The Dissertation Committee for Moggan Motamed
Certifies that this is the approved version of the following dissertation:

Developing A Real-time Freeway Incident Detection Model using Machine Learning Techniques

Committee:

Randy Machemehl, Supervisor

Michael Walton

Zhanmin Zhang

Stephen Boyles

Carlos Caldas

Wesley Barnes

Developing A Real-time Freeway Incident Detection Model using Machine Learning Techniques

by

Moggan Motamed, B.S.C.E, M.S., M.S.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2016

Acknowledgements

First, I would like to thank my supervisor, Dr. Machemehl for his leadership, support, enthusiasm, attention to detail, hard work, patience, and knowledgeable advice during my graduate studies. He and other faculty and staff at the University of Texas at Austin have been great mentors in my life. I am grateful for Lisa Smith' help and inspiration over the past few years.

I also want to thank my committee: Dr. Michael Walton, Dr. Zhanmin Zhang, Dr. Stephen Boyles, Dr. Carlos Caldas, and Dr. Wesley Barnes for their advise on dissertation writing and career development. I owe thanks to friends and colleagues have helped me through graduate school and my research: Ayush Talwar, Yiyi Wang, Dan Fragnant, Shadi Hakimi, Marisol Castro, Migdalia Carrion, Aidin Massahi, and Carolina Baumanis.

I would like to acknowledge the financial and technical support of the Federal Highway Administration and the State of Texas to the Southwest University Transportation Center for project #161326 and #600451-0083. This research was also supported by the Austin District of the Texas Department of Transportation for project #14-2XXIA002 and #14-3XXIA003. I am especially grateful for their staff researcher, Joseph Hunt, who provided the Dallas database for the study. Many thanks are also given to Prof. Mohammed Hadi and Research Assistant Tao Wang (Florida International University, USA) for providing the Miami dataset. This study would not have been successful without their assistance.

Also, thanks to my family and friends for their support throughout this process. Special thanks to my brother and lifelong mentor, Arash Motamed.

Developing A Real-time Freeway Incident Detection Model using Machine Learning Techniques

Moggan Motamed, Ph.D.
The University of Texas at Austin, 2015

Supervisor: Randy Machemehl

Real-time incident detection on freeways plays an important part in any modern traffic management operation by maximizing road system performance. The US Department of Transportation (US-DOT) estimates that over half of all congestion events are caused by highway incidents rather than by rush-hour traffic in big cities. An effective incident detection and management operation cannot prevent incidents, however, it can diminish the impacts of non-recurring congestion problems. The main purpose of real-time incident detection is to reduce delay and the number of secondary accidents, and to improve safety and travel information during unusual traffic conditions. The majority of automatic incident detection algorithms are focused on identifying traffic incident patterns but do not adequately investigate possible similarities in patterns observed under incident-free conditions. When traffic demand exceeds road capacity, density exceeds critical values and traffic speed decreases, the traffic flow process enters a highly unstable regime, often referred to as "stop-and-go" conditions. The most challenging part of real-time incident detection is the recognition of traffic pattern changes when incidents happen during stop-and-go conditions.

Recently, short-term freeway congestion detection algorithms have been proposed as solutions to real-time incident detection, using procedures known as dynamic time warping (DTW) and the support vector machine (SVM). Some studies have shown these procedures to produce higher detection rates than Artificial Intelligence (AI) algorithms with lower false alarm rates. These proposed methods combine data mining and time

series classification techniques. Such methods comprise interdisciplinary efforts, with the confluence of a set of disciplines, including statistics, machine learning, Artificial Intelligence, and information science. A literature review of the methodology and application of these two models will be presented in the following chapters. SVM, Naïve Bayes (NB), and Random Forest classifier models incorporating temporal data and an ensemble technique, when compared with the original SVM model, achieve improved detection rates by optimizing the parameter thresholds. The main purpose of this dissertation is to examine the most robust algorithms (DTW, SVM, Naïve Bayes, Decision Tree, SVM Ensemble) and to develop a generalized automatic incident detection algorithm characterized by high detection rates and low false alarm rates during peak hours. In this dissertation, the transferability of the developed incident detection model was tested using the Dallas and Miami field datasets.

Even though the primary service of urban traffic control centers includes detecting incidents and facilitating incident clearance, estimating freeway incident durations remains a significant incident management challenge for traffic operations centers. As a next step this study examines the effect of V/C (volume/capacity) ratio, level of service (LOS), weather condition, detection mode, number of involved lanes, and incident type on the time duration of traffic incidents. Results of this effort can benefit traffic control centers improving the accuracy of estimated incident duration, thereby improving the authenticity of traveler guidance information.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

The US Department of Transportation (US-DOT) estimates that over half of all congestion events are caused by highway incidents rather than by rush-hour traffic in big cities [1]. Incidents cause traffic congestion and lead to the loss of human and economic capital. This gives high priority to Traffic Incident Management (TIM) strategies in FHWA planning systems. Unlike traffic demand management (TDM) policies which try to change travelers long term behavior [79,80], Incident Management mainly focuses on optimizing the real-time traffic operation by reducing duration and impacts of incidents. Under medium to heavy traffic conditions, the promptness of response after an incident is a direct function of the detection time. Accurate and fast incident detection is essential for subsequent management plans that aim to reduce incident based congestion.

Background

Incident detection processes are closely tied to the sensing systems that provide real time traffic data. With respect to sensing systems, incident detection models can be grouped as spatial measurement-based models (video image processing) and point/link-based models (automatic incident detection (AID)) [12]. Some video processing techniques have been adapted to incident detection, but the accuracy of the models is sensitive to environmental factors, such as shadows, snow, rain, fog. [7] Point-based data collection, such as inductive loop detectors (ILD) and microwave radar, are common types of existing sensor technologies whereas link-based data collection systems use individual vehicles as probes. AID algorithms can be categorized as either macroscopic or microscopic. Most AID algorithms are macroscopic, and are designed to use point traffic quantities (speed and occupancy). Common AID algorithms include: [1] comparative or pattern recognition based (e.g. California #2 algorithm) [2], Bayesian algorithm [3], catastrophe theory and patterns based (e.g. McMaster algorithm) [4], time-series models based [5], and AI based. The most widely deployed AID algorithms are traditional “pattern recognition” based methods that compare the detector outputs of different parameters to a threshold value. Tuning the thresholds to properly detect incidents requires substantial effort and expert traffic engineering judgment [66].

To improve incident detection data, it is reasonable to expect that using multiple data sources, e.g., fixed detectors (collecting point data) and probe vehicles (collecting spatial data), could enhance the input data reliability and completeness, hence improving the performance of an incident detection system. Building a microscopic model to mimic driving behavior is extremely difficult, since drivers consider the current movement of the traffic stream ahead, not just one vehicle in front of them. To make microscopic modeling possible, vehicle trajectory data, that are not generally available, are required.

The trajectory data generated by vehicles in a vehicle-infrastructure integration (VII) [12] network have shown the potential to provide faster traffic condition detection and lower false alarm rates. Existing infrastructure-based incident detection systems that typically use inductive loop detectors, magnetometers and magnetic detectors, microwave radar, infrared, ultrasonic, acoustics and video image processing systems do not always detect traffic conditions quickly or correctly. Since 2003, the Federal Highway Administration (FHWA) has sponsored a variety of efforts to reduce the number of incidents that have led to the national development of the VII Architecture and Functional Requirements, which improves communication between vehicles and surface transportation infrastructure [13]. Two large states, California [14] and Michigan (MDOT, 2005) are also testing various methods for implementing these programs and have shown excellent results [15]. VII can improve incident detection models by communicating data faster than before, but currently it is not widely available. An important consideration for any incident detection process is the feasibility of its application in typical existing traffic control centers.

Once the sensing system is in place, the choice among incident detection algorithms is the next appropriate step which depends upon the type and reliability of the sensed data. Most incident detection algorithms are simple in theory and practical in operation. However, during peak hour stop-and-go conditions they often fail to deliver both high detection rates and low false alarm rates. Improving incident detection data using multiple data sources, e.g., fixed detectors (collecting point data) and probe vehicles (collecting spatial data), could enhance the input data reliability and completeness, hence improving the performance of any incident detection system.

Most traditional automated incident detection algorithms use ILD-based spot or point speed data because they are most readily available [8,9,10,11]. These types of sensing systems produce reliable spot speeds and reliable flow values. However, during highly variable stop-and-go traffic conditions, only use of spot speeds to estimate reliable space mean speeds is not sufficient. During incident occurrence, occupancy increases upstream and decreases downstream while speed and volume decrease upstream. These differences between up and downstream traffic volume measurements have been the basis of pattern recognition based freeway automated incident detection (AID) algorithms such as the widely-used California and the Minnesota [6] methods. The California algorithm only utilizes current time occupancy information of upstream, which may produce high false alarm rates (FAR) because of dynamic traffic fluctuations. To decrease the high FAR, the Minnesota algorithm employs a cumulative sum of differences between up and downstream conditions. Both of these models have not yet been able to successfully detect incidents during peak hour "stop and go" traffic.

Several algorithms have been described as having high detection rates but they had either high false alarm rates or long detection rates (ex. ARIMA) [43]. Earlier time-series algorithms such as ARIMA lost their favor because they tend to underestimate the variation of the traffic measurements, which causes high false alarm rates [43]. Later, Hi-ri-o-toppa (2012) used an advanced time-series classifier dynamic time warping (DTW) technique, for the first time, in incident detection. This study showed a 90% detection rate with a 5 minute and 40 second mean time to detection (MTTD). However, DTW has a unacceptably high likelihood of overfitting because of the small rate of incidents in their sample size (16 incidents over the course of six months).

The classical incident detection methods (comparative and statistical algorithms) apply current understanding of statistical methods to deal with data by following a series of rules that do not consider the non-linear nature of the data. These classical algorithms are simple in theory and easy to implement, but usually fail to deliver high detection rates and low rates of false alarm. On the other hand, the more powerful artificial intelligence (AI) algorithms can solve more complex problems through trial and error to improve detection performance over time.

The most commonly applied AI algorithms, Artificial Neural Network (ANN) models, have not been as successful as the newer Support Vector Machine AI counterparts. Support Vector Machine (SVM) models have greater learning and prediction potentials in comparison to ANN because they can provide faster results and more customization options for the modeler. In addition, SVM requires less computational cost, which is vital for real time incident detection and avoids the over-fitting problems of ANN [48]. Recently, Liu et al. (2014) conducted an incident detection study with a multiple Naïve Bayes classifier model. The Naïve Bayes classifier ensemble removes the burden of choosing the optimal threshold for the observed parameters. Their experiment on a Singapore freeway using the I-880 California data as the training data showed significantly better and more stable results than the standard Naïve Bayes (NB) experimental results. These advanced algorithms are more successful because they consider incident detection as a pattern recognition problem of incident versus non-incident conditions [16,17,18].

MOTIVATION

The Institute of Physics (2005) research shows that even tiny fluctuations in vehicle-road density can cause a chain reaction leading to a traffic jam. It is practically impossible to obtain coherent predictions from a macroscopic traffic flow model due to high occurrence of many small transient shockwaves under these conditions that can mistakenly be detected as an incident occurrence. Therefore, traditional incident detection efforts are not applicable under such conditions. This research focuses on developing a transferable incident detection algorithm that will not only detect incidents in a quick and effective manner, but will do so during peak hours with stop-and-go traffic flow. To meet this objective, understanding of stop-and-go traffic mechanisms is essential.

Recently, incident detection algorithms have been proposed using pattern classification models and advanced time series dynamic time warping (DTW) to mitigate this challenging issue of incident detection during all conditions, including stop-and-go traffic. It has been widely used in other disciplines (speech detection), and was recently introduced into transportation engineering. However, DTW requires a larger data set that could be very time consuming to develop for proper training. One of the most robust pattern classification models is known as the support vector machine (SVM), the foundations of which were developed in 1995 by Cortes and Vapnik [19,20,21]. Some studies suggest SVM has a higher detection rate than Artificial Intelligence (AI) algorithms with a lower false alarm rate. The proposed method has an interdisciplinary perspective, combining a set of disciplines, including statistics, machine learning, AI, and information science. In complicated real-world cases, such as traffic incident detection, data is not linearly separable. SVM has the power of solving non-linear classification problems by using the *kernel method* to transform the original input space into a higher dimensional feature space and constructing an optimal linear separating hyper-plane. Naïve Bayes and Random Forests are also gaining attention in the field of Transportation.

CONTRIBUTION

This study uses field traffic pattern data to provide a solution for incident detection during peak hours. The test data was first collected by the Dallas Texas Traffic Control Center and includes upstream and downstream speed, volume, and occupancy from two freeways: US-75 and I-635. The data from Miami was obtained to test the transferability of the model.

Frequently, field data are subject to bias and noise from different sources and the incident detection model must take these factors into consideration. The majority of previous real-time incident detection models often work well in a laboratory simulation setting, but do not perform well during actual deployment using real incident data [8]. The issue with the majority of these models is that they have not been field-tested. Another alternative is to use the 1993 I-880 California dataset, which may possibly not reflect current driving behavior. Few recent studies focused on developing their incident detection model use recent field data [71,73,74]. This study takes heed of the design of the evaluation procedure to avoid bias and old training data. However, previous research indicates that DTW and SVM methods can provide an automated incident detection model.

This study develops a generalized incident detection model that can be used in any traffic control center. Initially, as discussed in the next chapter, two groups of experiments were performed to evaluate the most promising incident detection algorithms: DTW and SVM. The data used in this study was collected by the Dallas Traffic Control center and includes speed, volume, and occupancy. A small sample of five selected incident locations was prepared for this section. After our experiment, the SVM model was found to be most robust and a better fit for the type of data available. Subsequently, additional cases were extracted from the dataset and SVM was chosen for the more extensive generalized incident detection model experimentation. A variety of different scenarios were defined to examine the SVM model such as utilizing different training datasets and kernel functions. In this dissertation, we examined the transferability of the developed SVM model using field data from different freeways in the Dallas, Texas area. As a next step, this study was expanded by gathering more incident cases from I-95 Miami to validate the utility of a generalized SVM model and provide guidelines for the future work in this field. Incident management is closely tied to incident detection, however, post-incident activities were considered outside the scope of this study.

Incidents are not only influenced by special factors (i.e., upstream/downstream speed), but also are affected by time. To include other effects into the model, this study incorporated temporal data and a multi-kernel SVM model. SVM, Naïve Bayes (NB), and Random Forest classifier models incorporating temporal data and the ensemble technique were compared with the original SVM model to improve the detection rate by optimizing the parameter thresholds. A literature review, methodology, and application of these models will be presented in the following sections.

