periodicals Capes and Google Scholar as databases. Initially some bibliometric aspects were presented, followed by a qualitative analysis. As a result, the main methodological approaches and their characteristics, model performance and explanatory variables were identified. In this way, the mapping was important to draw the panorama of the area of research, to point out limitations and opportunities of investigation and also, to highlight the potential of the use of ML for analysis of crash accidents.

**Keywords:** Crash accidents. Road safety. Crash prediction. Machine learning. Mapping study.

---

## 1. INTRODUÇÃO

A acidentalidade no trânsito é uma das grandes preocupações a nível mundial. Conforme estudo da OMS, no ano de 2010 ocorreram 1,24 milhões de mortes por acidente de trânsito e ainda, cerca de 50 milhões de pessoas sobreviveram aos sinistros com algum tipo de lesão. Já em 2013, em todo o mundo foram registradas 1,25 milhões de mortes decorrentes do trânsito. Estes acidentes consomem mundialmente recursos de mais de 518 bilhões de dólares por ano (OMS, 2010; WHO, 2015).

Nodari (2003) acredita que a forma mais eficiente de tratar a problemática da segurança viária, sem desperdiçar os escassos recursos disponíveis, se dá por meio de um programa abrangente de gerenciamento da segurança viária (GSV).

Em se tratando do GSV, é fundamental conhecer o que motiva os acidentes, o que impacta significativamente na ocorrência dos acidentes de trânsito e qual o resultado pode ser alcançado através de determinadas medidas ou mudanças a serem adotadas, nesse sentido, a modelagem da segurança viária dá grande contribuição (Cardoso, 2006).

Lord e Mannering (2010) reiteram que, atualmente, pesquisadores têm se dedicado à investigação dos fatores que afetam o número de acidentes que ocorre em algum espaço geográfico (normalmente uma interseção ou segmento de via) durante um período de tempo especificado, o que resulta em dados de frequência de acidentes e/ou severidade destes, objetivos dos modelos de previsão de acidentes (MPA).

Chang (2005), Kim e Washington (2006), Lord e Mannering (2010), Cafiso *et al.* (2010) e Costa (2013) destacam a importância da modelagem da segurança viária como ferramenta fundamental para a promoção da mobilidade mais segura, especialmente através da redução de acidentes e sua severidade.

A modelagem da segurança viária é tradicionalmente estatística, no entanto, alguns pesquisadores (Mussone *et al.*, 1999; Li *et al.*, 2012; Chang, 2005) destacam que este tipo de modelagem requer suposição sobre a distribuição dos dados e ainda, estabelece uma forma funcional entre variáveis dependentes e explicativas. Diversas vezes essas premissas podem não ser verdadeiras, e em sendo violadas, conduzem a estimativas equivocadas e produção incorreta de inferências. Os autores ainda evidenciam que o uso de redes neurais artificiais (RNA) não requer este tipo de relação pré-definida entre as variáveis e em vez de elaborar uma forma funcional analítica, tarefa bastante complexa e laboriosa, é reconstruído um modelo, a partir do aprendizado dos dados reais de acidentes, com os pesos empregados para cada variável. De forma a reiterar, Abdelwahab e Abdel-Aty (2001) sugeriram o uso de RNA em detrimento de métodos de regressão, dadas as relações internas altamente não-lineares de variáveis que influenciam na ocorrência de acidentes.

Análogo ao uso de RNA, pesquisadores também têm envidado esforços para explorar a aplicabilidade de outras técnicas de aprendizado de máquina (AM) na modelagem da segurança viária. Mannering e Bhat (2014) destacam o potencial dessas técnicas como alternativa na análise de acidentes de trânsito, incentivando, inclusive, sua exploração.

Assim, o presente artigo busca apresentar um panorama das pesquisas em modelagem da segurança viária desenvolvidas com recurso às técnicas de AM.

## 2. MATERIAIS E MÉTODOS

Um estudo de mapeamento sistemático (*Mapping Study*) foi conduzido para obter o panorama das publicações relacionadas ao uso de AM na previsão de acidentes de trânsito em rodovias. Para tanto, empregou-se metodologia similar a adotada por Paula e Ilha (2016).

As plataformas selecionadas para execução da pesquisa foram o Portal de Periódicos CAPES e Google Acadêmico. O Portal de Periódicos CAPES foi eleito por abarcar diversas bases de dados consolidadas, a saber: Web of Science, Scopus, Journal Citation Reports, Engineering Village, MAS, ASTM International, SciFinder, ProQuest, Britannica Academic Edition, Thomson Reuters, Eighteenth Century Collections Online, Begell House. Douro lado, o Google Acadêmico é uma plataforma mais abrangente, capturando resultados não alcançados pelas outras bases de dados.

Fez-se a combinação de *strings*, conforme apresentado na Tabela 1, relativas a previsão de acidentes e aprendizado de máquina, de forma a capturar o maior número de publicações de interesse para o mapeamento. Isso porque, a designação da análise de acidentes de trânsito e de sistemas de aprendizado artificial apresentam diversas variações e a não consideração delas pode implicar em resultados enviesados.

Após busca inicial, foram adotados como critérios de exclusão: artigos escritos em idioma distinto de português, inglês ou espanhol; artigos não pertencentes à área de conhecimento Transportes; trabalhos que tipo de publicação (artigo de periódico revisado por pares e com Fator de impacto JCR ou índice SJR superior a 0,5). Aplicados os critérios de exclusão em cada base de dados, procedeu-se à leitura e análise de títulos e resumos,

permitindo a exclusão dos trabalhos sem relevância para o objetivo do estudo. Além disso, fez-se a exclusão dos artigos duplicados.

De seguida, procedeu-se à amostragem “bola de neve” (*snowball sampling*), que consistiu em analisar as referências citadas pelos artigos selecionados, buscando incorporar estudos que não foram captados nas buscas realizadas nas etapas anteriores. Por fim, buscando qualificar a análise dos artigos, para o conjunto final de publicações foram selecionados apenas os artigos de periódicos com Fator de impacto JCR ou índice SJR superior a 0,5.

### 3. RESULTADOS E DISCUSSÕES

Mediante a aplicação do protocolo de condução do mapeamento foram obtidos os resultados apresentados na Tabela 1. O horizonte de tempo empregado nas pesquisas foi de 1975 a 2017.

