Accident Analysis and Prevention 129 (2019) 350–361

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Young driver fatal motorcycle accident analysis by jointly maximizing accuracy and information

Dan Halbersberg, Boaz Lerner*

Department of Industrial Engineering & Management, Ben-Gurion University of the Negev, Israel

ARTICLE INFO

Keywords:
Young drivers
Fatal accidents
Motorcycle
Machine learning
Information measure
Prediction
Bayesian network
Key factors

ABSTRACT

While young drivers (YDs) constitute ~10% of the driver population, their fatality rate in motorcycle accidents is up to three times higher. Thus, we are interested in predicting fatal motorcycle accidents (FMAs), and in identifying their key factors and possible causes. Accurate prediction of YD FMAs from data by risk minimization using the 0/1 loss function (i.e., the ordinary classification accuracy) cannot be guaranteed because these accidents are only ~1% of all YD motorcycle accidents, and classifiers tend to focus on the majority class of minor accidents at the expense of the minority class of fatal ones. Also, classifiers are usually uninformative (providing no information about the distribution of misclassifications), insensitive to error severity (making no distinction between misclassification of fatal accidents as severe or minor), and limited in identifying key factors. We propose to use an information measure (IM) that jointly maximizes accuracy and information and is sensitive to the error distribution and severity. Using a database of ~3600 motorcycle accidents, a Bayesian network classifier optimized by IM predicted FMAs better than classifiers maximizing accuracy or other predictive or information measures, and identified fatal accident key factors and causal relations.

1. Introduction

Road accidents rank as a major cause of death worldwide, particularly among young drivers (YDs), for whom road injuries are the leading cause of death. YDs (ages 18–24) are greatly overrepresented in accident statistics, even after controlling for the number of drivers, miles traveled, and population size (Toledo et al., 2012). In most European countries, YDs make up 9–13% of the population; however, their percentage in driver fatalities is 18–30% (OECD, 2006). Besides the tragic human cost, these accidents place a heavy financial burden on society. Based on 2006 OECD figures, a fatal accident costs around \$1.5 M, and in the US alone, the cost of YD road accidents in 2002 was \$40 billion (NHTSA, 2015).

Motorcycle accidents are particularly deadly; a motorcyclist that is involved in an accident is 6–13 times more likely to be severely injured compared to other vehicle drivers, and, per mile, the numbers are even more frightening; in the United States, the chance of dying in an accident per mile traveled is 29 times higher for motorcyclists compared to car passengers (NHTSA, 2015). Moreover, since motorcycles are a more cost-effective mode of transport (Vlahogianni et al., 2012), they are very popular among the YD population, who usually have less disposable income. Numerous studies (Haque et al., 2009; Law et al., 2009; De Rome and Senserrick, 2011; Cheng et al., 2013; Mnzava, 2015) have tried to identify key factors from driver, road, vehicle, and accident variables that influence severity. The identification of key factors contributes not only to the attempts to prevent accidents, but also to those aiming to reduce accident severity (Delen et al., 2006). De Rome and Senserrick (2011) found crash type (rear-end/head-on), intersection, lane changing, and entering traffic from parking as the major patterns for motorcycle driver fault in multi-vehicle accidents, which suggests that road safety should focus on specific locations (such as at curves on high frequency motorcycle routes). Mnzava (2015) studied motorcycle accidents in developing countries (where up to 70% of drivers have no formal training) and found a connection to substandard motorcycle defects (e.g., tire blowouts, brake failure, etc.), non-use of helmets, and traffic violations. Haque et al. (2009) found that for not-at-fault motorcycle crashes, night driving at intersections and on expressways, lane of driving, and junctions with surveillance cameras are key factors, whereas for at-fault crashes, wet road at non-intersections, lane of driving, speed limit, and engine capacity in crashes on expressways are key factors. The number of lanes was also found as a key factor, in which a single lane road contributes to the fatality of the accident (De Rome and Senserrick, 2011; Naqvi and Tiwari, 2017). Naqvi and Tiwari (2017) suggested to upgrade roads that are prone to accidents to roads with a higher number of lanes.

Mainly, there are two principle methods used to analyze and predict

* Corresponding author.
E-mail addresses: halbersb@post.bgu.ac.il (D. Halbersberg), boaz@bgu.ac.il (B. Lerner).

https://doi.org/10.1016/j.aap.2019.04.016
Received 12 August 2018; Received in revised form 28 February 2019; Accepted 17 April 2019
Available online 12 June 2019
0001-4575/

accident severity from data; The first is by the traditional statistical tools, and the second is by machine learning (ML). One of the most common statistical approaches to accident analysis is regression (Al-Ghamdi, 2002; Abdel-Aty, 2003; Savolainen et al., 2011; Mujalli and de Oña, 2013; Naqvi and Tiwari, 2017). The exact type of regression applied depends on the target/response variable. For example, logistic regression will be applied to binary response (say two accident severities: fatal and non-fatal), ordinal regression to ordinal response, and linear regression to continuous response. Usually, regression involves a stepwise procedure that aims to reduce the model complexity by selecting part of the features. The selected features are the most relevant to the prediction of the severity (of a future accident) and, thus, can be considered as the key factors behind the accident. Nevertheless, these models have their own strong assumptions (e.g., about the data and error distributions, linearity between the independent variables and the target, etc.), and if these are violated by the data, then the results and conclusions are in doubt. In ML, there is no need to assume a specific (parametric) model or specific data/error distributions. For these reasons and others, ML is widely used in various fields including accident analysis (Chong et al., 2005; Delen et al., 2006; De Oña et al., 2011; Li et al., 2012). In ML algorithms [e.g., support vector machine (SVM), neural networks (NN), decision tree (DT), and random forest (RF)], key factor identification is usually performed based on ranking the feature importance and selecting the highest ranked features/variables.

Prediction and identification of key factors of accident severity from data using ML classifiers are challenging for several reasons. First, classifiers that maximize accuracy (minimize the 0/1 loss function) during learning do not account for error distribution and, thus, are not informative enough about the classification result to guide road-safety experts.[1] On the other hand, classifiers that maximize information with respect to the class variable by accounting for the error distribution during training are not accurate enough. Second, fatal accidents, which are the focus of road safety, account for only around 1% of all accidents; hence, classifiers of this imbalanced accident data have no chance to learn this minority class, and usually err and predict all fatal accidents as non-fatal. Upsampling the minority class (Provost, 2000; Chawla, 2005; Lerner et al., 2007) of fatal accidents or downsampling the majority class of minor accidents results in overfitting and domain deformation or loss of data, respectively, and are less accurate (see Section 4). Third, 0/1 loss function (accuracy)-based classifiers are not optimized in learning to tackle different error severities differently; thus, they consider misclassification of fatal accidents as severe, similar to misclassification of fatal accidents as minor. That is, these classifiers do not utilize the ordinal nature of the problem, which is that fatal accidents are more severe than severe accidents, and severe accidents are more severe than minor ones. Different costs to different misclassifications, if applied as part of learning the classifier, can arbitrarily be set by the user and, thus, are questionable and not recommended – let alone not always accurate (Section 4). Fourth, classifiers excel in prediction, but not in knowledge representation, which is needed for accident key factor identification. The Bayesian network classifier (BNC), however, excels in knowledge representation and, thus, is ideal for identifying key factors, but, on the other hand, it is not considered a supreme classifier. In addition, similar to other ML classifiers, it suffers from all of the above problems. Thus, prediction of accident severity using ML classifiers is challenging, but also challenging is the identification of key factors of accident severity because such factors extracted from a classifier that was trained without proper consideration for the error distribution, class imbalance, and error severity are doubtful, and their contribution to the analysis and understanding of YD motorcycle accidents may only be limited.

To meet this challenge in this study, we propose a measure that jointly maximizes accuracy and information and accounts for error distribution and severity in imbalance ordinal classification problems, and apply it in learning a BNC to benefit from these qualities. This BNC provides prediction of YD fatal motorcycle accidents (FMAs) and knowledge representation to identify key factors for these accidents, which improves upon those of classifiers that maximize only accuracy or information measures. Therefore, our contributions are the new measure and BNC for YD FMA prediction and key factors identification, and their application to a dataset of YD motorcycle accidents to derive insights that are, on the one hand, inclusive, accurate, and informative, and, on the other hand, not biased by the imbalance and ordinal nature of the dataset.

