MAXIMISING ACCURACY AND EFFICIENCY OF TRAFFIC ACCIDENT PREDICTION COMBINING INFORMATION MINING WITH COMPUTATIONAL INTELLIGENCE APPROACHES AND DECISION TREES

Tatiana Tambouratzis[1], Dora Souliou[2], Miltiadis Chalikias[3] and Andreas Gregoriades[4]

[1]Department of Industrial Management & Technology, University of Piraeus,
107 Deligiorgi St, Piraeus 185 34, Greece

[2]School of Electrical and Computer Engineering, National Technical University of Athens,
9 Iroon Polytechniou St, Zografou 15780, Greece

[3]Department of Business Administration, Technological Educational Institution of Peiraius,
250 Thivon and Petrou Ralli Av., 122 44 Egaleo, Greece

[4]Department of Computer Science & Engineering, European University Cyprus, Cyprus

Abstract

The development of universal methodologies for the accurate, efficient, and timely prediction of traffic accident location and severity constitutes a crucial endeavour. In this piece of research, the best combinations of salient accident-related parameters and accurate accident severity prediction models are determined for the 2005 accident dataset brought together by the Republic of Cyprus Police. The optimal methodology involves: (a) information mining in the form of feature selection of the accident parameters that maximise prediction accuracy (implemented via scatter search), followed by feature extraction (implemented via principal component analysis) and selection of the minimal number of components that contain the salient information of the original parameters, which combined bring about an overall 74.42% reduction in the dataset dimensionality; (b) accident severity prediction via probabilistic neural networks and random forests, both of which independently accomplish over 96% correct prediction and a balanced proportion of under- and over-estimations of accident severity. An explanation of the superiority of the optimal combinations of parameters and models is given, as is a comparison with existing accident classification/prediction approaches.

List of Abbreviations:
Accident classification/prediction (AC/P)
Adaptive Resonance Theory (ART)
Abbreviated Injury Scale (AIS)
Akaike Information Criterion (AIK)
Artificial neural network (ANN)
Association for the Advancement of Automotive Medicine (AAAM)
Back-Propagation (BP)
Bayesian Belief Network (BBN)
Bayesian Information Criterion
Computational intelligence (CI)

Cross validation (CV)
Decision tree (DT)
Ensemble Methods (EMs)
Fisher's linear discriminant analysis (LDA)
Fuzzy logic (FL)
Genetic algorithm (GA)
Information mining (IM)
Injury Severity Score (ISS)
Fatal (K), serious (A), moderate (B), minor (C), and none (O) (KABCO)
Machine learning (ML)
Maximum Abbreviated Injury Scale (MAIS)

Meta-learning algorithm (MLA)
Multi Layer Perceptron (MLP)
Na Bayes classifier (NBC)
Principal component (PC)
Principal component analysis (PCA)
Probabilistic neural network (PNN)
Random Committee (C)
Random Forest (RF)
Random Tree (RT)
Scatter Search (SS)
Republic of Cyprus police (RCP)
Traffic accident prevention (TAP)
Traffic accident record (TAR)

1 Introduction

Traffic accident prevention (TAP) is of paramount importance to traffic safety and monitoring. Usually expressed as the short-time prediction of accident location and severity, TAP is effectuated via the timely dispatch of the police[1] to locations of "interest" for traffic monitoring and/or diversion as well as - if deemed necessary - for casualty management. Owing to the crucial nature of TAP, extensive - and usually redundant - accident-related datasets are collected at various locations of interest. Following information mining (IM) via the selection/extraction of pertinent accident-related information, the reduced datasets are combined with appropriate prediction models that afford maximally accurate as well as efficient prediction of accident location and severity (also known as accident classification and prediction, AC/P) across the locations of data collection.

This piece of research constitutes a step towards the development of universal methodologies for the maximally accurate, efficient, and timely AC/P of accident location and severity. The 2005 accident dataset brought together by the Republic of Cyprus Police (RCP) has been employed for establishing the best combinations of

- sets of salient TAP-related parameters from the database that can predict accident location and severity, as determined via IM (including statistical and feature selection/extraction techniques);

- classification/prediction models drawn from existing and customarily used statistical, mathematical, machine learning (ML) and computational intelligence (CI) methodologies.

Concerning IM, the most advantageous sets of parameters have been found to be derived from (a) feature selection of the accident parameters, implemented via scatter search (SS) either sequentially (Glover, 1977) or in parallel (Fx GarcL et al., 2006), followed by (b) feature

---

[1]And the alertness, if so required, of ambulances.

extraction, implemented via principal component analysis (PCA) (Pearson, 1901) and the subsequent preservation of only the first few principal components (PCs), as given by the combination of the scree test (Cattell, 1966) and Kaiser rule (Kaiser, 1960). Overall, a reduction of 74.42% of the original TAP-related information is effectuated. As far as AC/P is concerned, probabilistic neural networks (PNNs) (Specht, 1998), and the decision tree (DT)-based (Breiman, et al. 1984; Quinlan, 1986) ensemble model of random forests (RFs) (Breiman, 2001) have been found maximally accurate for location- and condition-dependent accident severity classification, both of them accomplishing over 96% correct predictions, and a balanced proportion of under- and over-estimations of accident severity.

This paper is organised as follows: section 2 reports on related research in the area of AC/P; section 3 introduces the 2005 RCP dataset and describes IM implemented via parameter selection and extraction; section 4 describes the various accident prediction and severity classification models used for AC/P of the present dataset; section 5 reports on the prediction accuracy of the various classification models implemented here, and makes comparisons with existing research; finally, section 6 concludes the paper.