A significant incident management challenge for traffic operations centers is estimating freeway incident durations. This study examined the effect of traffic related factors (V/C (volume/capacity) ratio, level of service (LOS)), weather condition, detection mode, number of involved lanes, and incident type on the time duration of traffic incidents. Results of this effort can benefit traffic control centers in improving the authenticity of traveler guidance information.

Chapter 2: The Support Vector Machine Concept

INTRODUCTION

The first incident detection algorithms were developed in the early 1970s and the development process continues today. Incident detection algorithms are typically categorized into five major groups depending on the type of operations data used. These include:

- Comparative algorithms;
- Statistical algorithms;
- Time-series and filtering based algorithms;
- Traffic theory based algorithms; and
- Advanced algorithms.

Comparative algorithms compare the value of measured traffic parameters (i.e., volume, occupancy or speed) to a pre-defined threshold value. When the measured traffic parameter exceeds the threshold an incident alarm is indicated. The California algorithm is the most well-known and frequently used model of this group in traffic control centers [22], but needs improved peak-hour detection.

The statistical algorithms apply regular statistical methods to determine whether observed detector data deviate in a statistically significant manner from estimated or predicted traffic characteristics [23]. This is another classic approach, and usually has better results that the comparative algorithm in terms of false alarm rates and incident detection.

Time series algorithms predict normal traffic conditions and detect incidents when detector measurements differ significantly from model outputs. This is under the assumption that traffic normally follows a predictable pattern over time, however driver choices and reactions to the driving environment evolve and change [24,25]. Including all of these factors is not any easy task and requires that the "normal" pattern be updated periodically. This approach is newer and not as frequently used most likely because of data limitations.

Traffic-theory based modeling approaches apply traffic flow theory to describe and predict traffic behavior under incident conditions. This type of model uses trajectory data and is based on the comparison between observed traffic parameters and parameter values estimated by the models [26]. These types of models would have a lot of potential if vehicle trajectory data were widely available.

Advanced algorithms have been developed with sophisticated mathematical techniques that incorporate uncertainty into complex decision-making and data-analysis processes. Machine learning or artificial intelligence (AI) techniques are typically included in this category [27].

Most incident detection algorithms tend to perform well until peak hour stop-and-go conditions occur. When stop-and-go conditions prevail, such algorithms most often fail to deliver high detection rates and low rates of false alarm. Based on their learning capabilities, AI algorithms allow the model to improve detection performance over time by adapting to changes in traffic conditions. Most recently, support vector machine (SVM) models have been successfully used to improve incident detection algorithm reliability. SVM is a type of supervised machine learning model that has been successfully applied for real-time incident detection because of its capability to produce a computationally efficient nonlinear classifier with maximum generality. The SVM eliminates bias by including unrelated parameters to the predicted variable in the modeling process. Such parameters may be weakly informative individually, but produce strong predictors when used in concert with other parameters [66]. Previous evaluation studies have focused on advancing incident detection models by either using traffic simulation software or a 1993 data set collected from I-880, but those studies generally did not fully validate the models by testing them on different freeways. The purpose of this study is to fully validate an SVM incident detection model on different freeways. This study uses field traffic pattern data to overcome the problem of incident detection during peak hours.

BACKGROUND OF SVM IN TRANSPORTATION

SVM has had limited applications in the transportation field. Previous studies describe use of SVM for travel time, speed and flow predictions, and limited incident detection applications [20]. Yuon and Cheu first applied SVM in traffic engineering to detect incidents on freeways [28]. They used the I-880 freeway data set and simulated incident data to test their model performance. They obtained better performance (a higher detection rate and a lower false alarm rate) than with a traditional neural network algorithm [28]. Chowdhury et al. and Bhavsar et al. also applied SVM to real-time traffic management and found that SVM was suitable for hierarchical intelligence applications and generally required low memory and processing requirements [29,30].

Chen et al. (2009) indicated that SVM was a robust method, but was highly sensitive to the kernel function choice and parameter specifications. They presented an SVM ensemble algorithm consisting of several independently trained SVMs and a voting process to deal with sensitivities to the kernel function for incident detection. They found that the SVM ensemble improved the comprehensive performance of SVM in most cases [31,32]. The drawbacks of the SVM ensemble were possible negative effects of the ensemble approach, including an increased computational burden and the possibility of drawing individual training samples that could bias the ensemble voting process [33,34]. One of the solutions to address this problem is the multiple kernel support vector machine (MKL-SVM) proposed by Xiao et al. (2012) to detect incidents. MKL-SVM produced robust performance that avoided the burden of choosing the kernel function and its parameters [33]. To further improve the performance of MKL-SVM, Xiao et al. (2012) proposed an MKL-SVM ensemble [35]. Lessons learned from the aforementioned research were applied to the real-time incident detection model presented in the following sections to address the problem of high false alarm rates during the peak hours.

Using field data to improve model comprehensiveness has been neglected in the incident detection process development. Most previous incident detection application studies have relied on simulation data or the I-880 database. Ma et al. (2010) used SVM with vehicle-infrastructure integration (VII) to predict travel time and incident detection using incidents generated by simulation software. They found that only 15% of vehicles needed to be VII enabled to have a 100% incident detection rate (*36*). Qu et al. (2013) used standard SVM to predict sideswipe crashes using recent sideswipe crash data from Milwaukee. The results from this study show that SVM achieved better crash identification at lower false alarm rates than other commonly applied neural network models. The developed model by Qu et al. has limited application for a specific type of incident (sideswipe) [37]. Motamed and Machemehl (2014) developed a real-time incident detection model using SVM with actual incident data from Dallas, however, this model still requires testing on other freeways [38]. A transferable real-time incident detection model is still needed; this is the purpose of this study.

SVM METHODOLOGY

SVM is a relatively new machine learning pattern classifier model that uses a sample of "training data" to define an optimal boundary between classes. Training vectors are chosen to lie closest to the class boundary and are called support vectors. Given a training set of instance-label pairs (x_i; y_i); i = 1, …, ι, y_i is either -1 for a non-incident and 1 for an incident and indicates the class to which the point x_i belongs.

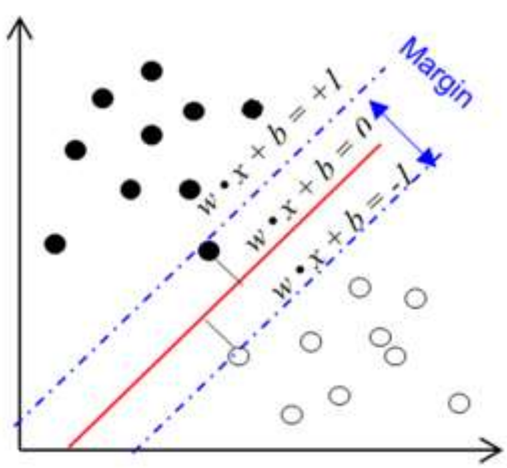


Figure 2- 1 Pattern classification concept

The general mathematical form of the linear SVM is $f(x) = w \cdot x + b$, where $w \cdot x$ corresponds to the dot product, w is the normal vector of the hyperplane, and b is a variable (Figure 2-1). The linear SVM function finds the two closest points in each data set and creates a hyperplane between them to separate the two data sets. The problem becomes more complicated when the data are not linearly separable. SVM achieves non-linear classification by mapping the input vectors into a higher-dimensional feature space through the kernel function ϕ until the data becomes linear, as shown in Figure 2-2. Because the kernel mapping only depends on the inner product of the input data vectors, the computational cost remains low.

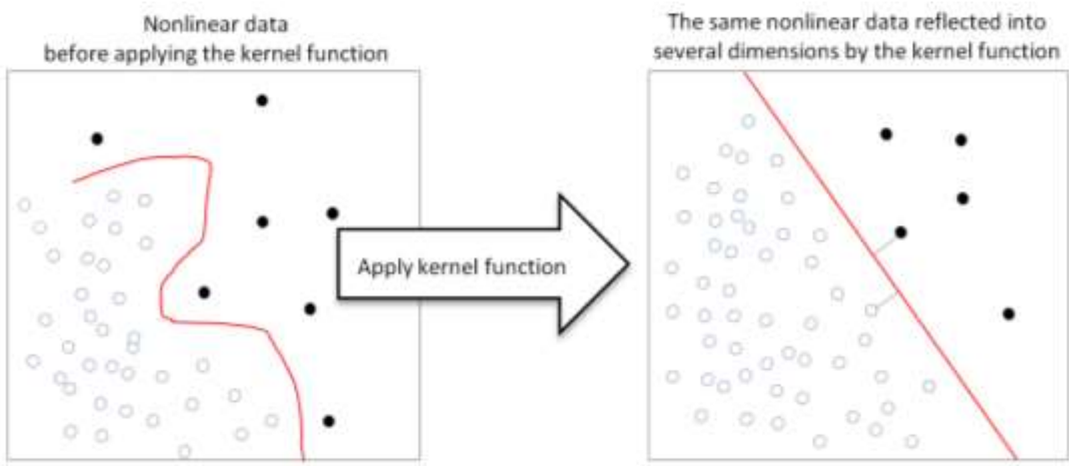


Figure 2- 2 Nonlinear classification created by applying a kernel function that reflects the data into higher dimensions until linear separation is possible

If a hyperplane to separate the two classes cannot be found, a Soft Margin method can be introduced to split the classes as cleanly as possible. The method introduces non-negative slack variables, ξ_i and C (error penalty), which measure the degree of misclassification of the data. The optimization becomes a trade-off between a large margin and a small error penalty. The prediction function objective can be achieved by solving the following optimization problem (*34*):

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i,$$

subject to $y_i(w^T\phi(x_i)+b) \geq 1-\xi_i$, and $\xi_i \geq 0$,

where C is a penalty parameter for the error term, ξ_i that is the measure of the degree of misclassification of instance i, x_i is the training/testing vector (input pattern), y_i indicates the class to which the point x_i belongs, $\frac{1}{||w||}$ is the distance from closest point to the hyperplane, and ϕ is a projection function from lower-dimension space into higher-dimension space. This is an optimization problem with a convex quadratic objective function and linear constraints, which can be solved using Quadratic Programming (QP). To make the model work efficiently in very high dimensional space, Lagrange duality has been implemented. The dual form typically improves the optimization problem. By applying Lagrange multipliers, this problem can be transformed into dual optimization problem:

$$0 \leq \alpha_i \leq C \max \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i\alpha_j y_i y_j (\phi(x_i) \cdot \phi(x_i)),$$

where $\sum_{i=1}^{N} \alpha_i y_i = 0$.

This yields a maximization problem with α_i parameters. The α_i's will be zero except for the α_i's belonging to support vectors. Therefore only the inner product of x and the support vectors is required in order to make prediction. The kernel function can be used to avoid $(\phi(x_i) \cdot \phi(x_i))$. Kernel functions are the mathematical transformation engines upon which the SVM process is built. Many kernel functions are available. The two most commonly functions used, the radial basis (RBF) and sigmoid, are shown below:

Radial basis function: $K(x_i, x_j) = (\phi(x_i) \cdot \phi(x_i)) = e□p(-\gamma|x_i - x_j|)^2$,

Sigmoid function: $K(x_i, x_j) = (\phi(x_i) \cdot \phi(x_i)) = \tanh(\gamma(x_i \,.\, x_j) + \delta)$,

where γ and δ are kernel parameters.

In general, the RBF kernel can be applied to non-linearly separable data as well as linear data because the linear kernel is a special case of RBF [39]. The RBF kernel has fewer numerical difficulties because of the number of hyperparameters that influence the complexity of model selection. For more information about SVM, see Jin et al. [40]. Both RBF and sigmoid kernel functions were used in the incident detection model described in the following sections.

APPLYING MODIFICATIONS TO THE SVM MODEL

SVM classification reliability can be improved by normalizing the data before calibrating the SVM model. Hsu [34] recommends normalizing data in the datasets to a range of [-1,1] or [0,1]. Normalizing the data assists the model by reducing numerical difficulties during calculation and prevents the effects of one variable from dominating the other variables. The same normalization method should be applied for both training and testing data sets.

After the dataset was normalized, two important tasks needed to be addressed. First, the optimal kernel function parameters (C, γ) must be determined, and second, the model validation method must be selected. Because of the significant effect of the (C, γ) model parameters on model accuracy, two methods were selected to select optimal parameter values: a grid search and a pattern search. A grid search finds the parameter value across the defined search range while a pattern search starts at the center and makes steps using trial and error to improve the model fit. Grid search is computationally more expensive than pattern search when there are a large number of parameters. The speed of the search is highly dependent upon the range of C and number of parameters. After creating the model, the next important task is to test and validate it. In this study, two different methods were applied and compared. The model was cross-validated with different numbers of folds and by holding aside a random percentage of the training dataset, validating the fold, and repeating the processes "n times" (n: number of folds). A sensitivity analysis was performed to find the optimal number of folds since the number of folds for cross-validation depends highly on the size of the dataset (this dataset was larger than previous studies). The cross-validation procedure can prevent over-fitting by indicating when further training will not achieve better generalization [41,42].

PERFORMANCE MEASURES

Incident conditions are rare in comparison to incident-free conditions. Due to the low ratio of incident to non-incident cases, the overall classification accuracy does not adequately represent model performance. The overall accuracy of all cases being properly predicted as non-incident reaches 85 percent. Unfortunately, an accuracy of 100 percent is useless if the time for detection is *unreasonably* slow. Therefore, other performance measures are necessary to fully evaluate the detection model.

To evaluate existing an incident detection algorithm's detection rates, false alarm rates and detection time have traditionally been used. Even though other algorithms (ARIMA, Bayesian and SSID) have reported cases of detection rates reaching 100 percent, they typically possess either a high false alarm rate or take too long to detect an incident [43]. To address the problem, a single performance measure for each metric is used to compare the different models that are studied.

The incident detection model performance is evaluated using the following metrics:

$$\text{Detection rate } (DR) = \frac{number\ of\ detected\ incident\ cases}{total\ number\ of\ incident\ cases} * 100,$$

$$\text{False alarm rate } (FAR) = \frac{number\ of\ false\ alarm\ cases}{total\ number\ of\ incident\ free\ cases} * 100,$$

$$\text{Classification rate } (CR) = \frac{number\ of\ correctly\ classified\ cases}{total\ number\ of\ cases} * 100,$$

$$\text{Mean time to detection } (MTTD) = \frac{\sum_1^m t_i}{m},$$

where

m is the number of incident cases detected,

t_i is the length of time to detect the *i*th incident.

