



# Avaliação de acidentes em travessias pedestres para identificação de fatores de risco utilizando redes neurais artificiais

Bertha Santos<sup>1</sup>, Jorge Gonçalves<sup>1</sup>, Shohel Amin<sup>2</sup>, Sandra Vieira<sup>3</sup> and Carlos Lopes<sup>4</sup>

<sup>1</sup> Universidade da Beira Interior

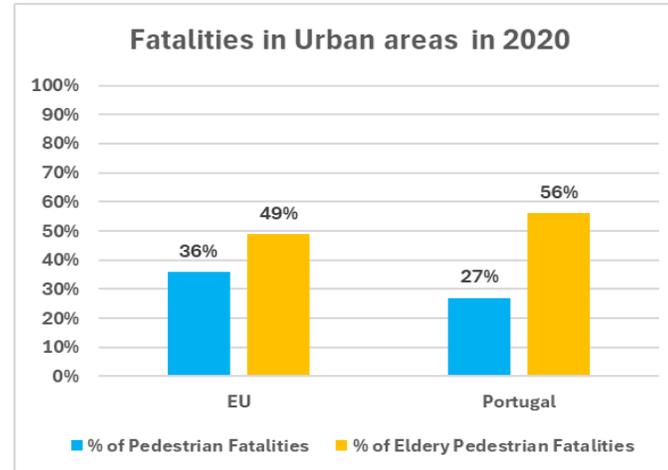
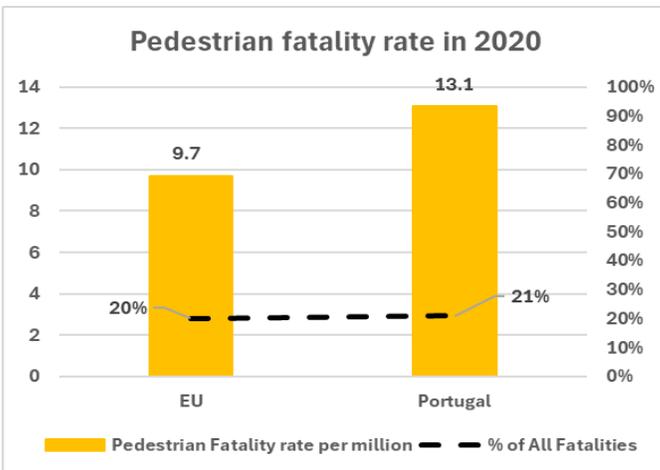
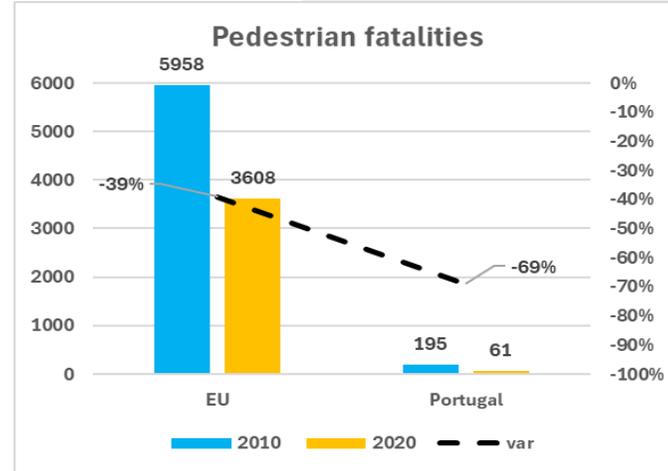
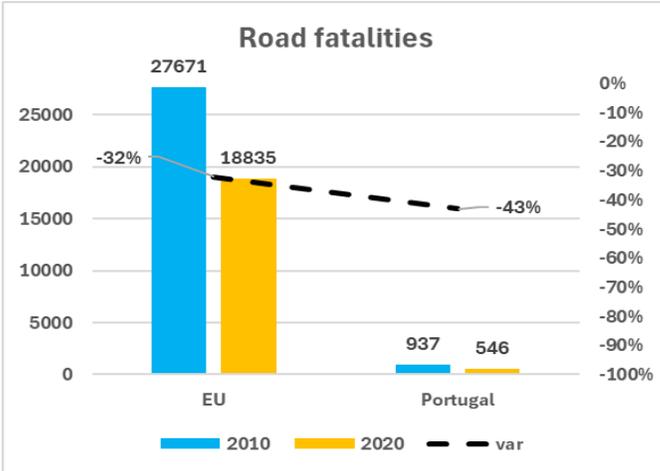
<sup>2</sup> Coventry University

