Toward Road Safety Recommender Systems: Formal Concepts and Technical Basics

Kaoutar Sefrioui Boujemaa, Ismail Berrada, Khalid Fardousse, Othmane Naggar, and François Bourzeix

Abstract—**Worldwide, traffic accidents are recognized as one of the leading causes of death. This phenomenon leads to significant daily losses affecting both road users and road authorities. Therefore, the need for effective dynamic road security systems is highly considered. Traffic accident data analysis is one of the promising approaches for improving road safety. By taking into account multiple factors (e.g., infrastructure, weather, driver behavior, etc.), it allows measuring the impact of traffic accidents on road security. However, reformulating this impact into practical road safety decisions remains limited and unstructured. To overcome the mentioned limitations, this paper proposes the first end-to-end recommendation framework for road safety. Our framework introduces a three-layered architecture, designed to handle data analysis and action recommendation tasks. For data analysis, we adopt a baseline of state-of-the-art machine and deep learning algorithms to build different traffic accident prediction models. For the action recommendation task, we developed a new approach involving model predictions, model interpretations, actions definition, and road-action interactions matrix annotation. The proposed framework has been successfully experimented and evaluated using two real-world datasets of historical traffic accidents of France (2006-2017) and Morocco (2010-2014), achieving interesting ROC-AUC scores of 0.93 and 0.96, respectively.**

Index Terms—**Traffic accident analysis, machine learning, deep learning, recommendation safety systems, traffic accident risk prevention, road safety countermeasures.**

I. Introduction

According to the World Health Organization "WHO" [37], the global road accidents damages are in a continuous climb, causing severe injuries, life-changing disabilities, and millions of deaths. WHO reports claim that on a yearly basis, 20 to 50 million people are injured and 1.35 million are killed. Traffic accidents are classified in these reports as the leading cause of death for children and young adults (ages between 5 and 28) and as the 8th cause of death for all age groups, surpassing HIV/AIDS, tuberculosis and diarrhoeal diseases. These losses have a strong impact on public physical and mental health. They also negatively affect the global economic growth with an estimated cost of 518 billion dollars per year [38]. Consequently, road safety remains an essential public subject that needs serious attention from all road users (e.g., drivers, pedestrians), road authorities (e.g., local transportation government), private sectors (e.g., infrastructure and construction companies) and researchers.

By adopting global strategies that meet the best road safety practices (e.g., road safety action decade strategy [1]), some middle and high-income countries have seen success in reducing the annual number of road deaths [37]. However, this progress is highly correlated with different factors, such as the geographical area, the victims' categories (e.g., driver, pedestrian, cyclist) and the victims' characteristics (e.g., age, gender). In fact, according to the 2018-2019 reports of the Metropolitan area of France, the global mortality rate has decreased by -4.5%, while it has increased by $+18\%$ for cyclist victims. In contrast, in the overseas regions of France, the global mortality rate has increased by $+11.1\%$, while it has decreased by -10% for cyclist victims [40]. Therefore, different indicators should be considered while measuring local and global road safety improvements.

Efficient road safety action plans require involving several components, namely [39]:

1) User behavior, including traffic law violation, non-use of seat-belt/helmet, aggressive driving, random street crossing, etc.
2) Vehicles, including equipment malfunctions such as brakes and tires, etc.
3) Environment, including inadequate infrastructure, intense traffic flow, bad weather conditions, etc.

The 3Es (Engineering, Enforcement, and Education) approach is one of the first proposed approaches to consider these components [64]. It allowed road safety strategies to evolve from considering road users (especially drivers) as the main risk source, to a wider thoughtfulness based on the 3Es elements. The Safe System "SS" approach, also known as Vision Zero [43], has extended the 3Es vision by introducing additional dimensions emphasizing the practicability of the approach [50]. In particular, the "Data, research and evaluation" component (see Fig. 1) encourages data-driven research toward building robust models that are designated for providing applicable road safety solutions. In literature, this aspect has been investigated through various studies: social

Manuscript received May 16, 2020; revised September 23, 2020 and January 4, 2021; accepted January 15, 2021. This work was supported in part by the Moroccan Ministry of Equipment, Transport and logistics and in part by the Moroccan Foundation of Advanced Science and Innovation. The Associate Editor for this article was X. Cheng. *(Corresponding author: Kaoutar Sefrioui Boujemaa.)*

Kaoutar Sefrioui Boujemaa and Khalid Fardousse are with the Dhar El Mehraz Science Faculty (FSDM), Sidi Mohamed Ben Abdellah University (USMBA), Fes 30050, Morocco (e-mail: kaoutar.sefriouiboujemaa@usmba.ac.ma).

Ismail Berrada is with the School of Computer and Communication Sciences (SCCS), Mohammed VI Polytechnic University, Benguérir 43150, Morocco.

Othmane Naggar and François Bourzeix are with the Moroccan Foundation for Advanced Science, Innovation and Research (MAScIR), Rabat 10100, Morocco.

Digital Object Identifier 10.1109/TITS.2021.3052771

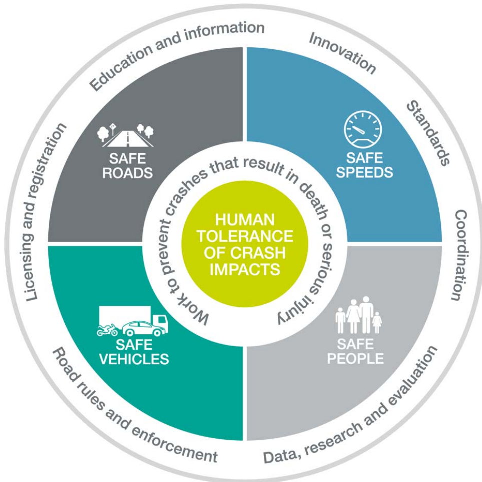

Fig. 1. The Safe System approach components [43].

studies (road safety educational systems [49], road user behavior analysis [45]–[48]), traffic accident analysis (predictions of accident occurrence/frequency, number of victims and accident severity level [25], [26], [41]), traffic flow management [51], emergency rescue systems [52], etc.

However, most of these studies do not provide a solid scientific discussion on the applicability of their findings in the decision-making procedure. Typically, road safety is addressed in two separate directions. The first one consists of the mentioned data-driven research, while the second one consists in defining counter-measures handbooks [13] based on observations of traffic accident records, general identification of causalities and opinions of road safety experts. Therefore, the research-practice relationship is still unstructured and ambiguous for real-world implementations [23]. Furthermore, serious attention needs to be harmoniously conducted by both researchers and stakeholders toward developing a purposeful structure for this relationship. A first attempt was initiated by the Safety CaUsation, Benefits and Efficiency "SafetyCUBE" project, introducing a road safety Decision Support System "DSS" as an online search-engine [14]. DSS is designed to help policy-makers to find adequate actions (a.k.a measures) for specific traffic accident risk-factors categorized in 4 domains: road users, vehicle, infrastructure and post-care. DSS is a valuable scientific basis for the road safety community. However, it does not support the analysis of features-interdependence for defining the risk-factors. It also suffers from a lack of formal representation for merging actions of different categories in one case study and most importantly, it does not support the automation of the search and the selection process of risk-factors and their related actions.

Within this context, we introduce the first automatized end-to-end recommendation framework for road safety. It is designed to recommend appropriate road safety actions by taking into account the dynamic changes and the contextual characteristics of the road-network. This is achieved through an Artificial Intelligence "AI" architecture tailored to analyse traffic accident data for risk prediction and significant related action recommendation. This architecture is composed of three main layers. The first one is responsible for data preprocessing using different techniques (e.g., missing data imputation, features encoding, sampling). The second one consists in building robust traffic accident prediction models trained using a baseline of Machine "ML" and Deep Learning "DL" algorithms. These models serve in detecting risk causalities at multiple levels. The last layer is considered as the innovative side of our framework. It is designed to develop a recommendation system (following the basics of recommendation systems in literature) that allows defining a learnable linking approach between the traffic accident predictions and the road safety actions that are appropriate for reducing the predicted risk-factors.

To validate the reliability and the robustness of our framework for real-world application, we carry out experiments using two challenging traffic accident datasets of France (2006-2017) and Morocco (2010-2014).

Thus, the main contributions of this paper are as follows:

1) Designing the architecture of the first end-to-end road safety actions recommendation framework.
2) Formalizing the research-practice relationship of road safety actions decision-making through graph theory.
3) Developing the first risk-factors and related actions data annotation approach.
4) Validating the framework through two real-word datasets: Moroccan dataset (a new unexplored dataset) and French dataset.

The rest of this paper is organized as follows. Section II discusses the state of the art of traffic accident data analysis. Section III gives an overview of the proposed framework. Section IV describes the developed approach for implementing the action recommendation layer in the framework. Section V presents the methodology for implementing the overall framework (dataset and method description, and result discussion). Finally, the paper is concluded in section VI, with a brief discussion of our contributions and future directions.

II. State of the Art

In literature, there are various case studies that investigate the development of proper road safety technologies and countermeasures [53]. Such studies are generally based on the analysis of traffic accident data. This last, commonly include features describing the occurred accident according to the environment, the victims, the involved vehicles and other characteristics. The variety of these features allow tackling traffic accidents from different angles. However, this quality can be exceptionally challenging to process because of some inherent difficulties of real-world data. For example, data heterogeneity [60] and the missing, aberrant and redundant values [30], [32].

Considering the proposed framework in this paper, we direct the following discussions (see Table I) towards two major research areas: "traffic accident occurrence prediction" and "other traffic accident related case studies".

A. Traffic Accidents Occurrence Prediction

Predicting traffic accident occurrence can be addressed as either a classification or a regression problem.

TABLE I
RELATED RESEARCH ON TRAFFIC ACCIDENT PREDICTION MODELS

Research	**Scoop**	**Dataset**	**Features**	**Algorithms**	**Results**
Abdel-aty et al [15]	Prediction of traffic accidents occurrence frequency and driver involvement (by gender and by age).	Traffic accidents (1992-1994) dataset, of State Road 50, Central Florida, USA.	AADT, degree of horizontal curvature, lane, shoulder and median widths, urban/rural, speed, section's length, driver age, driver gender.	NBR	R^2: 0.11 (0.27,0.16) (0.16,0.09,0.38)
Caliendo [4]	Annual traffic accidents count estimation by severity level	A four-lanes motorway, Italy (1999-2003)	Section length, Curvature, AADT, slope sight distance on curves,side friction coefficient, Presence of junctions.	PR NBR NMR	RMSE: (0.203,0.123) (0.236,0.146) (0.216,0.181)
Briz-Redon et al [10]	Traffic accidents occurrence prediction near school locations	Valencia, Spain (2014-2015)	No. of traffic accidents, non-pedestrian road length, no. of schools, distances to closest schools, no. of education-related services, no. of noneducational services from various sectors, no. of parking-zones, no. of bus-stops, land-use entropy, average-betweenness, complex-intersections, main road length, traffic-lights, no. of school-aged residents, percentage of high-end cars	CAR-NB MSR LR	$R^2 \in [0.12, 0.25]$
Ren et al [2]	Prediction of the hourly traffic accidents frequency	Beijing traffic accidents: (2016-2017)	Not available	LSTM Lasso Ridge SVR DTR RFR ARMA MLP	RMSE: 0.034 0.045 0.076 0.075 0.058 0.047 0.160 0.060
PARK et al [5]	Traffic accidents occurrence prediction	Gyeongbu road, Korea (2011-2013)	Time, day, position, death-casualty, weather, road-shape, road-alignment, accident type, number of line, traffic-volume, traffic-density, average-speed.	LR SVM	Accuracy: 42.7% 38.0%
Shiau et al [11]	Traffic accidents classification into person-automobile/motorcycle, vehicle-vehicle and automobile/motorcycle accident	Central Taiwan (January-December, 2011)	Road-type, location, road conditions, barriers, road-semaphore, road-separator, weather, light, accident-type.	BPNN LR FPRCA-BPNN FPRCA-LR	Accuracy: 84.37% 85.06% 85.89% 85.14%
Wahab et al [3]	Prediction of motorcycles accidents severity	Ghana (2011-2015)	Injury-severity, year, location type, settlement-type, time and type of collision, road description, collision-partner, road surface type, day of week, weather, road-separation, traffic-control, road-shoulder condition.	DT RF KNN MNL	Accuracy: 73.64% 73.91% 73.71% 52.04%
De et al [7]	Traffic accidents severity classification	Granada, Spain (2003-2005)	Accident type, day, month, time, weather, lightning, cause, lane-width, no. of injuries, occupants-involved, victim age and gender, paved-shoulder, pavement width and markings, shoulder-type, side-distance, vehicles involved	BN-BDe BN-MDL BN-AIC	Accuracy: 0.57% 0.58% 0.59%
Zhang et al [9]	Detection of impact factors of Pedestrians "red-light" running behavior at intersections	Pedestrian surveys, Hefei, China	Gender, age, education-level, profession, annual-income, trip-purpose, time-requirement, intersection-familiarity, easy time-segment to run a reg-light, tolerable waiting time, pedestrians attitudes at different crossing environments	BLM	Not available

1) Regression Modeling: For regression tasks, data is usually aggregated into accident counts or frequency rates. Early studies [16], successfully modeled these tasks, using traditional statistical algorithms, such as Multiple Linear Regression "MLR", Negative Binomial Regression "NBR" and Poisson Regression "PR". These algorithms are commonly performed by a step-wise training procedure for important features selection. For example, Abdel-aty *et al* [15] investigated in accident occurrence frequency using the NBR method. The Akaike Information Criterion "AIC" function was adopted for best model approximation, following a step-wise training procedure. The variables revealed having an increasing impact in their model are the length of roadway sections, the sharpness of horizontal curves and the Annual Average Daily Traffic "AADT" per lane. In contrast, lane, shoulder and median width variables were demonstrated to have a decreased impact on the accident frequency.