An efficient incident detection model should have high DR, low FAR, and short MTTD. These measures are interdependent. This makes it difficult to find an optimal incident detection model. To comprehensively evaluate classifier performance, all of the above metrics were combined to create a new performance index (PI) proposed by Chen et al. [31]. The PI can be computed as follows:

$$PI = w_{DR}.(1 - DR) + w_{FAR}.FAR + w_{MTTD}.\frac{MTTD}{THD_{MTTD}} + w_{CR}.(1 - CR),$$

where w_{DR}, w_{FAR}, w_{MTTD}, w_{CR} are normalized weights for corresponding measures, and THD_{MTTD} is the threshold of MTTD. For simplicity, the Chen et al. model weighted DR, FAR, and MTTD equally. THD_{MTTD} can be any positive value bounded above by the maximum possible value of MTTD. Since it is recommended for unbalanced data, CR weight value can be set to zero to take the effect of high detection out of the PI [31].

SUMMARY

The relatively new machine learning pattern classifier SVM model can solve a non-linear classification problem by mapping the input vectors into higher-dimensional feature spaces through the kernel function ϕ. The SVM benefits from low computational cost because the kernel mapping only depends on the inner product of the input data vectors. The Soft Margin method can be applied to split the classes for a non-separable data set. The soft margin allows data to be misclassified, but assigns a penalty cost prior to executions of the optimization objective function. To compare different models, a single performance measure is used across all methodologies.

Chapter 3: Data Collection

Technology advancements have made traffic data collection more convenient and efficient. Consequently, the number of studies assessing real-time incident detection has increased recently. As previously mentioned, incident detection processes are closely tied to the sensing systems providing real time traffic data. If the sensing system is in place, then the best incident detection algorithm is dependent upon the type and reliability of the sensed data.

One reason why previous model developments could not reliably detect incidents during peak hour conditions was because adequate before-and-after incident data was not incorporated into the models. These models rely on a single point measure of effectiveness, either downstream data or upstream data. If both patterns of up and downstream speeds are considered, then accurate detection of incidents during the peak hour increases [20,44]. Generally, vehicle speeds are expected to drop upstream of an incident and increase downstream of an incident. The pattern from the change in speeds can improve incident detection during peak hour stop-and-go conditions. This section presents the incident detection framework used in this study.

The best way to identify traffic patterns is by observing vehicle trajectories. However, trajectory data is not available everywhere. The next best option is traffic information from a series of point detectors. Currently traffic control centers across the world primarily depend on single point detector data. Therefore in this research, speed and volume information from point detectors provides the input data.

STUDY SITE, INCIDENT DATA, AND TRAFFIC DATA

For this study, data was collected from traffic control centers in Dallas and Miami (Figure 3-1). Dallas provided the training data from four segments on two directions of two different freeways.

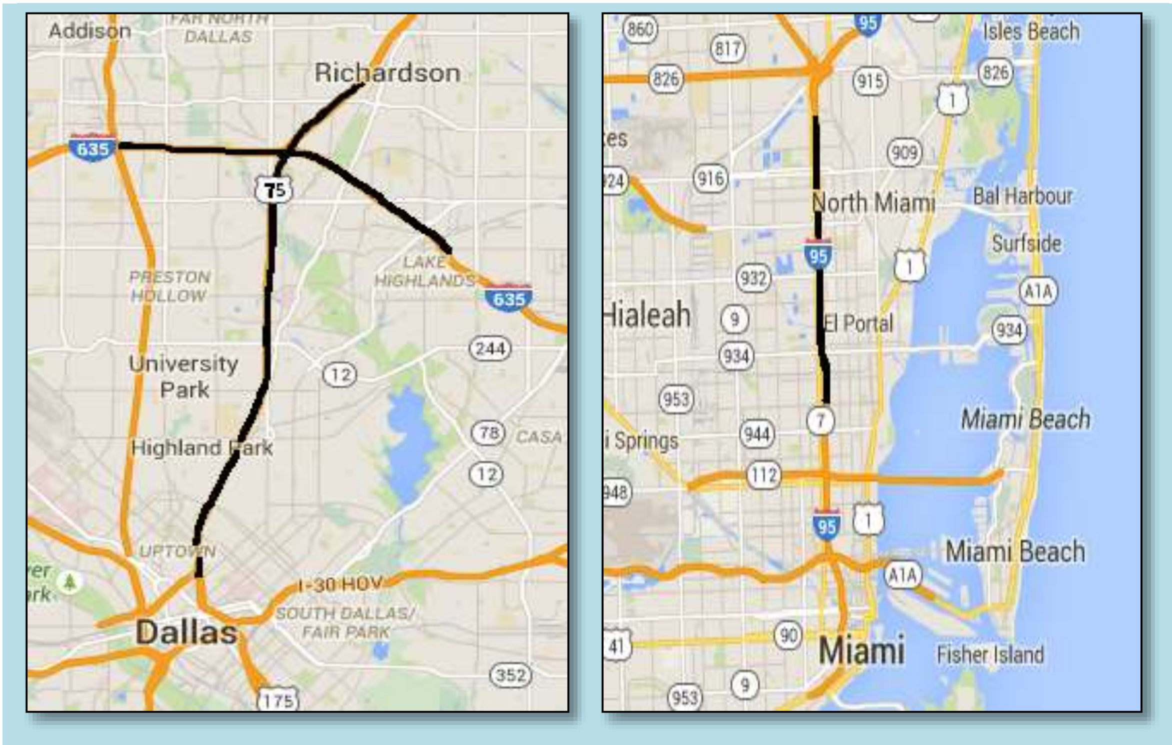


Figure 3- 1 Study sites

Data from four Dallas freeway sections were collected on US 75 North-South and I-635 East-West (Table 3-1). These sections are long so sometimes they vary from 2-lane to 4-lanes within the same section. The Dallas traffic control center, DalTrans, gathers real-time information from electronic sensors in the pavement, freeway call boxes, video cameras, 911 calls, officers on patrol, highway crews, motorist cellular calls, and commercial traffic reports 24-hours a day, seven days a week. The traffic data were collected by the Dallas, Texas Traffic Control Center and include upstream and downstream speed, volume, and occupancy from US-75 and I-635. The data were collected from July to September of 2012. The resolution of traffic information (speed, volume, and occupancy) is every 5 minutes on a per lane basis. Both directions of US-75 and I-635 were studied to build the model (see Figure 3-1).

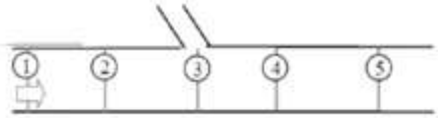
The collected incident information (Figure 3-2) included incident location, affected lane(s), time of incident detection, time cleared, and type of incident from 106 incidents in Dallas. For each Dallas (training) incident, the research team observed two hours both before and after the incident occurrence for incident detection.

The Dallas detection model was then applied to the Miami incident data retrieved from the FDOT D6 SunGuide for testing. The data was collected from May 2012 to June 2012 with the same traffic information and resolution as Dallas. The Miami incident information included location, start time, and end time. The Miami (testing) data were observed two hours before the incident. Traffic data were collected from a 4.5 mile segment of I-95 North in Miami by the Regional Integrated Transportation Information System (RITIS) to test the model.

Table 3- 1 Study sites and incident data

Freeway	Direction	Length(ml)	No. of Detectors	No. of Incidents
Dallas - US 75	Northbound	10	20	19
Dallas - US 75	Southbound	10	20	63
Dallas – I 635	Eastbound	7	13	12
Dallas – I 635	Westbound	7	13	12
Miami – I 95	Northbound	4.5	14	20

The two nearest detector stations (one downstream and one upstream) to the location of each incident were identified using a Google Earth. The traffic data were extracted two hours before the incident start to two hours after incident clearance for each incident.



No.	Road Name	Road Direction	CrossStreet Name	DetectedTime	ClearedTime	AffectedLanes	Type
1	US 75	North	Monticello Ave	8/10/12 17:06	8/10/12 17:34	Lane1	Disabled Vehicle
2	US 75	North	McCommas Blvd	8/31/12 23:23	9/1/12 1:47	Lane1	Accident
3	US 75	North	Mockingbird Ln	9/25/12 22:36	9/25/12 23:15	Lane1, Lane2	Accident
4	US 75	North	Caruth Haven Ln	9/14/12 17:20	9/14/12 17:57	Lane1, Lane2	Accident
5	US 75	North	Walnut Hill Ln	7/5/12 7:15	7/5/12 7:25	Lane1	Debris

A	B	C	D	E	F	G	H
Date & Time	Detector Name	Detector ID	Detector Status	Speed	Volume	Occupancy	Long Vehicle Volume
8/16/2012 2:03:25 AM	SB US75 @ Parker North Lane 1 SB	10132 300	3	50	100	7	10
8/16/2012 2:08:25 AM	SB US75 @ Parker North Lane 1 SB	10132 300	3	60	102	20	12
8/16/2012 2:13:25 AM	SB US75 @ Parker North Lane 1 SB	10132 300	3	45	122	19	13
8/16/2012 2:18:25 AM	SB US75 @ Parker North Lane 1 SB	10132 300	3	55	111	8	19

Figure 3- 2 Sample of the raw dataset

Chapter 4: Examining DTW and SVM

Incident detection can be viewed as a pattern classification problem. Therefore, any good classifier can serve as a potential tool to the incident detection problem. The recent research showed the dynamic time warping algorithm [51] and the support vector machine concept [31,51] are one the most successful solutions for pattern classification. These two solutions are examined in this chapter.

DYNAMIC TIME WARPING (DTW)

DTW algorithms were proposed around 1970 in the context of speech recognition to account for differences in speaking rates between speakers and utterances. For example, DTW can find a low distance score between the sound signals corresponding to utterances “look” and “loook” without being sensitive to the prolonged duration of the ‘o’ sound. Other applications of DTW have been found in genetics for gene sequencing and detection. DTW has also been applied for clustering and classification of Electro-cardiogram analysis [52,53].

Chandrasekaran et al. (2011) brought the concept of DTW to the transportation field for the first time to track vehicular speed variation [55]. Subsequently, there have been a few more studies in this field:

- “Tracking vehicular speed variations by warping mobile phone signal strengths” [55]
- “Traffic Event Automatic Detection Based on the OGS-DTW Algorithm” [56]
- “Traffic incident detection system using a series of point detectors” [50]

Some studies show that the procedure may produce higher detection rates than artificial intelligence algorithms with lower false alarm rates. For example, Hi-ri-o-toppa (2012) used upstream and downstream speed changes to develop a DTW incident detection algorithm, which achieved a 94% detection rate and a low false alarm rate. The method proposed in this research uses data mining and time series classification and is the confluence of a set of disciplines, including statistics, machine learning, artificial intelligence, and information science.

DTW Methodology

In the first step, we started with two inputs the observed test query, X, and the reference series, Y. To compare the two datasets, we measured similarity or likeness between X and Y and locally stretched or compressed portions of the series to get the smallest distance between the two. The series are hypothetical sequences of X and Y, with the x-axis showing the time index and the y-axis showing outcome measure. Although the series may possess different lengths, measurements were taken at equidistant time points.

An optimal warping path between X and Y is a warping path, p*, which has minimum total cost among all possible warping paths. The total cost $C_p(X,Y)$ of a warping path, p, between X and Y with respect to the local cost measure is:

$$c_p(X,Y) := \sum_{l=1}^{L} c(x_{n_l}, y_{m_l})$$

The DTW distance DTW(X, Y) between X and Y is then defined as the total cost of p*:

$$DTW(X,Y) := c_{p^*}(X,Y)$$

$$= \min\left\{c_p \quad (X,Y) \middle| p \text{ } is \text{ } an \text{ } (N,M) - warping \text{ } path\right\}$$

The optimal warping path between X and Y is represented graphically in Figure 4-1 below as Time Series A and Time Series B. The orange "diagonal" (the slanted band window) goes from one corner to the other of the possibly rectangular cost matrix, therefore having a slope of M/N. The computation is approximate: points having multiple correspondences are averaged, and points without a match are interpolated. This average between Time Series A and B is graphically represented by large red dots in Figure 4-1 and interpolated points without a map are represented by blue dots and a red directional arrow. The area is not normalized by path length.

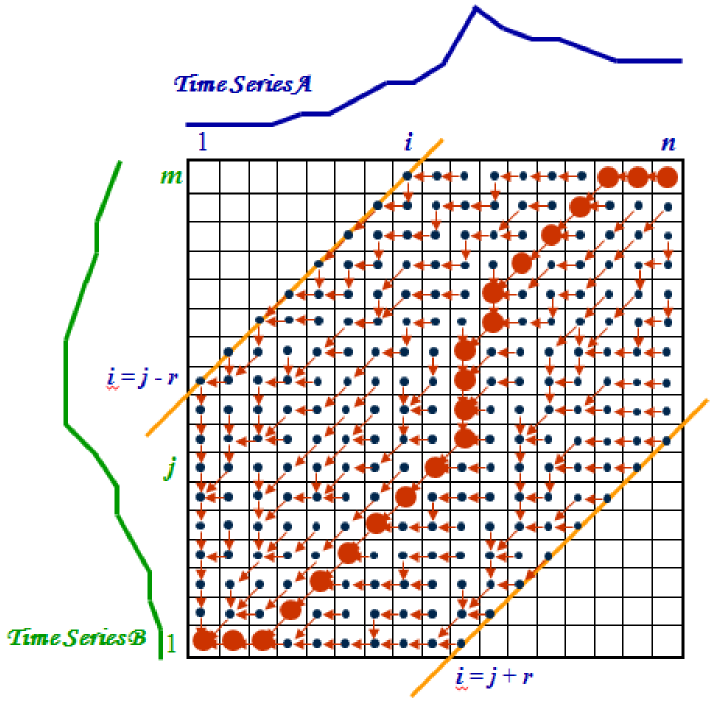


Figure 4- 1 Sequences of warping map matrix to find optimal warping path. Source: http://homepages.inf.ed.ac.uk/group/sli_archive/slip0809_c/s0562005/theory.html

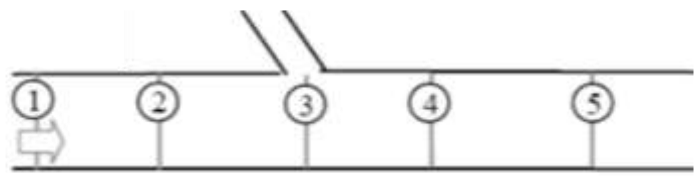
Modifications and constraints for DTW

Additional constraints can be applied to the model to produce specific results, such as introducing an additional weight vector to favor the vertical, horizontal, or diagonal direction (w_d, w_h, w_y) in the alignment. To constrain the slope of the admissible warping paths, the step size condition can be modified. However, while putting constraints on search windows can make the model faster, it can potentially prevent the model from finding a feasible solution.

