

# Real-time Classification of Aggregated Traffic Conditions using Relevance Vector Machines

Christos Katrakazas

PhD Student

School of Civil & Building Engineering

Loughborough University

Loughborough LE11 3TU

United Kingdom

Tel:+44 (0)7789656641

Professor Mohammed Quddus\*

School of Civil & Building Engineering

Loughborough University

Loughborough LE11 3TU

United Kingdom

Tel: +44 (0) 1509228545

Email: m.a.quddus@lboro.ac.uk

Professor Wen-Hua Chen

School of Aeronautical and Automotive Engineering

Loughborough University

Loughborough LE11 3TU

Tel: +44 (0)1509 227 230

Email: w.chen@lboro.ac.uk

\*Corresponding author

Word Count: 5,259 words + 4 Tables\*250=6,259

Paper submitted for presentation at the 95<sup>th</sup> Annual Meeting of Transportation Research Board (TRB),  
Washington D.C., USA

**ABSTRACT**

This paper examines the theory and application of a recently developed machine learning technique namely Relevance Vector Machines (RVMs) in the task of traffic conditions classification. Traffic conditions are labelled as *dangerous* (i.e. probably leading to a collision) and *safe* (i.e. a normal driving) based on 15-minute measurements of average speed and volume. Two different RVM algorithms are trained with two real-world datasets and validated with one real-world dataset describing traffic conditions of a motorway and two A-class roads in the UK. The performance of these classifiers is compared to the popular and successfully applied technique of Support vector machines (SVMs). The main findings indicate that RVMs could successfully be employed in real-time classification of traffic conditions. They rely on a fewer number of decision vectors, their training time could be reduced to the level of seconds and their classification rates are similar to those of SVMs. However, RVM algorithms with a larger training dataset consisting of highly disaggregated traffic data, as well as the incorporation of other traffic or network variables so as to better describe traffic dynamics, may lead to higher classification accuracy than the one presented in this paper.

## 1 INTRODUCTION

2 In many Intelligent Transport Systems and Services (ITSS), it is mandatory to estimate the  
3 probability of a traffic collision occurring in real-time or near real-time. Examples include various  
4 advanced driver assistance systems (1), autonomous or semi-autonomous driving (2) and operations  
5 of smart motorways in the UK (3). Previous research has established a fundamental theory that  
6 implies that a set of unique traffic dynamics (e.g. speed, flow, vehicle compositions) lead to a  
7 collision occurrence (4–6). Based on this concept, researchers have studied traffic conditions just  
8 before the collision occurrences, along with traffic conditions at normal situations, so as to establish a  
9 technique of estimating the risk of a collision in real-time. The correct and reliable identification of  
10 pre-collision traffic conditions for the purpose of real-time collision prediction are still in their  
11 infancy. More research is therefore imperative and timely as future vehicles and infrastructures will  
12 make intelligent decisions to avoid a collision without the need of any human input.

13 In most studies, the risk of a collision occurring in real-time has been estimated by comparing  
14 and contrasting collision-prone traffic conditions at a road segment prior to a collision with  
15 conditions on the same segment at different time points (7). The comparison between ‘safe’ and ‘  
16 ‘dangerous’ traffic conditions largely relies on selecting important variables from the dataset and then  
17 checking the influence of these variables on matched-case samples. In order for these models to be  
18 accurate enough for real-time applications, the temporal data aggregation should be carried out for a  
19 relatively short period and has to be noise-free. It is assumed though that roads are well instrumented  
20 to provide raw traffic data with adequate quality. In terms of motorway operations, it has been found  
21 that traffic conditions at 5-10 minutes before the collision would provide the required time for  
22 introducing an intervention designed by the responsible traffic agencies to avoid imminent traffic  
23 congestion or a collision (8).

24 In reality, not every road is instrumented and highly disaggregated quality traffic data is  
25 difficult to find. Moreover, because of the emerging technologies which aim to automate transport  
26 systems, faster real-time prediction and implementation are essential. Efficient handling and mining  
27 of low quality or highly aggregated data is therefore needed. Machine learning and data mining  
28 techniques could therefore prove an essential alternative or at least supplement traditional (e.g.  
29 logistic regression(9)) or more sophisticated techniques (e.g. Neural Networks (10), Bayesian  
30 Networks (11) or Bayesian Logistic Regression (12)) for collision prediction and traffic conditions  
31 classification. For example, adding an extra variable to a real-time collision prediction algorithm  
32 without compromising the computation time would enhance the prediction results.