The structure of the paper is as follows. In Section 2, we describe the accident data, while in Section 3, we present our approach of joint maximization during training of the accuracy and information provided by an accident severity classifier. In Section 4, we demonstrate the reduced performance of approaches that fail to perform joint maximization in comparison to the suggested approach with respect to the prediction of accident severity and key factor identification. Section 5 summarizes the paper and the advantages of the suggest approach.

2. Materials

Accident data were obtained from the Israeli Central Bureau of Statistics and include all accidents that occurred in Israel between 2002 and 2008 of YDs who received their driving license between 2002 and 2008. The accident data we used include 3664 motorcycle injury accidents, with each accident characterized by 73 variables. The class variable is Accident severity, and it consists of three values: Fatal, Severe, and Minor that are distributed as 1%, 12.5%, and 86.5%, respectively. Following Chang and Wang (2006), as well as De Oña et al. (2011), the injury severity of an accident is determined according to the level of injury to the worst injured occupant. A Spearman test (with a 0.05 confidence level) between each of the 73 variables and the class variable helped in preliminarily selecting 19 features that characterize the road, motorcycle, accident, environment, and driver (Table 1). Some of the variables were further refined and categorized (e.g., driving seniority originally had 40 distinct values that were condensed into 8 distinct values).

3. Learning by joint accuracy-information maximization

Although the Bayesian network (BN) is considered a powerful tool for knowledge representation, and thus may be superb for identifying key factors for FMAs, when it comes to achieving high classification accuracy, many researchers prefer using a “natural” classifier, e.g., SVM, NN, DT, or RF, instead of a BNC. The reason is that usually the structure learned for a BNC maximizes a likelihood-based (or other generative) score (Cooper and Herskovits, 1992), but it has already been claimed (Friedman et al., 1997) and demonstrated (Kontkanen et al., 1999; Grossman and Domingos, 2004; Kelner and Lerner, 2012; Halbersberg and Lerner, 2016) that to achieve high accuracy, a BNC should maximize a (discriminative) score that is specific to classification and not to general inference. Since we are interested in both FMA key factor identification (a task that is natural to a BN) and prediction (using a discriminative score), we selected as our classifier the most accurate BNC (Kelner and Lerner, 2012), which is learned to maximize accuracy (the 0/1 loss function). For this classifier, we modified its 0/1 loss function to a measure that jointly maximizes accuracy and information, and for ordinal classification also accounts for error severity.

[1] Although ML classifiers – in contrast to statistical classifiers – assume nothing about the error distribution, which is a virtue, they also do not exploit this distribution in order to augment their training, which is a flaw. We consider the error distribution a valuable source of information that can help a classifier focus on and reduce specific errors of concern, such as those in classifying a fatal accident as minor, and thus try in this study to maximize not only the accuracy in the classification, but also the information it provides (see Section 3).

Table 1
Characteristics of YD motorcycle accidents.

Category	Variable	Values	Minor	Severe	Fatal	Total
Motorcycle, accident, and road characteristics	Motorcycle type	50 cc	1574	179	9	1762
		51–125 cc	787	83	4	874
		126–400 cc	557	95	5	677
		> 400 cc	238	98	15	351
	Safety measures	Used helmet	2746	359	27	3132
		No safety measures	62	32	2	96
		Unknown	368	64	4	436
	Road type	City junction	1485	144	10	1639
		City non-junction	1447	217	12	1676
		Intercity intersection	105	27	2	134
		Intercity	139	67	9	215
	Road speed limit	Up to 50 kph	2870	331	17	3218
		60	93	26	3	122
		70	59	19	3	81
		80	74	52	4	130
		90	74	27	5	106
		100	6	0	1	7
	Accident type	Hitting a pedestrian	198	55	5	258
		Front-side	1933	186	8	2127
		Front-back	189	42	3	234
		Side-side	341	35	0	376
		Front-front	101	31	0	132
		Crash with a parked car	15	3	0	18
		Crash with an inanimate object	45	23	8	76
		Turned over	107	45	4	156
		Skid	182	26	4	212
		Other	65	9	1	75
	Road width	Unknown	3	0	0	3
		Up to 5 m	480	30	1	511
		5 to 7 m	1192	161	9	1362
		7 to 10.5 m	1249	196	17	1462
		10.5 to 14 m	193	37	5	235
		More than 14 m	59	31	1	91
	Traffic signs	Road poorly marked	168	10	0	178
		Poor signage	18	5	0	23
		Nothing is wrong	2762	397	32	3191
		No signs are needed	181	34	1	216
		Unknown	47	9	0	56
	Road single/multilane	Single	1959	258	14	2231
		Multi	1217	197	19	1433
Environment characteristics and contributing circumstances	Visibility	Day light	1929	239	18	2186
		Bad weather	5	0	0	5
		Night with lights	915	162	11	1088
		Light did not work	7	0	0	7
		Light does not exists	101	26	2	129
		Night, unknown light condition	207	24	2	233
		Night with lights, but limited visibility	6	2	0	8
		Day, unknown light condition	6	2	0	8
	Hour of accident	07:00–10:00	314	36	3	353
		10:01–16:00	1156	145	13	1314
		16:01–22:00	1259	185	8	1452
		22:01–06:59	447	89	9	545
	Day or night	Day	1936	241	18	2195
		Night	1240	214	15	1469
	Day of accident	Sunday	506	80	2	588
		Monday	520	56	4	580
		Tuesday	517	66	6	589
		Wednesday	501	73	8	582
		Thursday	513	57	8	578
		Friday	332	64	3	399
		Saturday	287	59	2	348

(*continued on next page*)

Table 1 (*continued*)

Category	Variable	Values	Minor	Severe	Fatal	Total
Driver characteristics	Gender	Male	2915	440	33	3388
		Female	261	15	0	276
	Ethnic group	Jewish	2873	386	28	3287
		Muslim	193	54	5	252
		Other	110	15	0	125
	Residential area	District 1	408	144	3	555
		District 2	2425	241	24	2690
		District 3	256	50	4	310
		District 4	87	20	2	109
	Residential type	City	3018	397	27	3442
		Non-Jewish village	43	31	3	77
		Jewish village	114	26	3	143
		Unknown	1	1	0	2
	Medical limitations	No	932	128	3	1063
		Yes	2244	327	30	2601
	Driving seniority	No license	911	162	11	1084
		1–24 months	1924	239	18	2181
		25–30 months	205	24	2	231
		31–36 months	101	26	2	129
		37–42 months	17	2	0	19
		43–48 months	7	0	0	7
		4–5 years	6	2	0	8
		> 6 years	5	0	0	5
	Age at accident	17–18	946	151	0	1097
		19	125	17	4	146
		20	834	116	7	957
		21	891	134	16	1041
		22–23	311	42	4	357
		24	58	6	2	66
Target/class variable	Accident severity	–	3176	455	33	3664

The advantages to road safety of this classifier after our modification over other state-of-the-art classifiers are: interpretable results (as opposed to SVM and NN, and more than DT) that enable efficient and accurate identification of FMA key factors and causal relations; high prediction accuracy for all accident classes (severities), particularly rare ones; and performance that is geared via learning to account for harsh errors more than mild ones (without using arbitrary, user-defined costs as in the DT).

3.1. Learning a BNC

Various ML classifiers can be learned to maximize the suggested information measure (IM) (see Eqs. (9 and 10) in Section 3.2 below), but for knowledge representation, we preferred to focus on the BNC. We chose the risk minimization by cross-validation (RMCV) classifier (Kelner and Lerner, 2012) that excels in both knowledge representation and classification. It is the most accurate BNC and on par with state-of-the-art ML classifiers.[2]

The RMCV algorithm is a greedy search and score algorithm that in each search iteration identifies the edge addition, deletion, or reversal that improves the current structure (BN graph) the most by minimizing the 0/1 loss function,

$$L(C_i, \hat{C}_i) = \begin{cases} 0, & C_i = \hat{C}_i, \\ 1, & C_i \neq \hat{C}_i \end{cases} \tag{1}$$

where C_i and $\hat{C}_i$ are, respectively, the true and estimated class labels for sample i from M labels (states) of the class variable C (in our case, $M = 3$ for $C_1 =$ fatal, $C_2 =$ severe, and $C_3 =$ minor accidents). Therefore, we applied the IM to the RMCV BNC.

