Journal of Forensic and Legal Medicine 84 (2021) 102256

Contents lists available at ScienceDirect

Journal of Forensic and Legal Medicine

journal homepage: www.elsevier.com/locate/jflm

Research paper

A pilot study for investigating the feasibility of supervised machine learning approaches for the classification of pedestrians struck by vehicles

Michelangelo Casali [a,d,*], Dario Malchiodi [b,c], Claudio Spada [d,e], Anna Maria Zanaboni [b], Rosy Cotroneo [d,e], Domenico Furci [d,e], Andrea Sommariva [d], Umberto Genovese [a,d], Alberto Blandino [d,e]

[a] Department of Oncology and Hemato-Oncology (DIPO) University of Milan, Via Luigi Mangiagalli 37, 20133 Milan, Italy
[b] Department of Computer Science & Data Science Research Centre, University of Milan, Via Celoria 18, 20133 Milan, Italy
[c] CINI National Laboratory of Artificial Intelligence and Intelligent Systems (AIIS), University of Rome, 00185 Rome, Italy
[d] Healthcare Accountability Lab, University of Milan, Via Luigi Mangiagalli 37, 20133 Milan, Italy
[e] Department of Biomedical Sciences for Health, Institute of Forensic Medicine, University of Milan, Via Luigi Mangiagalli 37, 20133 Milan, Italy

ARTICLE INFO

Keywords:
Traffic collision
Injury pattern
Autopsy
Classification
Supervised machine learning
Artificial Intelligence-based forensics

ABSTRACT

This research focuses on the application of Artificial Intelligence (AI) methodologies to the problem of classifying vehicles involved in lethal pedestrian collisions. Specifically, the vehicle type is predicted on the basis of traumatic injury suffered by casualties, exploiting machine learning algorithms. In the present study, AI-assisted diagnosis was shown to have correct prediction about 70% of the time. In pedestrians struck by trucks, more severe injuries were appreciated in the facial skeleton, lungs, major airways, liver, and spleen as well as in the sternum/clavicle/rib complex, whereas the lower extremities were more affected by fractures in pedestrians struck by cars. Although the distinction of the striking vehicle should develop beyond autopsy evidence alone, the presented approach which is novel in the realm of forensic science, is shown to be effective in building automated decision support systems. Outcomes from this system can provide valuable information after the execution of autoptic examinations supporting the forensic investigation. Preliminary results from the application of machine learning algorithms with real-world datasets seem to highlight the efficacy of the proposed approach, which could be used for further studies concerning this topic.

1. Introduction

One of the leading causes of death for people of all ages is represented by road traffic collisions, with circa 1.35 million road fatality deaths occurring globally each year. Pedestrians, accounting for 23% of road deaths, constitute the most vulnerable road users.[1–5] Forensic examination of vehicle–pedestrian collisions has become increasingly important in the detection, investigation and reduction of road casualties; however, post-collision identification of the striking vehicle based only on data from the pedestrian's autopsy (i.e., injuries and demographic information), is often challenging.[1,3–6] For this reason, any basic information can be useful to discriminate between striking car versus truck when the striking vehicle is unknown.

The pedestrian injuries caused by vehicles are complex and can affect almost the entire body surface, even if the most frequent ones are injuries involving the head and the lower extremities.[7–9] Available literature contributions mainly focus on the diagnostic meaning of suggestive single wounds (although the type and the distribution of external injuries may be affected by different variables,[10,11] such as the shape of the vehicle and the possibility of an impact against protruding surfaces of the vehicle), highlighting the limit for a practical feasibility for ex-post identification of the striking vehicle using injury pattern alone. But in recent years, several studies based on artificial intelligence (AI) have been conducted in the forensic field, posing new challenges and demonstrating the feasibility and advantages of using AI methodologies to solve forensic identification problems.[12]

The aim of this work was to evaluate and analyse the injury pattern of pedestrians deadly struck by vehicles, in order to generate a dataset to be processed by specific supervised machine learning (ML)

* Corresponding author at: Department of Oncology and Hemato-Oncology (DIPO) University of Milan, Via Luigi Mangiagalli 37, 20133 Milan, Italy.
E-mail addresses: michelangelo.casali@unimi.it (M. Casali), dario.malchiodi@unimi.it (D. Malchiodi), claudio.spada@unimi.it (C. Spada), annamaria.zanaboni@unimi.it (A.M. Zanaboni), rosy.cotroneo@unimi.it (R. Cotroneo), domenico.furci@unimi.it (D. Furci), andrea.sommariva@unimi.it (A. Sommariva), umberto.genovese@unimi.it (U. Genovese), alberto.blandino@unimi.it (A. Blandino).

https://doi.org/10.1016/j.jflm.2021.102256
Received 17 June 2021; Received in revised form 17 September 2021; Accepted 27 September 2021
Available online 9 October 2021
1752-928X/

algorithms for the post-autoptic distinction between pedestrians struck by a car and by a truck.

The application of AI methods focusing on classifiers directly induced from data via ML algorithms is relatively new. In general, the literature contains the following threads:

- Automated analysis of the textual content of an autopsy report, with the aim of predicting the corresponding outcome,[13–15] and
- Joint evaluation of crash data and environmental conditions in order to increase road safety.[16–20]

In the best of our knowledge, we could not find any work applying supervised ML to predict the type of striking vehicle starting from a record of numerical values describing the pedestrian struck by vehicles; a few papers use similar approaches, although focusing on different problems (e.g., inducing decision trees in order to discriminate between perimortem and postmortem cranial fractures[21]).