DTW Model Development and Results

In this algorithm, first the freeway incident data trains the DTW model. In the training process, the proposed system captures patterns associated with incidents in the training dataset. The patterns captured have common trends that can be described by categorizing each type of indicator (such as speed).

Five sequential locations were chosen in the incident data to detect patterns (the same five locations shown in Figure 4-2). In all cases, incidents happened Northbound on Lane 1 during different times of day. To have a basis for comparing incident with non-incident situations, we chose to characterize the non-incident situation as having no incident within 5 miles before and after the specific incident.



Road Name	Road Direction	CrossStreet Name	DetectedTime	ClearedTime	AffectedLanes	Type
US 75	North	Monticello Ave	8/10/12 17:06	8/10/12 17:34	Lane1	Disabled Vehicle
US 75	North	McCommas Blvd	8/31/12 23:23	9/1/12 1:47	Lane1	Accident
US 75	North	Mockingbird Ln	9/25/12 22:36	9/25/12 23:15	Lane1, Lane2	Accident
US 75	North	Caruth Haven Ln	9/14/12 17:20	9/14/12 17:57	Lane1, Lane2	Accident
US 75	North	Walnut Hill Ln	7/5/12 7:15	7/5/12 7:25	Lane1	Debris

Figure 4- 2 Locations, time, and impact of incidents on US 75

First we tried to identify the speed pattern for each location. Then, smoothing technique was implemented to remove some noise. Speed patterns for five locations during typical, non-incident flow, and during an incident were extracted. The speed pattern was different at each of the locations.

To develop DTW algorithm, "R" programing language was used. The first classic DTW models were developed with typical day speed data to describe what should be expected during non-incident cases. Two random incident-free days were considered within several miles radius of the test location. During a non-incident case, the two time series are compatible, so the cost matrix is expected to show green color (low cost) and a diagonal path. The cost model shows lowest cost or the best compatible match of two time series. If something unusual happens, the optimal path would deviate from the diagonal and the color will be more orange. Figure 4-3 shows the output of the DTW model using incident data showing incidents at various locations. The blue line is the shortest path found by the model. Deviation from the diagonal trend indicates higher cost to the user. The incidents patterns in Figure 7 show that the two time series are not well matched, so the cost matrices show more orange and yellow colors.

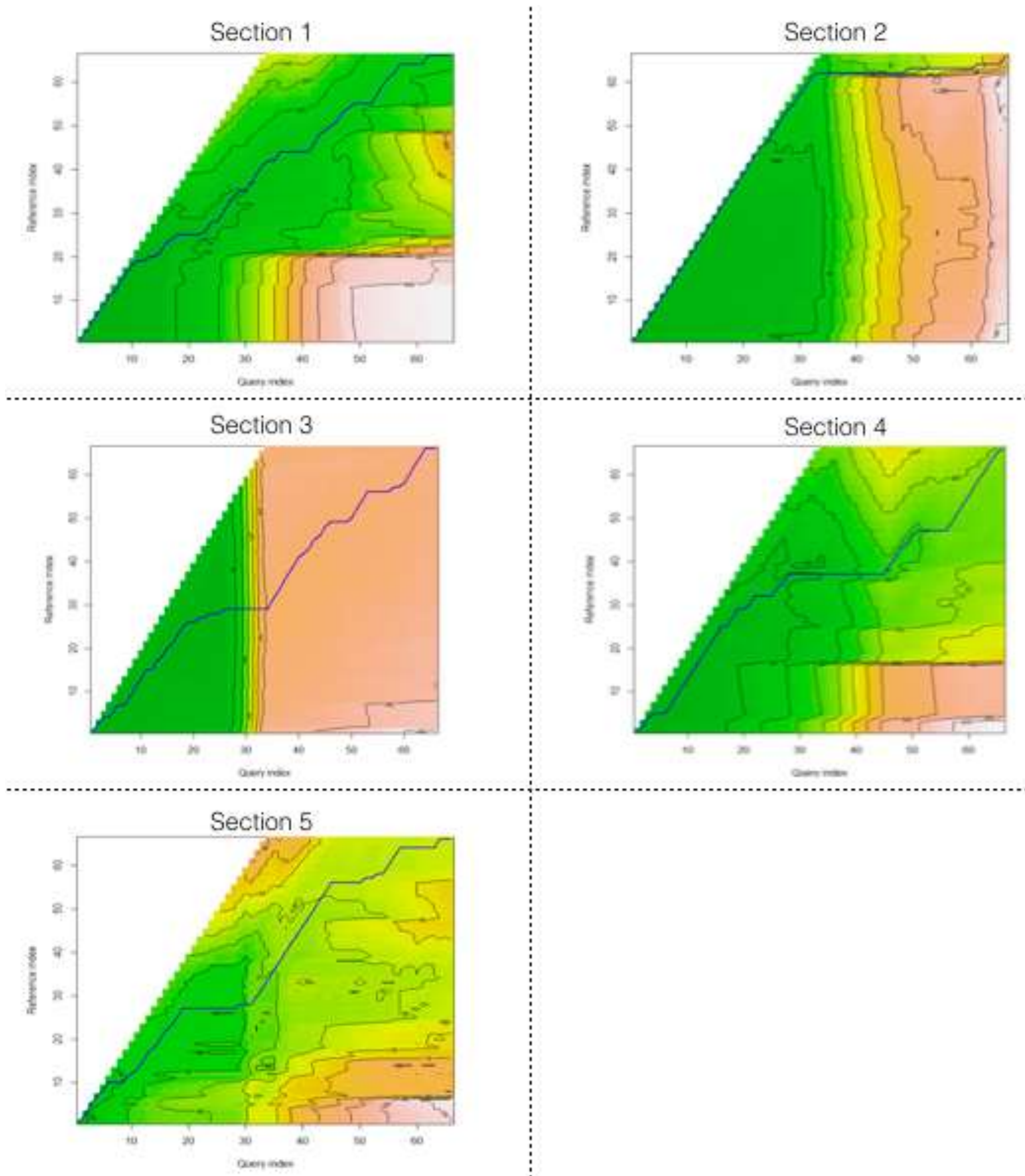


Figure 4- 3 DTW Cost Matrixes for Incident Detection

Figure 4-4 provides a better visual expression of the time warping concept. The red and gray lines represent traffic speed during non-incident and incident conditions respectively for the same location. The models are warped to show how one data point maps to the other time series. Time warping does not necessarily show that an incident happened, but it shows the pattern between the two time series. These patterns are saved in a library to be referenced in future tests. Therefore applying a classifier is necessary to identify new patterns as either an incident or a non-incident. To obtain a reliable estimate of the classifier accuracy a common classifier applied for DTW is the k-fold cross-validation technique, which will be used here.

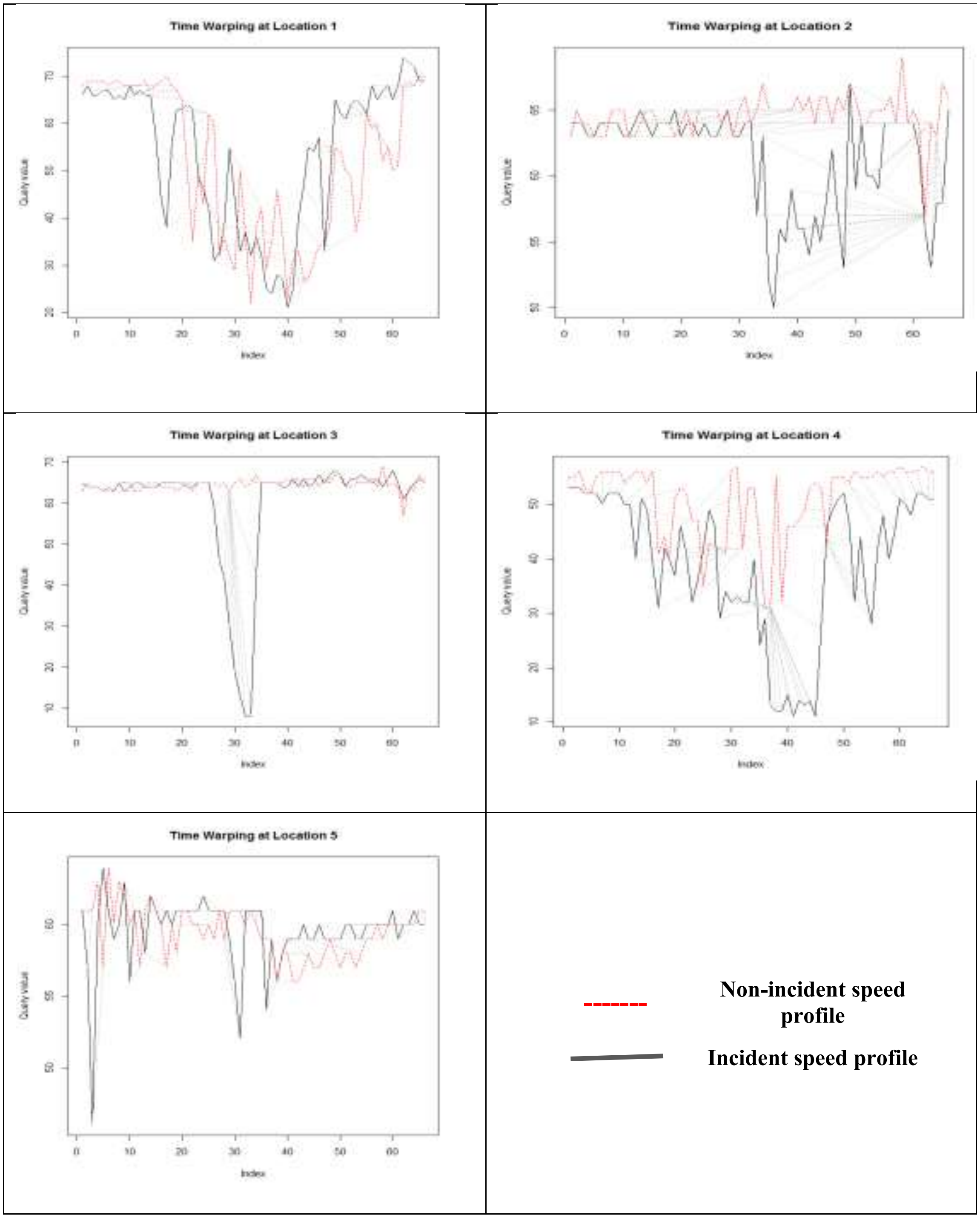


Figure 4- 4 Time Warping view of DTW output for incident

The DTW algorithm processes each test set and a classifier compares the similarity between normal or incident conditions and the new pattern. Because running DTW requires extensive computing resources, a threshold obtained from cross validation alerts the data collection program to start scanning for an incident. The threshold was calculated from the k-mean and is represented by red lines in the graphs in Figure 4-5. A higher speed threshold can be selected as the trigger to start collecting data. As speed decreases and meets the second threshold, the software starts scanning and collecting data points backwards from the current speed toward the free flow speed. It also starts collecting forward until it meets the second threshold. The second threshold is a trigger to stop collecting data and start the DTW algorithm.

Using dynamic thresholds based on historical traffic data, thereby accounting for typical variations of traffic throughout the day, can increase the accuracy of the algorithm. Therefore, this approach could recognize recurrent congestion and thereby reduce the incidence of false alarms.

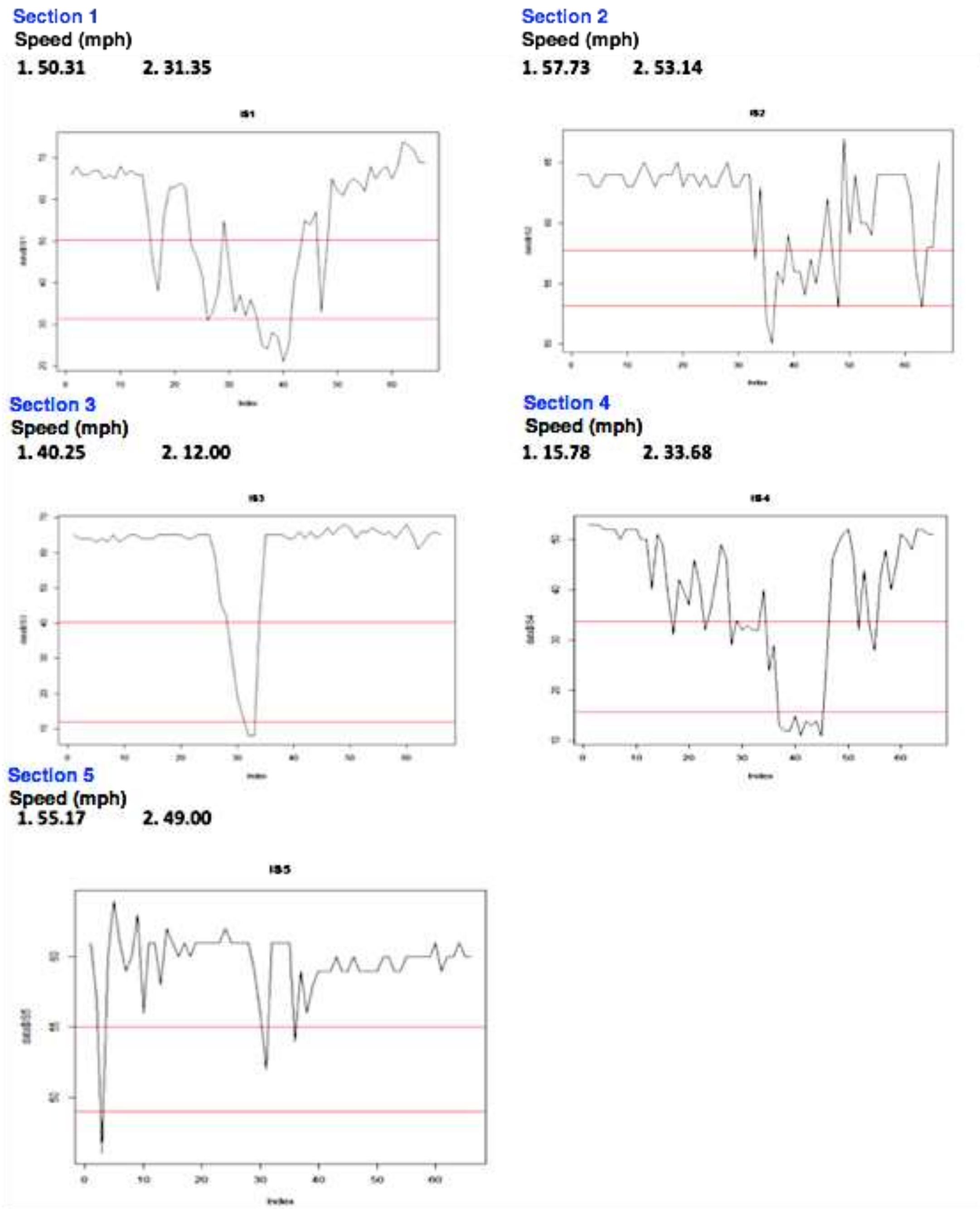


Figure 4- 5 K-means thresholds (1. First threshold speed, 2. Second threshold speed)

SUPPORT VECTOR MACHINE (SVM)

THE SVM CONCEPT HAS BEEN PRESENTED IN DETAIL IN CHAPTER 2. HERE WE FOCUS ON INITIAL MODEL DEVELOPMENT.