33 With this concept in mind, the current study explores the application of Relevance Vector  
34 Machines (RVMs) (13), a sparse Bayesian supervised machine learning algorithm, for classification  
35 of traffic conditions using highly aggregated measurements of speed and volume. A matched-case  
36 control data structure is used, in which 15-minutes traffic conditions before each collision (termed as  
37 a ‘dangerous’ condition) are matched with four 15-minute normal traffic conditions (termed as a  
38 ‘safe’ condition) using a RVM classification. The findings from the RVM classification is compared  
39 to that of a commonly employed Support Vector Machine (SVM) classification (14) in order to  
40 examine the viability of the RVM classification for real-time traffic conditions classification.

41 The paper is organised as follows: firstly, the existing literature and its main findings are  
42 reviewed. An analytic description of the RVM classification algorithm is described next. This is  
43 followed by a presentation of the data used in the analysis, the pre-processing method and the results  
44 of the classification algorithm. Finally, the last section summarises the main conclusions of the study  
45 and gives recommendations for future research.

## 46 LITERATURE REVIEW

48 Detecting incidents or predicting the probability of a collision occurring in real-time based on  
49 the evolution of traffic dynamics by using machine learning techniques such as Support Vector  
50 Machines (SVMs) has gained attention in the literature over the past decades. Yuan and Cheu (15)

pioneered using SVMs to detect incidents from loop detector data on a California freeway. Their work revealed that SVMs is a successful classifier with a small misclassification and false alarm error. For the same purpose of freeway incident detection, Chen et al. (16) used a modified SVM termed as SVM ensemble methods (i.e. bagging, boosting, cross-validation) to overcome the drawback of the traditional SVM with respect to the chosen kernel and its tuning parameters. The same approach of SVM ensembles was adopted by Xiao and Liu (17) who trained multiple kernel SVMs with the bagging approach to reduce the complexity of the classifier. Ma et al. (18) attempted to assess the traffic conditions based on kinetic vehicle data acquired by a Vehicle-to-Infrastructure (V2I) integration by comparing the performance of SVMs and artificial neural networks (ANNs). They showed that SVMs outperform ANNs in correctly identifying incidents but their work was limited to traffic micro-simulation data only. In Yu et al (19), a traffic condition pattern recognition approach using SVMs was employed to distinguish between blocking, crowded, steady and unhindered traffic flow using average speed, volume and occupancy rates. The classification results showed high accuracy in recognising traffic patterns. Furthermore, this study indicated that using normalised data can enhance the classification results, depending on the kernel function that is chosen.

Apart from the applications to detect incidents, SVM models have also been applied to real-time collision and traffic flow predictions. Li et al. (20) compared the use of SVMs with the popular negative binomial models for motorway collision prediction. Their results showed that SVM models have a better goodness-of-fit in comparison with negative binomial models. Their findings were in line with the study of Yu & Abdel-Aty (21) who compared the results from SVM and Bayesian logistic regression models for evaluating real-time collision risk demonstrating the better goodness-of-fit of SVM models. The prediction of side-swipe accidents using SVMs was evaluated in Qu et al. (22) by comparing SVMs with Multilayer Perceptron Neural Networks. Both techniques showed similar accuracy but SVMs led to better collision identification at higher false alarm rates. More recently, Dong et al. (23) demonstrated the capability of SVMs to assess spatial proximity effects for regional collision prediction.

SVMs have proven to be an efficient classifier as well as a successful predictor when utilised for traffic classification or collision prediction studies. However, a number of significant and practical disadvantages are also identified in the literature (13):

- 1) The number of Support Vectors (SVs) usually grows linearly with the size of the training set, and the use of basis functions<sup>1</sup> is considered rather liberal;
- 2) Predictions are not probabilistic and classification problems which require class membership posterior probabilities are sometimes intractable;
- 3) SVMs require a cross-validation procedure which can lead to misuse of data and computational time;
- 4) The kernel function must satisfy the Mercer's condition (i.e. it must be a continuous symmetric kernel of a positive integral operator).

Given these limitations of SVMs, an alternative method should be sought to examine whether such disadvantages pose a critical threat to the reliability and validity of real-time collision prediction algorithms. After reviewing data mining and pattern recognition methods employed in other disciplines, it seems that Relevance Vector Machines (RVMs) which is a sparse Bayesian supervised machine learning algorithm could be a potential candidate. RVMs have been applied in many different areas of pattern recognition and classification including channel equalisation (24), feature selection (25), hyperspectral image classification (26), as well as biomedical applications (27, 28). This paper is a first attempt to investigate the suitability and capability of Relevance Vector Machines to classify real-time traffic conditions using aggregated traffic measurements of average speed and volume collected from major UK roads.

---

<sup>1</sup> In SVMs and RVMs, a basis function is defined for each of the data points using a kernel. More explanation is given in the following section, which describes the RVM algorithm.