2 Existing Research for Accident Severity Classification/Prediction

Extensive research has been carried out on AC/P, mainly concerning accident location and severity (for a detailed reference review the author is referred to Wang et al., 2011; Savolainen et al., 2011 and references therein). The following subsections cover the three major characteristics of AC/P, namely accident datasets, prediction/classification models, and IM.

2.1 Accident Datasets and Characteristics

The datasets used in the various pieces of research differ significantly in terms of size, duration of data collection, number and nature of the collected input (independent) parameters, number of collection points (locations), number of output (dependent) classes expressing accident severity, proportion of records in each class etc., thus rendering a direct comparison between the various approaches (and the produced results) far from straightforward. A further complication arises from the fact that the independent parameters collected in the various datasets tend to be distinct (i.e. not to be repeated across datasets). A factor that is common in practically all datasets however is that, in order to ensure that the parameters affecting AC/P are included, some of the col-

lected independent parameters represent redundant, repeated, and/or irrelevant information.

The larger accident datasets usually comprise detailed traffic accident records (TARs) collected at a variety of locations over a significant number of years (e.g. Sohn and Shin, 2001; Sohn and Lee, 2003; Chong et al., 2004); as such, they demonstrate the inherent problems of changes in driving patterns, safety laws, and road/traffic conditions over the years of data collection. Conversely, the smaller datasets contain fewer - though not necessarily less detailed - records that cover shorter time intervals of data collection; although not always as representative of general accident characteristics as their larger counterparts, these smaller datasets tend to be more accurate locally.

Another important factor in data collection is the choice of accident severity classification format used for representing the different accident severity classes (dependent variable). As shown in Table I, the formats appearing in the literature involve anything from two to seven classes (for a review the interested reader is referred to Mujalli and Ona, 2012):

a The most detailed seven-class abbreviated injury scale (AIS) (AAAM 1985) constitutes an anatomical-based coding system of injury description and severity classification. Created by the Association for the Advancement of Automotive Medicine (AAAM), it expresses the threat to life associated with the injury (rather than the actual severity of the injury, Khattak and Rocha, 2003). The six-level maximum AIS (MAIS), constitutes a variant of AIS that (i) does not consider the no-injury/property loss levels of AIS, and (ii) expresses the AIS rating of the most severe injury of an individual involved in an accident; is is widely used by insurance companies for setting life insurance.

b Unlike AIS, the five-class KABCO scale (MCMV-TAR 1976) is used by law enforcement officials for documenting injuries according to their severity rather than their potential threat to life. Combined with MAIS, accident severity can be related to crash costs in terms of (i) the implementation of safety analyses, and (ii) the willingness to invest in road infrastructure and car development in order to reduce the risk of traffic accident fatalities.

c Various four-class formats, e.g. the injury severity score (ISS)[2] (Baker et al. 1974), which divides the human body into six regions, calculates the sum of squares of the AIS score of the (up to) three most serious injuries of an individual involved in an accident, and subsequently determines the care required

for preserving life. Other less frequently used four-class formats appear in (Shanthi and Geetha 2012; Beshah et al., 2012).

d An assortment of three-class formats, as described in (Rezaie et al., 2011; Abdelwahab and Abdel-Aty, 2001; Tavakoli et al. 2012).

e Binary classification formats, e.g. (Sohn and Shin, 2001; Shon and Lee, 2003).

It becomes clear from Table I that the different formats - even those with similar numbers of classes - are not always compatible. This is due not only to the objective of each format (as described above), but also to the different groupings of the accident severity classes, with both factors complicating the comparison of the various formats. For instance:

Table 1. The accident classes according to different classification formats; the grey cells denote class(es) not used in a classification format.

classification / classes	fatality	critical injury	severe injury	serious injury	moderate injury	minor injury	no injury	property loss
AIS	7	6	5	4	3	2	1	
MAIS	6	5	4	3	2	1	▓	▓
KABCO	5	4 (incapacitating)		3 (non-incapacitating)		2 (possible)	1	
ISS	75	16-74		9-15	4-8	1-3	0	
Four [52]	4	3		2		1 (possible)		
Four [7]	4	3 (injury)					2	1
Three	3	2						1
Two	2 (injury)						1	

– The various classes of injury are grouped in distinct ways under the various formats; the critical and severe injury classes sometimes constitute distinct classes, sometimes jointly express the incapacitating

---

[2]Although referred to the relevant literature as a four-class accident classification format, ISS can be considered as a six-class format, as the extreme values of 0 and 75 of the ISS scale actually constitute independent classes (no injury and automatically assigned fatality, respectively).

injury class, and are - at other times - grouped together with the serious, moderate (and under one format also with the minor) injury classes.

– Although the distinction between no-injury and property-loss is rather hard, the two classes surprisingly constitute separate classes in classification formats with a small (three or four only) number of classes.

– Minor injury sometimes constitutes a class of its own, is sometimes grouped with the moderate or non-incapacitating injury class, while it is - at other times - merged with the no injury/property loss class.

– Fatality is grouped with all levels of injury in two-class classification.

Taking into account that it is quite hard for the police officers who arrive on the site of the accident to accurately establish the exact accident class for the more detailed formats in an unbiased fashion (Popkin et al., 1991), it is suggested that it might be preferable to have a relatively small but discriminating number of classes rather than many overlapping and/or hard-to-distinguish classes.

2.2 AC/P

Various approaches to AC/P appear in the literature, including parametric and non-parametric models.

As far as the former models are concerned, statistical binary logit and probit, multinomial logit and probit, ordered logit and probit (Hardin and Hilbe, 2007) as well as regression (McCullagh and Nelder, 1989)) are implemented in Sohn and Lee, 2003; Quddus and Ison, 2011; Savolainen et al. 2011 and references therein.

