Road Accidents: Information Extraction from Clinical Reports

Enrico Mensa[1,†], Daniele Liberatore[1,†], Davide Colla[1,†], Matteo Delsanto[1,†], Marco Giustini[2,†] and Daniele P. Radicioni[1,*,†]

[1]*Dipartimento di Informatica, Università degli Studi di Torino*

[2]*Istituto Superiore di Sanità, Roma*

Abstract

In this paper we present a novel approach for the computation of statistical insights on road traffic accidents (RTAs). Instead of relying on numerical and categorical reports, we propose a system for the extraction of crucial information from clinical reports concerning RTAs, that can then be used to enrich more traditional data with relevant information associated to the injuries reported by accidents victims. After having tested and evaluated the system, we also illustrate and discuss the information extracted automatically from over 30, 000 medical records.

Keywords

Information Extraction, Road Traffic Accidents, Clinical Reports, BERT

1. Introduction

In recent years Electronic Health Records (EHRs) have become more and more common, leading to the collection of increasing amounts of health and clinical data. Among these, Emergency Room records are particularly interesting since they can provide structured data concerning the patient, combined with unstructured free-text data describing the events that led to the hospitalization. These records are usually categorized under the type of event that caused the patient injuries, allowing for multiple analysis centered on the specific events. This paper focuses in particular on road traffic accidents.

A road traffic accident (RTA) is defined by the French National Institute of Statistics and Economic Studies as *an accident that occurs when at least one road vehicle is involved in an accident which happens on an open public road, and at least one person ends up being killed or injured* [1]. Analyzing road accidents is a relevant task since, according to the World Health Organization (WHO) [2, 3] RTAs *i*) are likely to become the seventh leading cause of death by 2030, and *ii*) they cause death to vulnerable road users since more than half individuals died on the roads are either cyclists, motorcyclists, or pedestrians. For these reasons the study and prevention of RTAs is a priority in transportation management. Studies have been carried out on RTAs [4, 5, 6], mostly employing data mining techniques on structured data (e.g., traffic

crash data), often paired with further specific information concerning road quality or weather conditions. This sort of data is naturally suited for the analysis of RTAs, although often lacking of information on the specific injuries reported by the patients.

In this work we investigate a novel approach for the study of RTAs: we extract relevant information concerning road accidents from medical records, so to develop a system providing detailed statistical insights on the injuries possibly caused by a given accident; to the best of our knowledge this task was never previously addressed in literature. The feasibility of such a system requires however to determine if current state-of-the-art language models (e.g., BERT) are at least able to identify and extract which vehicles are involved in a medical report. More specifically, we are interested in the extraction of the vehicles involved in the accident, and in which vehicle the patient was in when the accident occurred. This precious information may then enrich patients' data, typically including age, gender, and the reported injuries. The whole extraction process needs to be completely automated, since asking the medical operators to promptly collect such data (maybe by choosing a vehicle in a complex list of tens of options) is completely unfeasible, due to the conditions of high time pressure and workload customarily afflicting emergency departments.

The paper is structured as follows: after describing the state of the art in Section 2, we introduce the adopted dataset and illustrate the annotation process in Section 3. We then propose and evaluate an extraction algorithm based on BERT contextual embeddings in Section 4. Finally, we report some aggregated statistics by running the system on the entire dataset, focusing on the axis of age and gender (Section 5), deferring the more complex study of injuries to future work, described in Section 6.

2. Related Work

Many works have been published on the topic of RTAs, showing that the problem can be analyzed from different points of view and with different aims. A great deal of effort has been naturally invested in accident severity prediction [4, 7, 8, 9, 10] and traffic accident anticipation [11], aimed at the understanding of which factors such as weather conditions, time, road quality, etc. contribute the most to the severity of the accidents. Other efforts have been spent in comparing different methods to understand which ones are better suited for the modeling of RTAs. In particular, a very recent study [5] compares many machine learning approaches including naïve Bayes, logistic regression, K-nearest neighbors, AdaBoost, support vector machines, and random forests finding the latter one as the best suited for the task. Particular emphasis has also been put on the explainability of the adopted algorithms: in [12] random forests and decision trees have achieved good results, allowing for the demonstration that weather conditions are very related to car accidents. In [13] a substantial sub-category of accidents has been examined, specifically addressing those against poles and trees. The authors rely on text mining and on other interpretable machine learning techniques to analyze available crash narratives, so to complement the results with explanations.