## 1 CLASSIFICATION ALGORITHM: RELEVANCE VECTOR MACHINES (RVM)

2 The main objective of this study is to classify traffic conditions between *dangerous* (collision-prone) and *safe*. For that purpose a learning algorithm to discriminate between these two traffic conditions and solve this binary classification problem needs to be trained.

5 Relevance Vector Machines (RVMs) (13) are employed in this study and are compared with the popular Support Vector Machines (SVMs) (14). Both techniques belong to the greater group of supervised learning algorithms as well as kernel methods.

8 In supervised learning, there exists a set of example input vectors  $\{x_n\}_{n=1}^N$  along with corresponding targets  $\{t_n\}_{n=1}^N$ , the latter of which corresponds to class labels. In this study, the two classes are defined as dangerous when  $t=1$  and safe when  $t=0$ . The purpose of learning is to acquire a model of how the targets rely on the inputs and use this model to classify or predict accurately future and previously unseen values of  $x$ .

13 For SVMs and RVMs these classifications or predictions are based on functions of the form:

$$14 \quad y = f(x; w) = \sum_{i=1}^N w_i K(x, x_i) + w_0 = w^T \varphi(x) \quad (1)$$

15 where  $K(x, x_i)$  is a kernel function, which defines a basis function for each data point in the training set,  $w_i$  are the weights (or adjustable parameters) for each point, and  $w_0$  is the constant parameter. The output of the function is a sum of  $M$  basis functions ( $\varphi(x) = [\varphi_1(x), \varphi_2(x), \dots, \varphi_M(x)]$ ) which is linearly weighted by the parameters  $w$ .

19 SVM, through its target function, tries to find a separating hyperplane to minimize the error of misclassification while at the same time maximise the distance between the two classes (21). The produced model is sparse and relies only on the kernel functions associated with the training data points which lie either on the margin or on the wrong side. These data points are referred to as "Support Vectors" (SVs).

24 RVMs use a similar target function as in Equation 1, but introduce a prior distribution over the model weights, which are governed by a set of hyperparameters. Every weight has a corresponding hyperparameter and the most probable values of those are estimated from the training data during each iteration. RVMs, in most cases, use fewer kernel functions compared to SVMs, without compromising the performance.

29 In the binary classification problem ( $\{t_n\}_{n=1}^N = \{0,1\}$ ), a Bernoulli distribution is adopted for the prior distribution  $p(t|x)$ . The logistic sigmoid function  $\sigma(y) = \frac{1}{1+e^{-y}}$  is applied to  $y(x)$  so as to combine random and systematic components. This leads to a generalised linear model such that:

$$32 \quad f(x; w) = \sigma(w^T \varphi(x)) = \frac{1}{1+e^{-w^T \varphi(x)}} \quad (2)$$

33 It should be noted here that there is no constant weight (e.g. noise variance). By making use of the Bernoulli distribution, the likelihood of the training data set is defined as:

$$35 \quad p(t|x, w) = \prod_{n=1}^N \sigma(w^T \varphi(x_n))^{t_n} (1 - \sigma(w^T \varphi(x_n)))^{1-t_n} \quad (3)$$

36 Using a Laplace approximation to calculate the weight parameters and for a fixed value of hyperparameters ( $\alpha$ ), the mode of the posterior distribution over  $w$  is obtained by maximizing:

$$\begin{aligned} \log(p(w|x, t, \alpha)) &= \log(p(t|x, w) p(w|\alpha)) - \log(p(t|x, \alpha)) = \\ &= \sum_{n=1}^N (t_n \log f(x_n; w) + (1 - t_n) \log(1 - f(x_n; w))) - \frac{1}{2} w^T A w + c \end{aligned} \quad (4)$$

39 where  $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$  and  $c$  is a constant.

The mode and variance of the Laplace approximation for  $w$  are:  $w_{MP} = \Sigma_{MP} \Phi^t B t$  and  $\Sigma_{MP} = (\Phi^t B \Phi + A)^{-1}$ , where  $B$  is an  $N \times N$  diagonal matrix with:

$$\beta_{nn} = f(x_n; w)(1 - f(x_n; w))$$

Using this Laplace approximation the marginal likelihood is expressed as:

$$p(t|x, \alpha) = \int p(t|x, w) p(w|\alpha) dw = p(t|x, w_{MP}) p(w_{MP}|\alpha) (2\pi)^{M/2} |\Sigma_{MP}|^{1/2} \quad (6)$$

By maximising the previous equation, with respect to each  $\alpha_i$ , the update rules are obtained as shown below:

$$\alpha_i^{new} = \frac{1 - \alpha_i \Sigma_{ii}}{m_i^2}$$

$$(\beta^{new})^{-1} = \frac{\|t - \Phi m\|^2}{N - \sum_{i=1}^N (1 - \alpha_i \Sigma_{ii})} \quad (7)$$

where  $m_i$  is the  $i$ -th element of the estimated posterior weight  $w$  and  $\Sigma_{ii}$  is the  $i$ -th diagonal element of the estimated posterior covariance matrix  $\Sigma_{MP}$ .