Usually, count data is over-dispersed, though, certain statistical methods can provide poor results. To overcome this issue, some studies analysed the accidents counts by subdividing the data into homogeneous clusters. Traditionally, these clusters are manually generated based on specific features (e.g., injuries' severity, the roadway curvature, nearby schools locations) [4], [10]. However, the random variation of the counts values, can still be challenging to handle. Therefore, recent studies generally focus on adopting ML and DL based approaches [17]. Such methods, can adequately fit the non-linearity and the over-dispersion of the data. Additionally, important features selection can be implicitly performed during the training process.

Most of recent approaches are based on Decision Tree "DT" [20] algorithms and Artificial Neural Networks "ANNs" [57]. Note that, these algorithms have interestingly contributed to the development of the traffic accidents analysis' research field. For example, Ren *et al.* [2], addressed the prediction of traffic accident frequency risk, using a baseline of different ML algorithms. Namely, Long Short Term Memory "LSTM", Support Vector Regression "SVR", Lasso and Ridge advanced regression, Random Forest Regression "RFR", Multilayer Perceptron "MLP" [57] and AutoRegressive Moving Average "ARMA" [22]. Their findings revealed the LSTM model to have the overall best results.

Another approach for handling serial traffic accident data, is by adopting time-series methods. A follow up, can be seen in the work of Quddus *et al* [6]. They performed two time-series models, Integer-valued AutoRegressive Poisson "NARP" and The AutoRegressive Integrated Moving Average "ARIMA". Data was analysed on an annual vs a monthly basis, to measure the impact of data aggregations on the models performance. The INARP model was found to have the best performance for predicting the monthly accident number, while the ARIMA model outperformed in predicting the annual accidents number.

2) Classification Modelling: Traffic accident occurrence classification can be modeled as a binary or a multi-class task. In fact, the binary task is rarely addressed by researchers. This is due to the possible data complexities that can be encountered after generating samples of the negative class ("Non-accident") from the positive class. Accordingly, a severe imbalance and high similarity between samples of both classes can be caused [21].

Park *et al* [5] investigated in binary traffic accident prediction using a data mining approach. To prepare their data for classification, the authors considered and compared two well-known techniques for the class imbalance problem: *over-sampling* the minority class and *under-sampling* the majority class. The k-means algorithm [18] was also adopted to reduce the data variability by creating homogeneous clusters for training. Two classification algorithms were experimented, Logistic Regression "LR" and "SVM", where best results were obtained by the LR model, trained on the over-sampled data.

Traffic accident occurrence can also be predicted based on specific features. For example, Shiau *et al* [11], introduced an approach to classify traffic accidents into: person-automobile/motorcycle, vehicle-vehicle and automobile/motorcycle. They applied the Recursive Feature Elimination "RFE" preprocessing method for selecting important features. As a result, seven variables were retained for the classification task. They are mostly related to road characteristics and the accident site. For the classification task, four models were constructed and compared using an ANN architecture and the LR algorithm. The algorithms were evaluated separately and in combination with the Fuzzy Robust Principal Component Analysis "FRPCA" method. Their results showed a high performance for all models trained with the FRPCA.

B. Other Traffic Accidents Related Case Studies

There are numerous traffic accident case studies that have significantly contributed in enhancing road safety, besides frequency and occurrence predictions. These studies aim to give a more detailed understanding of fundamental accident causes and consequences. Researchers handle such studies by focusing on precise subjects related to either vehicle attributes (the type of involved vehicles, level of vehicles' damage), victims characteristics (e.g., number of victims per injury level or age category) or other elements sharing the accident's environment (e.g., accidents rates at road intersections, accidents caused by pedestrians) [61]–[63].

For example, Wahab *et al* [3], studied motorcycle crash severity based on four levels (fatal, hospitalized, injured and damaged). They drew a performance distinction between three ML algorithms (DT, RF and K-Nearest-Neighbor "KNN" [24]) and one statistical method (MNL). The ML algorithms were evaluated using the K-Fold cross-validation technique. Results show that the RF model achieved the highest performance among all. Important feature detection was also addressed in their work, using the Gain ratio in DT based methods. The discovered critical determinants of their study are, date features, location type, road characteristics (separation, surface type and shoulder condition), settlement type and collision features (partner and type).

De *et al* [7], on the other hand, investigated the use of Bayesian Networks "BNs" for traffic accident severity classification in rural highways. Two severity levels were defined, slight injury and severe injury. The best BN model was evaluated based on three different scoring functions (BDe score metric, Minimum Description Length "MDL" and the Akaike Information Criterion "AIC" score). The hill-climbing search algorithm [8] was also adopted for fast and accurate training. Experiments were assessed using k-fold cross-validation. Performance of the different models varied according to the evaluation metric, where the DBe score outperformed based on accuracy, sensitivity and AUC, while the AIC and MDL scores similarly outperformed in terms of specificity and HMSS. Accident type, driver age, lighting and number of injuries, were also defined by inference as impact factors in accident severity classification. The shortcomings of their experiments consisted in the size and the class imbalance of the training data (a large training set can improve results).

Another sensitive subject outlined by safety researchers is pedestrian behavior analysis. The standard approach is based on two steps, first the identification of the impact factors (can be from different sources) and second the modeling of proper solutions in terms of prevention and restriction. For example, Zhang *et al* [9] conducted a study to define the factors impacting pedestrians' red-light running behavior at intersections. Their objective is to prevent traffic accidents caused by such behavior, by determining the intersections where it is more likely to happen and pedestrian groups that might act so. Data was collected from pedestrian surveys. They included three question categories, trip characteristics (e.g., trip purpose, tolerable waiting time), individual characteristics (e.g., age, gender, education level) and pedestrian attitude (e.g., whether to "red-light" run when traffic is low). Their methodology is designed to use the Binomial Logistic Model "BLM" for the "red-light" behavior classification task and the step-wise forward selection technique to exclude insignificant variables from the training process. Four features were revealed to have a crucial impact in the model estimation, which are time period of the day, trip purpose, whether to run a "red-light" when in a hurry and whether imperfect road facility affects crossing behavior.

Despite this interest, the application of traffic accident prediction models in the development of accurate road safety strategies is still ill-defined [23]. The findings of these models need to be directly exploited for road safety decision-making. This can open new enhancement opportunities in this field.

III. Global Framework Overview

As already mentioned, the contribution of traffic accidents data analysis in developing road safety strategies is not well grounded. Traditionally, most strategies for road safety define the appropriate countermeasures based on high risk-factors inferred using manually computed statistics (e.g., number of dead/injured victims in last year, number of motorcycles accidents, number of accidents occurred at night).

Such methodologies are not mathematically founded. Decisions are basically taken by opinion with the help of road safety experts. Consequently, possible future changes in the road-network are ignored. Thus, we end up with a strategy that is not evidence-based, incapable of evenly addressing all road-users categories and incompatible with the road safety conditions at the application time [23].

Integrating traffic accident prediction models in road safety strategies can have a significant importance in straightening and optimizing the decision-making process.

Following this aspect, we present an automatized end-to-end strategic framework for road safety actions' recommendation. It is designed to analyze traffic accident data using a set of analysis and prediction models, to detect high risk-scenarios, to learn and recommend the road safety actions that can reduce the risk rate in the road-network. Our objective is to provide a reinforced mechanism for the action decision-making process that carefully takes into account the dynamic characteristics of the road-network.

The architecture of our framework is composed of three major processing components (see Fig. 2):

- Data acquisition and storage.
- AI core.
- Validation and application.

The framework aims to offer a vast choice of best practices, techniques and approaches for each layer's implementation, allowing its adaptation to the data it will be trained on. The framework's methodology is presented in what follows.

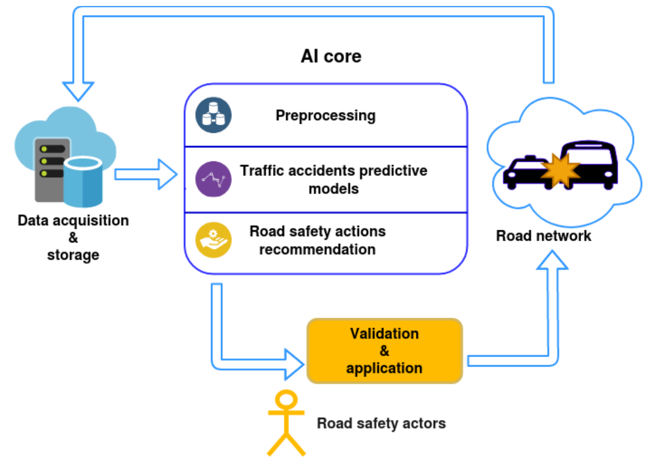

Fig. 2. The proposed end-to-end recommendation framework for road safety.

A. Data Acquisition and Storage

Traffic accident data is one of the fundamental data sources for road safety research. Therefore, collecting good quality data is essential for evaluating the effectiveness of road safety systems and for developing robust prediction models as well [28]. Traffic accident data contains recorded road crashes in a form that answers the five "W" questions (i.e., What, When, Why, Where and Who) for each incident by a range of variables of different types (e.g., numerical, continuous, categorical). Additional explanatory data (e.g., road design, traffic flow, weather conditions, implied driving law) can be afterward linked to the collected accidents base using location and time features.

Traffic accidents' reports provided by public authorities (e.g., police and hospitals) are the primary materials used for traffic accident data acquisition. However, there are multiple other sources that can also be adopted for this task such as insurance agencies, vehicles event data recorders, road sensors (e.g., surveillance cameras, radars) and media (e.g., posts of occurred traffic accidents on social media [29], weather websites [27]). The available tools for the collection and storage of this data are numerous. In case of real-time traffic accident, it is recommended to adopt a flexible streaming pipeline of three major components [66]. First, a fast data collector such as Kafka, Apache Flume and Amazon Kinesis. Second, a processing platform to clean and structure the collected data such as Apache Spark, Apache Storm and Yahoo S4. Third, a NoSQL storage database like MongoDB, Elasticsearch, Cassandra and Cloud Bigtable.

The core of our framework does not depend on any specific data streaming pipeline. However, the datasets we used for experiments in this paper are manually collected from police traffic accidents' reports.

B. AI Core

The AI core is considered as the engine of our proposed framework. It takes as input the collected traffic accident data and outputs a recommendation set of road safety actions for specific road-segments. It contains three successively stacked processing layers: preprocessing layer, traffic accidents predictive models layer and road safety actions recommendation layer.

1) Preprocessing Layer: Data preprocessing is a mandatory step of any knowledge discovery process. It consists in preparing good quality and well-structured data from the input raw data, for a forward pertinent analysis and modeling. Data preprocessing can be achieved at (or through) different steps, starting with data cleaning to feature engineering and data sampling. There is a vast amount of literature on specific methods and approaches to carry out these steps [44].

In what follows, we suggest the best practices needed to be thoroughly considered while handling traffic accident data.

- **Data cleaning:** Traffic accident data is collected from real-world events so it tends to include a lot of erroneous, redundant and missing values, especially when it is manually recorded. Environmental and victim descriptive features (e.g., roadway design, weather condition, address, road-id, age, gender, injury, category, etc.) are most likely to contain such values and can be very challenging to process, though requiring careful cleaning. There are two essential tasks to perform data cleaning.

  The first task is noise-cleaning. It consists of identifying and smoothing/removing outliers [32]. For example, they can be detected by a clustering technique and then replaced with a global constant value (e.g., mean, median) or removed if there are few noisy values. Roadway design and condition features are among the categories of variables that tend to contain the most number of outliers.

  The second task is missing data imputation. It can be handled by either traditional methods such as the "single imputation approach" (replacement with most-frequent, mean, median value, etc.) or by using a data mining algorithm (e.g., KNN, DT) [30], [31]. In the case of location data such as address (avenue name, city, etc.), road-ID or Global Positioning System "GPS" coordinates, the best approach to deal with their missing values is by using map-matching algorithms [33], [36]. For example, GPS coordinates can be used to fill their corresponding road-ID or other location missing values and vice versa.
- **Feature engineering:** This step concerns transforming raw-data features into a presentable and a meaningful format. This can be achieved at multiple levels.

  First, "feature encoding" can be performed to deal with categorical features. The choice of the encoding technique should consider the cardinality of the categorical feature. For low cardinality features (e.g., weather condition, victims gender, injury level), methods such as One-Hot-Encoding "OHE", Binary encoding, Mean encoding, Weight of Evidence encoding "WoE" are suggested. As for large features (e.g., roadway characteristics, address), it is recommended to use methods that transform categories into a lower dimension level without losing too much information. Hashing and word-embedding techniques are recommended in this case [42].

  Second, "feature scaling" can be executed to handle numerical/continuous features such as roadway dimensions, number of driving paths, number of involved vehicles and others. There are different techniques in literature (Min-Max scaling, Z-score, etc.) and their effectiveness depends on the training algorithm used (for ANN it is recommended to scale attributes between 0 and 1).