Each work typically examines a different dataset which is released by a particular State or Region of the world. A plethora of such datasets can be found, for instance the National Road Network dataset from Canada [14], the Road Safety Data dataset from the UK Department of Transport [15], the Montreal Vehicle Collisions dataset from the City of Montreal [16], the

Setùbal (Portugal) reports [7], the Gauteng (South Africa) reports from the Gauteng Department of Community Safety [5], etc. The amount of information provided by these datasets is highly variable. The crash reports from Victoria (Australia) [8] also include the gender and age of the drivers. Among the various surveyed datasets, the one from Louisiana treated in [13] is a rare instance of a dataset listing some insight concerning the gravity of the injuries incurred by the drivers. A complete review of many of these datasets can be found in [17].

Our work is hardly comparable with the current state-of-the-art, since we do not rely on categorical and numerical data in our analysis, but rather focus on text extraction from free-text narratives included in clinical data. The application of NLP techniques in this field is however not completely new, since in recent years some works have been carried out trying to extract and classify RTAs from social media [18, 19]. In summary, the novelty of our approach stems from the radically different dataset that we are employing, which requires specific NLP techniques to be dealt with. With this preliminary work we are allowing for the future extraction of detailed information regarding injuries in accidents, which can be paired with the data computed through more traditional approaches aimed at a better understanding of the impact of RTAs.

3. Dataset and Annotation

The data employed in the present study are real-world Emergency Room Reports (ERRs) collected in Italian Hospitals, and then made available by the Italian National Institute of Health in the frame of the SINIACA project [20]. The SINIACA project[1] is the Italian branch of the European Injury Database (EU-IDB), an EU-wide surveillance system concerned with accidents, collecting data from hospital emergency department patients according to EU recommendation [21]. The SINIACA-IDB is a data collection on injuries, based on a sample of hospital emergency departments, in implementation of the recommendation of the Council of the European Union no. C 164/2007/01 on injury prevention and safety promotion.

The original dataset consists of $153,826$ clinical records from the SINIACA-IDB, which were originally annotated by hospital staff as referring to road traffic accidents or not. In this work we only consider the $35,952$ records that concern RTAs. It is important to note that these reports are very challenging to process. ERRs are compiled by medical staff under huge time pressure, which leads to typos and disjointed fragments of text, at times resembling bullet lists rather than actual sentences. By randomly sampling and analyzing 592 records of the dataset we measured that the over 10% tokens contain either typos, abbreviations, or acronyms (on average 2.25 per record) [22, 23]. All of these elements significantly increase the difficulty of extracting relevant data from the records.

Data annotation. In this preliminary work we focus on the extraction of two types of information. Namely, given a record, we want to retrieve which kind of vehicles were actually involved in the incident (*Task 1*), and in which vehicle the patient was when the accident occurred (*Task 2*). The dataset was annotated separately for these two tasks, that will be referred to as T1 and T2, respectively. The label collection used to annotate vehicles for both tasks has been

---

[1]'Sistema Informativo Nazionale sugli Incidenti in Ambiente di Civile Abitazione', National Information System on Accidents in Civil Housing Environment.

Table 1

Label distribution for the medical records annotated, before (left - $1,000$ records) and after (right - 987 records) pruning the under represented labels.

Label	# in T1	# in T2
NS (not specified)	405	412
CAR	392	284
MOTORCYCLE	186	173
PEDESTRIAN	78	77
BUS	30	27
BICYCLE	26	24
VAN	5	1
TRUCK	5	0
TRACTOR	1	1
AUTO-RICKSHAW	1	1
ARMORED-CAR	1	0
Total	1,130	1,000

Label	# in T1	# in T2
NS	405	412
CAR	386	279
MOTORCYCLE	185	172
PEDESTRIAN	74	73
BUS	30	27
BICYCLE	26	24
Total	1106	987

designed based on the Wikipedia page related to vehicle categories,[2] which in turn relies on the UNECE (United Nations Economic Commission for Europe) categories and regulations [24]. Some classes were merged for simplicity, while the PEDESTRIAN class was added to account for the annotation of pedestrians. The final taxonomy is illustrated in Figure 2 (Appendix A).