## DATA DESCRIPTION AND PROCESSING

The datasets utilised in this study are: a) collision data from January 2013 to December 2013, obtained from the National Road Accident Database of the United Kingdom (STATS19) and b) link-level disaggregated traffic data (from loop detectors and GPS-based probe vehicles) obtained from the UK Highways Agency Journey Time Database (JTDB). Link-level traffic data include average travel speed, volume and average journey time at 15-minute intervals.

A 56-km section of motorway M1 between junction 10 and junction 13 and two A-class roads (i.e. A3 and A12) were selected as the study area. These segments were selected randomly and are a part of the UK strategic road network (SRN). In 2013, 132 injury collisions occurred on these segments. Traffic conditions just before these collisions are termed as *cases*. For the development and testing of RVM algorithms, traffic conditions related to non-collision cases (i.e. normal driving) needs to be extracted. The number of collision and non-collision cases was derived using the following process:

15-minute aggregated traffic data (i.e. 96 unique observations per day) from 2012 – 2014 were available for the entire SRN. In order to obtain traffic conditions for each of the collision cases, traffic data associated with the two unique observations were extracted: (i) the observation that coincides with the time of the collision and (ii) the previous observation of (i). These traffic conditions would represent ‘dangerous’ situations. Similarly, in order to represent ‘safe’ traffic conditions, the JTDB measurements for the same two 15-minute intervals representing traffic conditions at one week before and after the collision, as well as two weeks before and after the collision, were extracted. For example, if a collision happened at 14:08 on the 25/06, the traffic conditions from the 15-minute interval beginning at 14:00 and 13:45 were matched to the collision case, while traffic conditions on 11/06 and 18/06 (i.e. before the collision date) and 02/07 and 09/07 (after the collision date) at 14:00 and 13:45 were matched to the non-collision case. As a result each collision case was matched with two 15-minute intervals which indicate the traffic conditions immediately before the collision, and eight 15-minute intervals which show ‘safe’ traffic conditions at the same time on a similar day.

In order to have one value for each of the traffic variables (e.g. volume) a weighted average for the two 15-minute intervals was calculated using the same aggregation technique as in (29):

$$Volume_w = \beta_1 \cdot Volume_t + \beta_2 \cdot Volume_{t-1} \quad (8)$$

Where  $\beta_1$  and  $\beta_2$  are the weighting parameters that satisfy the following conditions:

$$\beta_1 = \frac{t}{T}; \quad \beta_1 + \beta_2 = 1; \quad T = 15$$

Where  $t$  is the time difference between the reported collision time and the beginning of the corresponding 15-minute interval;  $Volume_w$  is the weighted 15-minute volume,  $Volume_t$  is the 15-minute volume at the interval when the collision has occurred,  $Volume_{t-1}$  is the 15-minute volume at the preceding interval.

By using the matched-case control structure indicated, the influence of road geometry on the collisions is assumed to be eradicated, because each collision case is matched with variables related to the entire length of the link and not to a limited area of it (e.g. neighbouring loop detectors). The data corresponding to J10-J13 of the M1 Motorway (430 collision and non-collision cases) and the AL634 link of the A3 road (119 collision and non-collision cases) were chosen to be the training datasets, while the validation dataset was the one corresponding to AL2291 on the A12 road (105 collision and non-collision cases). The explanatory variables of average speed (km/h) and total volume (vehicles per hour) were chosen and average travel time was omitted because it was not considered as an important indicator for collision-prone conditions.

The total collision and non-collision cases which were taken into account for the training and validation datasets are shown in Table 1.

**Table 1: Collision and non-collision cases for each of the training and test datasets**

Road	Dataset	Non-collision Cases	Collision Cases	Total
M1 (Junctions 10-13)	Training_1	344	86	430
A3 (Link AL634)	Training_2	95	24	119
A12 (Link AL2291)	Validation	83	22	105
Total		521	132	

## RESULTS AND DISCUSSION

The above described classification methods of RVMs and SVMs have been applied to the datasets in order to solve the binary classification problem of distinguishing between safe and collision-prone conditions.

As mentioned before, both SVMs and RVMs rely on kernel functions to perform regression or classification. The most popular kernels used are the linear, polynomial and Gaussian or radial basis function (RBF) kernels. In this study the Gaussian kernels have been used, because the literature suggests that they provide more powerful results (21).