This paper is organized as follows: Section 2 describes the dataset and the ML models and algorithms used in the study, while Section 3 focuses on the overall analysis and processing phases of the performed experiments. The results are discussed in Section 4 followed by conclusions in Section 5, while two detailed tables are provided as supplementary material.

2. Materials and methods

Cases of lethal pedestrian struck by car or truck were collected from the 2000–2019 archive of the Institute of Legal Medicine of the Milan University, Italy.

The case selection was retrospectively continuous and stopped after getting a 60 cases sample for the truck subgroup. The inclusion criteria were defined as follows:

- Pedestrian age ≥ 18;
- No post-traumatic survival period;
- Clearly identified striking vehicle;
- Single-vehicle collision;
- Impact on front side of the vehicle;
- Urban-setting.

The striking vehicles were divided into two subgroups: cars versus commercial trucks. Being a pilot study evaluating the feasibility of applying supervised machine learning approaches, it was decided to evaluate initially whether there was a possibility of a generic distinction between pedestrians hit by car and pedestrians hit by truck, without further subcategories of vehicles that would have complicated the picture. It is believed that as a first result, in the investigation phase with unknown vehicle, bringing useful elements to discern the category of vehicle (car vs truck) can be of address for the investigation. Vehicles having a maximum mass not exceeding 3.5 tonnes were categorized as cars, while vehicles having a maximum mass exceeding 3.5 tonnes and showing a flat front-side were classified as trucks. The collected experimental variables are age, sex, height, weight, BMI, and injury score.[4,5,22] Autoptic data were summarized through the use of a synoptic visceral score already published for fatal falls from heights[23] (see supplementary material S1) and of an independent detailed table for bone fractures (see supplementary material S2). Both these autoptic tools show a marked descriptive structure, in contrast with previously available clinical injury scores.

2.1. Learnt classifiers

The supervised ML models used in the experiments detailed in Section 3 are briefly listed here below. A more detailed description is contained in the supplementary material.

- *Naive Bayes classifiers*, exploiting the assumption of probabilistic independence between attributes and the Bayes' theorem in order to find the class to which observations belong with maximal probability.[24]
- *Linear Discriminant Analysis* (LDA), assuming a multivariate Gaussian distribution for each class and processing observations by thresholding a linear combination of their attributes.[24]
- *Decision trees* (DT), implementing a series of queries hierarchically organized in a tree[25] and *symbolically* explaining the classification process through *decision rules*.
- *Random forests* (RF), basically consisting in sets of decision trees induced by repeatedly sub-sampling data[26] and queried by aggregating the results of the single trees, e.g., through majority vote.
- *Multi-layer perceptrons* (MLP), inspired by the human central nervous system[27] and organized as a feed-forward hierarchy of fully-connected layers of *artificial neurons*, each computing a nonlinear function of a weighted sum of their inputs. In this architecture, proven to act as a universal approximator under specific conditions[28], the weights of the various artificial neurons are inferred from data via the backpropagation algorithm[29], or one of its variants.
- *Support vector classifiers* (SVC), optimizing the minimal distance between data and decision surface, so as to increment generalization.[30] The related optimization process exploits the so-called *kernel trick* in order to tackle nonlinearities in the learning process.

3. Case study

The study is based on a retrospective dataset of 130 lethal pedestrians struck by vehicles (i.e. car or truck). Each observation is described by 367 features: date of collision, demographic data, suffered traumas, and what is to be considered the outcome in our study, namely the class of hitting vehicle. The descriptions of traumas are organized at different granularities: at a higher level, they are coded through integer values (expressing a severity score in the 0–4 range) organized within the head, thorax, abdomen, and skeleton macro-districts (see supplementary material); at a more detailed scale, they are stored using boolean values (presence/absence of trauma).

3.1. Empirical research results

Given the high ratio of number of features to number of observations, in order to learn a robust model it was necessary to reduce the considered number of attributes in our problem. For this reason, in this study we only considered demographic and higher-level severity features, leaving for a future work the exploitation of features carrying information about lesions at a finer-grained level. As a result, in all our experiments we initially considered less than 30 features.

3.1.1. Descriptive analysis results

The dataset was well balanced with respect to the outcome: 70 (53.8%) people were struck by a car and 60 (46.2%) by a truck. There were 77 males (59.2%) and 53 (40.8%) females. The youngest pedestrian was 19 years old and the oldest was 93. Demographic data are shown in Table 1, summarizing information by mean and standard deviation when normality hypotheses for the involved distributions hold, using median and inter-quartile distance otherwise; there were not statistically significant differences in demographic data between the two groups (cars and trucks), except for sex (33% women in the car group and 50% women in the truck group, p-value 0.05). The severity of traumas is illustrated in Table 2. We compared the severity score distribution in the two groups by Mann–Whitney U test. The following comparisons where statistically significant: *HEAD/Facial skeleton* (p-value 0.04), *THORAX/Lungs* (p-value 0.02), *THORAX/Trachea and Bronchi*

Table 1
Demographic variables.

	Position	Dispersion
Age (years)	67	31
Height (m)	1.66	0.13
Weight (kg)	73.4	14.4
BMI	26.53	4.66

Table 2
Lesion severity (score value in 0–4 range). Symbol ⋆ highlights a significant difference (Mann–Whitney U test for equality of distributions, at 0.05 significance level) between the two groups of cars and trucks.