Tabela 1 – Resumo dos resultados obtidos no mapeamento

BASE DE DADOS	Número inicial de artigos selecionados			Artigos selecionados após aplicação dos critérios de exclusão
	Grupo 1	Grupo 2	Grupo 3	
Periódico Capes	5478	10335	6439	118
Google Acadêmico	5293	4766	7756	195
<i>Snowball sampling</i>	-	-	-	4
<b>Total</b>				317

As seguintes expressões-chave foram utilizadas em cada grupo: a) Grupo 1: (“Accident prediction” OR “Crash accident” OR “Injury severity” OR “Accident model\*” OR “Road traffic accident” OR “Motor vehicle collision\*” OR “Crash injury accident\*” OR “Crash frequenc\*” OR “Crash frequenc\* by severity” OR “Crash prediction”) AND “Artificial intelligence”; b) Grupo 2: (“Accident prediction” OR “Crash accident” OR “Injury severity” OR “Accident model\*” OR “Road traffic accident” OR “Motor vehicle collision\*” OR “Crash injury accident\*” OR “Crash frequenc\*” OR “Crash frequenc\* by severity” OR “Crash prediction”) AND “Expert system\*”; e, c) Grupo 3: (“Accident prediction” OR “Crash accident” OR “Injury severity” OR “Accident model\*” OR “Road traffic accident” OR “Motor vehicle collision\*” OR “Crash injury accident\*” OR “Crash frequenc\*” OR “Crash frequenc\* by severity” OR “Crash prediction”) AND “Machine learning”

A lista de referências bibliográficas de cada artigo foi analisada, sendo selecionados artigos com conteúdo aderente ao tema e que não constavam no conjunto de artigos já selecionados, sendo o processo encerrado quando do atingimento da saturação teórica, situação em que novas referências não adicionam mais informação relevante ao fator estudado. Este procedimento gerou a inclusão de 4 artigos aos 313 trabalhos iniciais, totalizando 317 artigos.

Com relação a este total, para fins de garantir artigos de maior qualidade para o mapeamento, foram selecionados apenas artigos de periódicos com Fator de impacto JCR (Thomson Scientific’s Journal Citation Reports) ou índice SJR (Scimago Journal Rank) superior a 0,50, donde, finalmente, foram escolhidos 21 artigos.

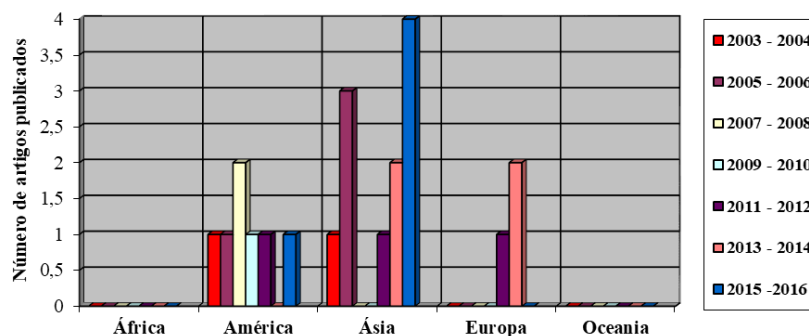
Na Tabela 2 estão apresentados os detalhes de publicação desses trabalhos. Quanto aos periódicos, é perceptível que o *Accident Analysis and Prevention* tem a maior expressividade dentro do tema, até porque trata-se de uma revista vocacionada para acidentes - em especial de trânsito - e a dotada dos melhores indicadores de qualidade. Outro periódico de reconhecida qualidade na área é o *Safety Science*, onde mais de 20% dos artigos foram publicados. Os demais periódicos não apresentaram destaque.

Tabela 2 – Relação de artigos publicados em periódicos, por ano

Periódico	Número de artigos/ano					Subtotal	%
	(2/2006)	(1/2007)	(1/2008)	(1/2010)	(1/2011)		
<i>Accident Analysis and Prevention</i>	(2/2013)	(1/2014)	(1/2015)	(2/2016)		12	57,14
<i>Analytic Methods in Accident Research</i>			(1/2016)			1	4,76
<i>Journal of Safety Research</i>		(1/2005)				1	4,76
<i>Journal of Transportation Engineering</i>			(1/2004)			1	4,76
<i>Promet - Traffic &amp; Transportation Safety Science</i>			(1/2015)			1	4,76
	(1/2003)	(1/2005)	(2/2011)	(1/2013)		5	23,82
<b>Total</b>						<b>21</b>	

Na Figura 1 foram agrupadas as publicações por período temporal e continente. Verificou-se que as publicações estão concentradas em apenas sete países. Os Estados Unidos lideram o ranking geral de publicações, com um terço dos artigos selecionados. Essa proporção aumenta quando se considera a origem dos dados utilizados na modelagem, saltando para quase 43%, uma vez que pesquisadores chineses e coreanos desenvolveram modelos a partir de dados americanos. E ainda, no continente americano somente os Estados Unidos possuem publicações. Ainda em termos de países, só China e Espanha também apresentam relativa expressividade.

Figura 1 – Relação temporal e continental de publicações



Já em termos continentais, é a Ásia que possui o maior número de publicações, concentrados essencialmente no período de 2015 a 2016. Estes trabalhos estão distribuídos entre em cinco países, a saber: China, Coreia do Sul, Irã, Taiwan e Turquia. No continente europeu apenas a Espanha figurou no rol de países com publicações selecionadas, ao passo que África e Oceania não possuem nenhum trabalho dentre os artigos selecionados. Destaque-se que o Brasil não possui nenhum trabalho publicado na temática.

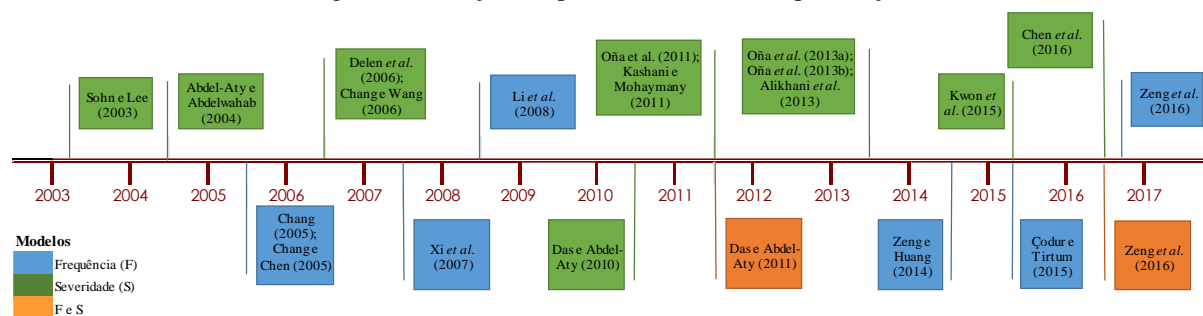
Considerando o número acumulado de artigos publicados por continente, nota-se tendência crescente de estudos na Ásia, especialmente nos últimos anos. Na América verifica-se um padrão de crescimento-estabilização-crescimento de publicações, tendo saído em 2016 da estagnação. Já a Europa, embora tenha apresentado crescimento em anos anteriores, tem a última publicação registrado em 2013.