The Gaussian kernel is calculated through the equation  $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ , where  $\gamma$  determines the width of the basis function. The coefficient  $\gamma$  was set to the value 0.5 because the targets of the classification lie in the interval  $\{0, 1\}$ .

In order to test the performance of RVMs for classifying traffic conditions data, two MATLAB implementation algorithms were employed, namely SparseBayes v1 and v2 (30, 31). Although both algorithms perform the same task, the difference lies in the fact that v1 has a built-in function to develop RVMs, while the second version is more 'general-purpose' and requires that the user defines the basis functions to be used. Furthermore, the hyperparameters  $\alpha_i$  are updated in v1 at each iteration using the formula  $\alpha_i = \frac{\gamma_i}{\mu_i^2}$ , where  $\mu_i$  is the i-th posterior mean weight and  $\gamma_i \equiv 1 - \alpha_i \Sigma_{ii}$  with  $\Sigma_{ii}$  being the i-th diagonal element of the posterior weight covariance used. This update

technique, although simplistic, is not the most optimal (30). On the contrary in v2, the marginal likelihood function with regards to the hyperparameters is efficiently optimised continuously and individual basis functions can be discretely added or deleted as described in (32). In that way, algorithm v2 converges faster but can prove greedy regarding the classification results.

For the RVM models the maximum iterations were set to 100000 with monitoring every 10 iterations, the Gaussian kernel width was set to 0.5 and the initial  $\beta$  value was set to zero. The first version of the RVM algorithm was initialised with  $\alpha = \frac{1}{N^2}$ , where N is the size of the dataset. The algorithm terminates if the largest change in the logarithm of any hyperparameter  $\alpha$  is less than  $10^{-3}$ . On the other hand, the second version of RVM initialises with an  $\alpha$  value which is automatically calibrated according to the size of the dataset used. The v2 algorithm terminates if the change in the logarithm of any hyperparameter  $\alpha$  is less than  $10^{-3}$  and the change in the logarithm of  $\beta$  parameter is less than  $10^{-6}$ . SVMs were developed using the Statistics and Machine Learning Toolbox™ of MATLAB, with kernel width for the Gaussian kernel set to 0.5 and Box constraint level set to 1. The linear kernel for SVMs ( $K(x_i, x_j) = x_i \cdot x_j$ ), is also tested for comparison reasons.

In order to test the performance of the three different algorithms (i.e. RVM\_v1, RVM\_v2 and SVMs), the execution time (using a laptop with i7 processor & 8 GB RAM), the classification error and the decision vectors during the training of the model, as well as when tested on the validation dataset of traffic conditions for link AL2291, on the A12 road was tested. The training datasets consist of the 430 traffic conditions of the M1 (J10-J13) and the 119 traffic conditions regarding link AL634 on the A3 road. These training and validation datasets are relatively small. However, collision occurrence is a rare event and it is not unusual for traffic safety experts to deal with small samples (21). Furthermore, other works on RVM classification (such as (26)) have tested training datasets of a same size.

Results for the training datasets are summarised in Table 2, while results for the validation dataset are summarised in Table 3.

**Table 2: Classification Accuracy during Training and Number of Decision Vectors for RVMs and SVMs**

Method	Kernel	Training Sample Size	Training Time	Training Error	Decision Vectors
RVM_v1	Gaussian	430	7.2 minutes	4.88%	358
RVM_v2	Gaussian		9.3 seconds	19.53%	93
SVM	Gaussian		4 seconds	15.80%	204
	Linear		4 seconds	20.00%	203
RVM_v1	Gaussian	119	1.5 minutes	0.84%	116
RVM_v2	Gaussian		1.08 seconds	18.49%	6
SVM	Gaussian		2 seconds	11.80%	64
	Linear		2 seconds	15.1%	42

As can be seen from Table 2, training for RVMs is slower than SVMs, something which agrees with the results of other RVM classification works (13, 26). The delay in training for RVMs is triggered by the iterated need for calculating and inverting the Hessian matrix and which leads to more computational time, as sample size increases. The best classification is performed by the RVM\_v1 algorithm, with a large margin (of about 10%) to the next more successful algorithm which is the SVM with a Gaussian kernel. However, it is noticeable that this successful rate of classification by the RVM\_v1 algorithm is due to the large number of decision vectors, which is about 1.5 to 2 times higher than the decision vectors used by RVM\_v2 and SVM. The efficient RVM\_v2 is about 4% less accurate than SVMs but the interesting fact is that it uses less than half of the decision vectors utilised by SVMs to perform the training classification and in a non-critical time interval which can be utilised in real-time (8 seconds). In the smaller sample size, it can also be seen that RVM\_v2 uses



only 6 vectors to perform the classification, while the other two approaches require a much larger number. Comparing training classification results between the small and the bigger sample size, it can be seen that all three algorithms perform better on the small sample, which also agrees with the literature (26, 27). It should be noted here that training time for SVMs is notably faster because of the fact that SVMs are quite a popular classification algorithm and the corresponding toolboxes or other software packages have undertaken a lot of attention in order to accelerate the algorithm's outputs.