			Mean severity			
			ALL	CAR	TRUCK	*p*-value
HEAD		Skull	1.61	1.49	1.75	
	⋆	Facial skeleton	0.71	0.43	1.03	0.04
		Cerebrum	1.18	1.11	1.25	
		Cerebellum	0.45	0.34	0.58	
		Brainstem	0.62	0.41	0.85	
THORAX	⋆	Lungs	1.21	0.9	1.57	0.017
	⋆	Trachea and Bronchi	0.3	0.01	0.63	<0.001
		Heart	0.53	0.36	0.73	
		Thoracic Aorta	0.65	0.46	0.88	
		Diaphragm	0.42	0.31	0.55	
ABDOMEN	⋆	Liver	1.08	0.8	1.42	0.029
	⋆	Spleen	0.59	0.26	0.98	<0.001
		Abdominal Aorta	0.12	0.04	0.2	
		Kidneys	0.28	0.23	0.35	
		Mesentery	0.26	0.19	0.35	
SKELETON		Cervical Spine	0.42	0.44	0.4	
		Thoracic Spine	0.87	0.7	1.07	
		Lumbar Spine	0.07	0.09	0.05	
		Pelvis	1.25	1.17	1.35	
	⋆	Complex sternum/clavicle/ribs	2.88	2.41	3.42	<0.001

(*p*-value < 0.001), *ABDOMEN/Liver* (*p*-value 0.03), *ABDOMEN/Spleen* (*p*-value <0.001), *SKELETON/Complex sternum/clavicle/ribs* (*p*-value < 0.001) (see Table 2).

The fine-grained detailed mapping of skeletal lesions consists of 298 boolean features, whose aggregate description is shown in Table 4.

As far as the collision date is concerned, we are aware that our data suffer from some bias, due to the length of the time span considered, as shown in Table 5 (t-test or U-test as appropriate). In particular, after 2008 there were fewer cases of trucks. However, as already mentioned, in order to have a well balanced dataset, we preferred to keep older observations rather than applying data augmentation techniques to more recent data.

3.1.2. Dimensionality reduction results

The first step in the used ML pipeline is devoted to dimensionality reduction, considering the following techniques:

- Principal Component Analysis (PCA)[31], based on the definition of an alternative coordinate system for the observation space, whose axes correspond to the eigenvectors of the data covariance matrix, sorted w.r.t. eigenvalues: this ensures that data projected in, say, ith component has a variance higher than in all components $j > i$;
- Truncated Singular Value Decomposition (SVD)[32], obtained by setting to zero the less relevant singular values in a special eigendecomposition of the data matrix;
- Fast Independent Component Analysis (FastICA)[33], aiming at mapping observed data to lower-dimensional vectors maximizing a proxy for the statistical independence of the latter components.

Each technique has been applied to all models, testing the extraction of 2, 5, 10, 15, and 20 components.

Table 3
Percentage of cases with lesion severity equal or greater than one. Significative differences (Wald Z test for equality of proportions, at 0.05 significance level) between the two groups of cars and trucks are highlighted by symbol ✓.

			Severity score ≥ 1			*p*-value
			ALL %	CAR %	TRUCK %	
HEAD		Skull	64	67	62	
		Facial skeleton	31	24	38	
		Cerebrum	47	51	42	
		Cerebellum	16	14	18	
		Brainstem	23	21	25	
THORAX		Lungs	60	56	65	
	✓	Trachea and Bronchi	10	1	20	<0.001
		Heart	31	28	35	
		Thoracic Aorta	21	16	28	
		Diaphragm	16	13	20	
ABDOMEN		Liver	41	34	50	
	✓	Spleen	21	10	35	<0.001
		Abdominal Aorta	3	1	5	
		Kidneys	11	7	15	
		Mesentery	12	13	12	
SKELETON		Cervical Spine	18	20	15	
		Thoracic Spine	36	30	43	
		Lumbar Spine	5	6	3	
		Pelvis	51	53	50	
	✓	Complex sternum/clavicle/ribs	83	76	92	0.016

Table 4
Detailed skeleton lesions. Top: mean number of lesions; significative differences (at 0.05 level) between the two groups (Mann–Whitney U test for equality of distributions) are flagged by the symbol ⋆. Bottom: percentage of cases with one or more lesions; significative differences (at 0.05 level) between the two groups (Wald Z test for equality of proportions) are flagged by the symbol ⋆.

	Mean number of lesions			
	ALL	CAR	TRUCK	*p*-value
Skull	4.3	3.26	5.52	
Rachis	0.95	0.73	1.2	
Thorax–rib cage ⋆	29.9	17.7	44.1	<0.001
Pelvis	1.7	1.49	2.0	
Upper limb	0.6	0.68	0.53	
Lower limb ⋆	1.0	1.2	0.87	0.025
	At least one lesion(%)			
	ALL	CAR	TRUCK	*p*-value
Skull	72	74	70	
Rachis	45	42	48	
Thorax–rib cage ⋆	83	75	91	0.008
Pelvis	51	53	50	
Upper limb	30	32	27	
Lower limb ⋆	50	54	31	0.005

Table 5
Comparison of observations between before and after 2008.

	<2008 ($n = 39$)	≥2008 ($n = 91$)	*p*-value
Age (years)	67.3	62.2	
Height (m)	1.66	1.66	
Weight (kg)	67.9	75.8	0.001
BMI	24.6	27.4	0.002
Women	51%	36%	
Truck	79%	32%	< 0.001

3.1.3. Learning results

As highlighted in the introductory section, the performed experiments have the purpose of investigating the feasibility of ML methods for the problem of classifying pedestrians struck by vehicles. This investigation was done comparing the performances of different classifiers induced from the dataset described in Section 2. As customary when dealing with ML, we carefully handled the phases of model selection

Table 6
Values considered for the model selection phase of each considered model.