Do ponto de vista de principais centros de investigação e pesquisadores, constatou-se que as publicações americanas, à exceção de dois trabalhos, são oriundas de duas universidades, em que, pelo menos um dos autores está presente em todos os artigos publicados pelo respectivo grupo de pesquisa. De forma similar, verificou-se que todos os artigos espanhóis têm o primeiro autor em comum (Oña, J.), caracterizando uma sequência de estudos realizados por aquele grupo de pesquisa. O fato se repete na China, onde dois autores

(Zeng, Q. e Huang, H.) estão presentes em todas as publicações selecionadas daquele país. Mesma situação ocorre em Taiwan, onde o primeiro autor (Chang, L.) está presente em todos os artigos. Contrariamente, nos demais países (Coreia do Sul, Irã e Turquia) não existe autoria comum entre os artigos.

Após análise do resumo e estrutura geral dos artigos selecionados, foram identificados três tópicos principais para agrupamento dos resultados obtidos, são eles: modelos de previsão de acidentes por severidade, modelos de previsão de frequência de acidentes e, modelos de previsão de frequência e severidade de acidentes. Na Figura 2 está apresentada a linha cronológica das publicações, destacando-se o tipo de abordagem empregada.

Figura 2 – Relação temporal e continental de publicações



De modo complementar, na Tabela 3 são apresentadas as referências, país do primeiro autor, tipo de modelagem e técnicas utilizadas. Nota-se a predominância (62%) de modelos de classificação de acidentes por nível de severidade, seguido de modelos de previsão do número esperado de acidentes, representando cerca de 30% dos casos e, apenas dois artigos que abordam de forma combinada a frequência e severidade de acidentes.

Tabela 3 – Tipo de modelagem e técnicas empregadas nos estudos

Referência	País	Tipo de modelagem	Técnica(s) utilizada(s)
Sohn e Lee (2003)	Coreia do Sul	Severidade	RNA; AD
Abdel-Aty e Abdelwahab (2004)	EUA	Severidade	RNA; ARTMAP; MP
Chang (2005)	Taiwan	Frequência	RNA; NB
Chang e Chen (2005)	Taiwan	Frequência	CART; NB
Delen <i>et al.</i> (2006)	EUA	Severidade	RNA
Chang e Wang (2006)	Taiwan	Severidade	CART
Xie <i>et al.</i> (2007)	EUA	Frequência	NB; RNA; RNB
Li <i>et al.</i> (2008)	EUA	Frequência	NB; SVM
Das e Abdel-Aty (2010)	EUA	Severidade	PG
Oña <i>et al.</i> (2011)	Espanha	Severidade	RB
Kashani e Mohaymany (2011)	Irã	Severidade	CART
Das e Abdel-Aty (2011)	EUA	Frequência e severidade de acidentes	PG
Oña <i>et al.</i> (2013a)	Espanha	Severidade	CART; AD
Oña <i>et al.</i> (2013b)	Espanha	Severidade	LCA; RB
Alikhani <i>et al.</i> (2013)	Irã	Severidade	RNA; ANFIS; AK; SOM
Zeng e Huang (2014)	China	Severidade	RNA; ML
Kwon <i>et al.</i> (2015)	Coreia do Sul	Severidade	CNB; AD; ML
Çodur e Tortum (2015)	Turquia	Frequência	RNA
Chen <i>et al.</i> (2016)	EUA	Severidade	CART; SVM

*continua*

Tabela 3 – Tipo de modelagem e técnicas empregadas nos estudos (continuação)

Zeng <i>et al.</i> (2016a)	China	Frequência	RNA
Zeng <i>et al.</i> (2016b)	China	Frequência por severidade	RNA; <u>MMP</u>
LEGENDA			
Técnicas de aprendizado de máquina			
<b>AD:</b> Árvores de decisão <b>AK:</b> Algoritmos k-means <b>ANFIS:</b> Sistema neuro-fuzzy <b>ARTMAP:</b> Redes neurais com teoria da ressonância adaptativa fuzzy <b>CART:</b> Árvores de classificação e regressão <b>CNB:</b> Classificadores Naive Bayes <b>LCA:</b> Análise de classes latentes <b>PG:</b> Programação genética <b>RB:</b> Redes Bayesianas <b>RNA:</b> Redes neurais artificiais <b>RNB:</b> Redes neurais Bayesianas <b>SOM:</b> Self-Organizing Maps <b>SVM:</b> Máquina de vetores de suporte			
Técnicas estatísticas			
<b>ML:</b> Modelo logit <b>MP:</b> Modelo probit <b>MMP:</b> Modelo multivariado Poisson-lognormal <b>NB:</b> Modelo de regressão com distribuição binomial negativa			

O Aprendizado de Máquina (AM), uma subárea da Inteligência Artificial, é bastante utilizado e se coloca como uma poderosa ferramenta para resolver problemas em vários domínios de aplicação. Os algoritmos de AM envolvem conhecimento de diferentes áreas, tais como Probabilidade e Estatística, Complexidade Computacional, Teoria da Informação, Psicologia, Neurobiologia e Teoria de Controle. De maneira geral, os grandes grupos de técnicas de AM identificadas para a tarefa de modelagem da segurança viária são: Árvores de Decisão, Algoritmos Evolutivos, Suporte de Vetor de Máquina e Redes Neurais Artificiais.

O uso de Árvores de Decisão (AD) é bastante útil nas tarefas de classificação. Na construção de AD é utilizado um conjunto de treinamento formado pelas entradas e saídas, estas últimas são as classes. A estrutura de uma AD contém um nó raiz (que inicia a árvore), nós de decisão (que dividem um determinado atributo e geram as ramificações) e folhas (que contém as informações de classificação). Cada nó indica o teste de um atributo, sendo a utilidade do atributo para a classificação utilizada como critério de ramificação. Dessa forma, o atributo escolhido, que será um nó da árvore, é aquele que gera maior ganho de informação (entropia), ou seja, melhor qualidade de classificação. Note-se que um percurso na árvore (da raiz à cada nó-folha) corresponde a uma regra de associação (Quilan, 1986; Perez, 2012). Os algoritmos de indução de árvores de decisão buscam, em meio a um conjunto de atributos, aqueles que separam da melhor forma os exemplos, gerando sub-árvores.

Algoritmos evolutivos (AE) são métodos estocásticos de busca baseado nos mecanismos de seleção natural em que os indivíduos mais aptos sobrevivem (Holland, 1975). Cada indivíduo corresponde a uma solução candidata para um determinado problema e é avaliado por uma função fitness (que mensura a qualidade da solução). A cada geração (iteração) os melhores indivíduos têm maior probabilidade de serem escolhidos para se reproduzirem. Os indivíduos selecionados sofrem operações de crossover (partes do material genético de dois indivíduos são trocadas) e mutação (parte do material genético de um indivíduo é substituída por material genético formado aleatoriamente), gerando novos indivíduos (filhos) que irão substituir os pais e formar uma nova geração de população. O processo descrito é iterativamente repetido até que um critério de parada seja satisfeito (Floreano e Mattiussi, 2008; Yu e Chen, 2010; Martino, 2015). Os dois tipos principais de AE são Algoritmos Genéticos (AG) e Programação Genética (PG), utilizados especialmente em problemas de otimização, geralmente associados a outras técnicas.