[2] The code for the RMCV algorithms is available online.

Table 2
Example confusion matrix for accident severity.

	True class (Y)			
Predicted class (X)		Fatal	Severe	Minor
	Fatal	C_{11}	C_{12}	C_{13}
	Severe	C_{21}	C_{22}	C_{23}
	Minor	C_{31}	C_{32}	C_{33}

3.2. Measures of accuracy and information

In this subsection, we first survey several existing performance measures that are potentially suitable for the highly imbalanced motorcycle accident data with its different severities before introducing two measures of our own.

The 0/1 loss function (Eq. (1)) used by the RMCV (i.e., classification accuracy, ACC) is biased according to the class prior probabilities and considers all misclassification types equally, similar to all accuracy-based ML classifiers. That is, it neither considers imbalanced data (Provost et al., 1998; Lerner et al., 2007; García et al., 2010) nor the distribution of errors. Tackling data imbalance by downsampling the majority class, upsampling the minority class, or applying different costs to different misclassifications provides an optimistic accuracy estimate, and thus is not recommended (Provost, 2000). Concerning accident severity prediction, the data is highly imbalanced (minor accidents are almost 90 times more frequent than fatal ones, see, e.g., Table 1), misclassification of fatal accidents as minor is harsher than their misclassification as severe, and information about the distribution of errors is needed for road-safety experts.

Thus, we consider here six measures (two are our own) that may tackle the above concerns as an alternative to replace the ACC measure in the RMCV algorithm (that is the classification accuracy computed by Eq. (1)). In one way or another, they all account for the entire confusion matrix and not just to its diagonal like ACC. A confusion matrix summarizes the classifier performance, where matrix entry C_{xy} counts the samples that are predicted as C_x, while their true class is C_y. For example, Table 2, which represents prediction of accident severity in our case, shows that C_{13} is the number of accidents that have been predicted as fatal, but actually are minor. The matrix also shows the error distribution of the misclassified samples per class (i.e., per column in the matrix). For example, if the errors for the fatal severity were distributed uniformly, then $C_{21} = C_{31}$.

1. *Mean absolute error* (MAE) is defined as the average deviation of the predicted class vector (X) from the true class vector (Y) (Baccianella et al., 2009; Hyndman and Koehler, 2006),

$$MAE = \sum_x \sum_y P(x, y)\cdot|x - y|, \tag{2}$$

which is the sum of all possible errors. Each is the (x, y) element in the confusion matrix C_{xy}, weighted by their relative prevalence according to the confusion matrix [$P(x, y)$]. For example, the (x, y) elements in MAE for $x = 1$ ("Fatal"), and $y = 1$ ("Fatal"), 2 ("Severe"), and 3 ("Minor") are $P(Fatal, Fatal)\cdot 0$, $P(Fatal, Severe)\cdot 1$, and $P(Fatal, Minor)\cdot 2$, respectively, where each probability is that of predicting "Fatal", where the true class is "Fatal", "Severe", and "Minor", respectively.

2. *The Matthew correlation coefficient* (MCC), known also as the Pearson correlation, has been used in the binary classification case (Baldi et al., 2000). Its generalization to the multiclass problem was introduced by Gorodkin (2004), where MCC is the correlation between the true (**U**) and predicted (**V**) class matrices (Jurman et al., 2012),

$$MCC = \frac{COV(\mathbf{U}, \mathbf{V})}{\sqrt{COV(\mathbf{U}, \mathbf{U}) \times COV(\mathbf{V}, \mathbf{V})}}. \tag{3}$$

U and **V** are $N \times M$, and N and M are the numbers of samples and classes, respectively, and $COV(\mathbf{U}, \mathbf{V})$ is

$$COV(\mathbf{U}, \mathbf{V}) = \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{N}(u_{im} - \bar{u}_m)(v_{im} - \bar{v}_m), \tag{4}$$

where u_{im} is either 1 or 0, if sample i is predicted as class m or not, respectively, v_{im} is either 1 or 0, if sample i's true class is m or not, respectively, and the average prediction and true value of class m over all samples are $\bar{v}_m = \frac{1}{N}\sum_{i=1}^{N} v_{im}$ and $\bar{u}_m = \frac{1}{N}\sum_{i=1}^{N} u_{im}$, respectively.

The next four measures are related to the entropy (of information theory), which measures the amount of uncertainty associated with the values of a discrete random variable X when only its distribution $P(X)$ is known. We call these measures information-based because they measure how informative the classifier predictions (X) are with respect to the true classes (Y).

3. *Mutual information* (MI) can be defined for two M-dimensional vectors X and Y holding predictions and true values for M possible classes, respectively. It measures the reduction in uncertainty for the true class $Y = y$ due to the prediction $X = x$ (Baldi et al., 2000),

$$MI = \sum_x \sum_y P(x, y)\cdot\log\left(\frac{P(x, y)}{P(x)P(y)}\right), \tag{5}$$

where $P(x, y)$ is the joint probability distribution between X and Y. If X and Y are uniformly distributed, i.e., all values in the confusion matrix are equal, then X and Y are independent (the predictions are not related to the true labels), whereas if all off-diagonal values in the confusion matrix are zero, then X and Y are completely dependent, and the predictions are very informative about the true labels. In classification of the severity of motorcycle accidents, X and Y hold predictions and true values for fatal, severe, and minor, and to calculate their MI, we should sum nine value combinations based on Table 2. For example, for $x = fatal$ and $y = minor$, we should calculate,

$$P(x = fatal, y = minor) = \frac{C_{13}}{N},$$

$$P(x = fatal) = \frac{\sum_{y=1}^{3} C_{1,y}}{N},$$

$$P(y = minor) = \frac{\sum_{x=1}^{3} C_{x,3}}{N},$$

where N is the number of samples (accidents). Since the accident severity is measured in an ordinal scale, we refer to fatal, sever, and minor as 1, 2, and 3, respectively.

4. *Confusion entropy* (CEN) (Wei et al., 2010) exploits the distribution of misclassifications of a class as any other of $M - 1$ classes and of the $M - 1$ classes as that class

$$CEN = \sum_{m=1}^{M} P_m \times CEN_m, \tag{6}$$

where P_m refers to the confusion probability of class m,

$$P_m = \frac{\sum_{k=1}^{M}(C_{mk} + C_{km})}{2\sum_k \sum_l C_{kl}}, \tag{7}$$

where C_{mk} is the (m, k) element of the confusion matrix between X and Y. CEN_m refers to the confusion entropy of class m,

$$CEN_m = \sum_{k=1,k\neq m}^{M}\left(P_{m,k}^{m} \times \log_{2M-2}(P_{m,k}^{m}) + P_{k,m}^{m} \times \log_{2M-2}(P_{k,m}^{m})\right), \tag{8}$$

where $P_{k,m}^{m}$ is the probability of misclassifying samples of class k to class m subject to class m (that is, the misclassification is normalized by the sum of all samples that belong to class m and those that were classified as class m).

5. *The information measure* (IM) (Halbersberg and Lerner, 2016) combines MI (Eq. (5)) to evaluate the error distribution measured between predictions and true values and the total error severity (ES) that is based on MAE (Eq. (2)).

$$\begin{aligned} IM &= -MI(X, Y) + ES(X, Y) \\ &= \sum_x \sum_y P(x, y)\left(-\log\left(\frac{P(x, y)}{P(x)P(y)}\right) + \log(1 + |x - y|)\right). \end{aligned} \tag{9}$$

ES measures the "classification distance" $|x - y|$, which we call the error severity between x and y, that is transformed by the joint distribution, $P(x, y)$, and the *log* to MI "units". Both $P(x, y)$ and $|x - y|$ are measured using the $M \times M$ classifier confusion matrix. The motivation to IM is a measure that its optimiaztion[3] leads to maximization of both the information with respect to error distribution, as measured by the MI between predictions and true classes; accuracy, as measured by both MI and ES; along with minimization of the error severity, as measured by ES.

6. *The information measure with alpha weight* (IM$_\alpha$) (Halbersberg and

Table 3
Properties of different measures.