Existing research on AC/P using non-parametric models is mainly centred on DTs and their combinations from ML, as well as artificial neural networks (ANNs) (Haykin, 1999) and fuzzy logic (FL) (Zadeh, 1965) from CI.

Concerning DTs, an interesting divide-and-conquer approach is implemented in (Sohn and Lee, 2003[3]) for improving DT-based AC/P accuracy, where the accident dataset is first submitted to clustering and a classification model is set up independently for each accident cluster. Prediction accuracy is reported as 91.7% in Chang & Wang (2006), 99.73% in Shanthi & Geetha Ramani (2012), and near 100% in Mujalli & Ona (2012), with a substantial collection of DT-based classifiers (C4.5, CRT, ID3, CS-CRT, CS-MC4, Na Bayes and Random Trees (RTs) (Breiman, 2001) implemented, including combinations of RTs and the Arc-X4 Meta classifier (Breiman, 1998), C4.5 (Quinlan,1993), Na Bayes (Langley et al., 1992) and RTs.

Concerning ANN-based AC/P, the back propagation (BP) ANN - also known as multi-layer perceptron (MLP) (Rumelhart et al., 1986) - has been used consistently for performing accident severity classification (Mussone et al. 1999; Abdelwahab and Abdel-Aty 2001; Sohn and Shin 2001; Rezaie et al. 2011), despite its slow training and limited prediction accuracy of around 65% during testing (Abdelwahab and Abdel-Aty, 2001; Sohn and Shin, 2001). It has been demonstrated in Delen et al. (2006) that BP ANN prediction accuracy can be significantly improved (in their case to 89%) by employing a series of ANN combinations that implement all-versus-all as well as one-versus-all binary classifications; clearly in this case, improved prediction accuracy comes at a non-negligible increase in computational complexity. Other ANNs, such as adaptive resonance theory (ART) ANNs (Grossberg, 1987) are combined with FL, demonstrating superior performance over the probit model (Abdel-Aty and Abdelwahab, 2004).

Alternative CI-based approaches to AC/P include FL classifiers which - when optimised using genetic algorithms (GAs) - reach 94.2% prediction accuracy (Beshah et al., 2012). Overall, it can be concluded that the performance of DTs for AC/P is at least comparable (Sohn and Shin, 2001; Chong et al., 2004; Chang & Chien, 2013) to that of ANNs.

2.3 The Significance of Information Mining

Further to dataset partitioning into clusters with similar input and output parameter characteristics (e.g. Sohn and Lee, 2003; Delen et al., 2006), IM in the form of parameter selection and/or extraction has proved to be effective at improving both classification accuracy and efficiency (e.g. Chang & Wang 2006; Delen et al. 2006; Milton et al., 2008; Mujalli & Ona, 2012; Chang & Chien, 2013; Ma et al., 2008; Worku et al. 2013, and the review paper by Savolainen et al., 2011). Especially concerning feature selection and ranking for AC/P, between six and 11 input parameters have been determined as salient in the related literature, for example:

– driver gender, vehicle speed, seatbelt use, vehicle type, point of impact, and type of accident location (Abdel-Aty & Abdelwahab 2004);

– driver age, seating position, drug involvement, manner of collision, harmful event, protection system (Shanthi & Geetha, 2012),

– frontal crash, 18-22 metre-wide highways, type of vehicle at fault, ignoring length space, ignoring width space, inability to control the vehicle, violating the confidence speed and leftward deviation (Mussone

---

[3]This piece of research also tests ANN-based in the same divide-and-conquer manner.

et al., 1999, focusing upon accidents in urban high-ways);

– alcohol usage, seatbelt use, vehicle type, collision type, contributing circumstance, impact position, number of vehicles involved, accident location (Chang and Chien, 2013);

– driver age, driver gender, alcohol usage, restraint system, eject, vehicle body type, vehicle role, vehicle age, rollover, road surface condition, and light conditions (Chong et al., 2004).

It is interesting that, despite the significant variety in the collected independent parameters, factors such as driver age and gender, substance abuse, seatbelt/protection system, vehicle type and speed, time/light conditions, accident location (subsuming manner of collision, rollover, and many of the causes of accident found in Mussone et al. (1999)) constitute important independent variables for AC/P across datasets.

3   Proposed AC/P Methodology

Similar to existing work (Shanthi and Ramani, 2012), this piece of research is focused upon selecting the most accurate and efficient models for AC/P. The tested models include statistical models, DTs, ANNs, and meta-learning algorithms (MLAs) (Vilalta and Drissi, 2002), the latter combining multiple classifiers for improving performance. In more detail:

a The most significant parameters that cause - or contribute to - accident occurrence and severity are initially determined, and only these are retained. Such a procedure improves efficiency, while it can also increase accident prediction accuracy by eliminating redundant and/or irrelevant information. Since different feature selection techniques (Molina et al. 2002) tend to produce different - though perhaps overlapping - sets of accident-related parameters, the Akaike and Schwartz (or Bayesian) information criteria (AIC (Akaike, 1974; 1980) and BIC (Schwarz and Gideon, 1978), respectively) have been used for deriving the optimal subsets of independent parameters.

b A transformation of the retained parameters into independent (orthogonal) components is subsequently implemented via PCA. Only the first few PCA components that contain the pertinent information of the retained parameters are used, thus boosting the computational efficiency of accident prediction while also preserving prediction accuracy. The combination of the two criteria scree test and Kaiser's rule

has been employed for selecting the number of retained principal components. The scree test operates by creating the curve of the cumulative information contained in the principal components, and selecting the point at which the curve changes slope as the principal component cut-off point; Kaiser's rule, on the other hand, uses the eigenvalues of the components in order to only retain those whose eigenvalues are greater than 1. This combination has been implemented as it has been found capable of overcoming the disadvantages of each methodology when used alone, namely a not always clear change in slope for the scree test, and the excessive stringency of the Kaiser rule that more often than not results in the cutting off of more principal components than is appropriate.