  Third, "important feature selection" can be performed to keep only the pertinent features for the study. It significantly helps in reducing the computation time and improving the learning performance. For traffic accident data, it is possible to select features based on interpreting their distribution according to specific accidents characteristics (e.g., number of occurrence, severity level). Otherwise, there are multiple selection techniques to adopt, namely filtering, wrapping and embedding methods. Filtering methods provide a set of ranking techniques to select informative features independently of the learning algorithm (e.g., Pearson correlation criteria [73]). Wrapping methods, on the other hand, use the learning algorithm to iteratively select or eliminate features based on the model accuracy (e.g., Sequential Forward Selection "SFS"). Embedded methods refer to algorithms that can implicitly integrate the selection of important features in the learning process (e.g., DT, RF). However, when it is possible, we recommend preserving features of the five following categories since they commonly impact road safety and their interdependence needs to be considered for robust analysis:
  - Road characteristics (design, surface condition, etc.).
  - Geographical features (city, country, GPS coordinates, road-ID, etc.).
  - Climatic features (weather, etc.).
  - Temporal features (date, hour, etc.).
  - Accident characteristics (cause of accident, collision type, the involved vehicles and victims information, etc.).
- **Data sampling:** This is usually the last step of the preprocessing phase. Sampling traffic accident data into homogeneous clusters is widely adopted in literature to reduce its heterogeneity. It can be handled by a clustering algorithm like K-means or manually achieved based on specific selected features like accidents severity, number or type of involved vehicles, etc. However, we recommend to sample the data into a set of m road-segments $R = \{r_1, r_2, \ldots, r_m\}$ using constant features. In this case, a road-segment $r_i \in R$ represents a roadway section with fixed dimensions and can be located by one or multiple location features including a road-ID, an avenue name, a town name, GPS coordinates, etc. In addition, r_i can be assigned a set of dynamic environmental and traffic accident descriptive data. This set is dedicated for modeling various case studies to capture and learn the traffic accidents causalities in r_i. As for sampling the training and testing sets, it is mandatory to split data based on temporal features (e.g., year, month) to preserve the chronological characteristic of accidents.

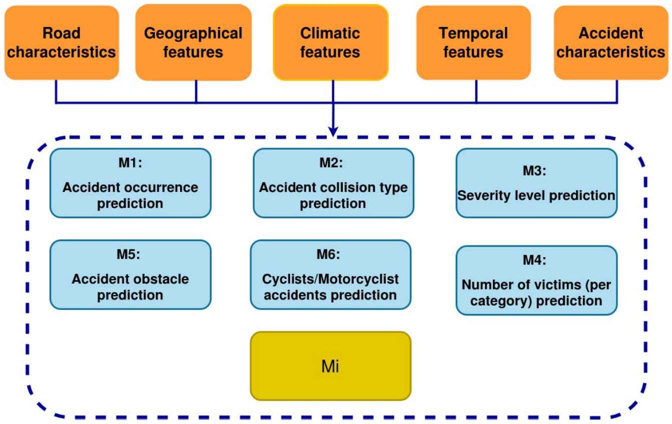

Fig. 3. Framework traffic accidents predictive models.

2) Traffic Accidents Predictive Models: This layer is dedicated for analyzing the preprocessed traffic accident data via different predictive models $M = \{m_1, m_2, \ldots, m_q\}$. The main objective of these models is to provide a profound understanding and solid conclusions concerning the impact of traffic accidents outcomes on road safety. Additionally, the predictions of these models allow recommending actions with high specificity and adeptness to the needs of the studied road-network.

Fig.3 illustrates an example of 6 traffic accidents predictive models that can be trained in this layer. Again, the design of this last does not necessitate a particular set of models. It can be adapted to the training data and particularly to the type of information available for analysis. However, we recommend to select model features based on the five categories as shown in Fig.3, and to sample the data into m road-segments R. As seen above, this type of preprocessing is advantageous to ensure analyzing the data from different angles and to be able to link predictions of the various models in M to a single road section. It is also highly important to build M by levels for a gradual knowledge discovery, where early level models can be related to pre-crash predictions (e.g., occurrence, obstacle) and latest models to post-crash predictions (e.g., number of victims, injuries severity).

To formally structure and optimize the possible dependencies between these models, we propose to adopt conceptual representations similar to Oriented Linear Trees "OLTs" or to Directed Acyclic Graphs "DAGs" [12]. We intend by these representations to design unified global traffic accidents predictive models, called a "Dependency-Models" and referenced as follows: $DM = \{m_1, m_2, \ldots, m_p\}$. A DM takes as input a road-segment $r_i \in R$ and increasingly assign to it predictions at each level to be forwarded as input data to the next levels. Fig. 4 illustrates an example of a Dependency-Model DM structured using an OLT and composed of 3 predictive models $DM = \{m_1, m_2, m_3\}$. First, m_1 is trained to predict traffic accident occurrence in R with a binary target y_1. Then, m_2 receives only road-segments with potential risk and predicts the type of accident's collision with a multi-class target y_2. Last, m_3 takes the outputs of m_2 and predicts the accident's severity level with a multi-class target y_3. Thus, DM outputs three additional descriptive features (y_1, y_2, y_3) for each r_i.

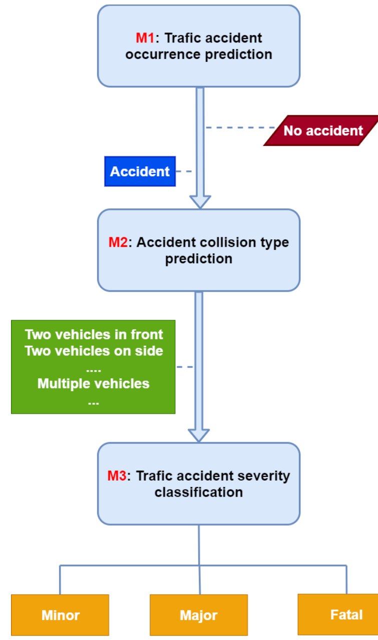

Fig. 4. An example of a dependency-model $DM = \{M_1, M_2, M_3\}$ composed of three traffic accidents prediction models structured using an oriented Linear Tree "OLT".

The DAG hierarchical structure can introduce some erroneous predictions. In fact, the predictive performance of each model is highly dependent on the structure of the DAG and this can make the predictive result be sensitive even erroneous to the model sequence selected. Furthermore, if the upstream model does not learn and include the factors that are essential for the downstream model's prediction, the downstream performance can be limited and therefore poorly performed. Therefore, the models' training and evaluation should be carefully handled. Otherwise, we recommend adopting attention-mechanism approaches [76] to build DM, since they can handle the models inner-dependencies during the training phase.

3) Action Recommendation Layer: This layer is designed to formally structure and learn the relationship between the road-segments R and the appropriate road safety actions. This relationship is modeled in form of a recommendation system that can significantly suggest actions based on the contextual data of R. The previous layer is considered as the groundwork for implementing this system.

We developed an innovative conceptual approach for implementing this layer. It is carefully highlighted in a separate dedicated section (see section IV).

C. Validation and Application

Generally, the definition and the application of a road safety action plan necessitates the integration of multiple road safety actors of leveled authorities (see Fig. 5) [65], [67]. High-level actors that are also recognized as road safety experts (e.g. traffic governmental committees) have the eligibility to make decisions related to the design and operation of the road-network

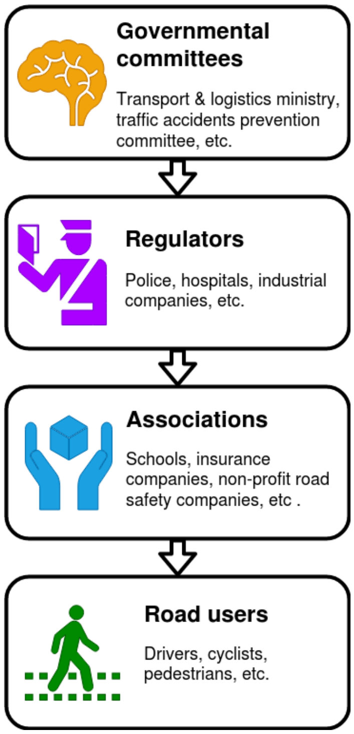

Fig. 5. Road safety actors categories [65], [67].

(e.g., speed limit, flow direction, roadways construction) and to assign/validate the responsibilities (i.e., actions) of low-level actors (e.g. police, road safety associations) as well.

In our framework, a road safety actor has two independent interactions. The first one consists in initializing the learning process in the third layer of the AI core. This interaction is only performed once during the implementation of the framework. The second one consists in acting as a monitor of the validation and application component. For each recommended action, the monitor is responsible for approving its validity and assigning its application in the road-network to the appropriate actors as well.

In summary, the proposed framework can serve as a tool for road safety actors to monitor and optimize the action decision-making mechanism of specific geographical areas. The developed analysis approach in the AI core allows fitting the framework to suit the best the dynamic changes of any road-network by learning its characteristics (i.e., traffic accident data). Therefore, this framework can be adopted to design personalized road safety strategies at international, national and local levels.

IV. Action Recommendation Layer

Applications of recommendation systems in the road safety domain have almost exclusively focused on two research areas: Intelligent Transportation Systems (ITS) and Advanced Drivers Assistance Systems (ADAS). These areas contribute in enhancing road safety via different implementations such as public transportation recommendation for road users, dynamic traffic lights management, speed and best route recommendation for drivers, and many others [71]. The advantage of using recommendation systems in the road safety field resides in their ability to learn recommendations with the awareness of the dynamic context of the road-network.

In this section, we discuss details of the developed approach for implementing the actions recommendation layer in the AI core of our proposed framework (see Fig. 3). As we previously mentioned in III-B.3, this layer consists of building a recommendation system that can recommend the appropriate road safety actions for specific road segments R based on their traffic accidents descriptive features and predicted risk-factors.

A. Background of Recommendation Systems

Recommendation systems are models designed to suggest suitable and meaningful items (product or service) to users (individual or group) within a particular informative environment (e.g., websites, mobile applications) to optimize their decision-making and enhance their experience. This type of systems has been widely investigated by researchers in multiple domains such as E-commerce, E-learning and E-tourism [70].

The recommendation learning in these systems is based on the historical interests of users in specific items. These interests can be implicitly concluded from the users' actions (e.g., clicks, views, purchase) or explicitly provided by them (e.g., products and services evaluation rates). For a set of m users and n items, interactions between users and items are represented by tuples of user/item and linked to their corresponding rating values V (i.e., interests). This set is translated into a sparse Interaction matrix I of size $m \times n$ (see eq. 1). I is essential for training any type of recommendation systems. In some cases, additional data can be adopted for training along with I. For example, using descriptive data profiles of users and items can serve in better learning and analyzing users' interests.

$$I = \begin{bmatrix} v_{11} & v_{12} & \dots \\ \vdots & \ddots & \\ v_{m1} & & v_{mn} \end{bmatrix}, \begin{cases} m : \text{number of users.} \\ n : \text{number of items.} \\ v_{mn} : \text{ rating value.} \end{cases} \quad (1)$$

In the literature, there are multiple approaches for modeling and training recommendation systems. The commonly adopted are Collaborative "CF", Content-Based "CBF" and Hybrid "HF" Filtering [69].

The CF approach requires only the interaction matrix I to predict new ratings. To implement this approach, there are two main techniques that are widely adopted in literature, namely KNN algorithms and Matrix Factorization "MF". KNN techniques can be either user-based or item-based. For KNN user-based CF approaches, similarities are computed between the target user and all the other users. As for KNN item-based CF approaches, predictions are calculated from ratings' similarities of the target user's items and all the other items. MF techniques, on the other hand, consist in decomposing I into two lower dimension matrices containing latent features of users and items respectively. A new rating for the target-user and a specific item can be predicted by multiplying their latent features [59]. The CF approach doesn't require a knowledge of the application domain which can be a significant advantage

TABLE II
EXAMPLES OF ROAD SAFETY ACTIONS PER CATEGORY

Awareness/information actions	Infrastructure actions	Regulatory/prohibitory actions
Crash caution messages on automatic panels	Upgrading or adding pedestrian crossings	Reducing speed limit
Reminding of nearby schools	Adding roundabout in dangerous junctions	Adding signs for hazardous curves
Educative campaign for road safety knowledge (street demonstration)	Split multi-way road with pavement	Prevent big tracks crossing (allow only at certain timing)
Automatic road panels for daily accidents predicted risk rate	Add shelters for bicycles	Plan police barrages in high accidents risk time periods

for a lot of case studies. However, its main limitation remains in its disability to recommend new items (not included in I), known as the "cold start" problem.

The CBF approach, on the other hand, can recommend new items by analyzing their descriptive content features. Recommendations in CBF are user-specific. For each user u, a profile is built to store his historical interactions with items that are represented by a set of descriptive features. These features are used for learning new recommendations for u. This approach can be treated as a classification problem and successively achieved by any ML or DL algorithm. This allows the possibility to interpret the predicted recommendations. Yet, this method might be over-specialized where only items that are highly similar to the user profile are recommended (new interests of the user might be undiscovered). It can also be unreliable in case there isn't enough data to build a user profile.

Finally, the HF approach was specifically introduced in literature to combine two or multiple filtering methods in order to overcome their limitations. For example, CF and CBF approaches can be merged to solve the cold-start problem by using content-features and the over-specialization problem by using other users ratings to discover new interests for a particular user.

B. Our Approach

We introduce a hybrid context-aware recommendation system that is designed to recommend n road safety actions $A = \{a_0, a_1, \ldots, a_n\}$ (i.e., items) for m road-segments $R = \{r_0, r_1, \ldots, r_m\}$ (i.e., users).

Our approach is highly focused on the annotation process of the interactions data of A and R. A flowchart of this interaction is illustrated in Fig. 6. As it is shown, three inputs are necessary to successfully adopt this approach. First, a road safety expert to ensure the reliability of the data by helping in the definition of A. Second, the preprocessed traffic accidents data, that is sampled into m road-segments. Third, the Dependency-Model DM trained to detect specific risk-factors of R. Once these elements are retrieved, the interactions data annotation process can be initiated via three tasks:

1) The definition of the Risk-Scenarios and related Actions "RSA" model.
2) The definition of the Road-segments profiles.
3) The annotation of the Interaction matrix I.

Finally, the system can be trained using any algorithm of the state-of-art of hybrid context-aware recommendation systems [69].