Only the leafs of the taxonomy were adopted as labels throughout the annotation process; however, just a few were actually found in our dataset. Table 1 (left side) reports the label distribution on the two tasks for the $1,000$ randomly selected records. Provided that testing for the classification of labels with very few occurrences (e.g., one or zero for T2) is not meaningful, we decided to remove from the test bed the 13 entries labeled as VAN/TRUCK/TRACTOR/AUTO-RICKSHAW/ARMORED-CAR for either T1 or T2. Table 1 (right side) illustrates the final dataset of 987 records. Future work will focus on the annotation of entries containing less common vehicles, so to be able to evaluate the system for the classification of this type of labels. As illustrated in the Table, in the T2 task only one label per entry has been annotated (since the patient could only be in one vehicle), while T1 allowed for multiple labels since multiple vehicles can be involved in one accident. Moreover, the NS (not specified) tag allows for the classification of records that do not provide enough information to determine which vehicles were involved in the accident. The annotation process was carried out by two annotators on the text annotation tool Doccano [25]. The resulting Inter Annotator Agreement – calculated as Cohen's kappa coefficient– was of 0.9957 for T1 and 0.9930 for T2. Such high IAA shows that this task is quite easy for humans; however, we will show that for artificial systems some language nuances are still difficult to decipher, especially for T2.

---

[2]https://en.wikipedia.org/wiki/Vehicle_category.

Table 2
Precision (P), Recall (R) and F1 score on the T1 task (10 fold).

Label	Baseline			BERT		
	P	R	F1	P	R	F1
NS	0.64	1.00	0.78	0.98	0.96	**0.97**
CAR	0.99	0.44	0.60	0.95	0.96	**0.95**
MOTORCYCLE	1.00	0.82	0.90	0.96	0.98	**0.97**
PEDESTRIAN	0.90	0.26	0.38	0.89	0.73	**0.80**
BUS	1.00	0.93	**0.96**	1.00	0.70	0.82
BICYCLE	0.90	0.77	**0.82**	1.00	0.50	0.67
All (weighted)	0.86	0.72	0.72	0.96	0.93	**0.94**

4. System and Evaluation

Our system is based on a multilingual version of the BERT neural model [26].[3] We employed the same model for the resolution of both T1 and T2. More specifically, the tasks are defined as follows: in T1 the system is requested to retrieve all the vehicles involved in the RTA, while in T2 the system has to detect the mean of transportation used by the patient at the time of the accident. We fine-tuned the existing language model on the whole set of $153, 826$ records for ten epochs so as to specialize the model on the type of language used (Italian, deeply blended with medical jargon). Finally, we stacked a linear classifier from an off-the-shelf Transformers library from HuggingFace [27] on top of the pre-trained model, which exploits the dense representation of the input record to classify it with the appropriate label(s). The final model employs two different types of loss function according to the task definition: in the case of T1 we employed the Binary Cross-Entropy loss, that is the model is allowed to classify multiple labels for each entry (more specifically, we have N loss functions, one for each class), while for T2 we adopted the Cross-Entropy loss, since in this setting the model can only return one class. The evaluation was conducted on the 987 annotated entries described in Section 3, in a 10-fold setup, with a training of 10 epochs for each such fold.

The system has been also been evaluated against a simple baseline based on string-matching for both T1 and T2: for each label l a set of trigger vehicles and their synonyms S^l has been obtained from the Treccani Italian Dictionary [28]. For T1 the baseline classifies a record as l if the record contains a trigger word from the corresponding set S^l. For T2 the baseline works similarly, but the trigger word has also to be in the proximity of a *trigger expression*. Trigger expressions are a collection of words and sentences (e.g., *was driving, guiding, run over, passenger of*, etc.) often used in the records to indicate the fact that the patient was either the driver or a passenger in a vehicle. The baseline has been run on the same 10 folds as the BERT system, discarding the training portion of each fold, since no training was performed for this system.

Discussion.   Results for the T1 and T2 tasks are reported in Table 2 and Table 3 respectively. Concerning T1, both BERT and the baseline perform well overall. However, depending on

---

[3]https://huggingface.co/bert-base-multilingual-cased.

Table 3
Precision (P), Recall (R) and F1 score on the T2 task (10 fold).