In the following analysis, five-fold cross validation (CV) (Devijver and Kittler 1982) is used for evaluating prediction accuracy: the dataset is randomly divided into five practically equal parts, with four parts used for setting up the model, and the remaining part reserved for testing it. The training and testing of each model is repeated five times[4] , and prediction accuracy is evaluated by averaging the results of the five folds used for testing (with each fold being used exactly once as the test set). The results of each tested model are computed using the same parameter values over all folds.

3.1 Dataset Description and Pre-Processing

As shown in Table II, the 2005 RCP accident dataset comprises 1407 records, with each record consisting of 43 input parameters that have been collected on-the-spot by police officers, eye witnesses and the involved parties, and a single output parameter pertaining to accident severity.

---

[4]Using each part of the dataset four times in the training set and once in the test set.

Table 2. Input accident parameters grouped by type (global, local, temporal, accident-related, driver and car); grey cells correspond to the parameters retained following feature selection.

Description	Type	Nature	Values
area (town/rural)	Global	categorical	1:2
police district		categorical	1:2
residential area		categorical	1:2
traffic control (none/out of order, stop/give way/round about, police/traffic signs)	Local	categorical	1:3
road width		continuous	[0 44]
pavement/shoulder width		continuous	[0 7]
diagram code		categorical	1:11
conjunction type		categorical	1:8
routes permitted		categorical	1: 3
barrier type (none, single/broken, continuous, island, combination)		categorical	1:5
break lane		categorical	1:4
speed limit		continuous	[10 100]
bus stop		categorical	1:2
road description		categorical	1:3
pedestrian crossing (none, zebra, pedestrian traffic signal, pedestrian push button, pedestrian pelican crossing, other)		categorical	1:6
Month (January, ...)	Temporal	categorical	1:12
Day (Sunday, ...)		categorical	1:7
time (11am-1.59pm, 2-4.59pm, 5 to 7.59pm, 8-10.59pm, 11-1.59am, 2-4.59am, 5-7.59am, 8-10.59am)		categorical	1:8
light conditions		categorical	1:5
pavement status		categorical	1:2
road works (yes, no)		categorical	1:2
weather conditions		categorical	1:2
action before accident	Accident-related	categorical	1:2
number of vehicles		continuous	[1 6]
number of injured		continuous	[1 7]
hit and run		categorical	1: 2
main cause (mental state, inability, carelessness, vehicle fault, environmental)		categorical	1:5
second cause (as in main cause plus none)		categorical	1:6
third cause (as in second cause)		categorical	1:6
fourth cause (as in second cause)		categorical	1:6
accident location		categorical	1:3
type of event for main cause		categorical	1:4
time for ambulance		continuous	[0 55]
vehicle sequence		categorical	1:5
type of event for second cause		categorical	0:1
age	Driver	categorical	1:4
gender		categorical	1:2
driver license type		categorical	1:3
manufacture year	Car	categorical	1:4
CC		categorical	1:10
vehicle type		categorical	1:5
vehicle license		categorical	1:2
vehicle worthiness certificate		categorical	1:2

The 43 input parameters describe a wide variety of accident-related characteristics grouped into the following six classes: global (concerning the vicinity/locality of the accident), local (concerning the morphology and road layout at the location of the accident), temporal (concerning time-varying variables at the location of the accident), and purely accident-, driver- and car-related classes. The output parameter takes on three values corresponding to light, serious and fatal accidents i.e. belongs to the three-class accident-severity format of Section 2.1. The small number of the output (accident severity) parameter values facilities data collection on the site of the accident, thus providing superior differentiation over two-class classification, while simplifying classification compared to formats with more classes. Even though there is still uncertainty and room for error under this format[5], it has the advantage of (i) being straightforward for the purpose of further action (e.g. dispatch of police officers only, or of police officers and ambulance(s) to the spot), (ii) more robust to errors of collection and/or judgment when compared to the more detailed accident classification formats, and (iii) easier to analyse.

Having 55 (3.91%), 512 (36.39%), and 840 (58.70%) records corresponding to fatal, serious and light accidents, respectively, the 2005 RCP dataset is heavily imbalanced, with fatal accidents - which it is of paramount importance to predict - constituting a very small proportion of the records.

In order (a) not to interfere with the data collection process, (b) to simulate real operating conditions, and (c) to boost the generalisation and robustness-to-noise properties of the prediction models, dataset pre-processing is minimal: the entire dataset is retained for the ensuing analysis, with (i) parameter values that are missing or fall outside the range of valid values being substituted by the mean of the values of the other records for that parameter, and (ii) input categorical parameters that take on more than distinct 12 values having their values re-grouped into fewer categories.

A preliminary statistical analysis (Tambouratzis et al. 2010) of the collected parameters employing the Spearman correlation coefficient values (Spearman 1904), and the p-values (Goodman, 1999) shows that the input parameters "area" and "police district" provide practically the same information (correlation coefficient of 0.9952), while - despite their names - the input parameters "area" and "residential area" contain distinct information (correlation coefficient of 0.4276), as do the input parameters "police district" and "residential area" (correlation coefficient of 0.4302). Taking into account these findings, only the "police district" input has been

---

[5]For instance, a thin line exists between light and serious accidents; serious accidents may actually be fatal if mortality occurs after prolonged hospitalisation instead of on the accident site or during transfer to the hospital.

removed from the dataset and is not used in the ensuing analysis.