A técnica Suporte de Vetor de Máquina (SVM) é fundamentada na teoria do aprendizado estatístico (Scholkopf e Smola, 2002). A ideia principal é construir um hiperplano como superfície de decisão, de forma que a margem de separação entre exemplos positivos e negativos seja máxima. O modelo discrimina, usando o hiperplano, o conjunto de amostras de teste dividindo-os em dois grupos: amostras negativas e amostras positivas. Apesar de concebida como técnica de classificação, também foi estendida para resolver

problemas de regressão e dados não linearmente separáveis. (Borges, 1998; Borin, 2007; Mareto, 2011; Alves, 2012).

Formalmente, uma rede neural artificial (RNA) pode ser entendida como um processador altamente complexo, não-linear e paralelo que, tem uma propensão natural de armazenar conhecimento experimental e torna-lo disponível posteriormente (Haykin, 2009). Uma RNA *Multi Layer Perceptron* (MLP) é tipicamente composta por três tipos camadas: uma camada de entrada, uma camada de saída e uma ou mais camadas ocultas. A camada de entrada recebe os valores das variáveis explanatórias, ou seja, os dados de entrada. A camada oculta, composta por  $m$  neurônios, sumariza o peso dos valores de entrada das diferentes variáveis explanatórias e calcula os complicados padrões de associação. Geralmente apenas uma camada oculta é adequada para as aplicações na análise de acidentes, ao passo que a definição do número de neurônios nesta camada é geralmente alvo de experimentação. Já a camada de saída, soma os valores dos diferentes neurônios ocultos e, na sequência, apresenta os valores de saída da rede (Villiers e Barnard, 1993; Chang, 2005). Em termos de treinamento, o algoritmo de retropropagação de erros (backpropagation algorithm) é o mais utilizado. Nele, conforme Haykin (2001), busca-se a minimização dos erros, a partir do ajuste dos pesos da rede. Baseado no método de gradiente descendente, a função de custo está na direção e sentido em que a função tem taxa de variação mínima e garante que a rede caminha na superfície na direção que leva à maior redução do erro. Por fim, a principal função de ativação utilizada, relacionada à capacidade representativa das redes neurais e que introduz uma componente não linear, é do tipo sigmoide, geralmente função logística ou hiperbólica.

### 3.1 Modelos de previsão de acidentes por severidade

O desenvolvimento de modelos de previsão de acidentes por severidade consiste num problema de classificação. Doutro modo, ajusta-se um modelo que, a partir de características do acidente, dos ocupantes, dos veículos e viário-ambientais, seja capaz de prever qual classe de severidade resultará aquele acidente. Os trabalhos desenvolvidos tiveram como objetos de estudo: rodovias rurais, rodovias inseridas em contexto urbano e vias expressas.

Nesse tipo de modelagem predominaram os trabalhos americanos, seguidos dos asiáticos. E a técnica mais empregada para esse objetivo é RNA, seguida de CART (uma extensão da tradicional árvore de decisão). As duas técnicas, por seu mecanismo de funcionamento e concepção, são bastante adequadas para as características do problema estudado, o que é confirmado pelos resultados dos diversos artigos analisados. Ainda assim, Das e Abdel-Aty (2010) e Chen *et al.*, (2016) também exploraram o uso de PG e SVM, respectivamente, para tal finalidade. Os resultados obtidos foram satisfatórios, mas não tão promissores quanto aos resultantes do emprego de RNA e AD.

As medidas utilizadas para aferir a qualidade dos modelos são: acurácia, sensibilidade, especificidade, precisão e *F-measure*. A acurácia refere-se ao grau em que o teste é capaz de determinar o verdadeiro valor do que está sendo medido ou, doutra forma, estima a probabilidade do classificar acertar suas previsões. A sensibilidade mede a capacidade de se prever uma classe positiva cuja predição está correta, ou seja, ela indica quantos exemplos positivos foram previstos do total de exemplos. A medida de especificidade mensura a capacidade de se prever uma classe negativa cuja predição está correta, ou seja, quantos exemplos negativos foram preditos do total de exemplos. A medida de precisão calcula a probabilidade de a predição positiva estar correta em relação a todas as amostras. Por fim, a medida *F-measure* é a combinação balanceada das medidas de precisão e sensibilidade (Borges, 2012).

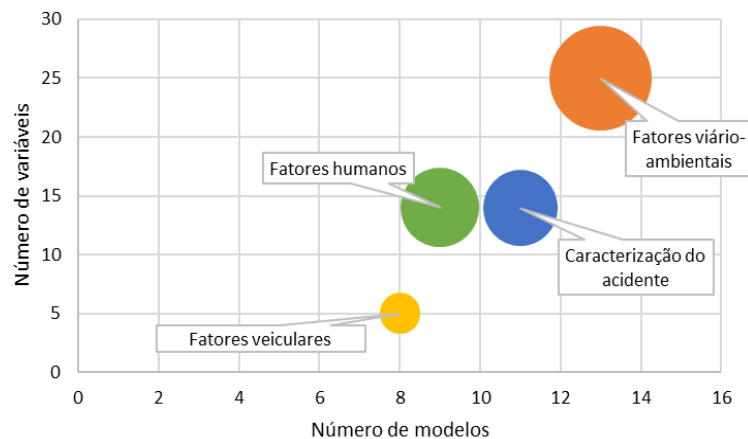
Dependendo do objetivo da modelagem, um ou mais indicadores podem ser mais importantes que os demais. No caso da classificação de acidentes em: com vítimas ou com vítimas fatais, por exemplo, a acurácia do modelo para acidentes com mortes pode ser

próxima de 100% enquanto que a sensibilidade pode ser nula. Isso porque, ainda hipoteticamente, se existirem apenas 3 acidentes com vítimas fatais (em um total de 100 acidentes) e o modelo prever todos eles como pertencentes à acidentes com vítimas, a acurácia é alta, mas o modelo não apresenta nenhum poder de classificação dos casos que envolvam vítimas fatais.

Em termos de número de classes adotadas, a maioria dos autores utiliza classificação binária, especialmente: lesão corporal ou danos materiais; lesão leve ou lesão grave/fatalidade. Ainda assim, autores também abordam o problema como sendo de multiclasse, se valendo de três (Chang e Wang, 2006; Kashani e Mohaymany, 2011; Chen *et al.*, 2016), quatro (Abdel-Aty e Abdelwahab, 2004; Zeng e Huang, 2014) ou até cinco (Delen *et al.*, 2006) classes. Ainda que o problema fosse multicategórico, após verificação inicial e constatação de resultados pouco razoáveis, alguns autores preferiram o decompor em classificadores binários. Dois métodos são mais comuns: *one-vs-all* (OVA) – onde cada classe é, separadamente, comparada com as demais classes; *all-vs-all* (AVA) – onde são criadas todas as combinações binárias possíveis.