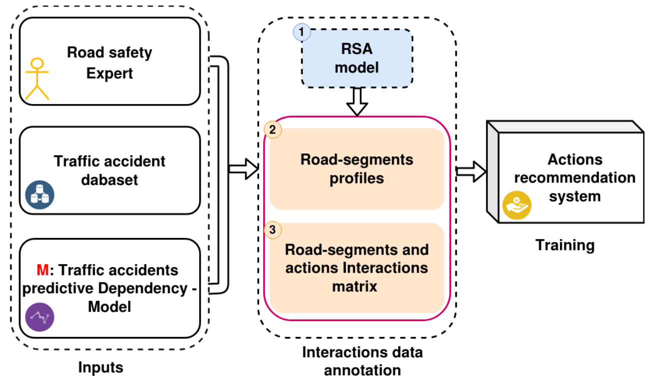

Fig. 6. The architecture of the developed approach for implementing the actions recommendation system of the third layer in the AI core.

1) Risk-Scenarios and Related Actions "RSA" Model: In order to define formally the relationship between A and R, the following three fundamental concepts are required:

- **Risk-scenarios.** A risk-scenario is defined by a context in the road-network, with a possibility of causing different traffic accident risk-factors. Formally, it is defined as propositional formula over a set of specific traffic accidents features F. In fact, an atomic proposition is a statement over F either True or False, and it is expressed as a logical condition of the form $x \sim c$, where $x \in F$, $\sim \in \{<, >, =, \leq, \geq\}$, and c is a constant. We will use the logical operators $\wedge$ (and), $\vee$ (or), and $\neg$ (not) to combine statements. For example, "Atmosphere=rain $\wedge$ $\neg$(Road-surface=wet)" is a risk-scenario that can cause a road crash occurrence. The set of risk-scenarios will be denoted by $RS = \{rs_1, rs_2, \ldots, rs_k\}$.
- **Actions.** We consider three road safety actions' categories to handle the possible risks of RS (see Table II). The first category ("information & awareness") represents actions aiming at positively influencing the road users' behavior by suggesting education campaigns, risk warning and driving safety tips. The other two categories correspond to environment-related actions, pointing at improving the road-network design and operation respectively. The set of all defined actions will be denoted by $A = \{a_0, a_1, \ldots, a_n\}$.
- **Risk-preconditions.** In our approach, the recommendation of road safety actions for road-segments is precisely conditioned by the positive predictions of their corresponding risk-factors (e.g., severe car crash, accident

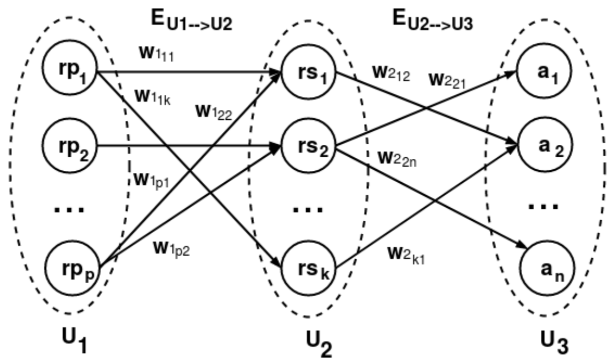

Fig. 7. The RSA model structured as a weighted Tripartite-Graph $TG = (U_1, U_2, U_3, E_{U_1 \Rightarrow U_2}, E_{U_2 \Rightarrow U_3}, W_1, W_2)$.

with involved pedestrians). Therefore, in the RSA model, we define a set of risk-preconditions, denoted by $RP = \{rp_1, rp_2, \ldots, rp_p\}$, whereas an $rp_i \in RP$ represents a risk-factor that is related to one or multiple risk-scenarios. Sequentially, these last are related to the appropriate actions capable of handling rp_i.

RS has a causative relationship with RP and a preventative relationship with A. Thus, the relationship between R and A can be formally modeled, using the three dependent sets RP, RS and A, through a model named "RSA" (Risk-Scenarios and related Actions). RSA is formally expressed as a weighted Tripartite-Graph $\mathbf{TG = (U_1, U_2, U_3, E_{U_1 \Rightarrow U_2}, E_{U_2 \Rightarrow U_3}, W_1, W_2)}$ (see Fig. 7), where:

- $U_1 = PR$, $U_2 = RS$ and $U_3 = A$ are the node partitions.
- $E_{U_i \Rightarrow U_{i+1}}$ is the set of edges connecting nodes of $U_i \times U_{i+1}$, for $i \in \{1, 2\}$.
- $W_i : U_i \times U_{i+1} \rightarrow \mathbb{R}^+$ is a weighting function that assign positive real values to edges of $E_{U_i \Rightarrow U_{i+1}}$, for $i \in \{1, 2\}$.

Next, $W_i(e_{n1,n2}) = w_{i_{n1,n2}}$ will denote the weight between (n_1, n_2), and $e_{n1,n2,n3} = (e_{n1,n2}, e_{n2,n3})$ will denote a valid composed edge linking 3 nodes $(u_{n1}, u_{n2}, u_{n3}) \in U_1 \times U_2 \times U_3$ such that $e_{n1,n2} \in E_{U_1 \Rightarrow U_2}, e_{n2,n3} \in E_{U_2 \Rightarrow U_3}$.

Fig. 7 illustrates an example of a TG. The three nodes rp_1, rs_1 and a_2 can have the following content in TG:

- rp_1: Traffic accident occurrence = "True"
- rs_1: (Atmosphere = "storm") $\wedge$ (Light = "night without public lightning") $\wedge$ (Road-type = "one way")
- a_2: "Limit speed to 40"

Given the composed edge $e_{1,1,2}$ connecting $(rp_1, rs_1, a_2) \in (U_1 \times U_2 \times U_3)$, both edges $e_{1,1}$ and $e_{1,2}$ are weighted by $w1_{1,1}$ and $w2_{1,2}$ respectively. $w1_{1,1}$ corresponds to how much rs_1 contributes in causing rp_1, while $w2_{1,2}$ corresponds to how much a_2 is effective in handling the damage of rs_1 (i.e, rp_1).

The weight functions are intended to identify high priority connections in TG. This will be helpful in the annotation of the interactions data of R and A. Several strategies can be adopted in order to define these weights. We restrict our self to two strategies, namely:

- Implicit Rating (IR). This technique suggests assigning the weight 1 to all edges.
- Explicit Rating (ER). This technique, on the other hand, suggests assigning weights in the interval [0, 1].

In the IR strategy, all the connections will have the same priority in TG, and interactions of R and A can have binary rating values $\in \{0, 1\}$. However, the ER strategy is more challenging as the definition of weights is based on an interpretation model that measure the impact of risk-scenarios in U_2. In fact:

- The Local Interpretable Model-Agnostic Explanations "LIME" model [72] is primarily designed to provide an interpretable explanation of a specific instance x according to its predicted target y by a black box model f, corresponding to eq. 2:

$$explanation(x) = argmin_{g \in G} \zeta(f, g, \pi_x) + \gamma(g) \quad (2)$$

where G is the class of explainer models (e.g linear or DT based models), $\gamma(g)$ is a complexity term (e.g., depth of tree, number of non-zero weights), π_x is a proximity measure (e.g., Euclidean distance) to define locality around x, and $\zeta(f, g, \pi_x)$ is a loss-like term to provide an optimal fit of g w.r.t f. This explanation is presented in form of contribution weights of each descriptive feature of x. In order to compute these weights, LIME, first perturbs the whole training set X to a new set Z, and make predictions y_Z of Z using f. Then, it assigns weights for instances $\in Z$ w.r.t their proximity with x. Finally, LIME trains a potentially interpretable model g on (Z, y_Z) to locally fit f.

In our case, we apply LIME on the traffic accidents training set, to obtain the local contribution weights of every feature to the prediction of u_{np} by a model m_i. These weights are then independently averaged to represent the overall contribution of each feature in m_i. In fact, for an edge $e_{np,nk} \in E_{U_1 \Rightarrow U_2}$, the weight $\mathbf{W_1(e_{np,nk})}$ is defined based on 1) a given traffic accident predictive model $m_i \in DM$, trained to predict $u_{np} \in U_1$, and 2) the "LIME" model. $\mathbf{W_1(e_{np,nk})}$ is then computed as the average contribution of atomic proposition of $u_{nk} \in U_2$, in the prediction of u_{np} in m_i. These values are then scaled to $[0, 1]$ and processed by an aggregation function F_1 (e.g, average, median) to provide a single value representing the final output of W_1, which will be $\mathbf{W_1(e_{np,nk}) = w1_{np,nk}}$.
- For an edge $e_{nk,nj} \in E_{U_2 \Rightarrow U_3}$, the weight $\mathbf{W_2(e_{nk,nj})}$ is computed as follows:

$$W_2(e_{nk,nj}) = F_2(\{W_1(e_{np,nk}) \mid e_{np,nk} \in E_{U_1 \Rightarrow U_2}\})$$

where, F_2 is an aggregation function (e.g, average, median). Note that all edges starting from U_2 have the same weights, and if there exits only one edge $e_{np,nk} \in E_{U_1 \Rightarrow U_2}$, then simply $w2_{nk,nj} = w1_{np,nk}$.

A concrete example of **TG** is illustrated in Fig. 12 (Appendix A), where W_1 and W_2 are defined by the ER technique, and the *average* functions F_1 and F_2. Given the following three nodes $(pr_1, pr_2, rs_1) \in U_1 \times U_1 \times U_2$, such that:

- rs_1: (School = "nearby") $\wedge$ (Holiday = "False") $\wedge$ (Day = "weekDay") $\wedge$ (Time slice = "school hours")
- pr_1: Traffic accident occurrence = "True"

- pr_2: Severity level of accident = "fatal"

We suppose that the two vectors $v_1 = [0.3, 0.4, 0.6, 0.7]$ and $v_2 = [0.3, 0.6, 0.8, 0.5]$, correspond to the scaled contribution weights of each atomic proposition of rs_1, in causing pr_1 and pr_2. Therefore, weights of edges $(e_{1,1}, e_{2,1}) \in E^2_{U_1 \Rightarrow U_2}$, are computed as follows:

- $W_1(e_{1,1}) = F_1(v_1) = 0.5$
- $W_1(e_{2,1}) = F_1(v_2) = 0.55$

Additionally, all edges $\{e_{1,i}\}, i \in \{1, 2\}$ connecting rs_1 to its related actions $a_i \in U_3$, hold the weight $W_2(e_{1,i}) = F_2(0.5, 0.55) = 0.525$.

2) Road-Segments Profiles: In recommendation systems, a user profile contains descriptive features that allow understanding why a user interacts with a given item. In our approach, we build profiles for R using the traffic accident data to reliably link the adequate actions of A. These profiles are denoted by $P = \{p_1, p_2, \ldots, p_m\}$. For a road-segment $r_i \in R$, its corresponding profile $p_i \in P$, expressed also as a propositional formula, can contain traffic accident descriptive features related to the environment, the involved victims, vehicles, etc. Note that, the risk-preconditions and risk-scenarios sets previously defined, are based on these features. Thus, P is crucial to allow implementing a context-aware recommendation framework.

3) Annotation of the Interaction Matrix I: In our approach, I contains interactions of m road-segments R and n defined actions A (see eq.1), and it is deduced from the RSA model. First I is initialized by 0. The annotation iterates over elements of R and A to deduce the rating values v_{ij} according to Algorithm 1. In fact, for $(r_i, a_j) \in R \times A$, v_{ij} will be updated if:

1) There exists a composed edge $e_{np,nk,nj} \in TG$ such as $a_j = u_{nj} \in U_3$
2) The precondition $u_{np} \in U_1$ is included in the profile p_i of r_i.
3) Each atomic proposition of the risk-scenario $u_{nk} \in U_2$ is included in p_i.

As several $e_{np,nk,nj} \in TG$, satisfying the three previous conditions, might exist, v_{ij} is updated using an aggregation function F_3 of the corresponding $W_2(e_{nk,nj})$ weights:

$$v_{ij} = F_3(\{W_2(e_{nk,nj}) \mid e_{np,nk,nj} \in TG\})$$

Note that, in this case, I supports both of the weight strategies previously defined (IR and ER).

Algorithm 1 Interaction matrix I annotation.

```
Input:
  • TG = (U1, U2, U3, E_{U1⇒U2}, E_{U2⇒U3}, W1, W2)
  • Profiles P = {p1, ..., pm} of R
Result: Complete annotation of I = (v_ij)_{i≤m, j≤n}
        summarizing all the interactions of R and A
Initialization: Initialize I to zeros I(1 : m, 1 : n) = 0
for i ∈ [1; m] do
    for j ∈ [1; n] do
        S = {}
        for e_{np,nk,nj} ∈ E(U1, U2, U3) where a_j = u_nj ∈ U3
        do
            if p_i match u_np ∈ U1 and u_nk ∈ U2 then
                Add W2(e_{nk,nj}) to S
            end
        end
        current v_ij = F3(S)
    end
end
```

V. Experiments

In this section, we focus on the implementation of the three layers of the AI core framework. Our objective is to give a clear demonstration of how research and practice in the field of road safety can be linked.

A. Traffic Accidents Datasets Description

As we previously mentioned, our framework can be considered as a strategic tool for building customized road safety action plans.

To demonstrate this, we adopt two real-world traffic accident datasets that have been collected from police reports. Both datasets have the same structure and descriptive features but are collected from different geographical areas: France and Morocco. They include accidents that have occurred on public roads and involved at least one vehicle and one injured victim.

- **The French dataset** is an open source traffic accident data provided by the National Inter Ministerial Observatory for Road Safety (OMNISR) of France [34]. It contains 900690 historical accidents' records from 2005 to 2017, covering the entire territory of France.
- **The Moroccan dataset** is a traffic accident data that we have been provided by the Moroccan Ministry of Equipment, Transport and Logistics (MMETL) to support our project. It contains 335899 road accidents' records from 2010 to 2014.