Model	Hyper-parameters	Considered values
DT/RF	Purity index	Gini, Entropy
	Max. leaf number	$2, 5, 10, 50, 100$
	Max. features[a]	$\sqrt{n}, \log_2 n$
	Max. tree depth	$2, 5, 10$
	Num. of trees	$5, 10, 50, 100, 200$ (only for RF)
MLP	Hidden layers[b]	$[2], [5], [10], [20], [4, 4], [10, 10]$
SVC	Tradeoff constant	$LS(-4, 3, 10)$[c]
	Kernel	linear, polynomial, RBF[d]

[a]Here, n denotes the number of used features.
[b]Each list of integer numbers specifies a layer configuration. These numbers describe the number of neurons in hidden layers, sorted according to their position in the feed-forward architecture.
[c]$LS(a, b, n)$ denotes the set containing n elements ranging from 10^a to 10^b evenly distributed in logarithmic space.
[d]Polynomial kernels depend on a parameter p whose values during model selection were taken from $\{2, 3, 5, 9\}$; analogously for RBF kernels, whose parameter σ was selected in $LS(-4, 3, 10)$.

(intended as the act of tuning the hyper-parameters of the used learning algorithm) and model assessment (defined as an estimation of the ability of learnt models to generalize to unseen data). Both issues were faced using a cross-validation (CV) procedure, consisting in partitioning the available data in k equally sized *folds* and running the learning algorithm on all folds except a fixed one, whose elements are subsequently used in order to measure the performance of the learnt classifier. The procedure is repeated k times, excluding at each iteration a different fold from the training phase and suitably aggregating (e.g., averaging) the k obtained performance measures.

All experiments were carried out using a common pipeline consisting in:

1. Scaling objects in the dataset;
2. Performing a nested CV exploiting an outer 7-fold CV devoted to assess the generalization ability, and an inner 5-fold CV used in order to perform a grid search-based model selection.[1]

The accuracy score was used as quantity to be optimized during model selection. Model selection concerned the scaler used in step 1 of the pipeline (precisely, we considered two standard normalization procedures either using mean and standard deviation, or maximal and minimal observations, a robust scaler detecting possible outliers, and a quantile-based scaler), and hyper-parameters specific to the tested models, as detailed in Table 6.

For the estimation of the generalization ability we preferred accuracy to more sophisticated metrics such as precision, recall or F1-score, since the classes under study are well balanced.

We also performed a manual feature selection, separately considering all subsets of the height, weight and BMI attributes (which are obviously interdependent). Reasonably, the use of the sole BMI turned out to be the best performing option. It is worth noting that in a companion experiment we also considered the collision date, obtaining better results than those illustrated here below. However, we prefer to exclude them in light of the considerations done in Section 3.1.1.[2]

The above mentioned pipeline has been executed considering all combinations of dimensionality reduction technique and inferred model. As stated in Section 3.1.2, we applied the PCA, truncated SVD and FastICA techniques and extracted 2, 5, 10, 15, and 20 components; besides that, we also processed the dataset in its original form (that is, without dimensionality reduction). The considered ML models are those detailed in Section 2.1. The results are summarized in Tables 7–9, showing the mean classification accuracy on the excluded folds in the outer CV (henceforth referred to as *test accuracy*). The results exhibit a remarkable variability, however accuracies are high enough to ensure that all inferred models perform at least better than guessing at random,[3]. Moreover, Table 10 shows the mean and standard deviation of test accuracy obtained in all experiments done within the above mentioned feature selection phase, supporting the hypothesis that the best results, attained by Naive Bayes classifiers relying on 10 and 15 FastICA components, were not obtained by chance. This is also in line with the fact, illustrated in Fig. 1 that extracting 12 principal components from the overall original dataset explains the 99% of its total variance.

Table 7
Accuracies obtained using PCA. Rows and columns correspond, respectively, to ML models and number of extracted components (∞ means no dimensionality reduction). Entries in boldface highlight the best results in each column.

	2	5	10	15	20	∞
SVC (linear)	0.64	**0.67**	0.63	0.65	**0.66**	0.62
SVC (RBF)	0.62	0.66	0.61	0.63	0.65	0.67
SVC (polynomial)	0.58	0.59	0.62	0.62	0.64	**0.69**
NB	0.63	0.62	**0.68**	**0.68**	0.65	0.68
LDA	**0.67**	0.67	0.66	0.64	0.63	0.61
DT	0.62	0.63	0.60	0.58	0.54	0.60
RF	0.56	0.55	0.64	0.61	0.65	0.66
MLP	0.65	0.65	0.65	0.65	0.62	0.60

Table 8
Accuracies obtained using FastICA. Same notations as in Table 7.

	2	5	10	15	20	∞
SVC (linear)	0.65	0.65	0.65	0.65	0.63	0.62
SVC (RBF)	0.64	0.54	0.67	0.62	**0.69**	0.67
SVC (polynomial)	0.58	0.61	0.62	0.65	0.68	**0.69**
NB	0.63	0.59	**0.72**	**0.73**	0.68	0.68
LDA	**0.67**	**0.67**	0.66	0.64	0.63	0.61
DT	0.55	0.58	0.61	0.56	0.54	0.60
RF	0.66	0.56	0.64	0.71	0.64	0.66
MLP	0.58	0.61	0.68	0.65	0.65	0.60

Table 9
Accuracies obtained using SVD. Same notations as in Table 7.