Por fim, é interessante destacar as principais variáveis explanatórias utilizadas nos modelos. Essas, para fins de comparação, foram agrupadas segundo: fatores viário-ambientais, fatores humanos, fatores veiculares e caracterização do acidente. Apresenta-se na Figura 3 a distribuição dos grupos de variáveis, considerando todos os modelos, em termos do número de variáveis por grupo, quantidade de modelos em que foram empregadas e total de ocorrência destas nos modelos. No eixo das abscissas consta o número de modelos que incorporaram cada grupo de variáveis, ao passo que o eixo das ordenadas corresponde ao número total de variáveis de cada grupo. Por fim, o tamanho das bolhas é dado pelo número total de ocorrências das variáveis de cada grupo.

Figura 3 – Distribuição dos grupos de variáveis explicativas na modelagem de acidentes por severidade



Denota-se que os fatores viário-ambientais, além de terem sido empregados em todos os modelos, possuem o maior número de ocorrência de suas variáveis, com quase 45% do total. As variáveis de caracterização do acidente estão presentes em 11 dos 13 modelos analisados e auferem a terceira maior proporção de ocorrência. Os fatores humanos foram incorporados em 9 modelos e figuram como segundo grupo de maior frequência de variáveis. Os fatores veiculares, por sua vez, foram utilizados em 8 modelos e o total proporcional de ocorrência de suas variáveis foi pouco relevante, menos de 7%. E ainda, especificamente tratando de variáveis explicativas (inputs), está apresentada na Figura 4 a distribuição geral das variáveis nos modelos.



Figura 4 – Distribuição das variáveis de entrada de todos os modelos de previsão de acidentes por severidade



Das 57 variáveis resultantes do agrupamento por similaridade, 18 tiveram apenas uma ocorrência, denotando serem variáveis insignificantes por constarem em só um modelo. Também foram identificadas 9 variáveis com somente duas ocorrências, o que não as torna expressivas. As variáveis de maior destaque foram: sexo, idade, condições climáticas, iluminação, uso do cinto de segurança, tipo de acidente, horário, dia da semana, trecho da via, tipo de acostamento e tipo de veículo. Os fatores humanos, por meio das variáveis sexo, idade e uso de cinto de segurança, figuraram nas primeiras posições do ranking de ocorrência de variáveis nos modelos analisados. Os fatores viário-ambientais, em convergência ao constatado individualmente nos modelos, são os principais na modelagem. A proporção média destes fatores nos modelos foi de 43,85% e dentre as 11 variáveis mais importantes aqui elencadas na análise geral, têm uma parcela de quase 37%. Isso confirma que essas variáveis não só estão em maior número, como ocorrem entre as principais. Características do acidente (tipo de acidente, horário e dia da semana) possuem a mesma proporção que os fatores humanos, no entanto, com menos expressividade que estes. Já no nível veicular, apenas uma variável foi verificada entre as principais, o tipo de carro, e na última colocação do ranking.

### 3.2 Modelos de previsão de frequência de acidentes

Os modelos de previsão de frequência de acidentes são típicos problemas de regressão nos quais a resposta é o número esperado de acidentes. Nos trabalhos analisados, foram desenvolvidos estudos abordando rodovias rurais de pista simples, rodovias rurais de pista dupla e diferentes tipos de rodovias (no caso de Zeng *et al.*, 2016), em séries históricas de 2 a 5 anos.

Os modelos tiveram como resposta: número de acidentes/segmento ou número de acidentes/segmento/ano. Diferente dos modelos de severidade, na previsão de frequência de acidentes é necessário segmentar a rodovia e atribuir o número de acidentes observados por segmento. No problema de classificação, cada acidente é contabilizado como uma observação a ser processada pelo modelo; no caso de regressão, por outro lado, cada segmento (com suas características viário-ambientais e acidentes agregados) é uma observação.

Diante disso, uma definição importante, mas não consensual, é o tipo de segmentação mais adequado. Os dois principais tipos são: segmentação em trechos com características homogêneas (Chang, 2005; Xie *et al.*, 2007; Li *et al.*, 2008; Çodur e Tortum, 2015; Zeng *et al.*, 2016) e segmentação em trechos de extensão fixa (Chang e Chen, 2005). Cada abordagem tem suas vantagens e fragilidades, cabendo ao pesquisador a análise da melhor alternativa que, na maioria das vezes é escolhida de forma pragmática.

Dentre as técnicas empregadas, RNA e CART novamente figuraram como as mais utilizadas. Isso porque tais técnicas têm a flexibilidade de lidar com problemas de classificação e regressão. Também houve a exploração SVM para este propósito. O uso de redes neurais, no entanto, é nitidamente predominante (cerca de 70%), dada a flexibilidade da técnica e aprimoramentos propostos ao longo dos anos. Em todos os casos, a base de dados foi dividida em duas parcelas: treinamento e teste, sendo os percentuais de 60%/40%, 70%/30%, 75%/25% ou 80%/20%.

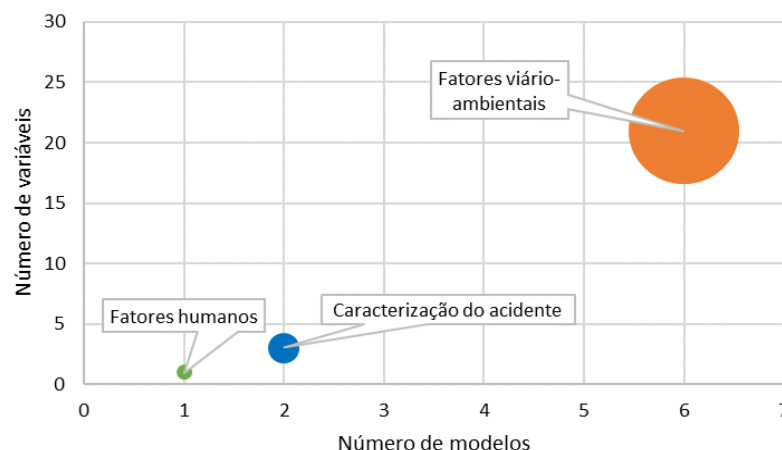
Em termos de desempenho, os modelos são avaliados basicamente por do desvio médio absoluto (MAD), erro quadrático (MSE) e raiz do erro quadrático médio (RMSE), medidas que estimam erros a partir da diferença entre o valor observado e o valor predito pelo modelo. As RNA, quando possível a comparação, apresentaram melhor desempenho. Além disso, de forma geral, sempre que houve comparação dos resultados dos modelos com os obtidos por meio de técnicas estatísticas, houve superioridade da abordagem com técnica de AM tanto na previsão de acidentes por severidade quanto da frequência de acidentes.