Model Development and results

This section presents the implications of SVM for incident detection using field data. The same database used in the DTW model has been used for this section. The first step is generating the cases required for developing and evaluating the SVM incident detection model. The idea of incident detection is based on the concept that when an incident happens, the kinetics of passing vehicles would be affected: the speed drops upstream and increases downstream, lane changing increases, and involved vehicles demonstrate large acceleration and deceleration rates. This part of study identified the speed profile and volume over a selected time step t_s (5 min for Dallas data) to recognize the patterns that indicate incident occurrence. An array of five values for each time slice has been chosen as the input file for the model (Table 4-1).

Table 4- 1 SVM sample input data and result

Time Instant	Kinetics					Decision
	Downstream Real-time Speed or DSpeed (mph)	Upstream Real-time Speed or USpeed (mph)	Upstream Typical Speed UTSpeed (mph)	Downstream Real-time Volume DVolume (vehicle/hr)	Upstream Real-time Volume UVolume (vehicle/hr)	yi
t	55	49	55	118	145	-1
t + ts	67	14	53	112	91	+1

The decision variable y_i can only have values of +1 representing an incident, or -1 representing a non-incident condition. The objective of SVM training is to find the prediction function:

$$f(x_i) = w * x_i + b,$$

where $\frac{1}{||w||}$ is the distance from closest point to the hyperplane, x_i is the training/testing vector (input pattern), and b is the offset from the origin.

This objective function optimizes the minimum distance between the classification hyper-plane and any sample of training data. Considering the complexity of traffic behavior, non-separable data must be allowed for training. As mentioned earlier scaling is important for the success of AI models such as ANN and SVM [34]. Before training, all the data were linearly scaled to a range of [0, 1].

Here v-fold cross-validation was used to maximize the use of training data and search for optimal parameters (C, γ). First, the data was divided into v subsets of equal size. Sequentially, one subset was tested using the classifier trained on the remaining (v - 1) subsets. Thus, each instance of the whole training set was predicted once so the cross-validation accuracy equals the percentage of data that are correctly classified. Different numbers of v-folds were tried to determine the sensitivity of data to the number of folds, and 5 and 6 groups produced the best results.

The optimal parameters were identified through grid searching of many combinations in the range of [C, γ] = [2^{-5}: 2^{2}: 2^{6}, 2^{-15}: 2^{2}: 2^{4}]. The experiment was performed by increasing parameters in exponential order, i.e. 2^{n}, in the range of -5 to 6 for C and -15 to 4 for γ within two steps. The identified optimal parameters were then used for the entire training set to generate a trained SVM algorithm.

This study used LIBSVM [41], an open source implementation routine for SVM to train and test the SVM model. The training time of the SVM model was less than five seconds in all the training cycles. The prediction time was quite short as well, which is a vital element for real-time application.

The Radial Basis Function (RBF) has been used as the SVM kernel function. The optimum values found for parameters using a grid search to minimize total error in the objective function are (C, γ) = (0.3125, 8). The overall accuracy of training and validation is presented in Table 4-2. The results prove the robustness of the model to predict incidents during peak hours.

Table 4- 2 Base model output

```
=============  Misclassification Tables  =============

 ---  Training Data  ---

------------Actual------------Misclassified---------
  Category           Count       Count   Percent   Cost
-----------------------------------------------------
 Non-Incident -1      208          1       0.481   0.005
 Incident      1       36          3       8.333   0.083
  ---------------------------------------------------
     Total            244          4       1.639   0.016

  Overall accuracy = 98.36%

 ---  Validation Data  ---

------------Actual---------------Misclassified-------
   Category          Count      Count     Percent  Cost
-----------------------------------------------------
 Non-Incident -1      208          2        0.962  0.010
 Incident      1       36          3        8.333  0.083
-----------------------------------------------------
     Total            244          5        2.049  0.020

  Overall accuracy = 97.95%

=======================================================
```

As shown by Table 4-2, overall accuracy of both training and prediction (validation) is high. Besides accuracy, the result table provides a misclassification count and cost of misclassified cases. Each data point represents a five-minute time window. Although the model may not detect the incident in the first five-minute time window, it succeeds in the second time window. In this example, there are only 3 incident cases in which the model did not detect an incident in the first five minutes, but all 3 were detected in the second five minutes. As expected, different penalty parameters are found for incident versus non-incident cases. There is a higher penalty (cost = 0.083) for misclassified incident cases. The findings show that the false alarm rate is low (around 2%). There was only one case in which an incident occurred and the model did not detect it. There were three cases in which non-incidents were falsely detected as incidents.

The next step was to define different scenarios and conduct sensitivity analyses to find the best-fit kernel function and associated parameter values. Five different scenarios were defined (Table 4-3). The validation accuracy is the checkpoint to choose the best model because achieving high validation accuracy from the unseen dataset more precisely reflects the prediction accuracy. These results suggested that using the RBF kernel function was not only faster but also more accurate. Different combinations of variables were considered to evaluate the model sensitivity. Most notably, the typical speed variable, representing non-incident historical speeds, did not have a significant effect on the real-time incident detection model (scenario 3 vs. 4). A potential explanation for this involves the way model is structured. Upstream and downstream speeds under incident conditions tend to be very different, providing a clear classification border while typical speeds tend to have little value in delineating that border. The results corresponding to volume showed using volume improved prediction accuracy (Scenario 1 vs. 3)

Table 4- 3 Scenario details and results

Scenario	Model	Variable	Model Accuracy	False Alarm
1	Base-RBF	• Downstream real-time speed • Upstream real-time speed • Upstream typical speed • Downstream real-time volume • Upstream real-time volume	• Training: 98.36% • Validation: 97.95%	• Training: 4 out of 244 • Validation: 5 out of 244
2	Sigmoid	• Downstream real-time speed • Upstream real-time speed • Upstream typical speed • Downstream real-time volume • Upstream real-time volume	• Training: 97.54% • Validation: 97.13%	• Training: 6 out of 244 • Validation: 7 out of 244
3	RBF	• Downstream real-time speed • Upstream real-time speed • Upstream typical speed	• Training: 97.54% • Validation: 96.72%	• Training: 6 out of 244 • Validation: 8 out of 244
4	RBF	• Downstream real-time speed • Upstream real-time speed	• Training: 97.13% • Validation: 97.13%	• Training: 7 out of 244 • Validation: 7 out of 244

The performance index (PI) of each scenario is reported in Table 4-4. The PI proposed by Chen was used here to compare different scenarios [14]. For simplicity, equal weights for DR, FAR, and MTD were applied as Chen et al. and Xiao et al. suggest [14,15]. The detection time is highly dependent upon the time resolution of the dataset. In this case, data was collected every five minutes. Traffic management centers collect data with even finer time resolution than 5-minute intervals, therefore the resolution requirement for the model can be easily met. In the presented methodology, incidents missed during the first interval were predicted for the second time slice. There is a time lag between the actual incident occurrence and incident detection by the proposed model partially because of the time resolution of the analysis intervals. A finer time resolution could reduce the time lag between occurrence and detection.

Table 4- 4 Evaluation of C-SVM for different scenarios

Scenarios	Brief details	DR (%)	FAR (%)	CR (%)	PI
1	RBF-Base	91.67	0.96	97.95	0.615
2	Sigmoid	91.67	1.92	97.13	0.935
3	RBF-all speeds	86.11	1.44	96.72	0.860
4	RBF-just real time speed	91.67	1.92	97.13	0.935

A lower PI represents better performance, so the base model has the best overall performance. If volume information is not available and speed is the only available variable, Scenario 3 provides the best results. However, these results are based upon application of equal weights to all PI terms. (Weight combinations that best suit the situation cold also be chosen.) These results suggest that real-time traffic volume improves the prediction accuracy compared to using only real-time speed data. They also suggest that historical typical speed data adds little value to the model. This part of study indicates that the technique is promising and should be further tested with more diverse and extensive data sets.

CONCLUSIONS

Two groups of experiments were performed to evaluate DTW and SVM. Evaluation revealed that both algorithms could successfully classify traffic conditions into two categories – incident and non-incident – during peak hours. Both model predictions were fast, however, the size of data set was small. DTW is computationally more expensive than SVM. Models were trained on a network based on freeway segment in Dallas, TX. Comparing these two methods, application of DTW in the field of transportation is quite new. Because of time, available data, and the nature of DTW, we did not have enough data to validate our DTW model. The advantage of using SVM is that it does not require a big dataset to train and validate the model. For the purpose having more accurate comparison, we recommend generating more data. For this dissertation, based on the available data, we found SVM to be a better choice for our incident detection model development.

Chapter 5: Developing the SVM Incident Detection Model

The majority of AID algorithms are focused on identifying traffic incident patterns, but do not adequately investigate possible similarities in patterns observed under incident-free conditions. When traffic demand exceeds freeway capacity, the operation enters a highly unstable regime often referred to as “stop-and-go” conditions, catagorized by the Highway Capacity Manual as Level of Service F. Under such conditions density, speed, and volume are highly variable – hence the descriptive name “stop-and-go”. The most challenging part of real-time incident detection is recognition of traffic pattern changes during incident versus stop-and-go conditions.

PART 1 FIRST STAGE OF MODEL DEVELOPMENT

The first step for model development was data preparation. According to recent studies [45,46] mean speed, standard deviation of speed, headway, and flow are the best indicators for incident prediction. Hi-ri-o-tappa (2011) conducted an evaluation of the best indicator based on a statistical comparison of two datasets. According to that study, mean speed and standard deviation of speed are the best indicators followed by occupancy and traffic flow rate [45]. Therefore for this part of study, these data items were extracted from Dallas incident database. The incidents occurred at different locations and different times of day along 10-miles of the US-75 freeway. Nineteen incident cases along US-75 northbound were extracted from the Dallas incident database. For each incident, the research team observed two hours before and after the incident began. Each time window of five minutes is represented as one data point for the SVM model. Therefore, the training data, shown in Table 5-1, contained 620 instances with 176 incident cases.

Table 5- 1 Input data counts

```
------ Training Data ------

---------------------------
Category              Count
---------------------------
Non-Incident -1         444
Incident      1         176
---------------------------
Total                   620
```

For SVM to work well, the dataset must contain enough samples from each of the classes [31,34]. In classification, the data is imbalanced when one class, such as the incident class, is relatively rare compared to the other classes. A C-SVM model is recommended for unbalanced data [41,47] because it introduces penalty values for different classes. This part of study applied C-SVM and other modification methods, including normalizing and cross-validation, to improve model accuracy.

The SVM model classifies traffic patterns into states of incident (+1) and non-incident (-1). The data is a vector of real numbers with a time window size of 5 minutes. The speed profile and volume data are collected at locations up/downstream of each incident over the selected time step (5 minutes for Dallas data). The SVM model recognizes the speed profile and volume data patterns that indicate incident versus non-incident conditions. In Figure 5-1, a schematic traffic pattern during an incident is presented. Speed and standard deviation of speed are expected to decrease upstream of the incident while remaining the same or increasing downstream of the incident. Volume will decrease upstream and downstream of the incident as well. Traffic flow changes upstream and downstream of the incident are not expected. Occupancy has an increasing pattern at the upstream detector and a constant pattern at the downstream detector as shown in the last line of Figure 5-1.

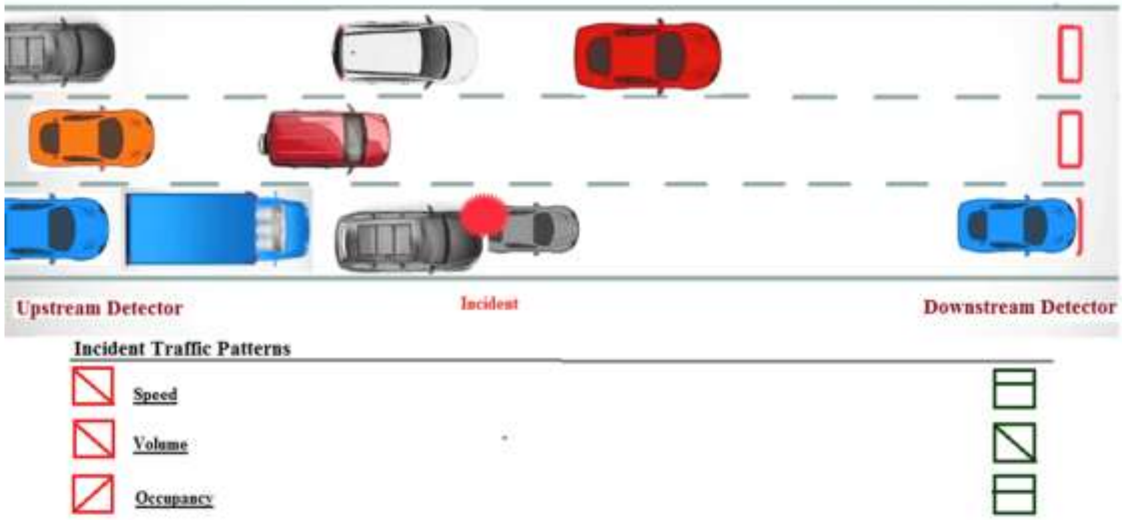


Figure 5- 1 Traffic Incident Patterns

Therefore, information from point detectors about speed, volume, and occupancy are the input data (Table 5-2).