Both datasets are structured using 4 separate files: accident-characteristics file (16 descriptive features), accident-place file (18 descriptive features), involved-vehicles file (9 descriptive features) and involved-victims file (12 descriptive features). Each one summarizes the descriptive data of a particular context related to the occurred accidents using categorical and numerical features. Each data row in these files is attached to an accident-ID to uniquely index accidents data (see Table III).

Unlike the France dataset, the Moroccan dataset doesn't entirely cover all Moroccan territory and can't be used to provide reliable predictions. This dataset is also considered very challenging to process due to its huge amount of erroneous and missing values. Some of the features in the data have more than 70% of missing values which makes them non-usable. However, to the best of our knowledge, our paper is the first to introduce and analyses any Moroccan related traffic accident data. We aim by this work to clarify and encourage the Moroccan authorities to provide good quality

TABLE III
CATEGORIES OF TRAFFIC ACCIDENT DESCRIPTIVE FEATURES IN THE FRENCH AND MOROCCAN DATASETS

Characteristics	Place	Vehicles	Victims
Accident-ID	Accident-ID	Accident-ID	Accident-ID
Day	Roadway-category	Vehicle-identifier	Vehicle-identifier
Month	Numerical road-id	circulation-direction	Victim-place-in-vehicle
Year	Alphanumeric road-letter	Vehicle-category	Victim-category
Hour	road-number	Moving- obstacle	Injury-level
Minutes	NB. of circulation paths	Fixed-obstacle	Gender
Departement	Traffic-regime	Initial-shock-point	Birth-date
Town	Type of reserved-lane if exists	Maneuver-before-accident	Traveling-reason
GPS	road-gradient profile	NB. of occupants	Existence of security-equipment
Address	Plan-layout		Security-equipment usage
Agglomeration	Roadway-length		Road-localization of pedestrian
Weather-condition	Roadway-width		Pedestrian-action
Day-light	Surface-condition		Pedestrian-accompaniment
Intersection-type if exists	Infrastructure-type		
Collision-type	Roadway-part where the accident occurred		
	Near-schools indication		

traffic accident datasets for researchers in order to significantly enhance road safety in Morocco.

B. Preprocessing Layer

Missing data imputation: We adopt three techniques to handle missing entries in both datasets:

- "Imputation by prediction" using the LightGBM algorithm.
- Imputation based on other dependent features' values. For example missing entries of feature "day light", can be deduced from the corresponding values of "month" and "hour" features.
- Location features imputation using map-matching algorithms (OpenStreet Map).

Note that, features having more than 60% of missing values were removed.

Features engineering:

- Categorical features: We adopted the OHE technique for low cardinality features. For large features such as road-id, department-id and town-id, we applied the "TargetEncoder" (TE) technique. TE consists in replacing features with posterior probability of the target given a particular categorical value, and the prior probability of the target over all the training data.
- Numerical features: The Min-Max scaling technique is used to scale numerical features, such as "roadway length", "roadway width", and "number of circulation lanes", to a range of [0, 1].
- New features generation: This task was considered to optimally transform some features to a more meaningful content. Old features are mapped by "binary" or "grouped by" transformations techniques to create new ones. For example, "day", "month" and "year" were used to create three new features: "is holiday" (binary), "season" (spring, summer, fall, autumn and Winter) and "is weekend" (binary). Additionally, "hour" and "minutes" were grouped into 3-hours ranges, to create a new feature called "time-slice".
- Important features selection: We aim by this step to identify the top k pertinent features using the Sequential Forward features Selection "SFS" technique. SFS consists of starting the training with the best selected feature. Pairs with remaining features are then formed to select the best pair. This procedure is repeated for next triplets, and so on, until a predefined number (k) of features is selected.

Data sampling:

- Road-segments clustering: For the French dataset, a road-segment is defined by three features: town-id, department-id and road-id. For the Moroccan dataset, only two features are used: road-id and province-id. Therefore, a unique road-segment id, can be assigned to each data sample, having a specific accident-id in the database.
- Training and testing sets: Data is sampled based on the "year" feature. We can keep around 80% for training and 20% for testing sets. Accordingly, the sampling timeline of each dataset is as follows:
  - French training and testing sets: from 2005 to 2014 and from 2015 to 2017, respectively.
  - Moroccan training and testing sets: from 2010 to 2013 and 2014, respectively.

C. Traffic Accidents Predictive Models Layer

In order to significantly achieve the objective of this layer, we design a Dependency-Model that is composed of 9 traffic accident predictive models, denoted by $DM = \{M_i \mid i \in [1, 9]\}$ (see Fig. 8).

These models are activated for a given $r_i \in R$ if M_1 positively predicts the occurrence of a traffic accident. In what follows, we present a baseline of state-of-the-art ML and DL algorithms, and evaluation metrics, that can be adopted for building DM. For the sake of simplicity, only experiments of M_1 and M_9 are illustrated and discussed in details (see V-C.1 and V-C.2). However, a discussion is tackled in section V-C.3 to highlight the impact of the hierarchical structure of DM on its performance.

Algorithms baseline: we recommend the application of three categories of learning algorithms, Logistic Regression "LR", Decision Trees "DTs" and Artificial Neural Network ANN:

- LR model: LR estimates the probability of an event Y being positive or negative based on an input data X with k descriptive features [19]. Dependencies between X and Y are modeled in a linear combination and probabilities $P(Y|X)$ are approximated using the Sigmoid function as

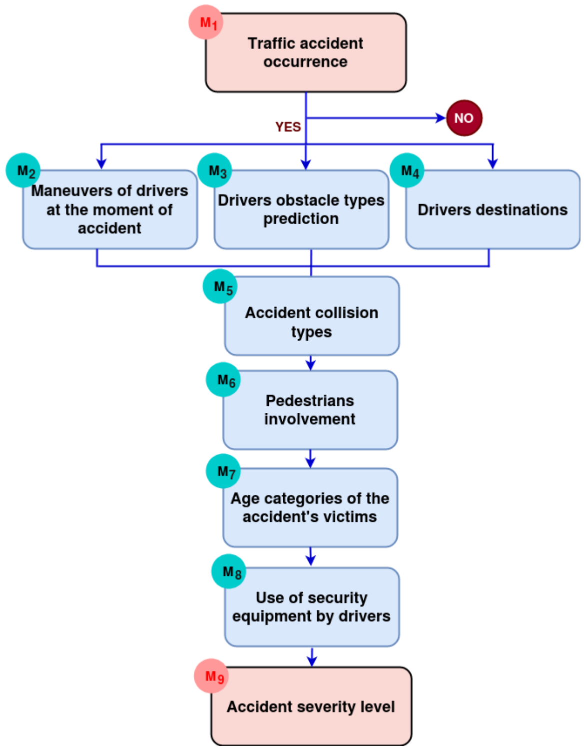

Fig. 8. Dependency-model $DM = \{M_i \,|\, i \in [1, 9]\}$, structured as a DAG, and designed for implementing and evaluating the second layer in the AI core of our proposed framework.

follows:

$$P(Y|X) = \frac{1}{1 + e^{-}(\beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k)} \tag{3}$$

β_0 is the bias and β_k are coefficients that the model learns for each column feature X_k in the input data X. In case of low dimensional data that is linearly separable, the LR model can be a perfect fit. In addition, it is easy to train and interpret, and less prone to overfitting.

- DT based models: We explore three DT based algorithms: DT, RF and LightGBM. DT model [20] builds a single tree using the entire training data. The tree is constructed in a greedy top-down recursive partitioning strategy. They are considered weak classifiers when compared to RF models and they are also very prone to overfitting. The RF model [54], on the other hand, are capable of reducing the high variance that can be seen in DTs by training on different data samples selected by the bagging technique. RF is an ensemble approach that builds in parallel a large number of trees with various features subsets selected randomly (bagging). The trees that perform the best are averaged at the end as a final robust predictor though achieving a higher accuracy than a single DT. LightGBM is a gradient boosting framework based on DT [55]. Learning in LightGBM is achieved recursively where each tree is trained to correct the errors of its predecessors. It can also use bagging and offers a lot of choice for parameter tuning when compared to RF and DT. In addition, it can be surprisingly faster than RF and adequate for large datasets but more likely to overfit.
- ANN models: ANN are combinations of computation and optimization functions that learn and model the dependencies between input and output data in a layered hierarchical process inspired from the human brain [57], [58]. In our case, we adopted three ANN based models to add to our comparative study baseline: MultiLayer Perceptron (MLP), 1D Convolutional Neural Network CNN [68] and Long Short Term Memory LSTM [56].
  MLP belongs to the category of feed-forward deep ANN that can be trained with "Backpropagation". Many variation of ANN were derived from the MPL model such as CNN. CNN introduced a new concept of layers called: convolution (performs dot products between a set of learnable filters and the input data on what they are slid) and pooling (down-sampling large data to avoid overfitting and for fast and pertinent training). This type of ANN was first designed to solve computer vision problems and then expended for processing 1D data. Neurons in CNN models are sparsely connected which makes it easier to process large data with high efficiency unlike MLP. In addition, 1D-CNN [68] requires less computations then MLP and are more suitable and highly recommended for real-time and low-cost applications.
  LSTM is a type of Recurrent Neural Nets (RNNs) that was first introduced in late' 90s to solve the vanishing gradient problem in RNN due to learning from long data sequences [56]. LSTM models are capable of learning long-term dependencies and are successfully applied for time-series data analysis. The main characteristic of LSTM neural nets is the memory cell in their hidden layers (instead of traditional neurons) which allows remembering only relevant information from previous sequences in the training phase.

Additional experiments were conducted to improve the models' performance, by using two Ensemble methods: Stacking and Hard-Voting.**Stacking** consists in combining predictions of multiple models to build a new input data for training a final model, called "combiner". In our case, we use LR as a combiner. **Hard-Voting**, on the other hand, consists in deciding on an instance label, by voting for the one being predicted by the majority of the different trained models.

Evaluation metrics: We adopt three evaluation metrics to measure the performance of models in DM: Precision, Recall and F1_score.

- Precision and Recall, measure the fraction of true positives among the total predicted positive cases and the fraction of true positives among the actual positive cases in the data respectively (see eq. 4 and eq. 5).

$$Precision = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \tag{4}$$

$$Recall = \frac{\text{True positives}}{\text{False negatives} + \text{False positives}} \tag{5}$$

- F1_score, is a weighted average of both Precision and Recall (see eq. 6).

$$F1\_score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

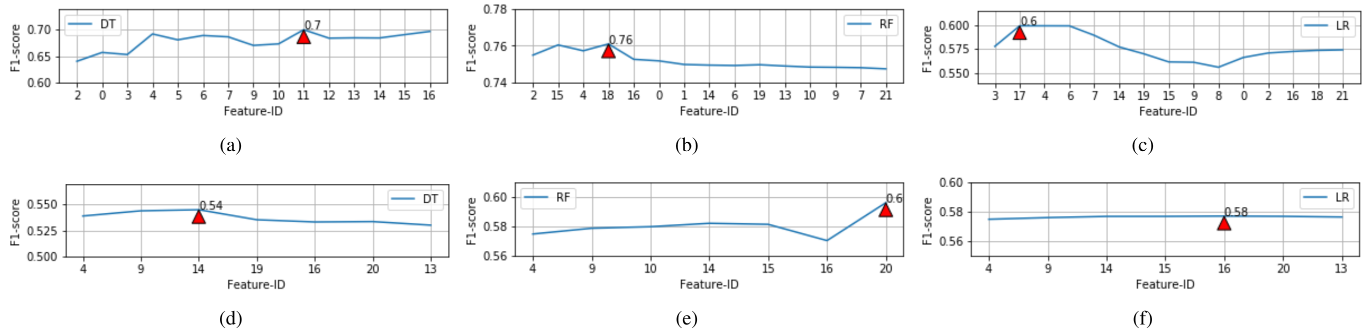

Fig. 9. Results of the SFS method to select the top k pertinent features of M_1 using DT, RF and LR for both the French (a), (b) and (c) and Moroccan (d), (e) and (f) datasets. Figures show the impact of each feature's addition on the F1-score, where the X-axis represents IDs the k selected features out of the 22 prior ones listed in Table IV.

Note that, the most appropriate approach to choose the best model is to consider an "optimizing" criterion while the rest remains "satisfactory". The choice of this criterion is totally task-dependent [74]. In our case, the "optimizing" criterion is precision, since our models' predictions will be employed for road safety actions recommendation. In fact, we need to carefully predict correct traffic accidents' occurrences, while False positive predictions might include unnecessary recommendations, which is time and money consuming for road safety actors.

1) **M_1** *Traffic Accident Occurrence Prediction Model:* We formulate the problem of traffic accident occurrence prediction, as a supervised binary classification task. M_1 estimates a binary target $y_i \in \{0, 1\}$ for a given $r_i \in R$ with a set of specific descriptive features (see Table IV), such that, 1 refers to an accident occurrence and 0 to no-accident. In order to generate data instances for the '0' class, we applied the process of the Algorithm 2. For each data sample $s_i, i \in [1, N]$, we randomly changes values of the feature $f =$ "hour" (resp. $f =$ "day"). Then, we correct the other features' values that might become erroneous after changing f:

- For $f =$ "hour", we correct "day light" feature, by defining for each of its categories the reasonably hours range.
- For $f =$ "day", we correct "is weekend" and "is holiday" features.

Finally, we check if the new s_i, denoted by s'_i, does not match any other sample in the whole database. If it true, we assign '0' class to s'_i and add it to the database.