	2	5	10	15	20	∞
SVC (linear)	0.62	**0.65**	0.62	0.61	0.66	0.62
SVC (RBF)	0.61	0.57	0.63	0.61	0.65	0.67
SVC (polynomial)	0.58	0.59	0.58	0.64	0.62	**0.69**
NB	0.63	0.60	0.61	**0.69**	**0.66**	0.68
LDA	**0.65**	0.64	**0.65**	0.62	0.63	0.61
DT	0.55	0.56	0.61	0.57	0.59	0.60
RF	0.56	0.62	0.59	0.63	0.65	0.66
MLP	0.58	0.61	0.61	0.58	0.63	0.60

4. Discussion

AI is contributing to accelerate the translation of research to clinical practice enhancing a real revolution in the biomedical domain, leading to improved clinical diagnoses achieving precision medicine results.[34–37] However, forensic applications of artificial intelligence and deep learning techniques have been adopted at relatively slow pace and largely focused on subfields other than forensic pathology.[38]

[1] The term *grid search* denotes an exhaustive search in the space of all configurations described by the Cartesian product of sets of tentative values for the hyper-parameters.

[2] Note also that in light of a future implementation of a decision support system to be used by forensic experts, it would not be reasonable to use collision date as a predictor.

[3] That is, tossing a coin governed by probabilities equal to the fraction of cars and trucks in the dataset.

Table 10
Mean and standard deviation of test accuracies obtained using all dimensionality reduction and feature selection procedures (same notations as in Table 7). The last line highlights the best results for a given number of extracted components.

	2	5	10
SVC (linear)	0.64±0.02	0.65±0.01	0.63±0.02
SVC (RBF)	0.60±0.02	0.58±0.04	0.62±0.02
SVC (polynomial)	0.58±0.03	0.60±0.02	0.63±0.03
NB	0.63±0.01	0.63±0.02	0.65±0.03
LDA	0.67±0.03	0.64±0.02	0.66±0.01
DT	0.57±0.03	0.58±0.03	0.60±0.03
RF	0.61±0.03	0.59±0.03	0.63±0.04
MLP	0.63±0.03	0.60±0.03	0.63±0.02
max	0.70	0.67	0.72
	15	20	∞
SVC (linear)	0.63±0.02	0.64±0.01	0.64±0.01
SVC (RBF)	0.64±0.02	0.66±0.02	0.67±0.02
SVC (polynomial)	0.64±0.03	0.67±0.02	0.68±0.01
NB	0.68±0.03	0.67±0.02	0.67±0.01
LDA	0.65±0.02	0.64±0.02	0.62±0.02
DT	0.58±0.03	0.57±0.04	0.63±0.02
RF	0.63±0.03	0.65±0.03	0.65±0.02
MLP	0.63±0.03	0.63±0.03	0.59±0.01
max	0.73	0.71	0.69

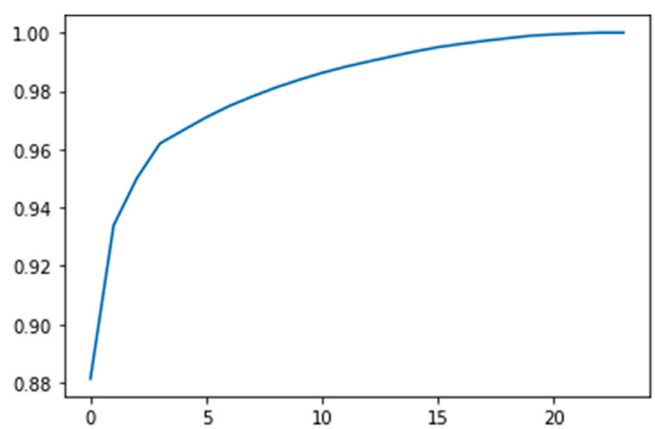

Fig. 1. Graph of the explained total variance of the original dataset when reducing dimensionality using PCA. The 99% of total variance is reached when using 12 components.

In this research, supervised ML has been applied in the field of forensic pathology as one of the first practical tests for AI-assisted diagnosis. The ex-post experimental diagnostic test aims at identifying the striking vehicle as car or truck based on the autoptic data.

The theoretical feasibility for such a distinctive diagnosis comes from preventable differences in both the kinetic energy of the striking vehicle ($E = \frac{1}{2}mv^2$) and the shape of the front side of the striking vehicle itself (pointed front-sides for cars prone to lift the hit pedestrian versus flat front-sides for trucks prone to simply project forward the hit pedestrian).

According to the injury score (see supplementary material S1), pedestrians struck by trucks show statistically worse injuries in facial skeleton, lungs, major airways (trachea and bronchi), liver and spleen, as well as in the sternum/clavicle/ribs complex (Table 3). In particular, it is noteworthy that in pedestrians hit by cars injuries to the trachea and bronchi are almost absent, unlike pedestrians hit by trucks where such injuries are significantly more frequent, although limited to one fifth of the total cases. Similarly, injuries to the spleen appeared significantly less present in pedestrians struck by cars. On the other hand, injuries to the sternum/clavicle/ribs complex are very frequent in both cases, with at least one fracture in 76% of pedestrians hit by a car and 92% of those hit by a truck: this difference resulted however statistically significant (p-value 0.016). The performed preliminary skeletal analysis (Table 4) showed a greater number of fractures in the truck subgroup [mean fracture values: 54.5 versus 25.0], with the widest gap coming from the rib cage [mean fracture values: 44.1 versus 17.7 – p-value < 0.001]. Interestingly, more fractured foci in the lower limbs were observed in pedestrians hit by cars (1.2 vs 0.87, difference resulted statistically significant – p-value 0.025), possibly explained by the different conformation of the vehicle. However, the detailed analysis of the skeletal injury pattern in the deceased pedestrians will be the aim of a further AI-guided work. Despite so, the authors believe that data shown so far could contribute to a better understanding of the effects of the type of the striking vehicle on the final injury pattern, although, as for previous data analysis[4,5], the issue is still controversial.