Table 5- 2 SVM Sample Input Data and Model Decision Variable

Time Instant	Kinetics (input)						Decision
	Downstream Speed (mph)	Downstream Volume (vehicle/hr)	Downstream Occupancy (sec)	Upstream Speed (mph)	Upstream Volume (vehicle/hr)	Upstream Occupancy (sec)	yi
t	47	137	9	23	59	60	1
$t + t_s$	51	127	7	55	137	13	-1

Different times of the day were chosen to develop a comprehensive incident detection model because traffic patterns change across the times of the day. To consider the impact of freeway geometric features in the model, different sites were selected from a 15-mile section where incident occurrence did not influence the assumed non-incident traffic pattern. One day was selected to represent the speed pattern of upstream traffic to be compared with the speed pattern belonging to the incident case. All incident cases were reviewed to guarantee that secondary accidents were not included.

Model Result

After preparing and normalizing the data for the SVM model, the next step identified methodological enhancements (kernel function selection, model parameter selection, and etc.) to find the optimal SVM model. It is necessary to point out that all parameters in both the training and testing datasets were normalized into the range of [0,1] to avoid possible bias caused by different scale.

To validate the model, two different methods were applied (the random percentage method and the cross validation (CV) method). A sensitivity analysis was performed to find the optimal method. For the CV case, the training data was divided into an equal number of folds and (ν-1) were used for training and one subset was reserved to test the model in the search space of $[C, \gamma] = [2^{-5}: 2^{2}: 2^{6}, 2^{-15}: 2^{2}: 2^{4}]$ as recommended by Lin et al. and Ma et al. [36,48]. For the random percentage method, a randomly selected percentage of data was held out from the model building process for model validation. Different percentiles were tried and the best results were chosen as shown in Table 3. Generally, CV is recommended unless it is not computationally feasible. In the next step, different kernel functions were applied. As expected, the linear kernel function was computationally fast with a low validation rate. The polynomial kernel function was too slow requiring hours of runtime without reliably completing building the model. Since real-time incident detection requires a fast model, the polynomial kernel was not chosen. The RBF and sigmoid kernel functions were chosen for different scenarios. Since here there were only two parameters, the grid search could function very quickly. First, the grid search found the optimal region, and then a pattern search found the global optimum. To have the most accurate parameter selection, the search used CV to maximize use of training data. Many combinations of grid searches were implemented to identify the optimal parameter values and the values found to minimize the total error in the objective function for each model are presented in Table 3. As mentioned earlier, CV was applied to both the model training and validation. Different numbers of folds were used to find the best fit. The algorithm was implemented using the open source LIBSVM code developed by Chang et al. [49]. Model training and execution times were dependent upon the type of kernel function. In the case of using the CV method for model validation, the number of folds affected modeling training and executions times. Different combinations of variables were considered to evaluate the model sensitivity. The findings showed including volume and occupancy into the model increased the accuracy. A potential explanation for this could be that this is an effect of the model structure. Moreover, upstream and downstream patterns under incident conditions tend to be very different and provide a clear classification border, which increases model accuracy. The results corresponding to the best four scenarios in terms of accuracy and runtime are presented in Table 5-3. All variables were included in all of these scenarios.

Table 5- 3 Descriptions and Outputs of the Best Scenarios

Scenario	Model Characteristics	Parameter Value	No. of SV	Model Accuracy (%)	False Alarm
1	C-SVM, RBF, CV Folds: 6-4	e = 0.001 C = 1.27967 Gamma = 37.7757	356	Training: 97.74 Validation: 92.9	Training: 0.97% Validation: 2.10%
2	C-SVM, RBF, CV Folds: 6-6	e = 0.001 C = 1.7896 Gamma = 11.4065	204	Training: 95.65 Validation: 93.55	Training: 1.61% Validation: 2.58%
3	C-SVM, Sigmoid, CV Folds: 6-6	e = 0.001 C =1.2599 Gamma= 0.00077 Coef0 = 1.668	352	Training: 88.71 Validation: 89.03	Training: 6.13% Validation: 5.0%
4	C-SVM, RBF, Random Sampling: 30%	e = 0.001 C = 1.2597 Gamma = 12.6992	89	Training: 95.45 Validation: 90.27	Training: 1.52% Validation: 0.88%

As shown by Table 5-3, overall accuracy of both training and validation for all scenarios is high. The findings suggests that RBF kernel function is a better choice for providing not only more accuracy, but also lower false alarm rates and faster running time - typically a matter of seconds. Generally, the false alarm rate is low (around 2%) except for in Scenario 3 in which the sigmoid kernel function was used. One criteria of choosing the best model is validation accuracy because achieving high validation accuracy from the unseen dataset more precisely reflects the prediction accuracy. In this table under model characteristics, the first fold value indicates validation and the second is used to find model parameters. Most notably, all the RBF scenarios have higher than 89% accuracy.

Comparing models is a very complex task. To make it more understandable, different researchers have introduced different performance functions to compare different scenarios. In this study, the PI proposed by Chen was used to compare different scenarios [31]. For simplicity, equal weights for DR, FAR, and MTD were applied as Chen et al. and Xiao et al. suggest [31,33]. The detection time is highly dependent upon the time resolution of the dataset. In this case, data was collected every five minutes and the model was able to detect the incident on average in the first time slice (the 1 to 5 min slice). Traffic management centers collect data with even finer time resolution than 5-minute intervals, which could improve detection time. In the presented methodology, incidents missed during the first interval were predicted during the second time slice. There is a time lag between the actual incident occurrence and incident detection by the proposed model partially because of the time resolution of the analysis intervals.

Table 5- 4 Evaluation of C-SVM for different scenarios

Scenarios	Model Characteristics (Folds: search, validation)	DR (%)	FAR (%)	MTTD (min)	CR (%)	PI
1	C-SVM, RBF, CV Folds: 6, 4	81.25	2.93	4	92.90	1.305
2	C-SVM, RBF, CV Folds: 6, 6	77.27	3.60	4	90.97	1.544
3	C-SVM, Sigmoid, CV Folds: 6, 6	78.98	6.98	5	89.03	2.731
4	C-SVM, RBF Random Sampling: 30%	71.43	1.28	4	90.27	0.789

Table 5-4 represents the results of the performance analysis for the same four scenarios. It should be pointed out that all the statistics were calculated from the validation data set, not the training data set, because earlier prediction accuracy is the key goal. The lower the PI, the better the model. Therefore, based on the applied performance function, Scenario 4 has the best overall performance. That could possibly mean that even though CV is highly recommended in large datasets it may not benefit the model as much as it could for smaller datasets. CV may cause over fitting of the model yielding a lower validation accuracy. However, these results are based upon application of equal weights to all PI terms, which can be chosen to best suit the situation in question.

Conclusions

The main goal of this research was using recent field data to evaluate the performance of SVM for real-time incident detection. Evaluation revealed that SVM can be applied successfully using real-time data during peak and off-peak hours. The model prediction meets the key requirements of real-time incident detection, high detection speed, and accuracy. Models were trained on data from a long freeway segment in Dallas, TX. The accuracy and speed of SVM is highly dependent upon the kernel function choice and empirically estimated parameter values. Applying the RBF kernel function yielded the best results in this study because it significantly improved the model speed and reliability. Using a random sampling technique to validate the model gives better prediction results compared to cross-validation. The results suggest that real time traffic volume and occupancy improves the prediction accuracy compared to using only real time speed data. The study described here yielded good results. However, the technique should be further tested with more diverse and extensive datasets.

PART 2 EXPANDING SVM INCIDENT DETECTION MODEL

In this part of study, the main focus is developing a generalized incident detection model for traffic control centers. Based on this evaluation, the proposed SVM model provides reliable results. To achieve the goal, more data points were collected from both directions of US-75 and IH-635. Incident cases were extracted from the Dallas incident database, and for each incident the research team observed two hours before and after the incident occurrence.

There are different possibilities for using this data set to build the model. As shown in Table 5-5, one way is to use three different freeways data sets (IH-635 East and West and US-75 Northbound) to train and validate the incident detection model, and then use the US-75 Southbound data set to detect incidents. In this case, the training data contains 1384 instances with 417 incident cases.

Table 5- 5 Training SVM data using IH-635E/w and US-75N data

```
------Training Data------

----------Actual---------
 Category          Count
-------------------------
 Non-Incident -1     967
 Incident      1     417
-------------------------
 Total              1384
```

Geometric features of freeways change from freeway to freeway, and those geometric features impact speed distributions, therefore these features are already present in the data in implicit ways. Other freeways with different geometric characteristics may have different distributions, however, since our approach includes available data from two different freeways, some freeway-to-freeway variability has been included. IH-635 is a typical loop freeway (East-West), while US-75 is typical radial freeway and a major connector of downtown Dallas to North Dallas.

From a general perspective, these two classes of freeways bring the implicit impact of geometric features, driver behavior, and vehicle type into account in the model validation. There is no evidence that these freeways are not typical of United States freeways.

The other approach with the available data would be to use only one freeway dataset (US-75 Southbound, Table 5-6) to train and validate the models. After finding the best model, the trained model could be used to detect incidents in other freeways separately.

Table 5- 6 Training SVM data using US75S

------Training Data------

----------Actual---------

Category		Count
Non-Incident	-1	1614
Incident	1	384
Total		1998

The process of building the model and its future application is presented in Figure 5-2.

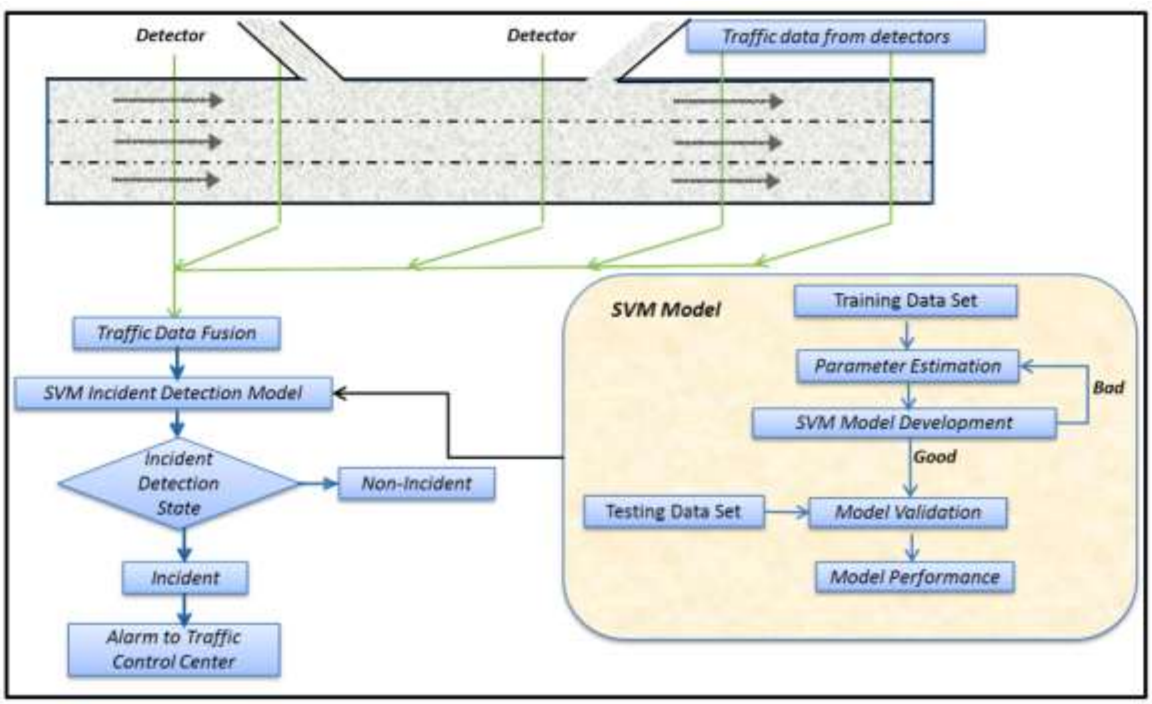


Figure 5- 2 Incident Detection Framework

First Model Analysis

After preparing and normalizing the data, the RBF kernel function was applied to build the model. A detailed description of how to build the model was presented in the previous section. RBF was selected as the kernel function, and different scenarios were tested to find the best-fit model. The search space was defined as [C, γ]= [2^{-5}: 2^{2}: 2^{6}, 2^{-15}: 2^{2}: 2^{4}]. To find the optimal parameter, first a grid search was used to find the optimal region, and then, a pattern search was performed to find the global optimum. To maximize the use of training data, cross-validation was applied. The objective is to minimize the total error for each model. The best model developed by C-SVM was compared with the Nu-SVM model (Table 5-7).

Table 5- 7 SVM different scenarios description and output

Scenario	Model Characteristics (Folds: search, validation)	Parameter Value	No. of Support Vector	Model Accuracy (%)	False Alarm Rate (FAR)
1	C-SVM, RBF, CV Folds: 10, 4	e: 0.001 C = 8 Gamma = 10.38998	355	Training: 94.44 Validation: 92.12	Training: 2.38 % Validation: 3.18%
2	C-SVM, RBF, CV Folds: 6, 4	e: 0.001 C = 8 Gamma = 10.38998	355	Training: 94.44 Validation: 91.91	Training: 2.38% Validation: 3.40%
3	C-SVM, RBF, CV Folds: 6, 6	e:.001 C = 1.10717 Gamma = 59.2693	675	Training: 95.59 Validation: 92.05	Training: 2.17% Validation: 2.96%
4	Nu-SVM, RBF, CV Folds: 6, 6	e: 0.001 Gamma = 28.79 Nu = 0.2282	488	Training: 94.8 Validation: 92.41	Training: 2.46% Validation: 3.03%

Studying these models revealed that the execution times were dependent upon the number of cross-validation folds mostly for finding model parameters. Interestingly, the number of support vectors increases when the number of folds increases for a parameter search. Nu-SVM can reduce training and execution time by one third (Scenario 4 vs. 3).

For a better comparison measure, the performance index (PI) proposed by Chen et al. [31] was used as presented in Table 5-8. The lower the PI, the better the model performs.