Quanto aos grupos de variáveis utilizados nos modelos, tem-se a Figura 5. Observa-se que os fatores viário-ambientais, além de terem sido empregados em todos os modelos, possuem quase a totalidade de ocorrência de suas variáveis, correspondendo a mais de 91% do total. Os fatores humanos e variáveis de caracterização do acidente mostraram-se insignificantes com 1,8% e 7,0% de representatividade, respectivamente. É notória, portanto, a imprescindibilidade dos fatores viário-ambientais nos modelos de previsão de frequência de acidentes.

desenvolvimento de modelos de previsão de acidentes por severidade consiste num problema de classificação. Doutro modo, ajusta-se um modelo que, a partir de características do acidente, dos ocupantes, dos veículos e viário-ambientais, seja capaz de prever qual classe de severidade resultará aquele acidente. Os trabalhos desenvolvidos tiveram como objetos de estudo: rodovias rurais, rodovias inseridas em contexto urbano e vias expressas.

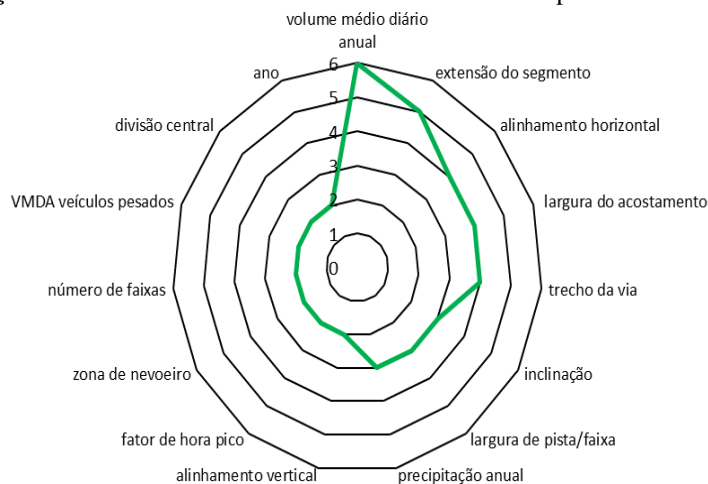
Nesse tipo de modelagem predominaram os trabalhos americanos, seguidos dos asiáticos. E a técnica mais empregada para esse objetivo é RNA, seguida de CART (uma extensão da tradicional árvore de decisão). As duas técnicas, por seu mecanismo de funcionamento e concepção, são bastante adequadas para as características do problema estudado, o que é confirmado pelos resultados dos diversos artigos analisados. Ainda assim, Das e Abdel-Aty (2010) e Chen *et al.*, (2016) também exploraram o uso de PG e SVM, respectivamente, para tal finalidade. Os resultados obtidos foram satisfatórios, mas não tão promissores quanto aos resultantes do emprego de RNA e AD.

Figura 5 – Distribuição dos grupos de variáveis explicativas na modelagem de frequência de acidentes



Especificamente tratando de variáveis explicativas (inputs), está apresentada na Figura 6 a distribuição geral das variáveis nos modelos. As variáveis de maior destaque foram as mesmas já identificadas entre os fatores viário-ambientais: volume médio diário anual, extensão do segmento, alinhamento horizontal, largura do acostamento, trecho da via, inclinação, largura de pista (ou de faixa) e precipitação anual. Os modelos estão, portanto, baseados nas variáveis relacionadas à via ou ao meio ambiente no qual está inserida.

Figura 6 – Distribuição das variáveis de entrada de todos os modelos de previsão de frequência de acidentes



A diferença entre os principais preditores observados nos modelos de severidade e frequência é esperada, já que a primeira abordagem está voltada para as características do acidente, envolvidos e veículos e a segunda, com o ambiente viário em que os acidentes ocorreram.

### 3.3 Modelos de previsão de frequência e severidade de acidentes

O acidente de trânsito é um fenômeno complexo e muitas vezes requer não somente a análise do ponto de vista de severidade ou de número esperado de acidentes, mas sim destes dois aspectos em simultâneo. Essa é uma abordagem que ainda carece de exploração e aprimoramento. Foram identificados apenas dois trabalhos nesta linha de abordagem.

Das e Abdel-Aty (2011) fizeram uma abordagem combinada de análise de frequência e severidade de acidentes de trânsito, visando ter uma visão mais ampla da segurança viária em rodovias. Embora não tenham desenvolvido um modelo de previsão conjunta, a partir do mesmo conjunto de dados os autores, por meio de programação genética (PG), desenvolveram modelos independentes para os dois propósitos: frequência de acidentes (número de acidentes/segmento) e severidade de acidentes (lesão não incapacitante, lesão incapacitante/fatalidade). No caso da previsão de frequência a rodovia foi segmentada e caracterizada. Além disso, as variáveis empregadas em cada tipo de modo foram ligeiramente diferentes, no entanto, essencialmente fatores viário-ambientais.

Os resultados de Das e Abdel-Aty (2011) revelaram a sobreposição de conjunto de fatores significativos (divisão central, resistência ao deslizamento e largura da via) tanto nos modelos de frequência quanto nos modelos de classificação; indício, segundo os autores, da presença de uma relação complexa entre os problemas aparentemente diferentes.

Zeng *et al.* (2016b), por sua vez, propuseram um modelo que fosse capaz de prever o número de acidentes segundo cada nível de severidade, sendo suas variáveis-resposta: número de acidentes com lesões leves/segmento/ano; número de acidentes com lesões graves ou fatais/segmento/ano. Para tanto procedeu-se à segmentação e caracterização de trechos homogêneos da rodovia, predominante as características relacionadas à via e ao ambiente do

entorno. Os autores utilizaram RNA para modelagem, buscando explorar a relação linear entre as variáveis explicativas e as dependentes, propondo um algoritmo de otimização da estrutura da rede e método de extração de regras. Os resultados, comparados aos provenientes do emprego de regressão multivariada Poisson-lognormal, indicaram melhor desempenho e poder preditivo das redes, especialmente a otimizada, frente ao modelo estatístico. O volume de tráfego, extensão do segmento e velocidade máxima regulamentada foram as variáveis tidas como mais importantes do modelo.

#### 4. CONCLUSÕES

Mediante execução do mapeamento foi possível verificar o crescente interesse pelo uso de técnicas de aprendizado de máquina na modelagem da segurança viária. Verificou-se a existência de três abordagens de modelagem, conforme o propósito: classificação de acidentes por severidade; previsão do número esperado de acidentes em um trecho da rodovia; previsão de acidentes por severidade. A primeira abordagem é a mais recorrente entre os trabalhos, podendo ter o modelo uma resposta binária ou multiclasse. No caso da frequência de acidentes, o passo inicial é segmentar a rodovia analisada em trechos, depois caracterizá-los e utilizá-los como dados para ajuste do modelo. O último tipo de modelo, ainda pouco explorado, aborda modelos independentes (severidade e frequência) a partir da mesma base de dados inicial ou, numa abordagem conjunta, modela o número esperado de acidentes segundo cada nível de severidade.