<sup>3</sup> Laboratório Nacional de Engenharia Civil

<sup>4</sup> Autoridade Nacional de Segurança Rodoviária

# Introdução

## Acidentes envolvendo pedestres - Portugal



- As vítimas mortais resultantes de acidentes de viação diminuíram significativamente entre 2010 e 2020
- Esta tendência é ainda mais acentuada no que respeita às vítimas mortais do tipo pedestre
- Em 2020, os pedestres vítimas mortais ainda representavam 20% de todas as mortes em acidentes de viação
- Nas áreas urbanas portuguesas, metade das vítimas mortais são pedestres, dos quais 56% são idosos

# Objetivo

Compreender os factores de risco que afectam a segurança dos pedestres nas travessias pedonais para definir medidas eficazes de proteção

- Definir um **Modelo de Previsão de Acidentes (MPA)** capaz de prever a **gravidade dos acidentes envolvendo pedestres e identificar os factores de risco** (características do ambiente rodoviário e dos pedestres) para os acidentes ocorridos em **travessias pedonais em meio urbano ou na sua proximidade**
- A obtenção de MPAs envolvem geralmente duas fases:
  - Definição de metodologias de modelação - Técnicas Estatísticas e de Inteligência Artificial (IA)
  - Identificação dos factores responsáveis pela ocorrência de acidentes envolvendo pedestres

# Método

Definir um modelo de RNA para prever a gravidade dos acidentes envolvendo pedestres e identificar factores de risco associados a acidentes ocorridos em travessias pedonais urbanas ou na sua proximidade

- Revisão de artigos científicos: **abordagens de modelação** e das **variáveis com impacto** na gravidade das lesões dos pedestres
- Definição da **técnica de *machine learning*** a utilizar
- Aplicação da técnica selecionada à base de dados de acidentes envolvendo pedestres de **2017-2021 da ANSR** (Autoridade Nacional de Segurança Rodoviária)

# Aplicação

## Dados de sinistralidade

- Algoritmo: Redes Neurais Artificiais (RNA)
- Portugal Continental e Arquipélagos dos Açores e da Madeira, 2017 a 2021 (5 anos)
- 24774 pedestres envolvidos
- 24211 acidentes com vítimas (mortais, feridos graves e leves)
- Organização de duas bases de dados para acidentes ocorridos em travessias pedonais em meio urbano ou na sua proximidade:
  - Para o período **Pre-covid19 (2017-2019) (n=9568)**
  - Para o período **Covid19 (2020-2021) (n=3310)**
- Variável dependente: **Gravidade da lesão sofrida pelo pedestre** (morte, ferido grave, ferido leve)
- Conjuntos de variáveis independentes (preditivas): **Temporal (3), Ambiente rodoviário (2), Características da estrada (12) e Características dos pedestres (4).**