3.2 Information Mining (Input Parameter Selection and Extraction)

In order to determine the minimal information that is pertinent for accident severity prediction into the three classes of fatal, serious and light accidents, the 42 input parameters resulting from pre-processing are initially submitted to feature selection. By removing redundant information, the number of input parameters required for AC/P is reduced; additionally, by removing irrelevant information, the input-output relationship between accident-related parameters and accident severity becomes clearer and, thus, easier to model.

Since searching the entire set of subsets of independent parameters for determining the subset that most accurately predicts the output parameter value constitutes an NP-hard optimisation problem (Kohavi and John, 1997), two distinct lines of approach are usually followed for parameter selection:

a) filter methods, which assess the parameters independently of the problem at hand and filter out the irrelevant and redundant attributes, and

b) wrapper methods, which employ machine learning techniques for evaluating the derived parameter subsets based on the output criterion.

Among filter methods, correlation- and probabilistic-based evaluators are used for finding sets of parameters that have good predictive ability but are not correlated with each other, e.g. CfsSubsetEval (Hall, 1998), and ConsistencySubsetEval (Liu and Setiono 1996). Both forward and backward subset selection (Devijver and Kittler, 1982) are tested here as well as combinations of the two, the latter involving starting with a subset of independent parameters, and either adding or deleting parameters at each step according to the criterion of maximum increase of prediction ability.

Wrapper methods generate a list of ranked parameters by estimating, usually via CV, the ability of each independent parameter to predict the dependent (output) parameter. Each wrapper method is based on a different estimation function, e.g. the chi-squared statistic with respect to the class attribute for ChiSquaredAttributeEval, the information gain with respect to the value of the class attribute by InfoGainAttributeEval etc. (Hall et al. 2009) etc. The consistency-based evaluator, combined with SS, has been found to produce the optimal subset of 18 retained parameters (highlighted in Table II), with optimality confirmed by both the Akaike and the Schwarz Information criteria. It is worth mentioning that these parameters agree to a considerable degree with the significant independent parameters discussed in section 2.3.

The final step of parameter extraction is implemented via PCA for further compressing the information contained in the 18 parameters which result from feature selection. Concerning the selection of only the first few PCA components which carry the most significant input parameter-related information, (a) a change in slope at the 11th PCA component is indicated by the scree test, while only the first nine PCA components are retained according to Kaiser's rule. Taking into account, however, (a) the tendency of the Kaiser rule to excessively "cut" PCA components, and (b) the eigenvalues of the 10th and 11th components assume values just below 1, the first 11 PCA components are retained. In order to verify this choice, all the models presented in the next section are also tested by using the first one, two, three, and so on up to the total of 11 PCA components, with the optimal results reported in each case.

4 AC/P Models

Four kinds of models are employed for AC/P in terms of severity:

I. Three statistical models:

- The well-known probabilistic Na Bayes Classifier (NBC) (Langley et al., 1992; Ramoni and Sebastiani, 2001), which is elegant and simple but can - due to its assumption of independence - perform in a suboptimal manner when the input parameters are highly correlated.

- The Bayesian Belief Network (BBN) (Heckerman, 1997), which constitutes a trainable probabilistic graphical model that improves over NBC by expressing parameters and conditional dependencies as a directed acyclic graph (DAG).

- Fisher's linear discriminant analysis (LDA) (Fisher 1936), which determines the optimal linear combination of the input parameters such that classification accuracy is maximised.

II. Two DTs:

- C4.5(J48) (Quinlan, 1993), where the input parametes are selected according to the maximisation of the information gain.

- CART (Breiman et al., 1984), where the Gini impurity function (GIF) (Gini C., 1909; Gini C., 1912) is used as the selection measure of the input parameters.

III. Two ANN architectures:

– The BP ANN.

– PNNs (Specht 1998), which are based on the Bayesian classification of categorical outputs. A single presentation of each training pattern is sufficient for training, with the created non-linear decision surface (boundary) between classes approaching the Bayes optimal for increasing sizes of the training set and an appropriate value of the smoothing parameter σ.

IV. Three MLAs:

– Bagging (also known as bootstrap aggregating, Breiman (1996)), where the repeated sampling of the dataset according to the uniform probability distribution reduces the variance of the base classifier, thus preventing overfitting while boosting generalisation.

– Random Tree (RT), which considers a fixed number of randomly chosen parameters at each node as candidates for splitting the dataset.

– Random Forest (RF) (Breiman, 2001), which combines the predictions made by multiple DTs created by introducing randomness in both the training dataset and the training algorithm.

– Random committee (RC) (Hall et al. 2009), which constitutes a combination of RTs each derived using a different random seed.

5   AC/P Results - Comparisons with Existing Research

The investigation into the most accurate and efficient combination of technique and model for AC/P proceeds by employing each model presented in section 4 to the following input parameter sets:

– The 42 input parameters resulting from pre-processing. This provides the measure of accuracy reached by each model when the original dataset is used.

– The set of 18 input parameters remaining after the application of feature selection. This determines whether prediction accuracy can be maintained - or even improved – when using an appropriately selected subset of the original parameters. An improvement not only supports that the removal of irrelevant and/or repeated parameters better exposes the relationship between input and output parameters, but can also be used to suggest a minimal set of parameters that needs to be collected in an accident dataset.

– The further reduced set of transformed parameters that result from PCA feature extraction of the 18 parameters used in the previous step. Each model is tested with the first one, two, three, ... up to 18 PCA components, and the prediction accuracy of each model is reported with the number of PCA components that maximise prediction accuracy; when this number equals 11, the combination of the scree test and Kaiser's rule for determining the optimal number of retained PCA components is further supported/conformed.