Nos modelos de severidade, o problema multiclasse ainda tem sido explorado de diferentes formas, especialmente mediante decomposição em problemas binários. Na frequência de acidentes não existe consenso a respeito da estratégia mais adequada de segmentação, recaindo numa adoção por conveniência. Na abordagem conjunta tem-se o desafio de abordar de forma simultânea a ocorrência de mais de um tipo de acidente. Isso restringe, inclusive, as técnicas passíveis de utilização uma vez que trata-se de um problema de resposta multivariada, tornando a RNA uma candidata natural. Já nos dois primeiros tipos de modelos, fica evidente a maior emprego de RNA e CART

Os fatores viário-ambientais predominam em todos os tipos de modelo, sendo essencialmente necessários aos modelos de previsão de frequência de acidentes. Ainda assim, no caso da modelagem de severidade, também têm lugar os fatores humanos e veiculares.

Em termos de resultados, todos os estudos apontaram para o potencial de utilização de técnicas de AM para o desenvolvimento de modelos de previsão de acidentes segundo diferentes propósitos. Inclusive, quando houve comparação, esses modelos mostraram-se de melhor desempenho e poder preditivo que os modelos estatísticos.

Por fim, é importante referir que o mapeamento sistemático da literatura é uma importante ferramenta para traçar o panorama de determinada área de pesquisa e apontar limitações e oportunidades de investigação. Diante disso, acredita-se que este artigo cumpriu com tal propósito e apresentou o contexto geral da modelagem da segurança viária com uso de técnicas de AM, destacando as principais publicações, abordagens metodológicas, desempenho de modelos, variáveis explicativas e avanços necessários.

#### 5. AGRADECIMENTOS

O primeiro autor agradece ao apoio financeiro do Instituto Federal Goiano (IF Goiano). Refere-se ainda que o presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e da Fundação para a Ciência e Tecnologia- Portugal - (FCT) por meio do projeto “Mobilidade Urbana Sustentável e Segura”, no qual este trabalho está inserido.

## 6. REFERÊNCIAS

ABDEL-ATY, M. A.; ABDELWAHAB, H. T. Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison. **Journal of Transportation Engineering**, v. 130, p.204-210, 2004. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2004\)130:2\(204\)](https://doi.org/10.1061/(ASCE)0733-947X(2004)130:2(204))

ABDELWAHAB, H.; ABDEL-ATY, M. Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections. **Transportation Research Record**, v. 1746, p.6-13, 2001. <https://doi.org/10.3141/1746-02>

ALIKHANI, M.; NEDAIE, A.; AHMADVAND, A. Presentation of clustering-classification heuristic method for improvement accuracy in classification of severity of road accidents in Iran. **Safety Science**, v. 60, p.142–150, 2013. <https://doi.org/10.1016/j.ssci.2013.06.008>

ALVES, Júlio Cesar Laurentino **Máquina de vetores de suporte aplicada a dados de espectroscopia NIR de combustíveis e lubrificantes para o desenvolvimento de modelos de regressão e classificação**. Tese (Doutorado). 247p. Programa de Pós-Graduação em Química, Universidade de Campinas, 2012.

BORGES, Helayne Bronoski **Classificador hierárquico multirótulo usando uma rede neural competitiva**. Tese (Doutorado). 165p. Programa de Pós-Graduação em Informática, Pontifícia Universidade Católica do Paraná, 2012.

BORIN, Alessandra **Aplicações de Máquinas de vetores de suporte por mínimos quadrados (LS-SVM) na quantificação de parâmetros de qualidade de matrizes lácteas**. Tese (Doutorado). 122p. Programa de Pós-Graduação em Química, Universidade de Campinas, 2007.

BURGES, C. J. C. A tutorial on support vector machines for pattern recognition. **Data Mining and Knowledge Discovery**, v. 2, n. 2, p.121–167, 1998. <https://link.springer.com/article/10.1023/A:1009715923555>

CAFISO, S.; DI GRAZIANO, A.; DI SILVESTRO, G.; LA CAVAA, G.; PERSAUD, B. Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables. **Accident Analysis and Prevention**, v. 42, p.1072-1079, 2010. <https://doi.org/10.1016/j.aap.2009.12.015>

CARDOSO, Gilmar **Modelos para Previsão de Acidentes de Trânsito em Vias Arteriais Urbanas**. Tese (Doutorado). 289p. Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, 2006.

CHANG, L.; CHEN, W. Data mining of tree-based models to analyze freeway accident frequency. **Journal of Safety Research**, v. 36, p.365-375, 2005. <https://doi.org/10.1016/j.jsr.2005.06.013>

CHANG, L.; WANG, H. Analysis of traffic injury severity: Na application of non-parametric classification tree techniques. **Accident Analysis and Prevention**, v. 38, p.1019-1027, 2006. <https://doi.org/10.1016/j.aap.2006.04.009>

CHEN, C.; ZHANG, G.; QIAN, Z.; TAREFDER, R. A.; TIAN, Z. Investigating driver injury severity patterns in rollover crashes using support vector machine models. **Accident Analysis and Prevention**, v. 90, p.128-139, 2016. <https://doi.org/10.1016/j.aap.2016.02.011>

ÇODUR, M. Y.; TORTUM, A. AN ARTIFICIAL NEURAL NETWORK MODEL FOR HIGHWAY ACCIDENT PREDICTION: A CASE STUDY OF ERZURUM, TURKEY. **Promet - Traffic&Transportation**, v. 27, n. 3, p.217-225. 2015.

COSTA, Jocilene Otilia **Desenvolvimento dum modelo de previsão de acidentes**. Tese (Doutorado). 296p. Doutorado em Segurança Rodoviária, Universidade do Minho (Portugal), 2013.

DAS, A.; ABDEL-ATY, M. A genetic programming approach to explore the crash severity on multi-lane roads. **Accident Analysis and Prevention**, v. 42, p.548-557, 2010. <https://doi.org/10.1016/j.aap.2009.09.021>

DAS, A.; ABDEL-ATY, M. A combined frequency-severity approach for the analysis of rear-end crashes on urban arterials. **Safety Science**, v. 49, p.1156-1163, 2011. <https://doi.org/10.1016/j.ssci.2011.03.007>

DELEN, D.; SHARDA, R.; BESSONOV, M. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. **Accident Analysis and Prevention**, v. 38, n. 3, p.434-444, 2006. <https://doi.org/10.1016/j.aap.2005.06.024>

FLOREANO, D.; MATTIUSI, C. **Bio-inspired artificial intelligence: theories, methods, and technologies**. Cambridge: The MIT Press, 2008. 674 p.

HAYKIN, S. **Redes Neurais: Princípios e prática**. Paulo Martins Engel (Trad.), 2ª Ed. Porto Alegre: Bookman, 2001. 898 p.