Table 5- 8 Performance Table

Scenarios	Model Characteristics (Folds: search, validation)	DR (%)	FAR (%)	MTTD (min)	CR (%)	PI
1	C-SVM, RBF, CV Folds: 10, 4	84.41	4.55	5	92.12	1.735
2	C-SVM, RBF, CV Folds: 6, 4	84.41	4.86	6	91.91	1.872
3	C-SVM, RBF, CV Folds: 6, 6	83.45	4.24	5	92.05	1.635
4	Nu-SVM, RBF, CV Folds: 6, 6	84.89	4.34	5	92.41	1.665

The performance table shows that six folds for both validation and parameter search produced the best performing models (Scenario 3 and 4). However, C-SVM performs slightly better than Nu-SVM, possibly because Nu-SVM parameters are more bounded (0<Nu<1 and C>0). In this performance measure, equal weight is used for all parameters. However, the parameters may be weighted differently depending on the type of application and its priority. For instance, in one case FAR might be the target to be minimized, and then, for that case, more weight should be assigned to FAR.

First Model Test Result

The best way to check the accuracy of a developed model is by using it on a new freeway (unseen dataset). For this part of the study, US-75 southbound was used as the unseen dataset. The incident data set is separated by collection date into two groups as shown by Table 5-9. Here we used

Table 5- 9 Testing SVM model using unseen data (US75S)

```
---------------Unseen Data-----------------

-----------------Count--------------------
 Category         US75S part1   US75S part2
-------------------------------------------
 Non-Incident -1    1614          976
 Incident      1    384           222
-------------------------------------------
 Total              1998          1198
```

The third scenario (with e: 0.001, C=1.10717, Gamma=59.2693) was chosen as the best model and the results are presented in Table 5-10.

Table 5- 10 SVM model evaluation results using unseen data

Testing with unseen data, Parameters:		e: 0.001, C=1.10717, Gamma=59.2693	
	Prediction accuracy (%)	False Alarm (%)	AUC
US75S 1	98.15	0.75	0.9967
US75S 2	99.25	0.25	0.9997

The results show the accuracy of the model is very high. The false alarm rate is below 1%. The area under the (ROC) curve (AUC) is another way to assess the overall performance of a classifier. In Table 5-10, AUC is very close to 1, implying very good results. This good result was expected because data from three different freeways were used to build the model. The only concern could be that only one freeway was used for testing. Therefore, to examine the other extreme, only US-75 southbound was used to generate the model.

Second Model Analysis

For this part of study, a different scenario is tested to improve the best model. The C-SVM model and RBF kernel function are present in all scenarios. Using training data from only one freeway, the number of folds for cross validation, cross validation versus random-holdback rows, and optimization of the number of misclassifications (MC) were examined.

Utilizing random-holdback rows is one way to examine model validation. A randomly selected percentage of data was held out from the model building process for model validation. Different percentages were examined, and the percentage with the best fit is presented in Table 5-11 (scenario 3).

Table 5- 11 SVM scenario descriptions and output for US75S data set

SVM US75S All variable					
Scenario	Model characteristics	Parameter Value	No. of Support Vectors	Model Accuracy (%)	False Alarm
1	C-SVM, RBF, CV folds: 10, MC: Equal	e:.001,C=19.1355, Gamma=2.7604	285	Training: 95.6, Validation: 94.79	Training: 1.85%, Validation: 2.15%
2	C-SVM, RBF, CV folds: 6, MC: Equal	e:.001, C=0.03125, Gamma=60.722	986	Training: 96.10, Validation: 93.49	Training: 1.25%, Validation: 1.8 %
3	C-SVM, RBF, Random Sampling (20%), MC: Equal	e:.001, C=109.115, Gamma=1.3195	231	Training: 95.24, Validation: 95.00	Training: 2.07%, Validation: 3.00%
4	C-SVM, RBF, CV folds: 6, MC: minimize total error	e:0.001, C=388.0234, Gamma=1.3195	267	Training: 96.25, Validation: 94.84	Training: 2.1%, Validation: 2.95%

When there is a large difference between the number of data rows in two categories, model training is more strongly effected by the category with more rows. The incident category with fewer training rows has a higher level of misclassification error. In previous models, misclassification cost was applied equally to both categories. The probability threshold that minimizes the total misclassification error is considered as one scenario (scenario 4).

The results show that model accuracies for all scenarios are similar; the distinct difference is false alarm rate and number of support vectors. Decreasing the number of folds on cross validation increases the number of support vectors and decreases the false alarm rate, as shown by Table 5-12 (scenario 1 versus 2). On the other hand, applying a threshold probability to minimize total misclassification error reduces the number of support vectors and increases the false alarm rate (scenario 2 versus 4).

Table 5- 12 Evaluation of different scenarios

Scenarios	Brief details	DR(%)	FAR(%)	MTTD (min)	CR(%)	PI
1	C-SVM, RBF, CV folds: 10, MC: Equal	84.11	2.66	-1	94.79	1.108
2	C-SVM, RBF, CV folds: 6, MC: Equal	75.52	2.23	-1	93.49	1.025
3	C-SVM, RBF, Random Sampling: 20%, MC: Equal	89.61	3.72	-1	95.00	1.440
4	C-SVM, RBF, CV folds: 6, MC: Minimize total error	88.54	3.66	-1	94.84	1.423

Based upon the performance index (PI) proposed by Chen et al. [31], the second scenario with equal misclassification is the best-fit model. The reason that scenario four does not perform better is minimizing the total error in this case means giving more priority to incident cases (minority), which increases the detection rate, but would increase the false alarm rate, as well. However, these results are based upon application of equal weights to all PI terms, so one can chose a weight combination that best suits one's situation.

Second Model Test Result

The data set to test the model developed by US-75 south bound is presented in Table 5-13, and the results are presented on Table 5-14.

Table 5- 13 Testing SVM model using unseen dataset

Unseen Data

Count					
Category		US75S part2	US75N	I635E	I635W
Non-Incident	-1	976	444	264	259
Incident	1	222	176	123	118
Total		1198	620	387	377

Table 5- 14 SVM model evaluation results using unseen data

Testing with unseen data			
e:.001, C=0.03125, Gamma=60.722			
	Prediction accuracy (%)	False Alarm (%)	AUC
US75S part2	98.91	0.67	0.9981
I635E	94.32	1.55	0.9891
I635W	95.49	1.06	0.9887
US75N	92.2	2.82	0.9665

First the model was examined by applying a different data set from the same freeway (US-75 SB). As was expected, it has the highest prediction accuracy and lowest false alarm rate. This result confirms the desirability of model examination on multiple freeways instead of only one.

The accuracy level for all chosen models is above 90% with a low false alarm rate. The Area Under the (ROC) Curve (AUC) is used to assess the overall classifier performance. The closer AUC is to one, the better the classifier. As shown in the Table 5-14, the AUC values averaged 0.9856, which is considered "very good." The question we needed to answer in this study was "Can the developed model be used for any freeway?" The answer to the question seems to be that there is no evidence to indicate it is not transferable. Because 3 months of traffic data and incident cases across different times at different locations on different freeways were used to develop the model, one can claim that this model is generalized and could be transferable.

Conclusion

During the course of this research study, SVM models were developed and applied to the freeway incident detection problem. Evaluation of the SVM algorithm revealed that SVM can successfully classify traffic conditions into incident and non-incident categories during peak hours and off peak hours. In the context of real-time incident detection, this model prediction is faster than the previously utilized models.

These models were trained and developed based on data from a freeway segment in Dallas, TX. More diverse and extensive data sets were tested for this study, as compared to the previous study, and the results show that the SVM incident detection technique presented here is promising.

PART 3 GENERALIZED REAL-TIME INCIDENT DETECTION MODEL USING SVM

The benefits of implementing automated incident detection (AID) are broadly accepted. However, the transferability of the model remains in question. In this part of study, we extend evaluation to additional test sites to strengthen confidence in the performance of the generalized incident detection model. Many thanks are given to Prof. Mohammed Hadi and Research Assistants Tao Wang and Aidin Massahi (Florida International University, USA) for providing the Miami dataset from FDOT, which allowed us to achieve the goal for this part of study. Data collected on two Dallas, Texas freeways (US-75 and I-635) and one Miami, Florida freeway (I-95 north bound) are used to test the algorithm. The new model was developed using loop and radar detectors (a summary of the salient characteristics of the study site is presented in Table 5-15). The complete detail of dataset is presented at chapter 3. The following table shows the summary of incident cases have been used for this part.

Table 5- 15 Study sites and incident data

Freeway	Direction	Length (mi)	No. of Detectors	No. of Incidents	Total No. of Data point for the Model
Dallas - US 75	Northbound	10	20	19	622
Dallas - US 75	Southbound	10	20	63	1811
Dallas – I 635	Eastbound	7	13	12	387
Dallas – I 635	Westbound	7	13	12	377
Miami – I 95	Northbound	4.5	14	20	706
Total		**38.5**	**80**	**126**	**3903**

Model Analysis and Testing

The model was trained using the Dallas dataset. In this SVM model, each data point corresponds to 5 minutes. After preparing and normalizing the data, the RBF kernel function was applied to build the model (for a detailed description of how to build the model, please see Model Development section in Chapter 4). To find the optimal parameter values for the model, a grid search and cross-validation were applied in the search space recommended by Lin et al. and Ma et al. (36,48) with the objective of minimizing the total error.

PREDICTION PERFORMANCE AND COMPARISON

Different scenarios were considered to examine the sensitivity of incident detection model development with regard to different study sites. For the purpose of examining transferability of the models, commonly used random separation of testing from training was not employed here. Since the data is collected from multiple freeways, cases from one specific freeway were completely removed from training data and were held for testing purposes. The first scenario does not include data from a different city. However, in all scenarios, the dataset includes cases from multiple freeways. In this case, cases from US75-S were held out for testing. The second scenario is using the same training data as first scenario but the testing data is from different city or state. The comparison of these two scenarios shows how the same model response to different texting data, aside from the transferability of the model. The third, training and validating the model with the Dallas data, was applied directly to another freeway in Miami to test the transferability of the model. This scenario shows the sensitivity of the model to the size of training data. In this scenario, all the cases from Dallas were implemented to build the model and the cases from Miami were used for testing the reliability of the incident detection model. The results of these scenarios are summarized in Table 5-16.

Table 5- 16 Generalized SVM incident detection model testing

Scenario	Training Dataset	Testing Dataset	DR	FAR	CR	MTTD
1	US75-N, I635-E, I635-W (Dallas)	US75-S (Dallas)	85.3%	22.9%	78.6%	-1min
2	US75-N, I635-E, I635-W (Dallas)	I95-N (Miami)	84%	19%	81.5%	-1 min
3	US75-N, I635-E, I635-W, US75-S (Dallas)	I95-N (Miami)	88.9%	6.8%	90.2%	-1.11 min

Comparing the results from scenario 1 and 2, the same training dataset implemented on different study sites, illustrates consistent results. The detection rate (DR) and the false alarm rate (FAR) in both scenarios are quite similar. In scenario 3, the

FAR is markedly less than in the other models. Another observation from the models 'comparison is that the developed model is more sensitive to the size of the training dataset (Scenario 2 vs. Scenario 3). The CR is also enhanced by increasing the size of training dataset.

The results from these scenarios indicate that the SVM incident detection model is stable and capable of maintaining good detection performance across different test sites.

Sensitivity analysis of the SVM model

In machine learning, sensitivity analysis is commonly applied to evaluate the relationship between features and output. The trained SVM model was used to test the sensitivity of traffic state (incident/non-incident) to the changes in one input at the time. For each feature, the changes of traffic state were observed by varying the value of that feature while all other features were unchanged. The sensitivity value, bounded between 0% and 100%, is more significant as the value approaches 100%. The results of sensitivity analysis for scenario 3 are illustrated in Table 5-17 where the most significant factor affected by incident is upstream speed.

Table 5- 17 Importance of features based on sensitivity analysis

	Importance of features %
Upstream Speed	100.00
Upstream Occupancy	64.71
Upstream Volume	60.30
Downstream Volume	44.35
Downstream Occupancy	27.64
Downstream Speed	0.00

Further, features related to upstream have more significant impact on detection incidents. A possible explanation is that during an incident more changes happen on traffic features upstream than downstream.

DISCUSSION AND CONCLUSION

Part of the challenge in developing an incident detection model using only traffic characteristics includes having to leave out potentially important environmental and vehicle factors. This can partially account for unexplained parts of the incident detection models. However, trying to account for environment and vehicle factors can conflict with the more important objective of model transferability for traffic control centers.

This study's goal was to utilize traffic characteristics to develop a transferable model that overcomes the problem of incident detection during peak hours and stop-and-go traffic. Data were collected from five freeway sections from two different states: Texas, Florida. The Dallas data trained the model and its transferability was tested by applying the model to Miami. The best performing SVM model developed with Dallas data can predict 68% of incidents on I-90 North in Miami with a 4.8% false alarm rate. Overall, the SVM incident detection technique presented shows promising results.

Chapter 6: An Ensemble Model with Temporal Data Development and Comparison to Other Models

INTRODUCTION

Research discussed in this chapter was conducted in two parts. First, the incident detection model was developed using temporal data. This part of the study investigates the impact of temporal data on the incident detection model and examines the importance of significant factors associated with incident state. The major contribution of the proposed incident detection method here is incorporating both spatial and temporal information in the detection model. Most incident detection models elaborate on either special information from road detectors over the course of one time interval (i.e. AI incident detection models) whereas incident state can be induced by disturbance of traffic flow. Multiple time interval time series traffic data were adapted to address this issue. Second, incident detection models were developed using competing techniques –e.g. Naïve Bayes, Random Forest, and SVM Ensemble – with and without temporal data.

TEMPORAL DATA STRUCTURE

Temporal data with multi-time-interval traffic data was used to capture the time influence on the model. Moreover, random variables that represent the incident state are in time sequence. Since the time intervals could be important influencing factors, finding the optimal influencing period of time before incident start needed to be determined. To answer this question, several time intervals before the incident start were applied in the incident prediction model. Table 6-1 presents the structure of using time interval t, which is the interval in question (0-5 minutes prior to t), time interval t-1 (5-10 minutes prior to t), and time interval t-2 (10-15 mins prior to t).