# Aplicação

## Análise exploratória dos dados

Variable	Description	Pre-Covid19 (N=9568)	Covid19 (N=3310)	%Δ
<b>Dependent variable</b>				
Injury severity	Unharmed	0.2%	0.1%	-0.1%
	Minor injury	92.6%	91.4%	-1.2%
	Serious injury	6.0%	7.1%	1.1%
	Fatal injury	1.2%	1.4%	0.2%
	Unharmed + Minor (1)	92.8%	91.5%	-1.3%
	Serious + Fatal (2)	7.2%	8.5%	1.3%
<b>Independent variables</b>				
<b>Temporal</b>				
Hour (H)	H: 8-10h*	20.8%	22.7%	1.9%
	H: 11-16h	31.1%	34.0%	2.9%
	H: 17-19h*	31.8%	29.9%	-1.9%
Day of the week (DW)	DW: Weekday*	83.8%	85.6%	1.8%
	DW: Weekend*	16.2%	14.4%	-1.8%
Month (M)	M: Jan-Apr*	36.5%	14.6%	-21.9%
	M: May-Ago	24.9%	31.8%	6.9%
	M: Set-Dec*	38.6%	53.6%	15.0%

# Aplicação

## Análise exploratória dos dados

Independent variables				
<b>Road environment</b>				
<b>Luminosity (L)</b>	L: Daylight	65.8%	72.2%	6.4%
	L: Blinding sun*	2.3%	2.7%	0.4%
	L: Dawn or twilight*	3.6%	3.6%	0.0%
	L: Night, no lighting*	1.7%	1.9%	0.2%
	L: Night, with lighting	26.5%	19.6%	-6.9%
<b>Weather (W)</b>	W: Clear	81%	85.6%	4.6%
	W: Rain*	18.4%	13.9%	-4.5%
Independent variables				
<b>Road characteristics</b>				
<b>Horizontal design (HD)</b>	HD: Straight	83.2%	84.1%	0.9%
	HD: Curve*	16.6%	15.9%	-0.7%
<b>Vertical design (VD)</b>	VD: Level	68.0%	69.3%	1.3%
	VD: Slope + Hump*	31.9%	30.6%	-1.3%
<b>Shoulder (S)</b>	S: Paved shoulder	63.7%	65.4%	1.7%
	S: Unpaved shoulder*	1.2%	1.3%	0.1%
	S: No shoulder or impassable*	34.9%	33.3%	-1.6%

# Aplicação

## Análise exploratória dos dados

Independent variables				
<b>Road characteristics</b>				
Circulation direction (CD)	CD: One way	29.2%	27.3%	-1.9%
	CD: Two ways	70.4%	72.3%	1.9%
Intersection (I)	I: Outside the intersection	72.3%	74.4%	2.1%
	I: 3-leg intersection*	13.8%	13.4%	-0.4%
	I: 4-leg intersection*	7.5%	6.8%	-0.7%
	I: Roundabout*	2.2%	2.5%	0.3%
	VS: Pedestrian crossing	59.3%	57.6%	-1.7%
Vertical signs (VS)	VS: Giving way*	1.4%	1.2%	-0.2%
	VS: Stop*	0.8%	0.8%	0.0%
	HS: With marks separating the direction and lanes of traffic	72.4%	73.7%	1.3%
Horizontal signs (HS)	HS: No marks or barely visible*	27.2%	26.1%	-1.1%
	Light signs (LS)	LS: Operating normally*	12.8%	13.0%
LS: Intermittent*		0.6%	0.7%	0.1%
LS: Off*		1.5%	2.6%	1.1%
LS: Non-existent		84.3%	83.4%	-0.9%
Pavement type (PT)	PT: Asphalt	90.1%	90.8%	0.7%
	PT: Cement*	1.2%	1.5%	0.3%
	PT: Stone*	8.4%	7.6%	-0.8%
Pavement grip (PG)	PG: Dry and clean	78.8%	83.9%	5.1%
	PG: Wet + Humid*	20.5%	15.7%	-4.8%
Pavement condition (PC)	PC: Good	47.8%	46.4%	-1.4%
	PC: Regular	51.4%	53.1%	1.7%
	PC: Poor*	0.5%	0.5%	0.0%
Local speed (LSp)	LSp: 0-30 km/h	2.3%	2.7%	0.4%
	LSp: 31-50 km/h	97.6%	97.2%	-0.4%
	LSp: 51-90 km/h*	0.1%	0.1%	0.0%