HAYKIN, S. **Neural Networks and Learning Machines**. 3rd edition. New York: PrenticeHall, 2009. 936 p.

HOLLAND, J. H. **Adaptation in Natural and Artificial Systems**. Ann Arbor: University of Michigan Press, 1975. 232 p.

KASHANI, A. T.; MOHAYMANY, A. S. Analysis of the traffic injury severity on two-lane, two-way rural roads based on classification tree models. **Safety Science**, v. 49, p.1314-1320, 2011. <https://doi.org/10.1016/j.ssci.2011.04.019>

KIM, D.; WASHINGTON, S. The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. **Accident Analysis and Prevention**, v. 38, p. 1094-1100, 2006. <https://doi.org/10.1016/j.aap.2006.04.017>

KWON, O. H.; RHEE, W.; YOON, Y. Application of classification algorithms for analysis of road safety risk factor dependencies. **Accident Analysis and Prevention**, v. 75, p.1-15, 2015. <https://doi.org/10.1016/j.aap.2014.11.005>

LI, H.; GRAHAM, D.J.; MAJUMDAR, A. The effects of congestion charging on road traffic casualties: A causal analysis using difference-in-difference estimation. **Accident Analysis and Prevention**, v. 49, p.366-377, 2012. <https://doi.org/10.1016/j.aap.2012.02.013>

LI, X.; LORD, D.; ZHANG, Y.; XIE, Y. Predicting motor vehicle crashes using support vector machine models. **Accident Analysis and Prevention**, v. 40, n. 4, p.1611-1618, 2008. <https://doi.org/10.1016/j.aap.2008.04.010>

LORD, D.; MANNERING, F. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. **Transportation Research Part A**, v. 44, p.291–305. 2010. <https://doi.org/10.1016/j.tra.2010.02.001>

MANNERING, F. L.; C.R. BHAT Analytic Methods in Accident Research: Methodological Frontier and Future Directions. **Analytic Methods in Accident Research**, v. 1, p.1-22, 2014. <https://doi.org/10.1016/j.amar.2013.09.001>

MARETTO, Damilo Althmann **Aplicação de máquinas de vetores de suporte para desenvolvimento de modelos de classificação e calibração multivariada em espectroscopia no infravermelho**. Tese (Doutorado). 113p. Programa de Pós-Graduação em Ciências, Universidade Estadual de Campinas, 2011.

MARTINO, Jarryer Andrade de **Algoritmos evolutivos como método para desenvolvimento de projetos de arquitetura**. Tese (Doutorado). 282p. Programa de Pós-Graduação em Arquitetura, Tecnologia e Cidade, Universidade Estadual de Campinas, 2015.

MUSSONE, L.; FERRARI, A.; ONETA, M. An analysis of urban collisions using an artificial intelligence model. **Accident Analysis and Prevention**, v. 31, n. 6, p.705–718, 1999. [https://doi.org/10.1016/S0001-4575\(99\)00031-7](https://doi.org/10.1016/S0001-4575(99)00031-7)

NODARI, Christine Tessele **Método de avaliação da segurança potencial de segmentos rodoviários rurais de pista simples**. Tese (Doutorado). 221p. Pós-Graduação em Engenharia de Produção, Universidade Federal do Rio Grande do Sul, 2003.

OMS. Organización Mundial de la Salud. **Informe sobre la situación mundial de la seguridad vial: es hora de pasar a la acción**. Ginebra, 2010. 287 p.

OÑA, J.; LÓPEZ, G.; ABELLÁN, J. Extracting decision rules from police accident reports through decision trees. **Accident Analysis and Prevention**, v. 50, p.1151-1160, 2013a. <https://doi.org/10.1016/j.aap.2012.09.006>

OÑA, J.; LÓPEZ, G.; MUJALLI, R.; CALVO, F. Analysis of traffic accidents on rural highways using Latent Class Clustering and Bayesian Networks. **Accident Analysis and Prevention**, v. 51, p.1-10, 2013b. <https://doi.org/10.1016/j.aap.2012.10.016>

OÑA, J.; MUJALLI, R. O.; CALVO, F. J. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. **Accident Analysis and Prevention**, v. 43, p.402-411, 2011. <https://doi.org/10.1016/j.aap.2010.09.010>

PAULA, H. M. D.; ILHA, M. S. D. O. Uso da Moringa oleifera no tratamento de águas residuárias de usinas de concreto: mapeamento sistemático. **REEC - Revista Eletrônica de Engenharia Civil**, v. 11, n. 1, p.50-60, 2016. <https://doi.org/10.5216/reec.V11i1.35149>

PEREZ, Pedro Santoro **Uma abordagem para a indução de árvores de decisão voltada para dados de expressão gênica**. Dissertação (Mestrado). 120p. Programa de Pós-Graduação em Bioinformática, Universidade de São Paulo, 2012.

QUINLAN, J. R. Induction of Decision Trees. **Machine Learning**, v. 1, p.81-106, 1986.

SCHOLKOPF, B.; SMOLA, A. J. **Learning with kernels**. Cambridge: MIT Press, 2002. 648 p.

SOHN, S.; LEE, S. Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accident in Korea. **Safety Science**, v. 41, n. 1, p.1–14, 2003. [https://doi.org/10.1016/S0925-7535\(01\)00032-7](https://doi.org/10.1016/S0925-7535(01)00032-7)

VILLIERS, J.; BARNARD, E. Backpropagation neural nets with one and two hidden layers. **IEEE Trans. Neural Netw.**, v. 4, n.1, p.136–141, 1993.

WHO. World Health Organization. **Global Status Report On Road Safety 2015**. Geneva, 2015. 323 p.

XIE, Y.; LORD, D.; ZHANG, Y. Predicting motor vehicle collisions using Bayesian neural networks: An empirical analysis. **Accident Analysis and Prevention**, v. 39, n. 5, p.922-933, 2007. <https://doi.org/10.1016/j.aap.2006.12.014>

YU, X.; GEN, M. **Introduction to evolutionary algorithms**. London: Springer, 2010. 401 p.

ZENG, Q.; HUANG, H. A stable and optimized neural network model for crash injury severity prediction. **Accident Analysis and Prevention**, v. 73, p.351–358, 2014. <https://doi.org/10.1016/j.aap.2014.09.006>

ZENG, Q.; HUANG, H.; PEI, X.; WONG, S. C.; GAO, M. Rule extraction from an optimized neural network for traffic crash frequency modeling. **Accident Analysis and Prevention**, v. 97, p.87–95, 2016a. <https://doi.org/10.1016/j.aap.2016.08.017>

ZENG, Q.; HUANG, H.; PEI, X.; WONG, S. C. Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks. **Analytic Methods in Accident Research**, v. 10, p.12–25, 2016b. <https://doi.org/10.1016/j.amar.2016.03.002>