Label	Baseline			BERT		
	P	R	F1	P	R	F1
NS	0.48	1.00	0.65	0.96	0.93	**0.95**
CAR	0.86	0.13	0.22	0.92	0.93	**0.92**
MOTORCYCLE	1.00	0.23	0.36	0.96	0.97	**0.97**
PEDESTRIAN	0.80	0.65	0.71	0.75	0.73	**0.74**
BUS	0.11	0.04	0.06	0.79	0.85	**0.82**
BICYCLE	0.20	0.04	0.07	0.76	0.92	**0.83**
All (weighted)	0.68	0.54	0.45	0.92	0.92	**0.92**

the class, the two systems show very diverse performances. The CAR and PEDESTRIAN classes appear to be more challenging compared to the MOTORCYCLE class: this is due to the linguistic variability adopted by the medical personnel when describing car or pedestrian accidents. As an example, PEDESTRIAN accidents are mostly described as 'hit by vehicle', however, some specific instances such as 'grazed by a rearview mirror', 'slided against a car', 'crushed by a tire', etc. can also be found. Such variance seems to be well dealt with by the neural model, while the baseline falls short in managing this complexity. On the other side, BUS and BICYCLE accidents are so few in the dataset that the neural model cannot properly generalize, whilst the baseline can deal with them by simply matching the words *bicycle* and *bus*, which are included in the respective trigger sets. The overall performance on the dataset (bottom row) shows that BERT mostly succeeds in this task with high accuracy (0.94 F1 score). BERT is also very precise on all classes (even on NS), which is a key feature for our goal: being able to rule out the records where no relevant information is available (thus limiting false negatives) is a priority.

The results on T2 show how much the language model is able to distinguish vehicles involved in the incident from those in which the patient was either a passenger or the driver. PEDESTRIAN is the most challenging class: by looking at the records, this is probably due to the fact that the notion of being a pedestrian is sometimes expressed in convoluted and less clear ways compared to the notion of being passenger or driver. Increasing the dimension of the training set may be helpful to solve this issue. Given the huge drop in the performances of the baseline, this task is evidently too complex for a string-matching approach. Once again, the overall performance on the dataset (bottom row) shows that BERT can successfully deal with the task (0.92 F1 score). By looking specifically at precision and recall, we observe that the system is more precise in finding T1 occurrences, but it lacks recall. This phenomenon is interestingly reversed in T2, which may be explained by the fact that the notion of driving or being a passenger is more complex to detect.

5. Road Traffic Accidents Insights

In this Section we report some insights on the whole dataset, obtained by running the BERT system on the $35,952$ entries concerning RTAs. In order to improve the stability of the system, we used all of the 987 entries as training set. We presently discuss the results on the T2 task
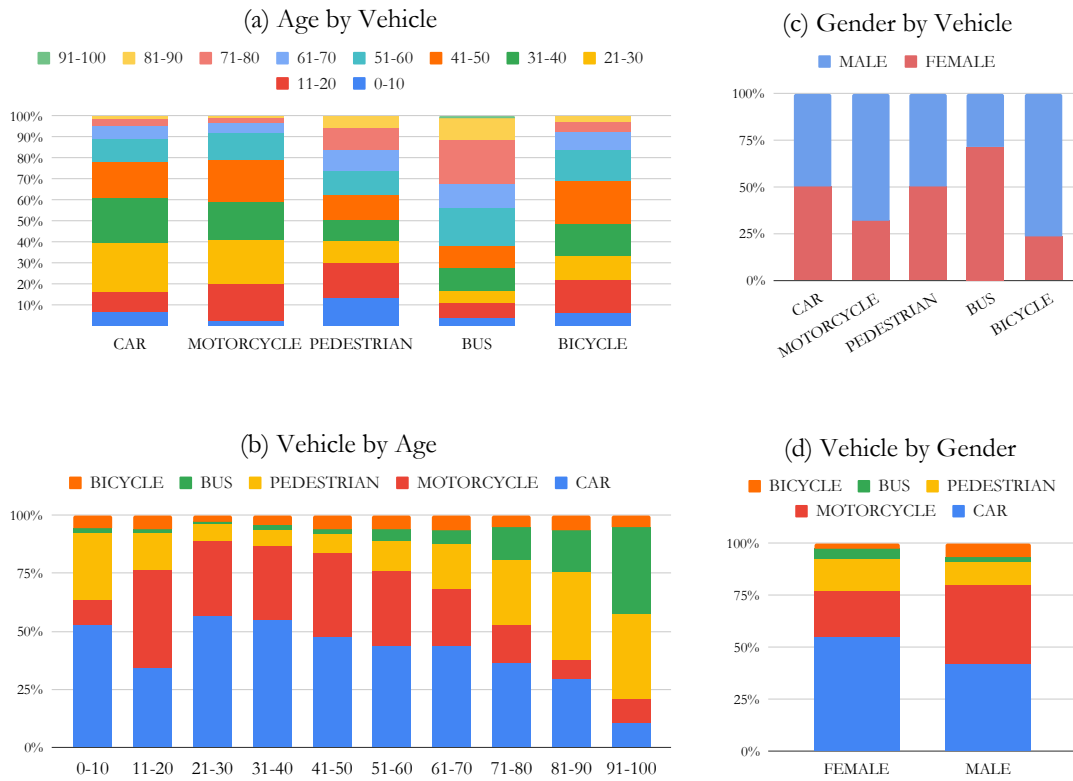
Figure 1: Statistical insights on road traffic accidents obtained by the full dataset. The reported statistics refer to the T2 task i.e. the vehicle in which the patient was in at the time of the accident.