# Aplicação

## Análise exploratória dos dados

Independent variables				
<b>Pedestrian</b>				
<b>Age analysis (A)</b>	A: 0-9 years old*	3.6%	4.1%	0.5%
	A: 10-17 years old*	12.3%	11.6%	-0.7%
	A: 18-24 years old*	10.3%	9.2%	-1.1%
	A: 25-49 years old*	22.1%	20.8%	-1.3%
	A: 50-64 years old*	21.6%	22.8%	1.2%
	A: >65 years old*	30.1%	31.5%	1.4%
<b>Sex (Sex)</b>	Sex: Female*	62.0%	60.4%	-1.6%
	Sex: Male*	38.0%	39.6%	1.6%
<b>Type (T)</b>	T: Isolated pedestrian*	86.6%	87.0%	0.4%
	T: Pedestrian in group*	12.4%	11.4%	-1.0%
	T: Moving on skate/scooter/others*	0.2%	0.4%	0.2%
	T: Pushing a bicycle/stroller/people with physical disabilities*	0.4%	0.7%	0.3%
<b>Pedestrian action (PA)</b>	PA: Crossing a marked crossing	76.7%	76.6%	0.3%
	PA: Crossing a marked crossing without respecting traffic lights*	3.7%	3.2%	-0.1%
	PA: Crossing outside a pedestrian crossing, at <50m from a crossing*	19.6%	20.1%	-0.5%

N is the number of cases; variables for which the sum of class percentages is not 100% have cases coded as 'Undefined' or classes not considered due to the reduced number of cases.

\* - Independent variables selected according to what is considered aggravate/contributing condition for accident occurrence.

# Aplicação

## Processo de modelação

- Procedimento MLP (*Multilayer Perceptron*) disponível na opção *Neural Network* do IBM SPSS Statistics 28.0.0.0
- Construção de **6 modelos preditivos para a gravidade das lesões dos pedestres e para a avaliação da importância das variáveis preditivas** (fatores de risco), 3 para o período pré-COVID19 e 3 para o período COVID19:
  - um modelo completo com **todas as variáveis preditivas**
  - um modelo para as variáveis **‘caraterísticas da estrada’**
  - um modelo para as variáveis **‘caraterísticas dos pedestres’**
- **70% do conjunto de dados** foi utilizado para o **treino do modelo** e **30% para o teste**, numa rede com uma camada oculta.

# Resultados

## Modelos

Table 2. Pre-Covid19 models

Model	Complete	Road characteristics	Pedestrian
<b>Training</b>			
% incorrect predictions	7.2%	7.3%	7.1%
% correct classification	92.7%*	92.7%*	92.9%*
<b>Testing</b>			
% incorrect predictions	7.3%	7.1%	7.4%
% correct classification	92.7%*	92.9%*	92.6%*
<b>ROC</b>	0.668 for class 1 0.668 for class 2	0.567 for class 1 0.567 for class 2	0.647 for class 1 0.647 for class 2
<b>Top 10 Independent variable importance</b>	LSp: 51-90 km/h PT: Cement LS: Off T: Moving on skate, scooter or others PC: Poor S: Unpaved shoulder PT: Stone L: Night, no lighting VS: Giving way LSp: 31-50 km/h	LSp: 51-90 km/h VS: Stop PT: Stone PT: Cement S: Unpaved shoulder LSp: 31-50 km/h HS: With marks separating direction / lanes of traffic LSp: 0-30 km/h LS: Intermittent PC: Poor	A: >65 years old T: Pushing a bicycle, stroller or people with physical disabilities A: 50-64 years old T: Pedestrian in group T: Moving on skate, scooter or others PA: Crossing a marked crossing without respecting traffic lights A: 0-9 years old A: 10-17 years old T: Isolated pedestrian PA: Crossing outside a pedestrian crossing, at less than 50m from a crossing