Focusing on the best performances of the various enrolled AI techniques, in our study the overall accuracy for the AI-based diagnoses turned out to range from 0.65 to 0.73. The best prediction was made using 10 to 15 experimental features extracted using the FastICA technique. Simply speaking, the AI-assisted diagnosis proved to be able to generalize, that is suggesting a correct prediction for unseen observations, 70% of the times.

As a starting point, such a success-rate can be considered at least as promising, although the distinction of the striking vehicle should clearly develop beyond autopsy evidence alone, considering also engineering aspects and the biological and non-biological traces transferred to the pedestrian by the vehicle. In the same way, the speed of the vehicle at the moment of impact could also be relevant, however in the present study, dealing with accidents where the type of vehicle involved is unknown, the speed itself must also be considered unknown: for this reason it has not been considered. However, analysis of the injury pattern would seem to be able to provide useful elements in the investigation.

Artificial intelligence applications may be subject to human-like bias through the way they process datasets and assess events.

From a legal perspective, to ensuring trustworthiness in fact-finding, it is too simplistic to introduce machine evidence as some form of documentary evidence or relate it to testimonial evidence, since the use of artificial intelligence to generate robust and comprehensible evidence remains problematic, and it is therefore challenging to examine results from AI application similarly to human witnesses in the courtroom.[38,39] With regard to this, we emphasize that models that only estimate outcomes, using either artificial or human intelligence, should be considered with caution and preferentially implemented in research studies to evaluate additional possible sources of information for investigations, rather than being considered reliable tools during a trial.[40]

There are, of course, some limitations in the present work, that we consider tolerable in a pilot study context.

First of all, in the selection of cases to consider for our analysis, we introduced some temporal bias, in order to achieve a fair label balancing, as mentioned in Section 3.1.1.

As a second weakness we mention the sample size, which *per se* is not so small, but, given the difficulty of the considered classification problem, is for sure inadequate; the possibility of collecting new data is likely to improve the obtained results, and could make it feasible to exploit a richer set of features.

In fact, a third drawback of the current study is that it only relies on a part of the information contained in the complete dataset, as it exploits only injury scores regarding macro districts rather than a much more detailed description of lesions already available. As briefly mentioned in Section 3, these richer set of data deserve a specific consideration, which shall however deal with an even more amplified ratio between features and observations.

The outlined drawbacks are related to the well known bias–variance trade-off: a large set of features ensure a large variance, which is desirable in order to have a good description of the phenomenon at study; on the other end a small bias is desirable, which directly depends on the ratio of the number of features to the sample size. Therefore, if the sample size is given, an increase in the number of features implies an increase also of bias. On the other side, if the number of features describing the problem is given, it is necessary to increase also the sample size in order to keep bias low. For these reasons we decided, only in the present feasibility study, to enlarge the data set as much as we could by the inclusion of some observations taken from an earlier point in time (which also guaranteed a well balanced sample), and to consider the largest number of features which was acceptable from the common sense point of view and did not penalize bias too much.

Lastly, most of the considered ML techniques belong to the so-called *black-box* methodologies. Simply put, they cannot provide any argumentation about the reasons behind the predictions given as output. Furthermore, the use of dimensionality reduction obfuscates also the predictions of decision trees and random forests, whose results are—at least in theory—comprehensible by human beings. Clearly, this represents a further limitation to the actual use of the proposed methodology by forensic experts, in view of a reasonable lack of trustworthiness.[41]

5. Conclusion

The aim of this work was to evaluate the possibility of distinction between pedestrians hit by a car or truck, using only autopsy information processed by specific supervised machine learning. Our result had a correct prediction of about 70%. This result can be considered at least promising for future similar developments. Clearly, however, this type of model can currently only be considered in pilot studies to obtain a guide for investigations rather than give a technically valid answer during a trial. 70% accuracy has not enough probative value, but it may have an investigative value, in order to direct investigations searching for the striking vehicle. Indeed, as a pilot study, this work could contribute to open the way for further experimental activities to explore the wider panel of the application of AI techniques in the forensic field.

Based on the evidence from this study, we recommend that further application of ML to predict the type of striking vehicle will be conducted to overcome the limits of the present research exploiting features carrying information about lesions at a finer-grained level.

Since uncertainty is an ineluctable issue in forensic decision making, learning algorithms that allow for *soft* (as opposed to *hard*) boundaries of decision surfaces could improve the overall performance of the system: *fuzzy*[42] classification methods should be investigated for their application in this domain.

Furthermore, an extension of the study, focusing on the theme of *explainable AI*[43], should be considered, in line with similar trends in the medical fields[44].

Human and animal rights and informed consent

The research described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). This study did not involve laboratory animals, living animals. No human were harmed for this study, as all of them had already died and were enrolled for this study only afterwards.