Table 3. Proportion of correct AC/P per accident severity class as well as overall; over- and under-estimations (42 original input parameters).

Prediction Models		Fatal	Serious	Light	Correct	Under-estimations	Over-estimations
Statistical	NB	49.09	40.23	83.3	66.31	20.54	13.14
	BN	35.55	53.71	80.48	68.94	18.62	68.94
	LDA	50.13	58.14	49.78	52.60	15.4	32.00
DTs	C4.5 (J48)	76.36	85.94	93.33	89.98	5.69	4.33
	CART	72.73	86.13	94.64	90.69	5.90	3.41
ANNs	BPN	70.91	86.33	92.62	89.48	5.76	4.76
	PNN	89.09	92.38	96.55	94.74	2.70	2.56
EM MLA	Bag-Ging	63.64	84.77	97.38	91.47	6.97	1.56
	RC	85.45	96.68	97.74	96.87	1.71	1.42
	RF	81.82	93.95	97.86	95.81	2.91	1.28

The proportions of correctly predicted (a) fatal, serious, and light accidents, (b) accidents independent of class, and proportions of (c) underestimated and overestimated predictions overall, are employed for evaluating and comparing prediction accuracy of the various combinations of techniques (i.e. sets of parameters) and models. For the evaluation, the accuracy of predicting fatal accidents, the percentage of correct predictions overall, and the balance of over- and underestimations (otherwise the bias towards overestimations rather than underestimations) are of particular interest, as these are crucial for the accurate (or - at worst - conservative/pessimistic) prediction of accident severity.

Table III illustrates the prediction results attained by the models when the 42 original parameters are used. The RC and RF DT-based ensemble models, together with the PNN, are found to be the most accurate, with prediction accuracy being around - or exceeding - 95%.

These are followed by the DT-based ensemble model of bagging, the single-DT methodologies (CART and C4.5) and the BPN, thus demonstrating a general agreement with the results produced by existing research as far as the superiority of DT- and ANN-based models for AC/P is concerned. The RC and PNN models have the additional advantage of being conservative[6], as shown by the balance between over- and under-estimations. Statistical models are not found satisfactory for AC/P.

Table IV illustrates the prediction results attained by the same models when only the 18 input parameters derived from feature selection are used. Classification accuracy improves for the PNN, which now constitutes the most accurate model for AC/P, remains unchanged for RF and falls slightly for RC, thus showing that IM in the form of parameter selection not only does not impair - but may actually improve - performance, also increasing efficiency. The same observations can be made for the bagging, CART, C4.5 and BPN models. Prediction accuracy of the BN and the NB further deteriorates, while LDA improves though not to a satisfactory level. As the accuracy of the statistical models remains well under 70%, these models are not considered further.

Table 4. Proportion of correct AC/P per accident severity class as well as overall; over- and under-estimations following feature selection (18 original parameters).

Prediction Models		Fatal	Serious	Light	Correct	Under-estimations	Over-estimations
Statistical	NB	38.18	34.57	83.57	63.97	22.74	13.29
	BN	14.55	46.48	81.67	66.24	21.54	12.22
	LDA	71.15	57.06	69.72	65.10	12.37	22.53
DTs	C4.5 (J48)	69.06	85.16	94.88	90.33	6.04	3.63
	CART	65.45	83.20	93.69	88.91	7.39	3.70
ANNs	BPN	54.55	82.03	95.24	88.84	8.17	2.99
	PNN	89.09	94.53	98.57	96.74	1.63	1.63
EM MLA	Bag-Ging	63.64	85.55	97.38	91.76	6.53	1.71
	RC	78.18	92.19	97.14	94.60	3.62	1.78
	RF	78.18	90.82	97.98	94.60	4.05	1.35

Concerning PCA information compression of the 18 input parameters, Table V illustrates prediction accuracy for the best results attained when testing each model with the first one, two, three, ..., through to 18 PCs; the number of PCs that are found optimal for each model are shown in the second column of the Table. It is clear that 10 or 11 constitutes the optimal number of PCAs retained for all models, thus confirming the combination of the scree test and Kaiser's rule. All the models show minimal variations in prediction accuracy, with only the PNN and RFs showing an actual improvement over feature selection (i.e. whrn using the 18 selected parameters), thus confirming that an increase in both efficiency and accuracy can be implemented via PCA. The PNN remains superior in terms of both accuracy and conservative prediction, as also supported by the balance in over- and under-estimations of accident severity.

Table 5. Proportion of correct AC/P per accident severity class as well as overall; over- and under-estimations following feature extraction (combination of scree test and Kaiser's rule on the selected and PCA-transformed parameters). The number of PCA components demonstrating optimal AC/P are shown in the second column of the Table.

Prediction Models		Fatal	Serious	Light	Correct	Under-estimations	Over-estimations
DTs	C4.5 (J48) 11	63.64	86.91	93.57	89.98	5.97	4.05
	CART 10	61.82	83.21	92.38	87.85	7.32	4.83
ANNs	PNN 10	83.33	97.99	97.41	97.01	1.35	1.63
EM MLA	RC 10	83.64	88.87	98.10	94.17	4.41	1.42
	RF 11	87.27	93.55	98.45	96.23	2.84	0.93

These results show that DTs, ensemble methods, and the PNN benefit from IM, combining efficiency with superior AC/P. These findings are in agreement with the results appearing in the relevant literature, especially the overall consensus that DTs are superior to other models, including BPNs. Furthermore, the good prediction capability of the PNN puts forward (for the first time) this kind of ANN as an appealing model for performing AC/P.