	Maximize accuracy and information	Prefer balanced class distrib.	Sensitive to error distrib.	Tackle error severity	Sensitive to number of classes
ACC	×	×	×	×	×
MCC	×	×	×	×	×
MAE	×	×	×	✓	×
CEN	✓	×	✓	×	×
MI	✓	✓	×	×	✓
IM	✓	✓	✓	✓	✓
IM$_\alpha$	✓	✓	✓	✓	✓

[3] We maximize -IM instead of minimizing IM.

Table 4
ACC, IM_α, MI, MAE, F-measure, and AUC (all should be maximized except MAE) of BNCs learned using seven measures.

	IM	IM_α	MI	CEN	MCC	MAE	ACC
ACC	84.37	84.38	*68.53*	**86.65**	84.37	85.37	85.39
IM_α	60.29	**60.45**	59.92	*59.31*	59.37	59.65	59.66
MI	0.029	0.030	**0.032**	*0.017*	0.018	0.022	0.022
MAE	0.167	0.166	*0.438*	**0.143**	0.166	0.155	0.155
F-measure	0.33	0.32	**0.40**	*0.31*	0.36	*0.31*	*0.31*
AUC	**0.66**	0.63	*0.59*	**0.66**	0.61	0.65	0.65

Table 5
Average confusion matrix of BNC-CEN.

		True class (Y)		
Predicted class (X)		Fatal	Severe	Minor
	Fatal	0	0.08	0.1
	Severe	0.02	0.12	0.1
	Minor	6.58	91	635

Lerner, 2016) allows trading between the two components of IM:

$$\begin{aligned} IM_\alpha = & \sum_x \sum_y - P(x, y) \cdot \log\left(\frac{\alpha \cdot P(x, y)}{P(x)P(y)}\right) \\ & + \sum_x \sum_y P(x, y) \cdot \log(\alpha \cdot (1 + |x - y|)) \\ = & \; IM - \log(\alpha) \times ACC, \end{aligned} \tag{10}$$

where $\alpha > 0$ is a user- or data-defined constant that balances accuracy, information, and error severity. When $\alpha = 1$, IM is a special case of IM_α.

Table 3 summarizes the most important properties of the seven measures (ACC plus the six outlined above), showing that our two suggested measures are those and only those to satisfy all needs from a measure. Meeting the property of jointly maximizing accuracy and information leads to two measures that maximize information (for which classification of the minority class also accounts) besides maximizing accuracy (which can be achieved by focusing only on the majority class). By meeting the other properties also, these measures can successfully address imbalance ordinal classification problems. The seven measures were used in training the RMCV structure learning algorithm to induce seven BNCs that were compared using the accident data.

4. Experimental results

4.1. Methodology

Each of ten random permutations of the accident data was split using five-fold cross-validation (CV5) into training and test sets (4/5 for training and 1/5 for testing), and was used five times to train and test seven BNC learning algorithms (based on the seven measures of Table 3) that were initialized by the naïve Bayesian classifier (NBC) (Langley et al., 1992). Each of the ten training sets was further divided using (another) CV5 (4/5 for actual training and 1/5 for validation), as the RMCV algorithm needs a validation set in order to select the best structure in each learning iteration. All the reported results here are averaged over the 10 × 5 = 50 test sets.

For each of the seven BNC structure learning algorithms, in each step of the greedy search, neighboring graphs (differing by edge addition/deletion/reversal) are compared with the current graph, and learning proceeds if the measure computed on the validation set is improved by any of the neighboring graphs, which then becomes the new current graph. When learning a classifier is completed, a confusion matrix is derived to evaluate the classifier using the test set. This matrix for each of the seven algorithms was evaluated using five performance measures: ACC, IM_α, MAE, F-measure, and area under curve (AUC). AUC is based on Hand and Till's (2001) extension for multiclass problems, and F-measure is an extension of precision and recall (Sokolova and Lapalme, 2009),

$$\text{Fmeasure} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

In the results reported, **bold** and *underlined italic* fonts indicate the best and worst algorithms/measures, respectively.

4.2. Performance evaluation

Table 4 shows the average values of our five performance measures: ACC, IM_α, MAE, F-measure, and AUC for the seven evaluated algorithms over 50 data sets. Table 4 shows that the CEN-based BNC has the best values of ACC, MAE, and AUC. However, as Table 5 shows, this classifier classifies almost all accidents as minor, which “pays off” with

Table 6
ACC, IM_α, MAE, F-measure, and AUC of five algorithms.

	DT-cost	DT-ord	BNC-IM_α	BNC-ACC	OR
ACC	82.64	*82.42*	84.38	**85.39**	83.67
IM_α	58.12	58.03	**60.45**	59.66	*54.92*
MI	0.022	0.019	**0.030**	0.020	*0.009*
MAE	0.183	*0.186*	0.166	**0.155**	0.172
F-measure	**0.39**	0.38	0.32	*0.31*	0.34
AUC	**0.70**	0.69	0.63	0.65	*0.62*

Table 7
Average confusion matrix of DT-cost.

		True class (Y)		
Predicted class (X)		Fatal	Severe	Minor
	Fatal	0.1	0.7	2.1
	Severe	1.62	20	47.42
	Minor	4.88	70.5	585.68

Table 8
Average confusion matrix of BNC-MI.

		True class (Y)		
Predicted class (X)		Fatal	Severe	Minor
	Fatal	3.68	38.24	121.1
	Severe	0.34	6.4	21.82
	Minor	2.58	45.46	492.28

Table 9
Average confusion matrix of BNC-ACC.

		True class (Y)		
Predicted class (X)		Fatal	Severe	Minor
	Fatal	0.28	0.76	1.94
	Severe	1.44	12.46	20.12
	Minor	4.88	77.98	613.14

Table 10
Average confusion matrix of BNC-IM_α.

		True class (Y)		
Predicted class (X)		Fatal	Severe	Minor
	Fatal	0.34	1.4	3.2
	Severe	2.18	21.96	35.8
	Minor	4.08	67.84	596.2

Table 11
Using upsampling (except for BNC-IM$_\alpha$) before reporting on the same measures and algorithms of Table 6.

	DT-cost	DT-ord	BNC-IM$_\alpha$	BNC-ACC	OR
ACC	75.35	75.75	**84.38**	67.63	*58.60*
IM$_\alpha$	52.26	52.48	**60.45**	49.15	*45.33*
MI	0.012	*0.011*	**0.030**	0.025	0.029
MAE	0.259	0.256	**0.166**	0.431	*0.525*
F-measure	0.38	0.38	*0.32*	**0.41**	**0.41**
AUC	0.54	0.55	0.63	**0.74**	0.71

Table 12
Average confusion matrix of DT-cost for the upsampled data (compare to Table 7).

		True class (Y)		
Predicted class (X)		Fatal	Severe	Minor
	Fatal	0.36	1.62	5.3
	Severe	1.68	23.92	98.92
	Minor	4.56	65.66	530.98

respect to the accuracy measures because the majority of the accidents are indeed minor, but it cannot tackle the class imbalance problem (as also the lowest value of IM$_\alpha$ suggests). The CEN-based BNC misclassifies almost all fatal and severe accidents as minor and, thus, is useless for road safety. Note that values in this, as well as other, confusion matrices below are averaged over all 50 test sets.

The classifiers that are augmented using the MAE and ACC measures (the two right-most columns in Table 4) also provides high values to the classification measures (ACC, MAE, and AUC), but they also classify most accidents as minor and, thus, are useless to predict an FMA. Although the IM- and IM$_\alpha$-based BNCs (the two left-most columns in Table 4) do not have the highest values of the classification measures, they achieve the best IM$_\alpha$ values, which is the most important measure for our purposes. In addition, they also provide relatively good values of the F and AUC measures, which are more balanced than ACC in evaluating the imbalanced classes.

We further extended the evaluation to other state-of-the-art algorithms suitable for ordinal classification and compared the ACC- and IM$_\alpha$-based BNCs with a decision tree (DT) and ordinal regression (OR) by the logit model. We choose to compare to OR since it represents a common statistical analysis method used in many road-safety studies (Al-Ghamdi, 2002; Abdel-Aty, 2003; Savolainen et al., 2011). Also, to avoid dependence on an expert, we exploited the $|x - y|$ values that were computed by IM and IM$_\alpha$ as costs for the DT and, thereby, established the DT-cost classifier. In addition, we extended the evaluation also to ordinal classification based on the DT (we called this classifier: DT-ord) (Frank and Hall, 2001).