CRediT authorship contribution statement

Michelangelo Casali: Conceptualization, Writing – original draft, Writing – review & editing. **Dario Malchiodi:** Conception and design of learning pipelines, Literature review, Training and validation of ML models, Analysis of results, Writing – review & editing. **Claudio Spada:** Conceptualization, Literature review, Data curation, Writing – review & editing. **Anna Maria Zanaboni:** Conception and design of learning pipelines, Literature review, Statistical analysis, Analysis of results, Writing – review & editing. **Rosy Cotroneo:** Literature review, Writing – review & editing. **Domenico Furci:** Literature review, Writing – review & editing. **Andrea Sommariva:** Literature review, Data curation. **Umberto Genovese:** Supervision. **Alberto Blandino:** Conceptualization, Literature review, Data curation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All relevant aggregated data as well as the results of the statistical analysis are included within the paper. In case of need, data and software applications are available upon reasonable request to the authors.

Acknowledgement

All authors have approved the final manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jflm.2021.102256.

References

1. Nogayeva S, Gooch J, Frascione N. The forensic investigation of vehicle–pedestrian collisions: A review. *Sci Justice*. 2021;61:112–118. http://dx.doi.org/10.1016/j.scijus.2020.10.006.
2. World Health Organization. *Global Status Report on Road Safety 2018*. Geneva: World Health Organization; 2018 https://www.who.int/violence_injury_prevention/road_safety_status/report/en/. [Accessed 22 May 2021].
3. Ni Y, Wang M, Sun J, Li K. Evaluation of pedestrian safety at intersections: A theoretical framework based on pedestrian-vehicle interaction patterns. *Accid Anal Prev*. 2016;96:118–129. http://dx.doi.org/10.1016/j.aap.2016.07.030.
4. Aleassa EM, Eid HO, Abu-Zidan FM. Effects of vehicle size on pedestrian injury pattern and severity: Prospective study. *World J Surg*. 2013;37:136–140. http://dx.doi.org/10.1007/s00268-012-1797-4.
5. Martin JL, Lardy A, Laumon B. Pedestrian injury patterns according to car and casualty characteristics in France. *Ann Adv Automot Med*. 2011;55:137–146. https://pubmed.ncbi.nlm.nih.gov/22105391/. [Accessed 22 May 2021].
6. Burke M. *Forensic Medical Investigation of Motor Vehicle Incidents*. Australia: CRC Press; 2007.
7. Halari MM, Shkrum MJ. Fatal motor vehicle-pedestrian collision injury patterns–A systematic literature review. *Acad Forensic Pathol*. 2020;10:144–157. http://dx.doi.org/10.1177/1925362120986059.
8. Arregui-Dalmases C, Lopez-Valdes FJ, Segui-Gomez M. Pedestrian injuries in eight European countries: An analysis of hospital discharge data. *Accid Anal Prev*. 2010;42:1164–1171. http://dx.doi.org/10.1016/j.aap.2010.01.005.
9. Moharamzad Y, Taghipour H, Firoozabadi NH, et al. Mortality pattern according to autopsy findings among traffic accident victims in Yazd, Iran. *Chin J Traumatol*. 2008;11:329–334. http://dx.doi.org/10.1016/s1008-1275(08)60067-x.
10. Ehrlich E, Tischer A, Maieiner H. Lethal pedestrian–Passenger car collisions in Berlin. Changed injury patterns in two different time intervals. *Leg Med (Tokyo)*. 2009;11(Suppl. 1):S324–6. http://dx.doi.org/10.1016/j.legalmed.2009.01.017.
11. Eid HO, Abu-Zidan FM. Biomechanics of road traffic collision injuries: a clinician's perspective. *Singapore Med J*. 2007;48:693–700. http://smj.sma.org.sg/4807/4807pe1.pdf. [Accessed 22 May 2021].
12. Fang YT, Lan Q, Xie T, Liu YF, Mei SY, Zhu BF. New opportunities and challenges for forensic medicine in the era of artificial intelligence technology. *Fa Yi Xue Za Zhi*. 2020;36:77–85. http://dx.doi.org/10.12116/j.issn.1004-5619.2020.01.016.
13. Yeow WL, Mahmud R, Raj RG. An application of case-based reasoning with machine learning for forensic autopsy. *Expert Syst Appl*. 2014;41:3497–3505. http://dx.doi.org/10.1016/j.eswa.2013.10.054.
14. Mujtaba G, Shuib L, Raj RG, Rajandram R, Shaikh K, Al-Garadi M. Automatic ICD-10 multi-class classification of cause of death from plaintext autopsy reports through expert-driven feature selection. *PLoS One*. 2017;12:e0170242. http://dx.doi.org/10.1371/journal.pone.0170242.
15. Mujtaba G, Shuib L, Raj RG, Rajandram R, Shaikh K. Prediction of cause of death from forensic autopsy reports using text classification techniques: A comparative study. *J Forensic Legal Med*. 2018;57:41–50. http://dx.doi.org/10.1016/j.jflm.2017.07.001.