Algorithm 2 Generating the no-accident class for M_1.

```
Input: Original traffic accident database of N samples.
Result: Generate new samples for the '0' no-accident
        class.
Initialization:
  • Add a new feature to store the target values ∈ {0, 1}.
  • Assign all original samples to the '1' class accident.
for a sample s_i, i ∈ [1, N] do
  for a feature f ∈ ("hour", "day") of s_i do
    Randomly change the value of f and set the new
    sample to s'_i
    Adapt values of the other features depending on
    the current value f
    if s'_i does not match any sample in the whole
    database then
      Add s'_i to the new database, with '0' target
    else
      Repeat
    end
  end
end
```

To overcome the issue resulting from class imbalance, we randomly removed instances from the dominant class using the under-sampling approach, to keep a 2 : 1 ratio. At the end, the sizes of the training and testing sets are: (1184886,87), (264876,87), (513943, 44), (131262, 44), for the French and Moroccan datasets, respectively.

For feature selection, we applied the SFS technique to discover the top k selected pertinent features of M_1. Fig. 9 illustrates the impact of each feature's addition on the model's performance using the F1-score. For the French dataset (see Fig. 9(a), 9(b) and 9(c)), the best performance was achieved using RF where F1-score=0.76 for $k = 4$ (see Fig. 9(b)). Accordingly, for the Moroccan dataset (see Fig. 9(d), 9(e) and 9(f)), the same algorithm achieved the highest F1-score=0.60 for $k = 7$ (see Fig. 9(e)). Yet, based on the overall findings, it is clearly noticed that an interesting part of the features was not selected and that would be a huge dropout for our case study. In other words, the action recommendation process requires the availability of a sufficient number of descriptive features of different categories. Taking $k = 7$ or $k = 4$, the implementation of the third layer would be inconvenient. Thus, we decide to keep all of the 22 prior selected features (see Table IV) to train M_1 (same for all models of DM) and we work on other strategies to maximize the quality of the training.

Training: M_1 is trained using LR, DT and ANN: for tuning LR and DT, grid-search was used with 6-folds cross-validation, and for ANN, small architectures were designed to avoid overfitting.

Results: Table V summarizes results of M_1 for both datasets. For the French dataset, we generally observe that

TABLE IV
SELECTED FEATURES FOR MODEL $M_1 \in DM$

Features of the traffic accident occurrence prediction model M_1	
0-Agglomeration	11-Month-season
1-Town	12-Type of reserved-lane if exists
2-Departement	13-Surface-condition
3-Nearby-schools	14-Road-gradient profile
4-Is-Holiday	15-Plan-layout
5-Roadway-length	16-DayLight
6-Roadway-Width	17-Traffic-regime
7-NB. of circulation paths	18-Roadway-category
8-Roadway-id	19-Weather-condition
9-Is-WeekEnd	20-Infrastructure-type
10-Time-slice	21-Intersection-type

TABLE V
RESULTS OF M_1 FOR THE MOROCCAN AND THE FRENCH DATASETS

Model	Dataset					
	France			Morocco		
	F1_score	Precision	Recall	F1_score	Precision	Recall
DT	0.72	0.71	0.74	0.54	0.58	0.50
RF	0.78	**0.94**	0.67	0.56	0.64	0.50
LightGBM	0.78	**0.94**	0.66	0.60	0.55	0.66
LR	0.71	0.65	0.79	0.66	0.99	0.49
MLP	0.71	0.76	0.67	0.65	**1.0**	0.48
1D-CNN	0.70	0.71	0.69	0.65	**1.0**	0.48
LSTM	0.68	0.70	0.64	**0.75**	**1.0**	0.61
Stacking	**0.81**	0.76	0.87	0.71	0.74	**0.68**
Hard-Voting	0.79	0.70	**0.91**	0.70	0.69	**0.72**

DT based models outperform the LR and ANN models. In particular, the overall best F1_score (0.78) among these last, is evenly achieved by the RF and LightGBM models. They additionally achieved the best precision (0.94) which is highly encouraged in our case study. This quality performance might be due to the ensemble learning technique in both models, where multiple trees are trained and averaged for final predictions. Also, these models are more likely to have low bias and variance and easy to tune for dealing with imbalanced data. Still, we can notice a good balance for the precision and recall trade-off in LR and ANN models. Moving to our attempt for increasing the F1_score, we can see that both Stacking and Hard-Voting achieved the highest F1_scores (0.81 and 0.79). However, we still consider RF and LightGBM as the most accurate models based on their precision. Experiments on the Moroccan dataset did not confirm the same findings. ANN models are found to outperform the LR model which in its turn surpassed DT based models. The quality and nature of the training data demonstrate their impact on the models learning. We can see that LSTM achieved the highest F1_score (0.75). Also, the best Precision value (1.0) is achieved by all NN models. In addition, it is clearly noticed that, for this dataset, maximizing the precision dramatically impact the recall. On the other hand, we can say that Stacking and Hard-voting did provide a good performance even F1_score was not improved, since the precision and recall trade-off was interestingly handled. Overall, we can select Stacking as the best model since its evaluation results are more reliable for our case-study (even though high precision is encouraged, recall still needs to exceed a minimum threshold of 50%).

TABLE VI
ADDITIONAL SELECTED FEATURES FOR MODEL $M_9 \in DM$

Additional features of the traffic accident severity level model M_9
Drivers obstacle types
Manoeuvres of drivers at the moment of the accident
Drivers destinations
If pedestrians are involved
Age categories of the victims
Vehicles collision type
If at least one driver was using a security equipment

TABLE VII
RESULTS OF M_9 FOR THE MOROCCAN AND FRENCH DATASETS

Dataset	Models	F1_score		Precision		Recall	
France	DT	0.68	0.58	0.77	0.50	0.61	0.68
	RF	0.68	0.66	0.69	0.65	0.66	0.68
	LightGBM	0.71	0.65	**0.79**	0.58	0.65	**0.73**
	Stacking	**0.75**	**0.67**	0.67	**0.79**	**0.84**	0.58
	Hard-Voting	0.74	0.65	0.66	0.77	0.83	0.58
Morocco	DT	0.97	0.98	0.95	0.96	**1.0**	**1.0**
	RF	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
	LightGBM	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**

According to the state-of-art, deep learning approaches are not always a good fit for tabular (structured) data [75], whereas, decision tree based algorithms are usually more efficient. In our case, this can be mainly explained by the heterogeneity of the features coming from various unrelated sources, where most of them are categorical. In other words, the processing of these features leads to a high-dimensional data space that is generally not dense and continuous, which it is quite difficult to exploit for ANN models. Having the sparsity of our data as the key factor for the performance of ANN models, we can understand why these last outperformed for the Moroccan dataset and not for the French one.

2) **M9***: Severity Level Classification Model:* M_9 is also formulated as a binary classification problem. It is modeled to classify the accident's severity level as minimal '0' or fatal '1'. Based on the architecture of DM, M_9 takes as input a given $r_i \in R$ with a set of descriptive features (see Table IV) and all predictions of its precedent models (see Table VI). The target of M_9 is created based on "the total number of injuries per category" of each accident. In both datasets, injuries are classified into 4 categories: minor, minimal, major and fatal. Accidents having only minimal and minor categories for their victims' injuries, are affected to class '0' and for others with at least one injury categorized as major or fatal are affected to class '1'. The class '0' is found to be highly dominated. To overcome this issue, we applied the random over-sampling technique by generating new samples for the minority class. This method has interestingly contributed in increasing the performance of the tested models.

Finally, the corresponding sizes of training and testing sets of M_9 are: (892086,117), (223022,117), (361628, 55), (93167,55) for the French and the Moroccan datasets, respectively.

Results: Table VII summarizes the results of M_9 for both classes. For the French dataset, we observe, among the DT based models, that LightGBM achieved the highest F1_score and precision for class '0' (0.71 and 0.79), and the highest recall score for class '1' (0.73). On the other hand, the RF model surpassed the other two models by obtaining the best F1_score and precision values for class '1' (0.66 and 0.65) and the best recall value for class '0' (0.66). The findings of the DT model are in line with the overall results. Mainly, all of the 3 models outperformed for class '0' even though both classes are equally balanced. This is can be explained by the random samples generated for class '1' which might not have been all useful or significant for learning. As for the Stacking and Hard-Voting models, our attempt for increasing F1_score was successfully achieved (0.75 and 0.74). However, in terms of precision, these models interestingly outperformed for class "1", whereas, for class "0", we can notice high recall values (0.84 and 0.83). Models for the Moroccan dataset, on the other hand, highlight superior results for both classes when compared to the French models. The training Moroccan data of M_9 is less complex in term of sparsity which has undoubtedly contributed in achieving almost 1.0 evaluation scores for both classes.

3) Robustness of **DM** *Vs Its Hierarchical Structure:* As we previously mentioned, structuring DM as DAG can weakness its performance or completely introduce erroneous predictions (see section III-B.2). This is most likely to happen if models of DM are trained separately. In our experiments, we followed this strategy by training each $M_i \in DM$ on DT based algorithms. Note that, we adopt the same parameters tuning and training/evaluation approach used for building M_1 and M_9. Table VIII, summarizes results of the best models trained on the French dataset. We base our discussion on this last, since it is sufficiently large and it includes all of the necessary descriptive features to construct DM. In order to evaluate the impact of the models' inner-dependencies on their performance, we consider three test sets. 1) raw data only, 2) raw data + predictions of the previous layer and 3) raw data + predictions of all previous layers. Starting from the third layer of DM (see Fig. 8), we can notice a slight decrease in the weighted F1_scores when previous predictions are considered. For example, M_8 achieved a weighted F1_score=0.84 for raw data and it maintained almost the same performance when only predictions of one previous model are considered. On the other hand, we can clearly notice that when predictions of all previous models are concatenated, an interesting impact on M_8 is shown with an F1_score=0.75. This new score is fair enough when we note the error probabilities of previous models. Overall, we can say that our DM is not highly sensitive to its hierarchical structure and it is capable of providing reliable predictions as long as it is carefully built. Yet, it can be enhanced by multiple other techniques and approaches in the state-of-art [76].

TABLE VIII

RESULTS OF MODELS $M_i \in DM$ FOR THE FRENCH DATASET. MODELS ARE EVALUATED BASED ON RAW DATA, ON CONCATENATION OF BOTH RAW DATA AND PREDICTIONS (ONE COLUMN) OF ONLY THE ANTECEDENT MODEL M_{i-1} AND FINALLY ON CONCATENATIONS OF RAW DATA AND ALL PREDICTIONS OF PREVIOUS MODELS $\sum_{n=1}^{8-i} M_{i-n}$

		Weighted F1-score		
M_i	Best model	raw data	+predictions of M_{i-1}	+Predictions of $\sum_{n=1}^{8-i} M_{i-n}$
M_1	Stacking	0.81	–	–
M_2	LightGBM	0.81	–	–
M_3	LightGBM	0.84	–	–
M_4	RF	0.75	–	–
M_5	RF	0.67	$+M_2$:0.65 $+M_3$:0.64 $+M_4$:0.66	0.63
M_6	LightGBM	0.77	$+M_2$:0.75 $+M_3$:0.74 $+M_4$:0.75 $+M_5$:0.72	0.74
M_7	RF	0.72	$+M_2$:0.71 $+M_3$:0.71 $+M_4$:0.72 $+M_5$:0.68 $+M_6$:0.70	0.67
M_8	RF	RF: 0.84	$+M_2$:0.82 $+M_3$:0.83 $+M_4$:0.81 $+M_5$:0.80 $+M_6$:0.81 $+M_7$:0.78	0.75
M_9	Stacking	0.71	$+M_2$:0.69 $+M_3$:0.70 $+M_4$:0.66 $+M_5$:0.68 $+M_6$:0.69 $+M_7$: 0.71 $+M_8$: 0.67	0.65

D. Road Safety Actions Recommendation Layer

The third layer, is implemented by collaborating with road safety experts from the Moroccan national road accident prevention community. Following the previously discussed approach (see section IV), our objective is to model a hybrid context-aware recommendation system, for road safety actions recommendation. This system is experimented for each of the preprocessed datasets.

As it is shown in Fig. 6, the interactions data for training this system is generated through three main steps:

1) RSA model definition as a Tripartite graph $TG = (U_1, U_2, U_3, E_{U_1 \Rightarrow U_2}, E_{U_2 \Rightarrow U_3}, W_1, W_2)$. U_1 is composed of the various predictions of $DM = \{M_i\}, i \in [1, 9]$ (38 precondition in total). For U_2 and U_3, we define 20 risk-scenarios and 50 road safety actions, respectively (see Table XI, Appendix B). Note that, the three sets are valid for annotating interactions data of both datasets. Moreover, the weighting functions W_1 and W_2 are defined based on the IR strategy.
2) Road-segments profiles P definition using the selected descriptive features for training DM (see Tables. IV and VI). Note that, categorical features are explicitly kept without preprocessing. They are afterwards handled in the training process.
3) We aim to experiment training the recommendation system with binary rates $\in \{0, 1\}$. Thus, we can choose to annotate the interaction matrix I by using $max()$

TABLE IX
STATISTICS OF THE ACTIONS' RATES DISTRIBUTION

		NB.rates/road-segment			NB.road-segments/action		
Dataset	NB.rates	Min	Average	Max	Min	Average	Max
France	2008776	1	2	15	24	40175	660925
Morocco	232228	0	2	10	1	10096	122639

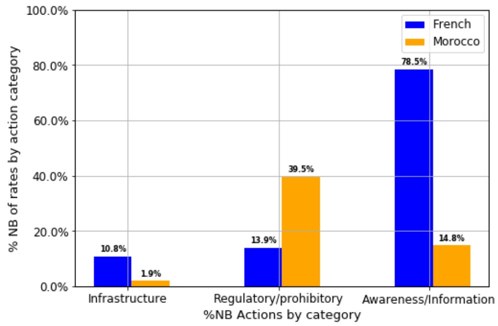

Fig. 10. Distribution of actions' rates % by category.

function as the aggregation function F_3. According to road safety experts, this can allow us to maintain the same level of priority for all the positively rated actions. However, other aggregation functions can be employed for this task such as $average()$, whereas a rate's threshold can be set to convert values into $\{0, 1\}$. Moreover, in order to reduce the sparsity of I, only actions having at least one positive interaction in the whole data, are considered. Though, sizes of I for the French and Moroccan datasets are (1115108,50) and (454795,24) respectively.