Table 6 shows values averaged over the 50 data sets of the ACC, IM$_\alpha$, MAE, F-measure, and AUC measures of the five classifiers, from which it can be seen that BNC-ACC is the most accurate classifier, and BNC-IM$_\alpha$ has the best IM$_\alpha$ measure. Following Friedman and Nemenyi tests (the non-parametric test that should be exercised in our case

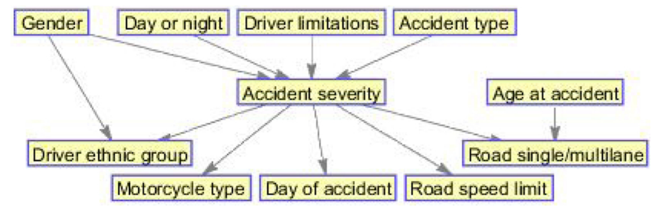


Fig. 1. Markov blanket of Accident severity learned by IM$_\alpha$-based BNC initialized by NBC.

(Demšar, 2006), we found that with respect to the ACC measure, BNC-ACC's superiority is significant to OR, DT-cost, and DT-ord, and that of BNC-IM$_\alpha$ is significant to DT-cost and DT-ord. With respect to the IM$_\alpha$ measure, the ACC and IM$_\alpha$ BNCs are superior to OR.

To demonstrate why IM$_\alpha$-based algorithms are more suitable to FMA prediction, we show in Tables 5, and 7–10, confusion matrices for five of the classifiers reported in Tables 4–6. The CEN-based classifier (Table 5) was already analyzed to show that it makes the harshest errors a classifier can make, classifying all accidents as minor. DT-cost (Table 7) was relatively good in predicting severe accidents, but was also the second worst in both fatal and minor accident prediction (which explains its poor accuracy in Table 6 although it received high F-measure and AUC values). The MI-based BNC (Table 8) predicted the highest number of fatal accidents, but it missed most severe accidents and was the worst in predicting minor accidents, predicting most of these as fatal (such imbalanced results are reflected in the low performance of this classifier in Table 4). The ACC-based BNC (Table 9) did very well on minor accidents (which explains its supreme accuracy in Table 6), but quite poorly on severe accidents. The IM$_\alpha$-based BNC (Table 10) had the third highest performance on minor accidents, the second highest performance on fatal accidents, and the highest performance on severe accidents, leading to good accuracy and MAE value and the best IM$_\alpha$ value in Table 6. These results show that the IM$_\alpha$-based BNC inherited the good ACC performance that an ACC-based BNC has on minor accidents, while jointly maximizing the information exposed by the MI-based BNC that excels in predicting fatal accidents. Such results are the type of performance we expected from an FMA classifier, i.e., high balanced level of accuracy for all classes.

In order to cope with the imbalance in the problem, we could have tried upsampling (downsampling performed very poorly and was neglected). For that, we upsampled severe and fatal accidents to be equal in number to the number of minor accidents in the training set, while keeping the true proportion among the severities in the test set. Then we compared the results of the IM$_\alpha$-based BNC from the previous experiment to those of the same classifiers in Table 6, which now were trained after balancing the data (Table 11). After upsampling, the minor classes (fatal and severe) were better predicted at the expense of the major class (minor accidents). Compared, for example, to the confusion matrix of DT-cost without upsampling (Table 7), upsampling increased the false alarm rate of this classifier, predicting way too many minor accidents as severe (Table 12). This is the trend in all classifiers, and Table 11 demonstrates that upsampling may be harmful in taking a classifier out of balance (as was already pointed out in Brodersen et al.

Table 13
Precision and recall lifts for different classifiers compared with BNC-IM$_\alpha$ (the baseline classifier) and three accident severities. Lift values for which BNC-IM$_\alpha$ is superior/inferior are colored in green (< 1)/red (> 1), if these values are higher/lower than the competitor in 10%.

		BNC-MI	BNC-ACC	BNC-IM$_\alpha$	DT-cost	DT-cost upsampling	OR	OR-upsampling
Fatal	Precision	0.3	1.4	1	0.5	0.7	0	0.5
	Recall	10.8	0.8	1	0.3	1.1	0	20.1
Severe	Precision	0.6	1	1	0.8	0.5	0.4	0.4
	Recall	0.3	0.6	1	0.9	1.1	0.2	1.1
Minor	Precision	1	1	1	1	1	1	1
	Recall	0.8	1	1	1	0.9	1	0.7

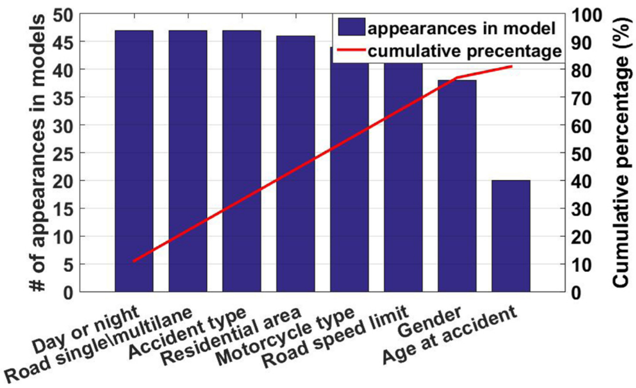


Fig. 2. Pareto analysis for YD motorcycle accident variables.

(2010)) for FMA prediction. Eventually, upsampling increased the advantage of IM$_\alpha$-BNC over the other classifiers even more (compare Table 11 with Table 6). In addition, recall that, practically, finding the best degree of upsampling needs some trial and error, whereas the method we propose requires no user-defined parameters.

Another view on the advantage of BNC-IM$_\alpha$ over the other classifiers can be seen from concentrating on the following performance scores: precision, recall, and lift, measured for each accident severity separately versus the other two severities. Precision, say for fatal accidents, is the percentage of fatal accidents we correctly predicted of all accidents we predicted as fatal, whereas recall (sensitivity) is the percentage of fatal accidents we correctly predicted of all fatal accidents. Although both measures are important (we want both of them to be high), recall is more important, as we want to correctly predict as many fatal accidents as possible, while predicting severe/minor accidents as fatal is less harsh although not desired. Lift, say for fatal accidents, measures either precision or recall compared to a reference (baseline) value for these measures, which in our case is the (precision/recall) value of BNC-IM$_\alpha$. Table 13 shows the lifts for precision and recall for the classifiers for which performances were provided above compared with that of BNC-IM$_\alpha$ (the "reference/baseline"). Values in the table (if they differ at least 10% from the competitor) are colored in green (< 1) to represent precision or recall for which BNC-IM$_\alpha$ is superior, and in red (> 1) to represent those cases for which BNC-IM$_\alpha$ is inferior. Table 13 shows that BNC-IM$_\alpha$ is superior (or at least not inferior) to all classifiers with respect to severe and minor accidents for precision and recall and to almost all classifiers with respect to fatal accidents. For fatal accidents, in the few cases in which the BNC-IM$_\alpha$ classifier is inferior to another classifier, it is only because the other classifier was not balanced well. For example, BNC-MI has better recall for fatal accidents only because it sees all accident severities evenly and, by that, fails in all scores (e.g., precision) except recall of fatal accidents. OR-

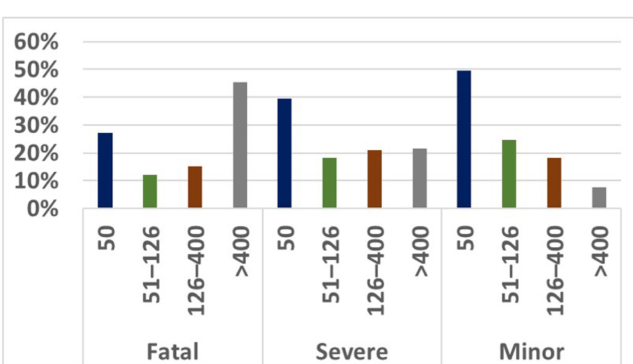


Fig. 3. Accident severity by Motorcycle type. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

Table 14
Frequencies of accident severities for heavy vs. non-heavy motorcycles.

	Heavy (%)	Non-heavy (%)
Fatal	4.3	0.5
Severe	28.0	11.0
Minor	67.7	88.5

upsampling upsamples fatal accidents and classifies different copies of the same accidents well, but then tends to miss all other accidents as the lift for its precision for fatal accidents shows. For road-safety experts, BNC-IM$_\alpha$ is not only more accurate and informative, but it also identifies a smaller but purer group of fatal accidents for which it is easier to find key factors.