(concerned with the vehicle on which the patient was traveling at the time of the accident), which it is definitely more interesting. To these ends, we provide aggregated data on gender and age, deferring the investigation on injuries to future work. Figure 1a reports the distribution of road accidents for each age interval with respect to the vehicles involved. According to such figures, the age of people involved in car and motorcycle accidents spans from 21 to 60. Differently, bus accidents mainly involve people over 60 years old, while 21% of bicycle accidents concern people from 41 to 50 years old. In contrast, accidents involving pedestrians seem to mainly concern lower and higher bounds of the age range: almost 50% of such collisions concern people under 30, and 25% involves over 70. In Figure 1b we illustrate the distribution of road accidents for each vehicle class with respect to the age interval of the patient. People aged between 0 and 10 are involved in accidents mainly as car passengers (53%) and pedestrians (29%), probably due to the fact that children are typically accompanied by parents. Conversely, people aged from 11 to 20 are often involved in motorcycle accidents (42%), consistently with the changes in travel habits during adolescence. The car becomes the main vehicle involved in accidents up to the 50 years limit, while only about the 30% of accidents concerns motorcycles. Progressively, after the age of 50, the percentage of car accidents decreases consistently with the aging of the patients; at the same time accidents involving pedestrians and buses increase accordingly to the

changes in movement habits. Figure 1c shows the distribution of road accidents for gender of the patient with respect to the vehicles involved. Accidents involving cars and pedestrians are equally distributed between males and females, while events where the patient travels by bus mainly concern females (71%). Motorcycle (68%) and bicycle (77%) accidents mainly involve males. Figure 1d reports the distribution of road accidents for each vehicle class with respect to the gender of the accident patient. The 55% of accidents involving females concern cars, while the 22% involve motorcycles. Consistently with such figures, the events involving males mainly concern cars (42%), followed by motorcycle accidents (38%). Accidents involving pedestrians are of the same order of magnitude for both genders, while males experience three times more bicycle accidents than females.

6. Conclusions and Future Work

In this work we presented a novel approach for the computation of statistical insights on road traffic accidents. We annotated a portion of the SINIACA dataset including Italian medical records concerning RTAs for the automatic extraction of vehicles, focusing on the extraction of those in which the patient was either the driver or a passenger. We observed that a simple baseline based on string-matching obtains .72 F1 score in the first task, but only .45 in the latter one. Conversely, the system based on modern contextual language models yields .94 and .92 F1 scores in the two tasks, respectively: this result suggests that such linguistic devices are mostly able to manage the language nuances and the medical jargon surrounding the relevant portions of text. We then reported and discussed some preliminary statistics extracted from our dataset through the system presented.

Future work includes the release of the dataset in order to allow other researchers to propose and compare different approaches on the extraction tasks. We also plan to develop a pipeline that relies on the T1 results to guide the system through the T2 task. Additionally, we will explore how to employ both static [29] and contextual sense-enabled embeddings [30] (along with their related proximity measures, [31]) to exploit their lexicographic precision and to find links and associations among the extracted pieces of information.