Dependent variable: 'Injury severity', class 1 = Unharmed+Minor, class 2 = Serious+Fatal

\*All cases of the dependent variable classified in class 1

Table 3. Covid19 models

Model	Complete	Road characteristics	Pedestrian
<b>Training</b>			
% incorrect predictions	8.6%	7.8%	8.5%
% correct classification	91.4%*	92.2%*	91.5%*
<b>Testing</b>			
% incorrect predictions	8.2%	9.9%	8.4%
% correct classification	91.8%*	90.1%*	91.6%*
<b>ROC</b>	0.630 for class 1 0.630 for class 2	0.604 for class 1 0.604 for class 2	0.626 for class 1 0.626 for class 2
<b>Top 10 Independent variable importance</b>	T: Moving on skate, scooter or others VS: Stop LS: Off* LSp: 51-90 km/h L: Blinding sun S: Unpaved shoulder PT: Cement PC: Poor L: Night, no lighting LS: Intermittent	PC: Poor LSp: 0-30 km/h I: Roundabout LSp: 51-90 km/h PT: Stone PT: Asphalt LS: Intermittent I: 4-leg intersection S: Unpaved shoulder VS: Stop	T: Pushing a bicycle, stroller or people with physical disabilities A: 18-24 years old PA: Crossing a marked crossing without respecting traffic lights T: Pedestrian in group T: Moving on skate, scooter or others PA: Crossing a marked crossing A: 10-17 years old A: 25-49 years old T: Isolated pedestrian A: 0-9 years old

Dependent variable: 'Injury severity', class 1 = Unharmed+Minor, class 2 = Serious+Fatal

\*All cases of the dependent variable classified in class 1

# Conclusão

## Principais resultados e Trabalho futuro

- Modelos com **elevada percentagem de classificações** corretas mas **fraco poder discriminativo**
- Possíveis **razões** para o fraco poder discriminativo:
  - As lesões dos pedestres são essencialmente do tipo 'ferido leve' (cerca de 92%) - conjunto de dados enviesado
  - Pouca diferenciação entre os fatores que contribuem para os acidentes envolvendo 'feridos graves + vítimas mortais' e 'ilesos + feridos leves'
- Coerência entre modelos sobre os principais tipos de intervenções que podem reduzir a gravidade deste tipo de acidentes
  - Melhorar o funcionamento da **sinalização luminosa nas travessias pedonais**
  - Melhorar a **iluminação noturna das travessias pedonais**
  - Melhorar as **bermas** das estradas
  - Melhorar o **estado do pavimento**
- Deve ser dada especial atenção às **zonas urbanas com limites de velocidade mais elevados** (51-90 km/h) e à sensibilização dos pedestres que se deslocam em **modos suaves/ativos**

# Conclusão

## Principais resultados e Trabalho futuro

- Comparação dos cenários pré-Covid19 e Covid19:
    - **Diminuição significativa do número anual de acidentes** (cerca de 30-35%) no período Covid-19
    - Classes etárias dos pedestres **vítimas mortais**:
      - As classes críticas no período **pré-Covid19** eram **50-64 e >65 anos**
      - As vítimas mortais das classes **18-24 e 25-49 anos** aumentaram no **período Covid-19** (os mais velhos ficaram em casa!)
  - Trabalho futuro
    - Avaliação de outros *softwares* e algoritmos de *machine learning* que acomodem melhor as variáveis categóricas e a distribuição desequilibrada das classes da variável dependente
- WEKA, Scikit-learn, R e Matlab**
- Decision tree (DT), Random forest (RF), Artificial neural network (ANN) e Support vector machine (SVM)**

# Avaliação de acidentes em travessias pedestres para identificação de fatores de risco utilizando redes neurais artificiais

# Obrigada!

Bertha Santos  
[bsantos@ubi.pt](mailto:bsantos@ubi.pt)



Programa