**Table 3: Validation results of the algorithms using an independent sample**

Datasets with sample size	Method	Kernel	Classification error (%)
Training dataset: 430 cases from a motorway; Validation dataset: 105 observations from A-class roads	RVM_v1	Gaussian	25.71
	RVM_v2	Gaussian	20
	SVM	Gaussian	18.09
		Linear	20
Training dataset: 119 cases from A-class road; Validation dataset: 105 observations from A-class roads	RVM_v1	Gaussian	21.9
	RVM_v2	Gaussian	20
	SVM	Gaussian	10.47
		Linear	20

Looking at the validation results of the classification algorithm that was trained with the larger sample (Table 3), it is shown that RVM\_v1 is no longer the most accurate classifier. SVMs with Gaussian kernel produce the most successful result, followed by RVM\_v2 and SVM with linear kernel. RVM\_v1 probably leads to worse results due to the fact that it requires a lot of decision vectors and these classifier vectors cannot perform well when applied to an unknown independent dataset. When the classifier was trained with a small sample and the algorithm was applied to a relatively small sample, the results show that SVMs with Gaussian kernel outperform RVMs with a classification error which is half of the classification error of both RVMs algorithms.

To further investigate the classification performance of the three algorithms<sup>2</sup>, the measures of sensitivity and specificity were employed. For that purpose, four commonly employed terms (as defined below) are employed.

- True Positive (TP): Dangerous (collision-prone) conditions ( $t_{\text{real}}=1$ ) correctly identified as dangerous ( $t_{\text{classified}}=1$ )
- False Positive (FP): Dangerous (collision-prone) conditions ( $t_{\text{real}}=1$ ) incorrectly identified as safe ( $t_{\text{classified}}=0$ )
- True Negative (TN): Safe traffic conditions ( $t_{\text{real}}=0$ ) correctly identified as safe ( $t_{\text{classified}}=0$ )
- False Negative (FN): Safe traffic conditions ( $t_{\text{real}}=0$ ) incorrectly identified as dangerous ( $t_{\text{classified}}=1$ )

By making use of the known formula for sensitivity and specificity (33), the performance of these algorithms for the larger training dataset and the validation dataset are presented in Table 4:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \text{ and } \text{Specificity} = \frac{TN}{TN+FP} \quad (9)$$

<sup>2</sup> SVM with linear kernel was excluded in Table 4 because it did not provide better results than the other algorithms

**Table 4: Sensitivity and Specificity of RVMs and SVMs (N= 430+105=535)**

Method	Kernel	TN	FN	TP	FP	Sensitivity (%)	Specificity (%)
RVM_v1	Gaussian	419	9	68	39	88.4	91.5
RVM_v2	Gaussian	427	1	2	105	66.67	80.26
SVM	Gaussian	417	11	31	76	73.8	84.6

It is noticeable from Table 4 that the RVM\_v1 algorithm performs well in identifying the traffic conditions that lead to a collision. RVM\_v1 is the best classifier with 88.4% sensitivity and 91.5% specificity implying minimum Type I and Type II errors. The RVM\_v2 algorithm underperforms in terms of sensitivity and specificity among the two datasets. This is probably a result of the greediness of the algorithm which converges fast but at the expense of a large number of false positives.

The reason for these misclassification rates associated with all the algorithms, especially with RVM may relate to the use of highly aggregated (i.e. 15-minute) traffic data, relatively small sample size and the use of only two variables for representing traffic. In addition, the algorithm was primarily trained with traffic data from a motorway (M1 J10 – J13) but validated with traffic data from A-class roads. Traffic dynamics between these two classes of roads are quite different.

## CONCLUSIONS

Machine learning approaches, especially support vector machines, have become a credible solution for real-time collision prediction and traffic pattern recognition over the recent decades. Their strengths lie in the use of a kernel function to control non-linearity and the efficient handling of arbitrarily structured data. The performance of a SVM however largely relies on the kernel function and the number of decision vectors that increases proportionately with the size of the training dataset.