4.3. Key factor identification

To identify key factors for FMAs, Fig. 1 shows the Markov blanket (MB) of the class variable—Accident severity—in one of the 50 BNCs learned by the IM$_\alpha$-based BNC. In the BN, the MB of the class variable is the set of variables (nodes) that includes the class variable's parents, children, and parents to common children. The importance of the MB concept is driven from the fact that, given the MB of the class variable, this variable is independent of all other variables in the graph (Pearl, 1988). That is, only the MB variables of the class variable affect it.

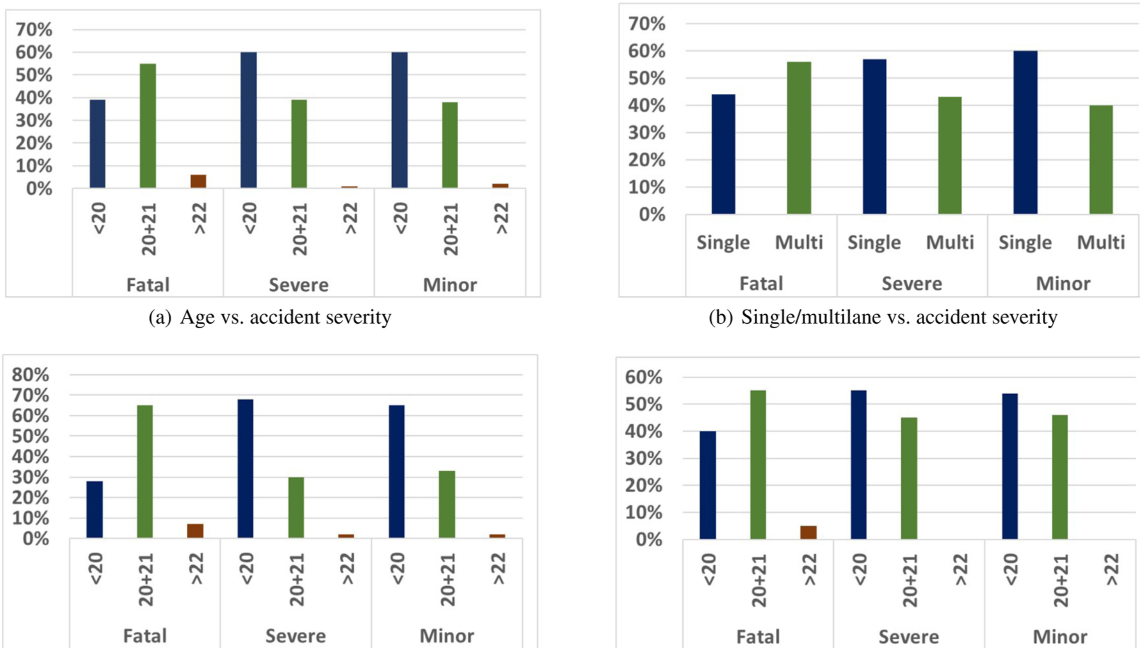
Learning the MB is usually based on statistical tests to identify edges that connect parent and child nodes. The presence of an edge between two nodes in an MB indicates that they are statistically (conditionally) dependent [or more accurately not statistically (conditionally) independent] with a specific confidence level. The equivalence in a statistical model such as regression is a p-value < 0.05 that indicates a significant impact of an independent variable on the dependent (class) variable. The second line of equivalence in the regression model is between its parameters (i.e., the β coefficients) and the BN conditional probability table (CPT) that probabilistically quantifies the strength of impact of a parent variable set on a child variable. However, compared with the regression model that assumes linear impact of independent variables on the dependent one, the BN allows nonlinear connections among the variables without assuming any roles (independent vs. dependent) to the variables.

As Fig. 1 indicates, the variables discovered in this MB—especially the driver variables: Gender and Age at accident, the road variables: Road speed limit and Road single/multilane, Motorcycle type, Accident type, and the environment variable Day or night—are extremely important in understanding accident severity.

Because the data is finite, small differences among MBs derived for the 50 training sets are natural. Therefore, Fig. 2 shows a Pareto analysis by which we counted the number of MBs among the 50 in which a variable appeared and, thereby, considered significant variables suspected as key factors. We found that Road speed limit, Motorcycle type, Residential area, Accident type, Road single/multilane, and Day or night appeared in around 45 of the MBs, Gender appeared in 38 MBs, and Age at accident in 20 MBs. These variables are more than 80% of the variables that appeared in the 50 models (as is seen on the right y-axis in Fig. 2), and thus can be considered FMA key factors. Related work (Yau, 2004; Chang and Wang, 2006; De Oña et al., 2011), and consultation with road-safety experts support this evidence.

Next, we discuss some of these variables that were found significant in most of the MBs of the accident severity variable (Fig. 2), and thus are considered key factors to FMAs. The analysis is based on the CPTs of the MBs:

Motorcycle type: Fig. 3 shows the distribution for each accident severity by motorcycle engine capacity. The figure reveals that 50 cc motorcycles (blue columns) are more probable as we move from fatal to minor accidents, as opposed to > 400 cc motorcycles (gray columns)



(a) Age vs. accident severity

(b) Single/multilane vs. accident severity

(c) Single lane: age vs. accident severity

(d) Multilane: age vs. accident severity

Fig. 4. Accident severity by Age at accident with respect to Road type. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

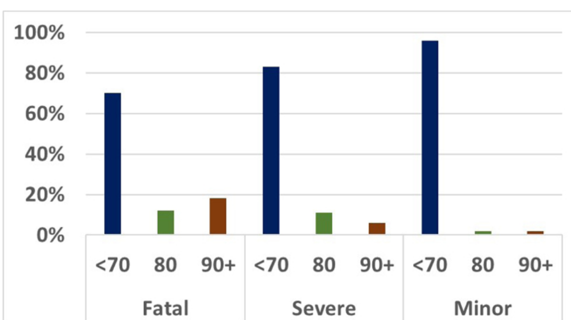


Fig. 5. Accident severity by Road speed limit. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

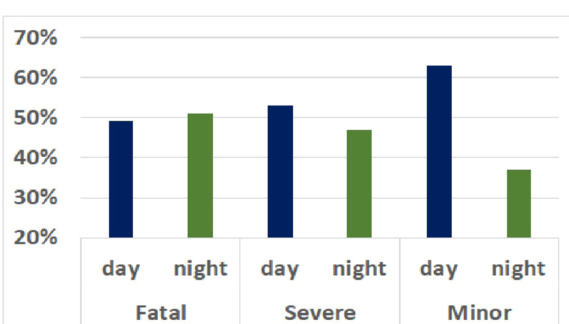


Fig. 6. Accident severity by Day or night. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

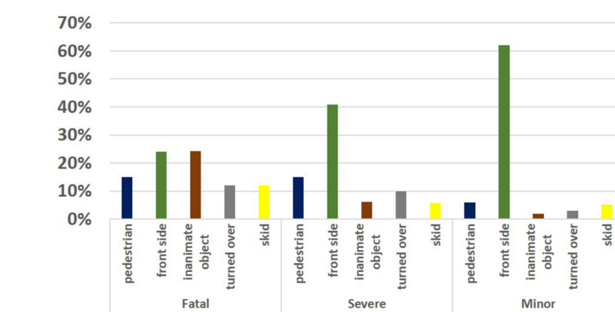
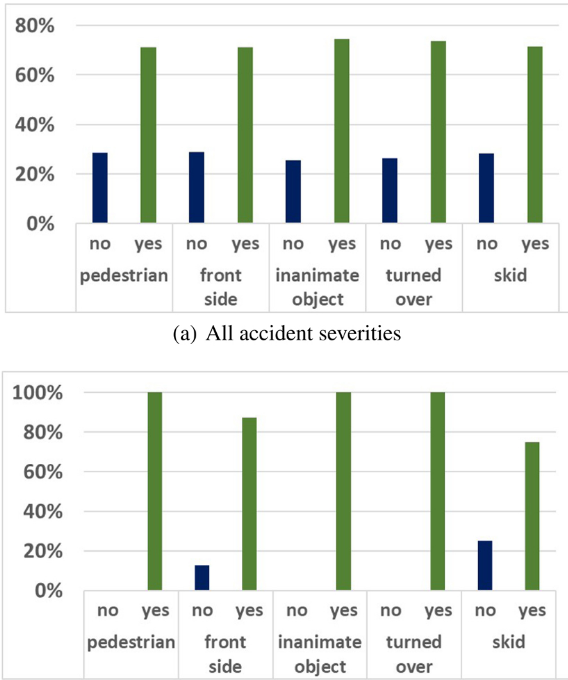


Fig. 7. Accident severity by Accident type (considering the five accident types that account for 90% of the fatal accidents). (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

that are less probable as we move from fatal to minor accidents. This is reasonable as heavy motorcycles drive faster, making an accident more likely to be severe or fatal (Haque et al., 2009). In addition, we summed all non-heavy motorcycle ($<$ 400 cc) accidents together and compared them to heavy motorcycle ($>$ 400 cc) accidents with respect to fatal vs. non-fatal accidents. This demonstrated that a YD accident that involves a heavy motorcycle is eight times more likely to be deadly than for a non-heavy motorcycle [4.3% compared with 0.5% (Table 14)].