---

[6]Owing to the significantly larger number of light accidents in the dataset and the small number of fatal accidents, underestimations would be expected to be more abundant.

6 Conclusion

The combination of information mining and parametric as well as non-parametric accident classification and prediction models has been investigated for predicting traffic accident severity on the 2005 accident dataset brought together by the Republic of Cyprus Police.

Despite the lack of any kind of standardisation of the accident datasets to date, the selected set of accident-related parameters for the accident dataset used here has been found to have many common points with the sets of independent parameters deemed as important in the relevant literature.

The application of parameter selection followed by feature extraction and the preservation of only the first few salient components (implemented by a combination of the scree test and Kaiser's rule), coupled with PNNs and RFs, has been found to cause a noteworthy (74.42%) reduction in the original accident-related information (i.e. a significant increase in computational efficiency), while still improving on accident severity prediction accuracy (97% for the PNNs, with a balanced proportion of over- and under-estimations of accident severity).

Future research will focus upon confirming the superiority of PNNs as accident severity classification/prediction models using existing datasets of the relevant literature, also putting forward a set of independent parameters that are not only salient, but also sufficient, for traffic accident severity prediction.

References

[1] AAAM, 1985. Abbreviated Injury Scale 1985. Des Plaines IL: Association for the Advancement of Automotive Medicine.

[2] Abdel-Aty M.A., Abdelwahab H.T., 2004. Predicting injury severity levels in traffic crashes: a modeling comparison, Journal of Transportation Engineering, vol. 130, pp. 204-210.

[3] Abdelwahab, H.T., Abdel-Aty, M.A., 2001. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalizes, Intersection Transportation Research Record issue 1746, pp. 6-13.

[4] Akaike, H., 1974. A new look at the statistical model identification, IEEE Transactions on Automatic Control 19(6): 716–723

[5] Akaike Hirotugu, 1980. Likelihood and the Bayes procedure, in Bernardo, J. M.; et al., Bayesian Statistics, Valencia: University Press, pp. 143–166.

[6] Baker S.P., O'Neill B., Haddon Jr W., Long W.B., 1974. The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care, The Journal of Trauma (Lippincott Williams & Wilkins), vol. 14, pp. 187–196.

[7] Beshah T., Ejigu D., Kromer P., Snasel V., Platos J., Abraham A., 2012. Learning the Classification of Traffic Accident Types, Fourth International Conference on Intelligent Networking and Collaborative Systems, Bucharest, Romania, September 19th-21st, 2012, pp. 463-468.

[8] Breiman L., Friedman J.H., Olshen R.A., Stone, C.J., 1984. Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

[9] Breiman Leo, 1996. Bagging predictors. Machine Learning 24(2): 123–140.

[10] Breiman, Leo., 1998 Arcing classifiers, The Annals of Statistics, vol. 26, pp.801-849.

[11] Breiman Leo., 2001 Random Forests. Machine Learning, volume 45, pp.5-32.

[12] Catell, R.B., 1966. The scree test for the number of factors. Multivariate Behavioral Research, 1,245-276

[13] Chang, L-.Y, Wang H.-W., 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques, Accident Analysis and Prevention, vol. 38, pp. 1019–1027.

[14] Chang L.Y., Chien J.-T, 2013 Analysis of driver injury severity in truck-involved accidents using a non-parametric classification tree-model, Safety Science, vol. 51, pp. 17-22.

[15] Chong M.M., Abraham A., Paprzycki M., 2004. Traffic accident analysis using decision trees and neural networks, IADIS International Conference on Applied Computing, Portugal, IADIS Press, Pedro Isaias et al. (Eds.), ISBN: 9729894736, Vol. 2, pp. 39-42.

[16] Delen D., Sharda R., Bessonov M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks, Accident Analysis and Prevention, vol. 38, pp. 434-444.

[17] Devijver P.A., Kittler J. 1982. Pattern Recognition: A Statistical Approach, Prentice-Hall, London, U.K.

[18] Fx GarcL, Miguel GarcTorres, BeleliBatista, Jos Moreno-Pz, J. Marcos Moreno-Vega: Solving feature subset selection problem by a Parallel Scatter Search. European Journal of Operational Research 169(2): 477-489 (2006)

[19] Fisher, R. A., 1936. The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics 7 (2): 179–188.

[20] Gini C., 1909. Concentration and dependency ratios (in Italian). English translation in Rivista di Politica Economica, 87 (1997), 769–789.

[21] Gini C., 1912. "Italian: VariabilitutabilitVariability and Mutability', C. Cuppini, Bologna, 156 pages. Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi (1955).

[22] Glover F., 1977. Heuristics for integer programming using surrogate constraints. Decision Sciences, vol. 8, pp. 156–166.

[23] Goodman, SN 1999. Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy. Annals of Internal Medicine 130: 995–1004

[24] Grossberg S., 1987. Competitive learning: from interactive activation to adaptive resonance, Cognitive Science, vol. 11, pp. 23-63.

[25] Hall M. A., 1998. Correlation-based Feature Subset Selection for Machine Learning. Hamilton, New Zealand

[26] Hall Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten 2009; The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

[27] Hardin J., Hilbe J., 2007. Generalized Linear Models and Extensions (2nd edition). College Station: Stata Press.

[28] Haykin S., 1999. Neural Networks: A Comprehensive Foundation (2nd Edition), Prentice-Hall, Upper Saddle River, NJ.

[29] Heckerman D. 1997. Bayesian Networks for Data Mining. Data Mining and Knowledge discovery, 1(1) : 79-119, 1997.

[30] Kaiser, H. F., 1960 The application of electronic computer to factor analysis. Educational and Psychological Measurement, 20, 141-151.