The novelty of this work relates to the application of Relevance Vector Machines, a Bayesian analogue to SVMs, for the task of classifying traffic conditions. RVMs have proven to be an efficient classifier when applied to other pattern recognition tasks, but had not been applied in transport studies. Classification was performed using highly aggregated 15-minute traffic measurements of average speed and volume while a matched case-control structure was adopted to remove spatial influence on the collision occurrence. Two different RVM classifiers were trained with a small (i.e. 119 collision and non-collision cases on an A-Road) and a relatively large training dataset (i.e. 430 collision and non-collision cases on a motorway) and were validated with an independent dataset of 105 collision and non-collision cases from a separate A-class road. In order to increase the understanding on the classification differences between RVMs and SVMs, specificity and sensitivity measures were calculated for each of the classification algorithms.

The advantage of RVM classification relates to the fact that the decision vectors it used to classify new data are significantly fewer than the decision vectors used by SVMs. This may lead to faster classification decisions which are crucial in real-time applications. Furthermore, the assignment of a posterior probability to each classified instance makes RVMs more useful to experts and more substantial than the non-probabilistic results of SVMs. However, attention should be given to the training of the algorithm, so as to overcome misclassification errors and assure robustness with new data. Classifying traffic conditions in order for the classification to be used in real-time collision prediction is a promising direction for improving the efficiency and computational speed of current collision prediction algorithms. RVMs with the small number of decision vectors they need in order to classify conditions can prove helpful in this direction. On the other hand, improvements need to be done in order for RVM classification to become more efficient. First of all, the incorporation of disaggregated data should give a much clearer picture on the strengths of RVMs for real-time traffic conditions classification. Furthermore, the inclusion of other traffic or network variables (e.g. traffic

density, vehicle compositions, variable speed limits) should be used to describe the collision-prone conditions in a more realistic way. Lastly, training could take place with a large amount of data, so as to more accurately describe traffic conditions and circumstantial precursors leading to traffic accidents.

## ACKNOWLEDGEMENT

This research was funded by a grant from the UK Engineering and Physical Sciences Research Council (EPSRC) (Grant reference: EP/J011525/1). The authors take full responsibility for the content of the paper and any errors or omissions. Research data for this paper are available on request from Mohammed Quddus ([m.a.quddus@lboro.ac.uk](mailto:m.a.quddus@lboro.ac.uk)).

## REFERENCES

1. Hummel, T., M. Kühn, J. Bende, and A. Lang. *Advanced Driver Assistance Systems: An investigation of their potential safety benefits based on an analysis of insurance claims in Germany*. Research report FS 03, German Insurance Association, 2011.
2. Thrun, S. Toward Robotic Cars. *Communications of the ACM*, Vol. 53, No. 4 (April), 2010, pp. 99–106.
3. Highways Agency, and Robert Goodwill. New generation of motorway opens on M25 - UK Government Press release. <https://www.gov.uk/government/news/new-generation-of-motorway-opens-on-m25>, Accessed July 29<sup>th</sup> 2015
4. Lee, C., B. Hellinga, and F. Saccomanno. Real-time crash prediction model for application to crash prevention in freeway traffic. In Transportation Research Board 82nd Annual Meeting, Washington DC, United States, 2003..
5. Hossain, M., and Y. Muromachi. A real-time crash prediction model for the ramp vicinities of urban expressways. *IATSS Research*, Vol. 37, No. 1, Jul. 2013, pp. 68–79.
6. Abdel-Aty, M., and A. Pande. The viability of real-time prediction and prevention of traffic accidents. *Efficient Transportation and Pavement*, Taylor & Francis Group, pp 215-227, 2009..
7. Abdel-Aty, M., A. Pande, L. Y. Hsia, and F. Abdalla. The Potential for Real-Time Traffic Crash Prediction. *ITE Journal on the web*, 2005, pp. 69–75.
8. Ahmed, M. M., and M. a. Abdel-Aty. The Viability of Using Automatic Vehicle Identification Data for Real-Time Crash Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 2, Jun. 2012, pp. 459–468.
9. Hassan, H. M., and M. Abdel-Aty. Predicting reduced visibility related crashes on freeways using real-time traffic flow data. *Journal of safety research*, Vol. 45, Jun. 2013, pp. 29–36.
10. Pande, A., and M. Abdel-Aty. Assessment of freeway traffic parameters leading to lane-change related collisions. *Accident; analysis and prevention*, Vol. 38, No. 5, Sep. 2006, pp. 936–48.
11. Hossain, M., and Y. Muromachi. A real-time crash prediction model for the ramp vicinities of urban expressways. *IATSS Research*, Vol. 37, No. 1, Jul. 2013, pp. 68–79.
12. Yu, R., and M. Abdel-Aty. Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes. *Accident; analysis and prevention*, Vol. 58, Sep. 2013, pp. 97–105.
13. Tipping, M. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, Vol. 1, 2001, pp. 211–244.
14. Vapnik, V. *The nature of statistical learning theory*. Springer-Verlag, New York, 1995.
15. Yuan, F., and R. L. Cheu. Incident detection using support vector machines. *Transportation Research Part C: Emerging Technologies*, Vol. 11, No. 3-4, 2003, pp. 309–328.
16. Chen, S., W. Wang, and H. van Zuylen. Construct support vector machine ensemble to detect traffic incident. *Expert Systems with Applications*, Vol. 36, No. 8, 2009, pp. 10976–10986.
17. Xiao, J., and Y. Liu. Traffic Incident Detection Using Multiple Kernel Support Vector Machine. in Proceedings of the 2012 15th International IEEE Conference on Intelligent Transportation Systems, 2012, pp. 1669-1673.