This system is formulated as a multi-label classification problem, in order to adopt AI methods capable of learning recommendations based on the descriptive profiles P. For a given $(r_i, p_i) \in R \times P$, the system outputs n binary rating values as a target, where $n = 50$ for the French system and $n = 24$ for the Moroccan system.

Before presenting results of our experimentation, we summarize in Table IX statistics of the actions' rate distribution over the road-segments, for both datasets. In addition, Fig. 10 illustrates the distribution of the action rates by category. As it is shown, the least rated category is "Infrastructure" which also has the least number of actions in the RSA model (see Appendix B). Fig. 11, on the other hand, shows the percentage of actions having rate frequencies $< 1\%$, in $]1\%, 5\%[$ or in $[5\%, 52\%]$. The French dataset has a balanced distribution, since 80% of its 50 actions have almost the same rate frequencies. For the Moroccan dataset, we can notice a very slight imbalance between its 24 actions, where 40% have frequencies $> 5\%$.

Training: We design an MLP-Embedding architecture composed of Embedding, Concatenation, Dense, Dropout, Activation and Fully-connected layers. Note that, for each categorical features in P, an Embedding layer was dedicated, and the sizes of the training and testing sets are the same as in M_9. For performance evaluation, three metrics are adopted:

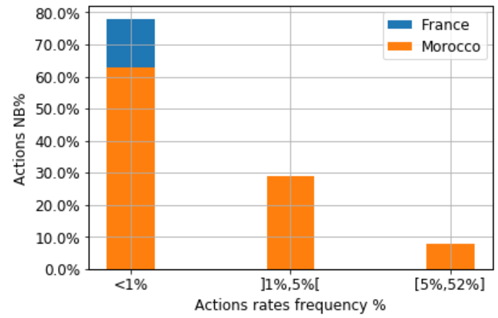

Fig. 11. Distribution of NB% actions according to three rate frequency intervals.

TABLE X
RESULTS OF THE ROAD SAFETY ACTIONS RECOMMENDATION SYSTEM. P@K AND R@K ARE EVALUATED FOR THREE K LEVELS: FOR FRANCE K=2, K'=15 AND N=50, WHILE FOR MOROCCO K=2, K'=10 AND N=24

Model	Dataset	ROC-AUC	P@2	R@2	P@k'	R@k'	P@n	R@n
MLP-Embedding	France	0.93	0.91	0.92	0.87	0.93	0.76	0.95
	Morocco	0.96	0.96	0.97	0.89	0.97	0.81	0.99

- Area Under the Receiver Operating Characteristic Curve (ROC-AUC) score. It is computed based on results of the n recommendations [35].
- Precision at k (P@k), to demonstrate the proportion of actual relevant items of a user, in the top-k recommendations (see eq. 7).

$$P@k = \frac{\text{Relevant actions Recommended in top-k}}{\text{Recommended actions}} \quad (7)$$

- Recall at k (R@k), to demonstrate the total number of relevant items appearing in the top-k recommendations (see eq. 7).

$$R@k = \frac{\text{Relevant actions Recommended in top-k}}{\text{Relevant actions}} \quad (8)$$

Results: The performances of the French and Moroccan models are interestingly good, achieving **ROC-AUC scores of 0.93 and 0.96**, respectively. In addition, a significant precision/recall trade-off balance is achieved by both models (see Table X). However, we can notice that, for both models, P@k has a considerable dropout for larger k values, precisely for k=n (where n is 50 for France and 24 for Morocco). This is quit expected when considering the statistics of the action rates in both sets (see Table IX). Overall, these findings prove that the recommendation system successfully learnt the linking relationship between the appropriate actions and the traffic accident latent features. Thus, we can conclude that our approach can certainly help in building action recommendation systems that are robust, reliable and highly precise.

VI. CONCLUSION

This paper introduces an innovative and a newly designed framework for recommending road safety actions based on traffic accident data analysis. It is intended to be used by road safety authorities as a tool for developing customized and

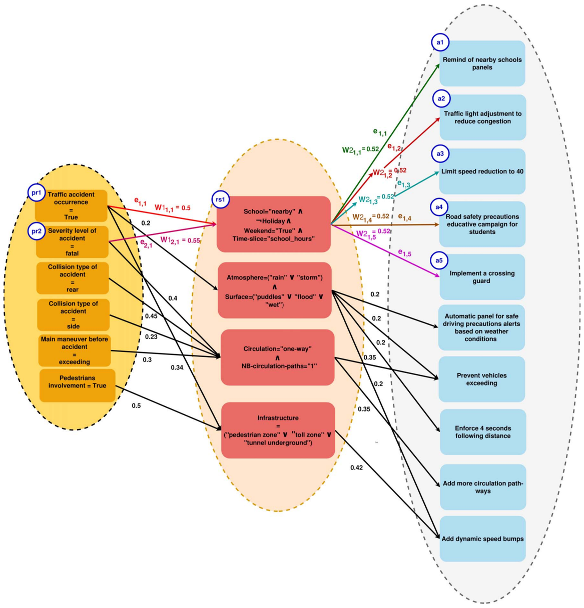

Fig. 12. A concrete example of an RSA model structured as TG, where W_1 and W_2 are defined based on the ER technique.

TABLE XI
COMPONENTS OF THE RSA MODEL DEFINED IN OUR EXPERIMENTS FOR BOTH MOROCCAN AND FRENCH DATASETS

Preconditions	Risk-scenarios	Actions
1. Traffic accident occurrence = True	1. ¬Weekend ∧ Time-slice="school-hours" ∧ ¬Holiday ∧ school="nearby" ∧ Season-month="winter∨spring∨fall"	1. Add road lights
2. Severity level of accident = fatal∨major	2. Atmosphere="snow" ∧ Surface="icy"	2. Add shelters for cyclists
3. Severity level of accident = minimal∨minor	3. Light="night-with-public-lighting-not-lit"	3. Remind of using vehicle lights using automatic panels
4. Pedestrians involvement = True	4. Light="night-without-public-lighting"	4. Opposite circulation direction for cyclists
5. Age category of victims = "children"	5. Surface="puddles∨flood∨wet" ∧ Atmosphere="rain∨storm"	5. Add automatic barrier
6. Age category of victims = "adolescent"	6. Agglomeration="in-city" ∧ Time-slice="rush-hours"	6. Separate road with a median strip
7. Age category of victims = "old"	7. Infrastructure="pedestrian-crossing"∨"toll-zone"∨"tunnel"	7. Safe driving precautions alert based on weather conditions using automatic panels
8. Use of security equipment = True	8. Circulation="bidirectional" ∧ Width of solid-central="0"	8. Add tunnel lights
9. Use of security equipment = False	9. Atmosphere="strong-wind"	9. Campaigns to promote safe crossing behavior
10. Accident obstacle = "fixed"	10. Surface="greasy"	10. Add stop sign
11. Accident obstacle = "human"	11. Intersection="T"	11. Widen the roadway
12. Accident obstacle = "vehicle"	12. Intersection="X"∨"4"∨"Y"	12. Provide on site equipped ambulances
13. Accident obstacle = "wild animal"	13. Circulation="with-variable-assignment"	13. Add dynamic speed bumps
14. Driver's destination = "home-work"	14. Reserved-lane="bicycle-bank" ∧ Agglomeration="in-city"	14. Turn on road lights
15. Driver's destination = "shopping"	15. Circulation="one-way" ∧ Agglomeration="in-city" ∧ NB-circulation-paths="1"	15. Add road studs
16. Driver's destination = "professional use"	16. ¬ (Infrastructure = "pedestrian-crossing")	18. Sign of pedestrian crossing
19. Maneuver of driver = "normal"	19. Circulation="two-way" ∧Atmosphere="rain" ∧ Agglomeration="in-city" ∧ "Time-slice=rush-hours"	19. Road safety precautions educative campaign for students
20. Maneuver of driver = "turning left∨right"	20. Day-light="night" ∧Season-month="summer" ∧ ¬Weekend ∧ Agglomeration="in-city"	20. Educative campaign road safety knowledge (street demonstration)
21. Maneuver of driver = "crossing the central reservation"		21. Stop sign in front of schools
22. Maneuver of driver = "deported to the left∨right"		22. Add pedestrian crossing
23. Maneuver of driver = "stopped (except parking)"		23. Intersection sign
24. Maneuver of driver = "against the direction"		24. Enforce 4 seconds following distance
25. Maneuver of driver = "crossing the road"		25. Inform nearby hospitals panels
26. Maneuver of driver = "avoidance"		26. Fix water channels
27. Maneuver of driver is "turning around on the road"		27. Road sign of assignment channels
28. Maneuver of driver is "change lane right/left"		28. Adding roundabout
29. Maneuver of driver is "exceeding"		29. Plan police barrages for alcohol test and driving licence control'
30. Maneuver of driver is "in the bus corridor"		30. Campaigns to promote the use of security equipments while driving'
31. Maneuver of driver is "backwards"		31. Sign of wild animals
32. Maneuver of driver is "parking"		32. Limit speed to 20/40 nearby schools
33. Collision type is "2 vehicles by side"		33. Crash caution messages on automatic panels
34. Collision type is "2 vehicles in front"		34. Provide salt spreader and snow ploughs machines
35. Collision type is "2 vehicles from back"		35. Tar road reconstruction
36. Collision type is "multiple vehicles in chains"		36. Reduce speed limit
37. Collision type is "without collision"		37. Enforce priority traffic signs
38. Collision type is "random"		38. Use of roads sweeper machine to remove grease/oil
		39. Traffic light adjustment
		40. Early traffic panels to remind of deviation road
		41. Separate bicycle bank from road with a median strip
		42. Add more circulation paths
		43. Sign of near curve
		44. Warn of predicted atmosphere using automatic panel
		45. fix road lights
		46. Crossing guards
		47. Traffic policeman
		48. Remind of nearby schools: Panel
		49. Road sign to remind of infrastructure type
		50. Road sign of bicycle bank

reliable actions plans. Subsequently, road safety can be significantly improved without neglecting the dynamic changes of the road-network. Our framework remarkably achieved high ROC-AUC scores $>= 0.93$ for the recommendation system and interesting F1-scores $\in [0.67, 1.0]$ for the experimented traffic accident predictive models. These findings show the possibility of promising real-world applications of the proposed approach. In addition, this work is considerably informative for filling the gap of the research-practice relationship in the road safety domain. We highly encourage road safety actors of all categories, to provide open-source datasets of the actions' application history. This type of data will certainly help researchers to fruitfully experiment the implementation and the enhancement of our framework.

APPENDIX A

See Fig. 12.

APPENDIX B

See Table. XI.

ACKNOWLEDGMENT

The authors would like to thank the Moroccan Ministry of Equipment, Transport and Logistics, the "Moroccan Foundation of Advanced Science, Innovation and Research" (MAsCIR) and the "Mohammed VI Polytechnic University" UM6P as well, for sponsoring this work. They would also like to thank the anonymous reviewers for their insightful suggestions and comments that definitely helped improving the clarity of our article.

REFERENCES

[1] World Health Organization. *Road Safety Actions Decade 2011–2020*. Accessed: Mar. 2010. [Online]. Available: https://www.who.int/roadsafety/decade_of_action/plan/plan_english.pdf

[2] H. Ren, Y. Song, J. Wang, Y. Hu, and J. Lei, "A deep learning approach to the citywide traffic accident risk prediction," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3346–3351.

[3] L. Wahab and H. Jiang, "A comparative study on machine learning based algorithms for prediction of motorcycle crash severity," *PLoS ONE*, vol. 14, no. 4, Apr. 2019, Art. no. e0214966.

[4] C. Caliendo, M. Guida, and A. Parisi, "A crash-prediction model for multilane roads," *Accident Anal. Prevention*, vol. 39, no. 4, pp. 657–670, Jul. 2007.

[5] S. H. Park and Y. G. Ha, "Large imbalance data classification based on mapreduce for traffic accident prediction," in *Proc. 8th IEEE Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput.*, Jul. 2014, pp. 45–49.

[6] M. A. Quddus, "Time series count data models: An empirical application to traffic accidents," *Accident Anal. Prevention*, vol. 40, no. 5, pp. 1732–1741, Sep. 2008.

[7] J. de Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks," *Accident Anal. Prevention*, vol. 43, no. 1, pp. 402–411, Jan. 2011.

[8] B. Selman and C. P. Gomes, "Hill-climbing search," in *Encyclopedia of Cognitive Science*. Hoboken, NJ, USA: Wiley, 2006. [Online]. Available: https://onlinelibrary.wiley.com/

[9] W. Zhang, K. Wang, L. Wang, Z. Feng, and Y. Du, "Exploring factors affecting pedestrians' red-light running behaviors at intersections in China," *Accident Anal. Prevention*, vol. 96, pp. 71–78, Nov. 2016.

[10] Á. Briz-Redón, F. Martínez-Ruiz, and F. Montes, "Estimating the occurrence of traffic accidents near school locations: A case study from valencia (Spain) including several approaches," *Accident Anal. Prevention*, vol. 132, Nov. 2019, Art. no. 105237.

[11] Y.-R. Shiau, C.-H. Tsai, Y.-H. Hung, and Y.-T. Kuo, "The application of data mining technology to build a forecasting model for classification of road traffic accidents," *Math. Problems Eng.*, vol. 2015, pp. 1–8, Jun. 2015.

[12] J. Pearl and T. Verma, "The logic of representing dependencies by directed graphs," Dept. Comput. Sci., Univ. California, Los Angeles, CA, USA, Tech. Rep., 1987, pp. 374–379.