Age at accident: At first glance, it seems that Accident severity is indifferent to Age at accident [Fig. 4(a)], as the distributions of accidents for Age at accident are different for different severities. However, the way in which age and severity are connected in the graph (Fig. 1) implies that they are dependent given the Road type (single/multilane)[4]. Fig. 4(b) shows that most fatal accidents are made on multilane roads, whereas most minor and severe accidents are made on single lane roads. Analysis of the interaction of the three variables demonstrates that for FMAs on single lane roads [Fig. 4(c)], YD are more likely to be 20–21 years old, while younger drivers ($<$ 20) are involved in minor and severe accidents. This is not the case on multilane roads [Fig. 4(d)], where the differences between $<$ 20 and 20–21-year-olds are small, regardless of the accident severity. An additional result is that regardless of the road type, most accidents of older drivers ($\geq$22) are

[4] In graph theory (Spirtes et al., 2000), two marginally independent variables X and Y become conditionally dependent given a third variable Z they collide into, e.g., X (say Age at accident) and Y (say Accident severity) are common parents of Z (i.e., Road type: single/multilane), as in Fig. 1.



(a) All accident severities

(b) Only fatal accidents

Fig. 8. Accident type (considering the five types as in Fig. 7) by Medical limitations. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

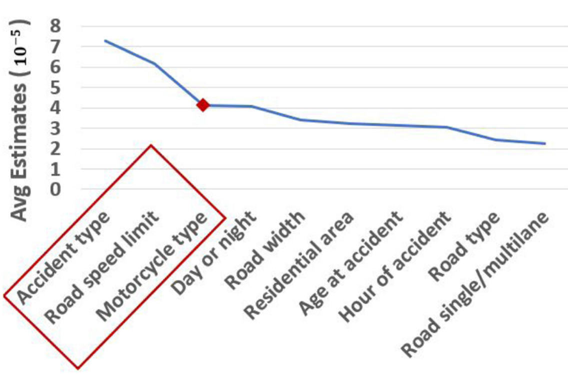


Fig. 9. Predictor importance estimates of the DTcost. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

fatal (we assume this is because they use heavier motorcycles than younger drivers).

Road speed limit: Fig. 5 shows that accidents on roads where the speed limit is high tend to be fatal (or severe), as found also in Delen et al. (2006). It does not mean that speed is the cause for the accident, but only that given that an accident occurred on a high speed road, it is more likely ended lethally. We can also see that 96% of minor accidents occurred on roads on which the speed limit was low (< 70 km/h).

Day or night: Minor accidents are more likely to happen during daytime than at nighttime (Fig. 6), while more than 50% of fatal accidents occurred at nighttime. This was also claimed by Yau (2004), as young drivers are less experienced driving at night. Also, at nighttime, we believe, YDs are exposed to alcohol more than other age groups, which again increases the accident rate and severity for this group.

Accident type: Fig. 7 shows prevalence by Accident severity for those types of accident that are responsible for 90% of the fatal accidents (Table 1): hitting a pedestrian, front-side, crashing with an inanimate object, turning over, and skidding. Several conclusions can be drawn from the figure: (1) front-side accidents account for most of the minor and severe accidents, far more than any other type, and a substantial type also of fatal accidents; (2) hitting a pedestrian usually does not result in a minor accident; (3) turn over and skid accidents are mostly fatal (or to a lesser degree severe) and have similar patterns to all accident severities; and (4) accidents with an inanimate object mostly are fatal. An interesting finding is that the majority of fatal accidents due to this type are of relatively mature YDs (> 22 year-olds); however, this is explained by the fact that 70% of them have low driving seniority (< 12 months) (it does not mean that all relatively mature young drivers have low driving seniority). That is, as expected, inexperienced YDs are more likely to lose control over their motorcycle and crash into inanimate objects with usually deadly results.

Medical limitations (not in Fig. 2, but a very important factor that is found in 12 of the MBs): Fig. 8(a) shows that the Accident type and Medical (driver) limitations are independent as the Medical limitations distribution is identical to all accident types (more or less 30–70%, which is also true for the five categories of Accident type that are not shown here). However, as can be seen in Fig. 1, these two variables are conditionally dependent given Accident severity (similar to Age at accident and Accident severity that are conditionally dependent given Road type; see the above analysis of Age at accident). Fig. 8(b) reveals that given that the accident is fatal, the relation between Accident type and Medical limitations is entirely different. It is shown that in most types of FMA (hitting pedestrians or inanimate objects and turning over), only drivers with medical limitations are involved, and also in the other accident types (front-side and skids), these drivers are the majority of drivers involved in the accidents [as opposed to all accident severities together [Fig. 8(a)], where drivers with no medical limitations are still responsible for around 30% of the accidents, which are mostly minor and severe]. One explanation could be that many YD motorcyclists feel uncomfortable wearing their glasses (one of the main YD medical limitations) together with an helmet and prefer to ride without them, which leads to hitting pedestrians or inanimate objects, turning over, etc. ending in fatal accidents [Fig. 8(b)].

Gender: All 33 fatal accidents were made by men, and in general, severe and fatal YD accidents are three times more likely in men.

Residential area: While District 1 constitutes only 15% of all YD motorcycle accidents in Israel, it includes 27% of the non-minor accidents, which is almost double the accident representation of this district nationally. If we further investigate the interaction of Residential area with the Residential type (Table 1), we see that given a YD motorcycle accident happened in a non-Jewish village in District 1, its likelihood to be non-minor is 91%.

Finally, we analyze key factors that were identified by other methods. Fig. 9 shows the top ten features that were found to be the most significant by the DT algorithm. The results in the figure are averaged over the 50 test sets, and the red point on the line represents the knee in which the importance estimates becomes less significant. According to Fig. 9, the features that were found important by the DT are consistent with the Pareto analysis (Fig. 2) of our proposed method.

According to the ordinal regression the variables that were found significant are: (1) Motorcycle type, (2) Accident type = [hitting a pedestrian|front-front|crash with a parked car|crash with an inanimate object] (3) Road width, (4) Residential type = [non-Jewish village], (5) Residential area = [District 1], (6) Traffic signs = [poor signage], and (7) Road speed limit. These variables appear with a p-value < 0.05 in at least 40 out of the 50 models (i.e., 50 test sets). According to this list, we can see that the ordinal regression important variables do not overlap with our proposed method as the DT does. The four variables that overlap are 1, 2, 5, and 7 while variables such as Day or night and Road single/multilane were not found as significant by the regression, and that could explain the model's relatively poor performance.

5. Summary and discussion

Young drivers' FMAs are of great concern to the road-safety community, which is struggling to identify key factors and causal relations that may explain these accidents and help point community efforts and resources to the most promising educational, infrastructural, and enforcement means, and to drivers who are more likely to be involved in FMAs.

Learning by only maximizing accuracy and ignoring the error distribution and severity for this class-imbalance problem results in accurate classification of only minor accidents at the expense of incorrect prediction of the most devastating fatal accidents. Such classification is useless in road safety.