[31] Khattak A., Rocha M., 2003. Are SUVs "supremely unsafe vehicles"? Analysis of rollovers and injuries with sport utility vehicles, Transportation Research Record 1840, pp. 167–177.

[32] Kohavi R., John G. H., 1997. Wrappers for feature subset selection, Artificial Intelligence 97 (1–2) 273–324

[33] Langley P., Iba W., Thompson K., 1992. An analysis of Bayesian Classifiers. In Proc. Of the 10th National Conf. on Artificial Intelligence, pages 223-228.

[34] Liu H., Setiono R., 1996. A probabilistic approach to feature selection - A filter solution. In: 13th International Conference on Machine Learning, 319-327

[35] Ma J., Kockelman KM., Damien P. 2008 A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis and Prevention 40:964-975 (2008).

[36] McCullagh P., Nelder J., 1989. Generalized Linear Models, London: Chapman and Hall, London, U.K.

[37] MCMVTAR 1976 Manual on Classification of Motor Vehicle Traffic Accidents-Revision of 016.11970, Third Edition, National Safety Council, Chicago, Illinois, 1976.

[38] Milton Jc, Shankar Vn, FL Mannering Fl Highway accident severities and the mixed logit model: An exploratory empirical analysis Accident Analysis & Prevention 40 (**??**), 260-266

[39] Molina L.C., Belanche L., and Nebot A., 2002. Feature Selection Algorithms: A survey and Experimental Evaluation. In Proc. Of the 2002 IEEE Intl. Conf. on Data Mining.

[40] Mujalli R.O., J. de Ona, 2012. Injury severity models for motor vehicle accidents: a review, Proceedings of the ICE - Transport, vol. 166, pp. 255-270.

[41] Mussone L., Ferrari A., Oneta M., 1999. An analysis of urban collisions using an artificial intelligence model, Accident Analysis and Prevention, vol. 31, pp. 705–718.

[42] Pearson, K., 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine, vol. 2, pp 559–572.

[43] Popkin C.L., Campbell B.J. Hansen A.R., and Stewart J.R., 1991. Analysis of the accuracy of the existing KABCO injury scale, Chapel Hill, NC: University of North Carolina Highway Safety Research Center e-archives scan.

[44] Quddus M.A., Ison S.G., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model, Accident Analysis & Prevention, vol. 43, pp. 1979-1990.

[45] Quinlan, J. R., 1986. Induction of Decision Trees. Machine Learning 1: 81-106, Kluwer Academic Publishers.

[46] Quinlan, J. R.,1993 C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[47] Ramoni M., and Sebastiani P., 2001. Robust Bayes Classifier. Artificial Intelligence, 125: 209-226.

[48] Rezaie Moghaddam F., Afandizadeh Sh., Ziyadi M., 2011. Prediction of accident severity using artificial neural networks, International Journal of Civil Engineering, vol. 9,pp. 41-49.

[49] Rumelhart D.E., Hinton G.E., Williams R. J., 1986. Learning internal representations by error propagation, Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1: Foundations, Rumelhart D.E., McClelland J.L., and the PDP research group. (eds), MIT Press, 1986

[50] Savolainen, P.T., Mannering, F.L., Lord, D., and M.A. Quddus, 2011. "The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives", Accident Analysis and Prevention, Vol. 43, No. 5, 2011, pp. 1666-1676.

[51] Schwarz, Gideon E. 1978. Estimating the dimension of a model. Annals of Statistics 6 (2): 461–464

[52] Shanthi S., Geetha Ramani, 2012. Feature relevance analysis and classification of road traffic accident data through data mining techniques, Proceedings of the World Congress on Engineering and Computer Science (WCECS 2012), October 24th-26th, 2012, San Francisco, U.S.A., Vol I, pp. 122-127.

[53] Shanti S., Geetha Ramani, 2012. Vehicle Safety Device (Airbag) Specific Classification of Road Traffic Accident Patterns through Data Mining Techniques. ACITY (2) 2012: 433-443

[54] Sohn, S.Y., Shin, H.W., 2001. Data mining for road traffic accident type classification, Ergonomics, vol. 44, pp. 107–117.

[55] Sohn S.Y., Lee S.H., 2003. Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accidents in Korea, Safety Science, vol. 41, pp. 1–14.

[56] Spearman C., 1904. General Intelligence, objectively determined and measured. Am J Psychol 15:202-93.

[57] Specht D. 1998. Probabilistic neural networks for classification, mapping, and associative memory, in Proceedings of the IEEE International Conference on Neural Networks, New York, U.S.A, pp. 525-532 (vol. 1).

[58] Tambouratzis Tatiana, Souliou Dora, Chalikias Miltiadis S., Gregoriades Andreas: Combining probabilistic neural networks and decision trees for maximally accurate and efficient accident prediction. IJCNN 2010: 1-8

[59] Tavakoli Kashani A., Shariat-Mohaymany A., Ranjbari A., 2012. Analysis of factors associated with traffic injury severity on rural roads in Iran, Journal of Injury and Violence Research, vol. 4, pp. 36-41

[60] Vilalta R., Drissi Y., A perspective view and survey of meta-learning, Artificial Intelligence Review, VOL. 18, PP. 77-95, 2002

[61] Wang C., Quddus M.A., Ison S.G., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model, Accident Analysis & Prevention, vol. 43, pp.1979–1990

[62] Worku Y.M., Deogratias E., Deo C., Maher Q. 2013, Exploring factors contributing to injury severity at freeway merging and diverging locations in Ohio. Accident Analysis & Prevention Volume 55, June 2013, Pages 202–210

[63] Zadeh, L.A., 1965. Fuzzy sets, Information and Control, vol. 8, pp. 338–353