16. Cuenca LG, Puertas E, Aliane N, Andres JF. Traffic accidents classification and injury severity prediction. In: *2018 3rd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*. IEEE; 2018:52–57. http://dx.doi.org/10.1109/ICITE.2018.8492545.
17. Dong C, Shao C, Li J, Xiong Z. An improved deep learning model for traffic crash prediction. *J Adv Transp*. 2018;2018. http://dx.doi.org/10.1155/2018/3869106.
18. Arteaga C, Paz A, Park J. Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. *Saf Sci*. 2020;132:104988. http://dx.doi.org/10.1016/j.ssci.2020.104988.
19. Srivastava S, Bagga A, Shekhawat RS. Review of the machine learning techniques in road crashes. In: *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)*. IEEE; 2020:337–373. http://dx.doi.org/10.1109/ICCAKM46823.2020.9051506.
20. Ghandour AJ, Hammoud H, Al-Hajj S. Analyzing factors associated with fatal road crashes: A machine learning approach. *Int J Environ Res Publ Health*. 2020;17:4111. http://dx.doi.org/10.3390/ijerph17114111.
21. Ribeiro P, Jordana X, Scheirs S, et al. Distinction between perimortem and postmortem fractures in human cranial bone. *Int J Legal Med*. 2020;134:1765–1774. http://dx.doi.org/10.1007/s00414-020-02356-3.
22. Matsui Y, Hitosugi M, Mizuno K. Severity of vehicle bumper location in vehicle-to-pedestrian impact accidents. *Forensic Sci Int*. 2011;212:205–209. http://dx.doi.org/10.1016/j.forsciint.2011.06.012.
23. Casali MB, Blandino A, Grignaschi S, Florio EM, Travaini G, Genovese U. The pathological diagnosis of the height of fatal falls: A mathematical approach. *Forensic Sci Int*. 2019;302:109883. http://dx.doi.org/10.1016/j.forsciint.2019.109883.
24. Friedman J, Hastie T, Tibshirani R. *The Elements of Statistical Learning*. New York: Springer series in statistics; 2001. http://dx.doi.org/10.1007/978-0-387-84858-7.
25. Quinlan JR. Learning efficient classification procedures and their application to chess end games. In: *Machine Learning*. Springer; 1983:463–482. http://dx.doi.org/10.1007/978-3-662-12405-5_15.
26. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32. http://dx.doi.org/10.1023/A:1010933404324.
27. Rosenblatt F. *Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms*. Buffalo NY: Cornell Aeronautical Lab Inc; 1961. http://dx.doi.org/10.1007/978-3-642-70911-1_20.
28. Cybenko G. Approximation by superpositions of a sigmoidal function. *Math Control Signals Syst*. 1989;2:303–314. http://dx.doi.org/10.1007/BF02551274.
29. Hecht-Nielsen R. Theory of the back propagation neural network. In: *Proceedings of International Joint Conference on Neural Networks*. New York: IEEE; 1989:593–605. vol. 1, http://dx.doi.org/10.1109/IJCNN.1989.118638.
30. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20:273–297. http://dx.doi.org/10.1007/BF00994018.
31. Pearson III KL. On lines and planes of closest fit to systems of points in space. *Lond Edinb Dubl Philos Mag J Sci*. 1901;2:559–572. http://dx.doi.org/10.1080/14786440109462720.
32. Stewart GW. On the early history of the singular value decomposition. *SIAM Rev*. 1993;35:551–566. http://dx.doi.org/10.1137/1035134.
33. Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw*. 1999;10:626–634. http://dx.doi.org/10.1109/72.761722.
34. Zhang L, Wang DC, Huang Q, Wang X. Significance of clinical phenomes of patients with COVID-19 infection: A learning from 3795 patients in 80 reports. *Clin Transl Med*. 2020;10:28–35. http://dx.doi.org/10.1002/ctm2.17.
35. Kosorok MR, Laber EB. Precision medicine. *Annu Rev Stat Appl*. 2019;6:263–286. http://dx.doi.org/10.1146/annurev-statistics-030718-105251.
36. Marraro GA, Casali M, Genovese U. Responsible precision medicine in pediatric acute respiratory distress syndrome: The challenge of searching for biomarker-driven earlier diagnosis, effective treatment, and stratified outcomes. *Crit Care Med*. 2018;46:172–174. http://dx.doi.org/10.1097/CCM.0000000000002809.
37. Lefevre T. Big data in forensic science and medicine. *J Forensic Leg Med*. 2017;57:1–6. http://dx.doi.org/10.1016/j.jflm.2017.08.001.
38. Oura P, Junno A, Junno JA. Deep learning in forensic gunshot wound interpretation-a proof-of-concept study. *Int J Legal Med*. 2021. Epub ahead of print. http://dx.doi.org/10.1007/s00414-021-02566-3.
39. Gless S. Ai in the courtroom: A comparative analysis of machine evidence in criminal trials. *Georget J Int Law*. 2020;51:195–253. https://ssrn.com/abstract=3602038.
40. Blandino A, Travaini G, Rifiorito A, et al. Prediction model for autopsy diagnosis of driver and front passenger in fatal road traffic collisions. *Forensic Sci Int*. 2021;324:110853. http://dx.doi.org/10.1016/j.forsciint.2021.110853.
41. Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artif Intell*. 2019;267:1–38. http://dx.doi.org/10.1016/j.artint.2018.07.007.
42. Zadeh L. Fuzzy sets. *Inf Control*. 1965;8(3):338–353. http://dx.doi.org/10.1016/S0019-9958(65)90241-X.
43. Adadi A, Berrada M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*. 2018;6:52138–52160. http://dx.doi.org/10.1109/ACCESS.2018.2870052.
44. Casiraghi E, Malchiodi D, Trucco G, et al. Explainable machine learning for early assessment of COVID-19 risk prediction in emergency departments. *IEEE Access*. 2020;8:196299–196325. http://dx.doi.org/10.1109/ACCESS.2020.3034032.