This work constitutes the first step in the development of a more complete system aimed at the integration of different sorts of information extracted from injury data; in fact, extracted information may be coupled with categorical and numerical data to improve the understanding of the impact of RTAs, and pinpointing details relevant to better understand the phenomenon of road injuries, and thus to assist the policies on injury prevention and transportation management.

References

[1] Institut National de la Statistique et des éstudies économiques, Road Accident definition, Accessed: 10-10-2022. https://www.insee.fr/en/metadonnees/definition/c1116.

[2] W. H. Organization, Global status report on road safety 2015, World Health Organization, 2015.

[3] W. H. Organization, Global status report on road safety 2018, World Health Organization, 2018.

[4] M. Chong, A. Abraham, M. Paprzycki, Traffic accident analysis using machine learning paradigms, Informatica 29 (2005).

[5] T. Bokaba, W. Doorsamy, B. S. Paul, Comparative study of machine learning classifiers for modelling road traffic accidents, Applied Sciences 12 (2022) 828.

[6] B. Kumeda, F. Zhang, F. Zhou, S. Hussain, A. Almasri, M. Assefa, Classification of road traffic accident data using machine learning algorithms, in: 2019 IEEE 11th international conference on communication software and networks (ICCSN), IEEE, 2019, pp. 682–687.

[7] D. Santos, J. Saias, P. Quaresma, V. B. Nogueira, Machine learning approaches to traffic accident analysis and hotspot prediction, Computers 10 (2021) 157.

[8] K. Assi, Traffic crash severity prediction—a synergy by hybrid principal component analysis and machine learning models, International journal of environmental research and public health 17 (2020) 7598.

[9] R. E. AlMamlook, K. M. Kwayu, M. R. Alkasisbeh, A. A. Frefer, Comparison of machine learning algorithms for predicting traffic accident severity, in: 2019 IEEE Jordan international joint conference on electrical engineering and information technology (JEEIT), IEEE, 2019, pp. 272–276.

[10] Z. Li, P. Liu, W. Wang, C. Xu, Using support vector machine models for crash injury severity analysis, Accident Analysis & Prevention 45 (2012) 478–486.

[11] W. Bao, Q. Yu, Y. Kong, Uncertainty-based traffic accident anticipation with spatio-temporal relational learning, in: Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 2682–2690.

[12] C. Parra, C. Ponce, S. F. Rodrigo, Evaluating the performance of explainable machine learning models in traffic accidents prediction in california, in: 2020 39th International Conference of the Chilean Computer Science Society (SCCC), IEEE, 2020, pp. 1–8.

[13] S. Das, S. Datta, H. A. Zubaidi, I. A. Obaid, Applying interpretable machine learning to classify tree and utility pole related crash injury types, IATSS research 45 (2021) 310–316.

[14] Government of Canada, National Road Network, Accessed: 10-10-2022. https://open.canada.ca/data/en/dataset/3d282116-e556-400c-9306-ca1a3cada77f.

[15] UK Department for Transport, Road Safety Data, Accessed: 10-10-2022. https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data.

[16] City of Montreal, Montreal Veichle Collisions, Accessed: 10-10-2022. http://donnees.ville.montreal.qc.ca/dataset/collisions-routieres.

[17] C. Gutierrez-Osorio, C. Pedraza, Modern data sources and techniques for analysis and forecast of road accidents: A review, Journal of traffic and transportation engineering (English edition) 7 (2020) 432–446.

[18] A. Salas, P. Georgakis, Y. Petalas, Incident detection using data from social media, in: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2017, pp. 751–755.

[19] A. Salas, P. Georgakis, C. Nwagboso, A. Ammari, I. Petalas, Traffic event detection framework using social media, in: 2017 IEEE International Conference on Smart Grid and Smart Cities (ICSGSC), IEEE, 2017, pp. 303–307.

[20] A. Pitidis, G. Fondi, M. Giustini, E. Longo, G. Balducci, Gruppo di lavoro SINIACA-IDB, Dipartimento di Ambiente e Connessa Prevenzione Primaria, ISS, Il Sistema SINIACA-IDB per la sorveglianza degli incidenti, Notiziario dell'Istituto Superiore di Sanità 27 (2014)

11–16.

[21] R. Lyons, R. Kisse, W. Rogmans, Eu-injury database introduction to the functioning of the injury database (idb), 2015. https://bit.ly/37FAKaB.