[13] R. Elvik *et al.*, "Vehicle design and protective devices," in *The Handbook of Road Safety Measures*. Bradford, U.K.: Emerald Insights, 2009, pp. 543–731. [Online]. Available: https://www.emerald.com/insigh

[14] H. Martensen *et al.*, "The European road safety decision support system on risks and measures," *Accident Anal. Prevention*, vol. 125, pp. 344–351, Apr. 2019.

[15] M. A. Abdel-Aty and A. E. Radwan, "Modeling traffic accident occurrence and involvement," *Accident Anal. Prevention*, vol. 32, no. 5, pp. 633–642, Sep. 2000.

[16] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives," *Transp. Res. A, Policy Pract.*, vol. 44, no. 5, pp. 291–305, Jun. 2010.

[17] X. Zhang, S. T. Waller, and P. Jiang, "An ensemble machine learning-based modeling framework for analysis of traffic crash frequency," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 3, pp. 258–276, Mar. 2020.

[18] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proc. 5th Berkeley Symp. Math. Statist. Probab.*, 1967, vol. 1, no. 14, pp. 281–297.

[19] H. Swaminathan and H. J. Rogers, "Detecting differential item functioning using logistic regression procedures," *J. Educ. Meas.*, vol. 27, no. 4, pp. 361–370, Dec. 1990.

[20] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, no. 3, pp. 660–674, May/Jun. 1991.

[21] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017.

[22] G. E. P. Box, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.

[23] H. Ezra, "On the relationship between road safety research and the practice of road design and operation," *Accident Anal. Prevention*, vol. 128, pp. 114–131, Jul. 2019.

[24] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy K-nearest neighbor algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 4, pp. 580–585, Aug. 1985.

[25] L. Kuang, H. Yan, Y. Zhu, S. Tu, and X. Fan, "Predicting duration of traffic accidents based on cost-sensitive Bayesian network and weighted K-nearest neighbor," *J. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 161–174, Mar. 2019.

[26] T. Chanpanit, N. Arkamanont, and N. Pranootnarapran, "Predicting the number of people for road traffic accident on highways by hour of day," in *Proc. 8th IEEE Int. Conf. Ind. Technol. Manage. (ICITM)*, Mar. 2019, pp. 285–289.

[27] A. Abdulhafedh, "Road traffic crash data: An overview on sources, problems, and collection methods," *J. Transp. Technol.*, vol. 7, no. 2, pp. 206–219, 2017.

[28] M. Imprialou and M. Quddus, "Crash data quality for road safety research: Current state and future directions," *Accident Anal. Prevention*, vol. 130, pp. 84–90, Sep. 2019.

[29] Z. Zhang, Q. He, J. Gao, and M. Ni, "A deep learning approach for detecting traffic accidents from social media data," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 580–596, Jan. 2018.

[30] R. Deb and A. W.-C. Liew, "Missing value imputation for the analysis of incomplete traffic accident data," in *Proc. Int. Conf. Mach. Learn. Cybern.* Berlin, Germany: Springer, 2014, pp. 275–286.

[31] A. N. Baraldi and C. K. Enders, "An introduction to modern missing data analyses," *J. School Psychol.*, vol. 48, no. 1, pp. 5–37, Feb. 2010.

[32] R. Deb and A. W.-C. Liew, "Noisy values detection and correction of traffic accident data," *Inf. Sci.*, vol. 476, pp. 132–146, Feb. 2019.

[33] M. Haklay and P. Weber, "OpenStreetMap: User-generated street maps," *IEEE Pervasive Comput.*, vol. 7, no. 4, pp. 12–18, Oct. 2008.

[34] *Traffic Accident Dataset of France*. Accessed: Jul. 8, 2013. [Online]. Available: https://www.data.gouv.fr/fr/datasets/base-de-donnees-accidents-corporels-de-la-circulation/

[35] *Computation of the ROC AUC Score by the Scikit-Learn Package*. Accessed: 2007. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html

[36] M. A. Quddus, W. Y. Ochieng, and R. B. Noland, "Current map-matching algorithms for transport applications: State-of-the art and future research directions," *Transp. Res. C, Emerg. Technol.*, vol. 15, no. 5, pp. 312–328, Oct. 2007.

[37] World Health Organization. (2018). *Global Status Report on Road Safety 2018*. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/277370/WHO-NMH-NVI-18.20-eng.pdf?ua=1

[38] World Health Organization. (2004). *World Report on Road Traffic Injury Prevention*. [Online]. Available: https://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/en/

[39] A. S. Hakkert and V. Gitelman, "Thinking about the history of road safety research: Past achievements and future challenges," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 25, pp. 137–149, Jul. 2014.

[40] National Interdepartmental Observatory of Road Safety. (Jan. 2019). *Barometer of the National Interdepartmental Observatory of Road Safety in France*. [Online]. Available: https://www.interieur.gouv.fr/Media/MI/Files/Communiques/Barometre-du-mois-de-janvier-2019-de-l-Observatoire-national-interministeriel-de-la-securite-routiere-ONISR

[41] B. P. Hughes, S. Newstead, A. Anund, C. C. Shu, and T. Falkmer, "A review of models relevant to road safety," *Accident Anal. Prevention*, vol. 74, pp. 250–270, Jan. 2015.

[42] C. Seger, "An investigation of categorical variable encoding techniques in machine learning: Binary versus one-hot and feature hashing," Dept. EECS, KTH, Stockholm, Sweden, Tech. Rep., 2018. [Online]. Available: https://www.kth.se/en/eecs/om-oss/elektroteknik-och-datavetenskap-1.769578

[43] International Transport Forum. (Nov. 2008). *Towards Zero, Road Safety Approach*. [Online]. Available: https://www.itf-oecd.org/content/free-2

[44] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, "Big data preprocessing: Methods and prospects," *Big Data Anal.*, vol. 1, no. 1, p. 9, Dec. 2016.

[45] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3017–3032, Dec. 2015.

[46] W. Ge, R. T. Collins, and R. B. Ruback, "Vision-based analysis of small groups in pedestrian crowds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 1003–1016, May 2012.

[47] T. Gandhi and M. M. Trivedi, "Pedestrian protection systems: Issues, survey, and challenges," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 413–430, Sep. 2007.

[48] E. Papadimitriou, G. Yannis, and J. Golias, "A critical assessment of pedestrian behaviour models," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 12, no. 3, pp. 242–255, May 2009.

[49] A. Lugmayr, "Children road safety training with augmented reality (AR)," in *Proc. Int. Inf. Syst. Manage. Creative eMedia (CreMedia)*, 2018, pp. 40–42.

[50] A. E. Mendoza, C. A. Wybourn, M. A. Mendoza, M. J. Cruz, C. J. Juillard, and R. A. Dicker, "The worldwide approach to vision zero: Implementing road safety strategies to eliminate traffic-related fatalities," *Current Trauma Rep.*, vol. 3, no. 2, pp. 104–110, Jun. 2017.

[51] A. M. de Souza, C. A. Brennand, R. S. Yokoyama, E. A. Donato, E. R. Madeira, and L. A. Villas, "Traffic management systems: A classification, review, challenges, and future perspectives," *Int. J. Distrib. Sensor Netw.*, vol. 13, no. 4, Apr. 2017, Art. no. 155014771668361.

[52] A. Parashar, V. Maheshwari, and A. Shrivastava, "Enhanced n-wheeler post accidental rescue system connecting to the correct person at the correct time," in *Proc. IEEE Region 10 Hum. Technol. Conf. (R10-HTC)*, Dec. 2016, pp. 1–4.

[53] R. Elvik, "To what extent can theory account for the findings of road safety evaluation studies?" *Accident Anal. Prevention*, vol. 36, no. 5, pp. 841–849, Sep. 2004.

[54] L. Breiman, "Random forests, machine learning 45," *J. Clin. Microbiol.*, vol. 2, no. 30, pp. 199–228, 2001.

[55] G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3146–3154.

[56] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[57] K. Gurney, *An Introduction to Neural Networks*. Boca Raton, FL, USA: CRC Press, 2014.

[58] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[59] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen, "Collaborative filtering recommender systems," in *The Adaptive Web*. Berlin, Germany: Springer, 2007, pp. 291–324.

[60] M. G. Karlaftis and A. P. Tarko, "Heterogeneity considerations in accident modeling," *Accident Anal. Prevention*, vol. 30, no. 4, pp. 425–433, Jul. 1998.

[61] G. Qi, H. Wang, M. Haner, C. Weng, S. Chen, and Z. Zhu, "Convolutional neural network based detection and judgement of environmental obstacle in vehicle operation," *CAAI Trans. Intell. Technol.*, vol. 4, no. 2, pp. 80–91, Jun. 2019.

[62] E. Javouhey, A.-C. Guérin, and M. Chiron, "Incidence and risk factors of severe traumatic brain injury resulting from road accidents: A population-based study," *Accident Anal. Prevention*, vol. 38, no. 2, pp. 225–233, Mar. 2006.

[63] E. A. Mohamed, “Predicting causes of traffic road accidents using multi-class support vector machines,” *J. Commun. Comput.*, vol. 11, no. 5, pp. 441–447, 2014.

[64] K. W. Ogden and S. Roads, “A guide to road safety engineering,” Dept. Civil Eng., Monash Univ., Melbourne, VIC, Australia, Tech. Rep., 1996.

[65] K. J. Parnell, N. A. Stanton, and K. L. Plant, “What’s the law got to do with it? Legislation regarding in-vehicle technology use and its impact on driver distraction,” *Accident Anal. Prevention*, vol. 100, pp. 1–14, Mar. 2017.

[66] G. Guerreiro, P. Figueiras, R. Silva, R. Costa, and R. Jardim-Goncalves, “An architecture for big data processing on intelligent transportation systems. An application scenario on highway traffic flows,” in *Proc. IEEE 8th Int. Conf. Intell. Syst. (IS)*, Sep. 2016, pp. 65–72.

[67] R. C. McIlroy *et al.*, “Who is responsible for global road safety? A cross-cultural comparison of actor maps,” *Accident Anal. Prevention*, vol. 122, pp. 8–18, Jan. 2019.

[68] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, “1D convolutional neural networks and applications: A survey,” 2019, *arXiv:1905.03554*. [Online]. Available: http://arxiv.org/abs/1905.03554

[69] R. Burke, “Hybrid recommender systems: Survey and experiments,” *User Model. User-Adapted Interact.*, vol. 12, no. 4, pp. 331–370, 2002.

[70] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: A survey,” *Decis. Support Syst.*, vol. 74, pp. 12–32, Jun. 2015.

[71] W. Alghamdi, E. Shakshuki, and T. R. Sheltami, “Context-aware driver assistance system,” *Procedia Comput. Sci.*, vol. 10, pp. 785–794, Jan. 2012.

[72] M. T. Ribeiro, S. Singh, and C. Guestrin, “Model-agnostic interpretability of machine learning,” 2016, *arXiv:1606.05386*. [Online]. Available: http://arxiv.org/abs/1606.05386

[73] J. Benesty, J. Chen, Y. Huang, and I. Cohen, “Pearson correlation coefficient,” in *Noise Reduction in Speech Processing*. Berlin, Germany: Springer, 2009, pp. 1–4.

[74] A. Ng. *Machine Learning Yearning*. Accessed: Dec. 20, 2018. [Online]. Available: https://www.goodreads.com/book/show/30741739-machine-learning-yearning

[75] J. A. Marais, “Deep learning for tabular data: An exploratory study,” Ph.D. dissertation, Dept. Statist. Actuarial Sci., Stellenbosch Univ., Stellenbosch, South Africa, 2019.

[76] A. Vaswani *et al.*, “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

Kaoutar Sefrioui Boujemaa received the bachelor’s degree in computer science and mathematics and the master’s degree in big data analytics and smart systems from Sidi Mohammed Ben Abdellah University (USMBA), in 2015 and 2017, respectively, where she is currently pursuing the Ph.D. degree. She is also pursuing the research internship with the Moroccan Foundation of Advanced Science and Innovation (MAScIR). Her research interests include intelligent transportation systems, road safety, and computer vision.

Ismail Berrada received the Ph.D. degree in computer science from The University of Bordeaux 1, France, in 2005. He worked five years as an Assistant Professor with the University of La Rochelle. He is currently an Associate Professor of Computer Science with the School of Computer and Communication Sciences (SCCS), Mohammed VI Polytechnic University (UM6P), Benguérir, Morocco. His research interests include cognitive radio networks, radio resource management, vehicular ad hoc networks, and software testing and verification.

Khalid Fardousse received the master’s degree in computer science and decision making and the Ph.D. degree in computer science from the Faculty of Science, Sidi Mohammed Ben Abdellah University (USMBA), Fes, Morocco, in 2002 and 2010, respectively. He is currently works as a Computer Science Professor with USMBA. His research interests include computer vision, recognition in natural images, IA for modeling, analysis of driving behavior and their application in intelligent transportation systems.

Othmane Naggar received the Engineering degree in telecommunications and information systems from the Institute of Posts and Telecommunications, Rabat, Morocco, in 2014. He is currently pursuing the Ph.D. degree in the use of satellite images for agriculture with the Department of Geomatics, Institute of Agriculture, Mohammed V University, Rabat. He is currently a Senior Research Engineer with MAScIR. His research interests include real-time image processing, AI, and yield crop estimation.

François Bourzeix received the DEA degree in applied mathematics from the University Paris IX Dauphine in 1995. He was an Engineer with the Ecole Polytechnique of Paris in 1994 and the Ecole Nationale Supérieure des Télécommunications, Paris, in 1996. He worked with several high-tech companies, such as Motorola Semiconductor, ST Microelectronics, and Amesys-Maroc. In 2009, he joined MAScIR. He is currently the Director of the Embedded Systems and AI Department with MAScIR, Rabat, Morocco. His research interests include image processing, satellite, deep learning, and embedded systems.