Thus, we were interested in learning classifiers that: (1) are both accurate and informative (about the distribution of misclassifications), (2) are interpretable with respect to the identification of FMA key factors and causal relations, (3) are sensitive to error severity (penalizing, e.g., a misclassification of a fatal accident as minor more than if it were severe), and (4) that utilize the ordinal nature of accident severities to improve performance.

We proposed to use an information measure, IM_α, that jointly maximizes accuracy and information, is sensitive to error distribution and severity, and is successful in identifying key factors when implemented in a BNC. We applied an IM_α to a BNC and showed its advantage compared to other measures and upsampling using a relative large database of YD motorcycle accidents.

A major advantage of the IM_α-based BNC is in its ability to accurately classify all classes, and not only the major one as other classifiers do, which is essential to FMA prediction and key factor identification. Another major advantage is the ability to identify key factors automatically from the BNC structure (graph), as revealed by the MB, and not to consider all possible variables or a list of unrelated variables (as provided by statistical models) as key factors. Not only may we use the MB to identify key factors to FMAs, but it also allows us to analyze value combinations of variables included in the MB with relation to accident severities of specific interest, e.g., fatal. That is, beyond a simple list of suspected key factors, the MB provides factors that (in the absence of latent variables in the domain (Spirtes et al., 2000)) have (or imply on) causal relations among themselves and with the target variable (i.e., accident severity), which using graph theory can be translated to essential knowledge about the domain [see, e.g., the analysis for Age at accident with reference to Fig. 4 and that for Medical limitations with reference to Fig. 8(b)]. With lack of intervention, which is of course impossible to exercise in road safety, this knowledge cannot be gained otherwise (e.g. from checking all possible interactions manually or using statistical tools).

Finally, we expect that for other problems with high imbalance and errors that account differently for different classes (which is common in road safety applications, particularly when fatal accidents are involved), the advantage of the IM_α measure over other accuracy and information measures and the methodology we suggest here will also be significant.

Acknowledgement

This research was supported by a grant from the Israeli Netional Road Safety Authority and a grant from the Ran Naor Foundation for the Advancement of Road Safety Research (2008-035).

References

Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. J. Saf. Res. 34 (5), 597–603.

Al-Ghamdi, A.S., 2002. Using logistic regression to estimate the influence of accident factors on accident severity. Accid. Anal. Prev. 34 (6), 729–741.

Baccianella, S., Esuli, A., Sebastiani, F., 2009. Evaluation measures for ordinal regression. In: Proceedings of the 9th International Conference on Intelligent Systems Design and Applications. IEEE. pp. 283–287.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A.F., Nielsen, H., 2000. Assessing the accuracy of prediction algorithms for classification: an overview. Bioinformatics 16 (5), 412–424.

Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M., 2010. The balanced accuracy and its posterior distribution. In: Proceedings of the 20th International Conference on Pattern Recognition. IEEE. pp. 3121–3124.

Chang, L.-Y., Wang, H.-W., 2006. Analysis of traffic injury severity: an application of non-parametric classification tree techniques. Accid. Anal. Prev. 38 (5), 1019–1027.

Chawla, N.V., 2005. Data mining for imbalanced datasets: an overview. Data Mining and Knowledge Discovery Handbook. pp. 853–867.

Cheng, L., Geedipally, S.R., Lord, D., 2013. The Poisson-Weibull generalized linear model for analyzing motor vehicle crash data. Saf. Sci. 54, 38–42.

Chong, M., Abraham, A., Paprzycki, M., 2005. Traffic accident analysis using machine learning paradigms. Informatica 29 (1).

Cooper, G.F., Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Mach. Learn. 9 (4), 309–347.

De Oña, J., Mujalli, R.O., Calvo, F.J., 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accid. Anal. Prev. 43 (1), 402–411.

De Rome, L., Senserrick, T., 2011. Factors associated with motorcycle crashes in New South Wales, Australia, 2004 to 2008. J. Transport. Res. Board (2265), 54–61.

Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accid. Anal. Prev. 38 (3), 434–444.

Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7, 1–30.

Frank, E., Hall, M., 2001. A simple approach to ordinal classification. In: Proceedings of the 12th European Conference on Machine Learning. Springer. pp. 145–156.

Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Mach. Learn. 29 (2–3), 131–163.

García, V., Ramon, A., Mollineda, J., Sanchez, S., 2010. Theoretical analysis of a performance measure for imbalanced data. In: Proceedings of the 20th International Conference on Pattern Recognition. IEEE. pp. 617–620.

Gorodkin, J., 2004. Comparing two K-category assignments by a K-category correlation coefficient. Comput. Biol. Chem. 28 (5), 367–374.

Grossman, D., Domingos, P., 2004. Learning Bayesian network classifiers by maximizing conditional likelihood. In: Proceedings of the Twenty-First International Conference on Machine Learning. ACM. pp. 46–53.

Halbersberg, D., Lerner, B., 2016. Learning a Bayesian network classifier by jointly maximizing accuracy and information. Proceedings of the 22nd European Conference on Artificial Intelligence 1638–1639.

Hand, D.J., Till, R.J., 2001. A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach. Learn. 45 (2), 171–186.

Haque, M.M., Chin, H.C., Huang, H., 2009. Modeling fault among motorcyclists involved in crashes. Accid. Anal. Prev. 41 (2), 327–335.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22 (4), 679–688.

Jurman, G., Riccadonna, S., Furlanello, C., 2012. A comparison of MCC and CEN error measures in multi-class prediction. PLoS ONE 7 (8).

Kelner, R., Lerner, B., 2012. Learning Bayesian network classifiers by risk minimization. Int. J. Approx. Reason. 53 (2), 248–272.

Kontkanen, P., Myllymäki, P., Silander, T., Tirri, H., 1999. On supervised selection of Bayesian networks. In: Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc. pp. 334–342.

Langley, P., Iba, W., Thompson, K., 1992. An analysis of Bayesian classifiers. AAAI, vol. 90 223–228.

Law, T.H., Noland, R.B., Evans, A.W., 2009. Factors associated with the relationship between motorcycle deaths and economic growth. Accid. Anal. Prev. 41 (2), 234–240.

Lerner, B., Yeshaya, J., Koushnir, L., 2007. On the classification of a small imbalanced cytogenetic image database. IEEE Trans. Comput. Biol. Bioinform. 4 (2), 204–215.

Li, Z., Liu, P., Wang, W., Xu, C., 2012. Using support vector machine models for crash injury severity analysis. Accid. Anal. Prev. 45, 478–486.

Mnzava, E.G., 2015. The Impact of Motorcycle Accidents in Tanzania: A Case Study of Morogoro Municipality. Mzumbe University PhD thesis.

Mujalli, R.O., de Oña, J., 2013. Injury severity models for motor vehicle accidents: a review. Proceedings of the Institution of Civil Engineers-Transport.

Naqvi, H.M., Tiwari, G., 2017. Factors contributing to motorcycle fatal crashes on national highways in India. Transport. Res. Proc. 25, 2089–2102.

NHTSA, 2015. Motorcycles Traffic Safety Facts.

OECD, 2006. Young Drivers: The Road to Safety. Organization for Economic Co-operation and Development.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Provost, F.J., Fawcett, T., Kohavi, R., 1998. The case against accuracy estimation for comparing induction algorithms. International Conference on Machine Learning 445–453.

Provost, F.J., 2000. Machine learning from imbalanced data sets. Proceedings of the AAAI'2000 Workshop on Imbalanced Data Sets 1–3.

Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accid. Anal. Prev. 43 (5), 1666–1676.

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Inf. Process. Manag. 45 (4), 427–437.

Spirtes, P., Glymour, C., Scheines, R., 2000. Causation, Prediction and Search, 2nd ed. MIT Press, New York.

Toledo, T., Lotan, T., Taubman-Ben-Ari, O., Grimberg, E., 2012. Evaluation of a program to enhance young drivers' safety in Israel. Accid. Anal. Prev. 45, 705–710.

Vlahogianni, E.I., Yannis, G., Golias, J.C., 2012. Overview of critical risk factors in power-two-wheeler safety. Accid. Anal. Prev. 49, 12–22.

Wei, J.-M., Yuan, X.-J., Hu, Q.-H., Wang, S.-Q., 2010. A novel measure for evaluating classifiers. Expert Syst. Appl. 37 (5), 3799–3809.

Yau, K.K.W., 2004. Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong. Accid. Anal. Prev. 36 (3), 333–340.