18. Ma, Y., M. Chowdhury, A. Sadek, and M. Jaihani. Real-time highway traffic condition assessment framework using vehicleInfrastructure integration (VII) with artificial intelligence (AI). *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, No. 4, 2009, pp. 615–627.
19. Yu, R., G. Wang, J. Zheng, and H. Wang. Urban Road Traffic Condition Pattern Recognition Based on Support Vector Machine. *Journal of Transportation Systems Engineering and Information Technology*, Vol. 13, No. 1, 2013, pp. 130–136.
20. Li, X., D. Lord, Y. Zhang, and Y. Xie. Predicting motor vehicle crashes using Support Vector Machine models. *Accident Analysis and Prevention*, Vol. 40, No. 4, 2008, pp. 1611–1618.
21. Yu, R., and M. Abdel-Aty. Utilizing support vector machine in real-time crash risk evaluation. *Accident; analysis and prevention*, Vol. 51, Mar. 2013, pp. 252–259.
22. Wang, W., X. Qu, W. Wang, and P. Liu. Real-time freeway sideswipe crash prediction by support vector machine. *IET Intelligent Transport Systems*, Vol. 7, No. 4, 2013, pp. 445–453.
23. Dong, N., H. Huang, and L. Zheng. Support vector machine in crash prediction at the level of traffic analysis zones : Assessing the spatial proximity effects. Vol. 82, 2015, pp. 192–198.
24. Chen, S., S. R. Gunn, and C. J. Harris. The relevance vector machine technique for channel equalization application. *IEEE Transactions on Neural Networks*, Vol. 12, No. 6, 2001, pp. 1529–1532.
25. Carin, L., and G. J. Dobeck. Relevance vector machine feature selection and classification for underwater targets. *Oceans 2003. Celebrating the Past ... Teaming Toward the Future (IEEE Cat. No.03CH37492)*, Vol. 2, 2003, p. 933-957.
26. Demir, B., and S. Ertürk. Hyperspectral Image Classification Using Relevance Vector Machines. *2007 IEEE International Geoscience and Remote Sensing Symposium*, Vol. 4, No. 4, 2007, pp. 586–590.
27. Phillips, C. L., M. A. Bruno, P. Maquet, M. Boly, Q. Noirhomme, C. Schnakers, A. Vanhaudenhuyse, M. Bonjean, R. Hustinx, G. Moonen, A. Luxen, and S. Laureys. “Relevance vector machine” consciousness classifier applied to cerebral metabolism of vegetative and locked-in patients. *NeuroImage*, Vol. 56, No. 2, 2011, pp. 797–808.
28. Wei, L., Y. Yang, R. M. Nishikawa, M. N. Wernick, and A. Edwards. Relevance vector machine for automatic detection of clustered microcalcifications. *IEEE Transactions on Medical Imaging*, Vol. 24, No. 10, 2005, pp. 1278–1285.
29. Imprialou, M., M. Quddus, and D. Pitfield. Exploring the Role of Speed in Highway Crashes : Pre-Crash-Condition-Based Multivariate Bayesian Modelling. In *Transportation Research Board 94th Annual Meeting*, Washington DC, United States, 2015.
30. Michael E. Tipping. A Baseline Matlab Implementation of “ Sparse Bayesian ” Model Estimation. *Baseline*, 2009.
31. Michael E. Tipping. An Efficient Matlab Implementation of the Sparse Bayesian Modelling Algorithm (Version 2.0). 2009.
32. Tipping, M. E., and A. C. Faul. Fast Marginal Likelihood Maximisation for Sparse Bayesian Models. *Ninth International Workshop on Artificial Intelligence and Statistics*, No. x, 2003, pp. 1–13.
33. Powers, D. Evaluation : From Precision , Recall And F-Measure To ROC , Informedness , Markedness & Correlation. *Journal of Machine Learning Technologies*, Vol. 2, No. 1, 2011, pp. 37–63.