[22] E. Mensa, D. Colla, M. Dalmasso, M. Giustini, C. Mamo, A. Pitidis, D. P. Radicioni, Violence detection explanation via semantic roles embeddings, BMC Medical Informatics Decis. Mak. 20 (2020) 263. URL: https://doi.org/10.1186/s12911-020-01237-4. doi:`10.1186/s12911-020-01237-4`.

[23] E. Mensa, G. M. Marino, D. Colla, M. Delsanto, D. P. Radicioni, A resource for detecting misspellings and denoising medical text data, in: J. Monti, F. Dell'Orletta, F. Tamburini (Eds.), Proceedings of the Seventh Italian Conference on Computational Linguistics, CLiC-it 2020, Bologna, Italy, March 1-3, 2021, volume 2769 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 1–7. URL: http://ceur-ws.org/Vol-2769/paper_48.pdf.

[24] UNECE, Definitions of Vehicles (pages 5-11), Accessed: 10-10-2022. https://unece.org/fileadmin/DAM/trans/main/wp29/wp29resolutions/ECE-TRANS-WP.29-78r6e.pdf.

[25] H. Nakayama, T. Kubo, J. Kamura, Y. Taniguchi, X. Liang, doccano: Text annotation tool for human, 2018. URL: https://github.com/doccano/doccano, software available from https://github.com/doccano/doccano.

[26] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[27] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, A. M. Rush, Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, Online, 2020, pp. 38–45. URL: https://www.aclweb.org/anthology/2020.emnlp-demos.6.

[28] Treccani, Treccani Italian Dictionary, Accessed: 10-10-2022. https://treccani.it.

[29] D. Colla, E. Mensa, D. P. Radicioni, Lesslex: Linking multilingual embeddings to sense representations of lexical items, Computational Linguistics 46 (2020) 289–333.

[30] D. Loureiro, A. M. Jorge, J. Camacho-Collados, Lmms reloaded: Transformer-based sense embeddings for disambiguation and beyond, Artificial Intelligence 305 (2022) 103661.

[31] D. Colla, E. Mensa, D. P. Radicioni, Novel metrics for computing semantic similarity with sense embeddings, Knowledge-Based Systems 206 (2020) 106346.
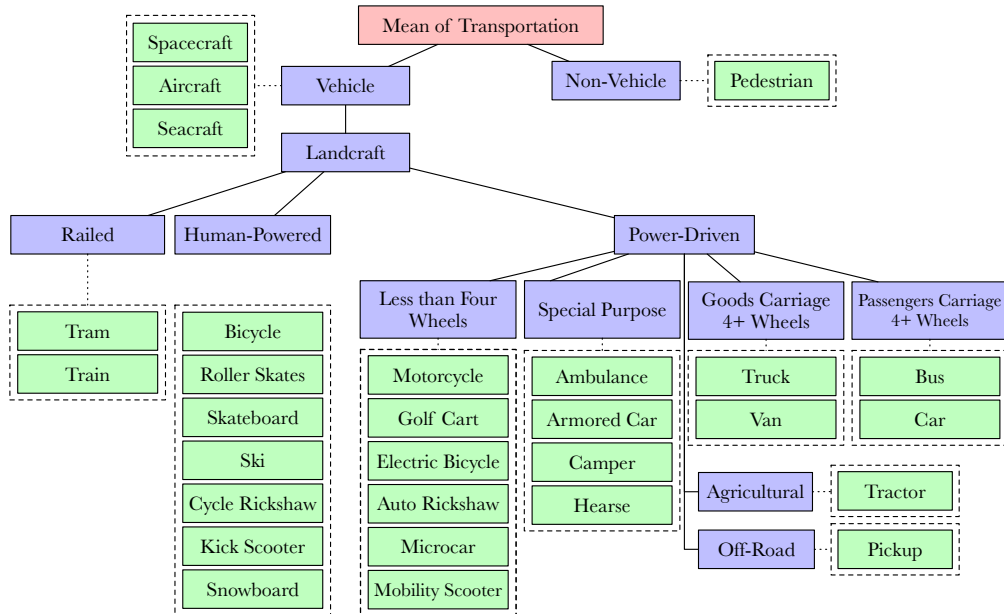
A. Appendix

Figure 2: Taxonomy of vehicles adopted for the annotation process.