Table 6- 1 Data structure for temporal model

Time Step	Downstream			Upstream			Traffic State
	Speed	Volume	Occupancy	Speed	Volume	Occupancy	
t (0-5 min interval in question)	S11	V11	O11	S21	V21	O21	yi
t-1 (5-10 min prior)	S12	V12	O12	S22	V22	O22	----
t-2 (10-15 min prior)	S13	V13	O13	S23	V23	O23	----

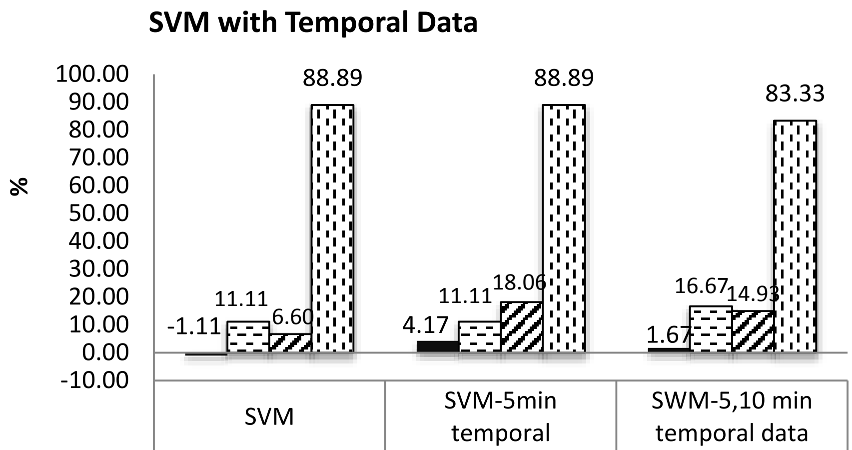
In Table 6-1, Sij is speed at location i in time step j (1 represents downstream and 2 upstream in time step j), Vij is volume at location i in time step j, and Oij is occupancy at location i in time step j.

SVM WITH TEMPORAL DATA

This part of the research attempted to find the effect of adding several time intervals to the SVM detection model. The proposed approach includes spatial-temporal data mining using the SVM algorithm. It should be noted that the reference base model is the currently presented SVM model using data collected in time interval t.

Prediction Performance and Comparison

Three scenarios were defined to examine the usefulness of including temporal data in the SVM incident detection model. For all scenarios, the training dataset and testing dataset were the same. The scenarios' differ in how the input data points were structured. The first scenario is the base model proposed in the last part of the Chapter 5. This model was built using the Dallas dataset and tested on the Miami dataset. The associated test results are presented in Figure 6-1. The second scenario was built using time step t and t-1 (interval in question and prior interval, respectively). Finally, the third scenario was built using t, t-1, and t-2. The bar charts provide a more intuitive understanding of the various results. The base scenario is the single SVM model without temporal data.

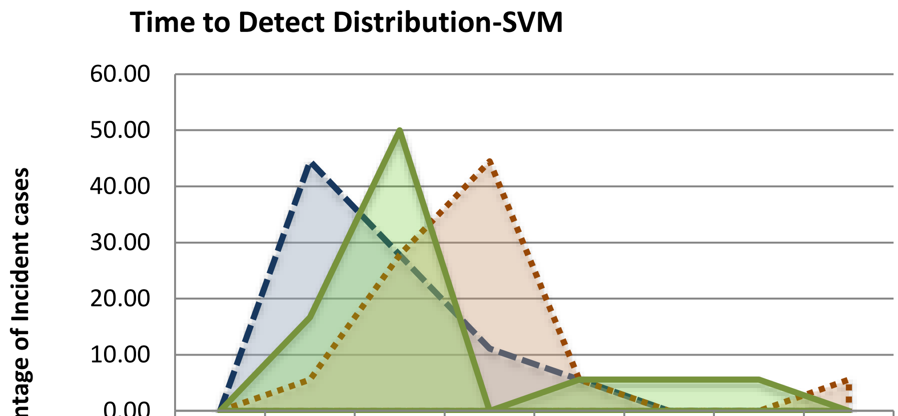


	SVM	SVM-5min temporal	SWM-5,10 min temporal data
Mean Time to Detect (MTTD) min	-1.11	4.17	1.67
False Negative (Percentage of Incident has not being detected by model) %	11.11	11.11	16.67
False Alarm Rate (FAR) %	6.60	18.06	14.93
Detection Rate (DR) %	88.89	88.89	83.33

Figure 6-1 Performance of SVM models with/without temporal data

Figure 6-1 shows that detection rate (DR) and mean time to detect (MTTD) slightly changed across the three models. However, the false positive (FAR) vary significantly across the models. The base SVM model tends to detect incidents on average 1.11 minutes before incident occurrence.

As mentioned earlier, each data point (here traffic information resolution was every 5 minutes) was classified. Sometimes incident detection models are able to predict incidents earlier/later than actual time, which is the reason why MTTD was considered as a performance measure. However early/later detection times are related to higher FAR/False Negative percentages. For example, if we decrease the MTTD, then the trade-off is FAR increases, because when an incident alarm goes off and the incident occurs in the next time interval, it is considered a false alarm. To further investigate this problem, time to detect distribution is presented in Figure 6-2.



	10 min before	5 min before	Exact time	5 min later	10 min later	15 min later	20 min later	30 min later
SVM	0.00	44.44	27.78	11.11	5.56	0.00	0.00	0.00
SVM-5min temporal	0.00	5.56	27.78	44.44	5.56	0.00	0.00	5.56
SWM-5,10 min temporal data	0.00	16.67	50.00	0.00	5.56	5.56	5.56	0.00

Figure 6-2 SVM Time Performance: Time to Detect Distribution

To comparatively evaluate models against each other, all are presented in same graph. Figure 6-2 clearly shows the advantage of using prior information. The goal is to have the model detect the incident at the exact time the incident occurs (i.e., MTTD=0). The SVM model with 5, 10 min temporal data is the model most centered on the exact incident occurrence interval, which implies that including more temporal data can increase the reliability of the model.

Sensitivity analysis of the SVM models

Sensitivity analysis was implemented to evaluate the relationship between the predictors and dependent variable. According to the sensitivity analysis (Table 6-2), upstream speed (i=2) at time step t (time interval in question), was the most important factor affecting incident detection across all models. Another insight from Table 6-2 is that for the temporal models, the upstream speed (i=2) at other time slices (j=2,3) is the second most important variable. A possible explanation is that speed variation happens quickly upstream of an incident, therefore observance speed over the time provides the best information to the incident detection model.

After speed upstream of incident location, occupancy and volume at the upstream location were the next important variables (65-60 % range).

Table 6- 2 Importance of predictors from sensitivity analysis

SVM Importance (%)		**SVM with 5 min temporal Importance (%)**		**SVM with 5,10 min temporal Importance (%)**	
S21	100	S21	100	S21	100
O21	64.71	S22	65.95	S23	74.45
V21	60.3	O21	62.73	S22	67.47
V11	44.35	V21	6218	O21	64.7
O11	27.64	O22	48.78	V21	61.24
S11	0	V11	42.88	O22	49.1
		V22	35.18	O23	48.94
		O11	25.56	V23	48.74
		V12	23.78	V11	45.87
		S12	6.65	V13	39.24
		O12	5.05	V22	37.81
		S11	0	O11	28.87
				V12	27.53
				O13	23.31
				S12	9.2
				O12	8.39
				S11	0.26
				S13	0

NAÏVE BAYES CLASSIFIER WITH TEMPORAL DATA 5, 10 MINUTES

In machine learning, Naïve Bayes (NB) classifiers are simple and powerful probabilistic classifiers. NB assumes features are independent, which simplifies the model drastically. (This assumption is often not valid.) It became popular because of its simplicity, good results, and low computational cost (fast). However, this simplification can cost accuracy reduction.

Naïve Bayes Classifier Methodology

NB can handle both continuous and categorical variables. Assume $X = [x_1, x_2, \dots, x_n]$ are independent variables and c_1, c_2 are possible outcomes (incident and non-incident). Given the variables $x_1, x_2, \dots, x_n$, we want to predict the posterior probability $p\left(C_j \middle| x_1, x_2, \dots, x_n\right)$ of outcome c_1, c_2. Based on Bayes rule:

$$p\left(C_j \middle| x_1, x_2, \dots, x_n\right) \propto p\left(x_1, x_2, \dots, x_n \middle| C_j\right) * p(C_j),$$

where $p(C_j)$ is the prior probability of each outcome, which can be estimated from a training dataset. The likelihood of x to have a particular outcome c_1 / c_2, or $p\left(X \middle| C_j\right)$, must be multiplied by the prior probability in order to get the posterior probability, $p\left(C_j \middle| x_1, x_2, \dots, x_n\right)$. The robustness of the classification is highly tied to the accuracy of $p\left(X \middle| C_j\right)$ estimation, which is *not* straightforward when there is correlation between the variables. However, the naïve Bayes variables independency assumption simplifies the complexity to some extent by allowing us to decompose the likelihood of the product terms to

$$p\left(X \middle| C_j\right) \propto \prod_{i=1}^{n} P(x_i \mid C_j).$$

This requires estimation of conditional probability for each feature in each class, thus eliminating joint probability calculations. The naïve Bayes employs these two estimates to predict the label.

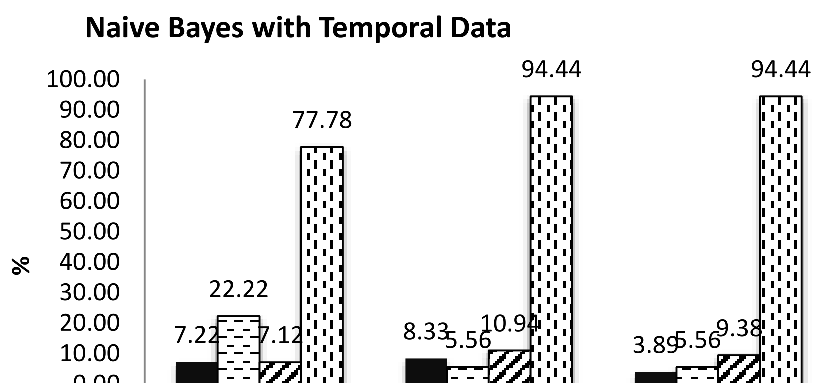
$$p\left(C_j \mid X\right) \propto p\left(C_j\right) * \prod_{i=1}^{n} P(x_i \mid \square_j)$$

Using this equation, Naïve Bayes will calculate the posterior probabilities so that a new case x will be assigned to the class C_j with the highest posterior probability.

Naive Bayes can be modeled in several different ways including normal (Gaussian), lognormal, gamma, and Poisson density functions. An "R" package for modeling Naïve Bayes has been used, which implemented a Gaussian function for modeling.

Prediction Performance and Comparison

The same training dataset (Dallas dataset) was used to build the NB model as a baseline case for this part of study. Two more scenarios were considered to evaluate the effect of temporal data on the accuracy of the incident detection model. The results are presented in Figure 6-3.

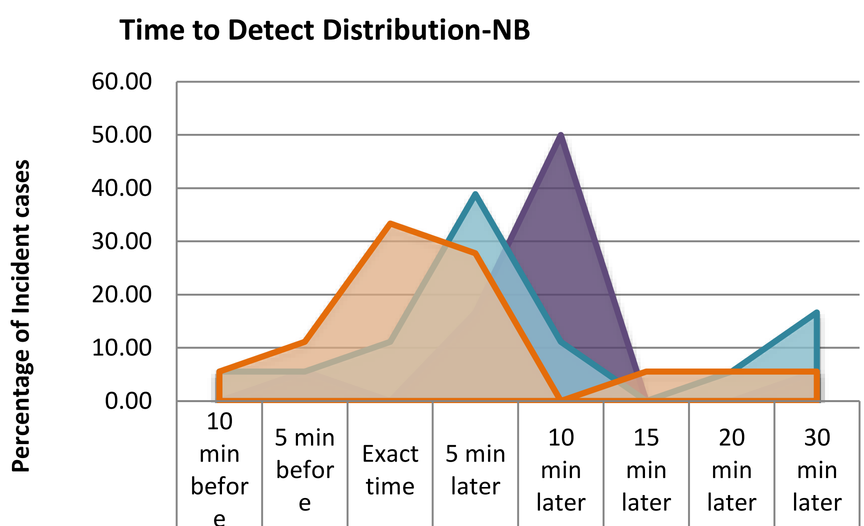


	Naïve Bayes	Naïve Bayes-5min temporal data	Naïve Bayes-5,10min temporal data
■ Mean Time to Detect (MTTD) min	7.22	8.33	3.89
False Negative (Percentage of Incident has not being detected by model) %	22.22	5.56	5.56
False Alarm Rate (FAR) %	7.12	10.94	9.38
Detection Rate (DR) %	77.78	94.44	94.44

Figure 6- 3 Performances of NB models with/without temporal data

Use of temporal data showed significant increase in detection rate (DR) for Naïve Bayes. Observing the false negative rates also indicates improvement on model accuracy. Lower detection time is expected with the use of 5 and 10-minute time steps. However, models with temporal data were expected to have slightly higher false alarm rates.

The results perceived from time to detect distribution (TTDD) in the previous section motivated its addition as part of model performance measurement (Figure 6-4). The time to detect distribution graph illustrated that the Naïve Bayes model tends to detect incidents 10 min after the actual incident start time. Using temporal data shifts the detection time toward the exact time that the incident starts. As for SVM models with temporal data, using 5 and 10 minute temporal data helped center the model detection time on the exact time of incident.



	10 min before	5 min before	Exact time	5 min later	10 min later	15 min later	20 min later	30 min later
Naïve Bayes	0.00	5.56	0.00	16.67	50.00	0.00	0.00	5.56
Naïve Bayes-5min temporal data	5.56	5.56	11.11	38.89	11.11	0.00	5.56	16.67
Naïve Bayes-5,10min temporal data	5.56	11.11	33.33	27.78	0.00	5.56	5.56	5.56

Figure 6-4 NB Time Performance: Time to Detect Distribution

Sensitivity analysis of the NB models

The results from sensitivity analysis for the importance of the variables are presented in Table 6-3. The results from this analysis were consistent with SVM sensitivity analysis. In all the models upstream speed at the current time step was the most significant variable.

Table 6-3 NB Importance of predictors from sensitivity analysis

NB Importance (%)		NB with 5 min temporal Importance (%)		NB with 5,10 min temporal Importance (%)	
S21	100	S21	100	S21	100
O21	63.58	S22	73.31	S23	75.49
V21	59.39	O21	69.61	S22	69.48
V11	42.03	V21	64.29	O21	66.61
O11	24.42	O22	57.97	V21	60.242
S11	0	V11	50.02	O22	52.61
		V22	42.47	O23	50.86
		O11	35.51	V23	46.36
		V12	32.66	V11	44.68
		S12	20.39	V13	36.25
		S11	13.45	V22	35.24
		O12	0	O11	29.31
				V12	24.37
				O13	21.41
				O12	8.93
				S12	7.95
				S13	1.44
				S11	0

RANDOM FOREST WITH TEMPORAL DATA

Random Forest, another competitive technique in the field of incident detection, was performed and compared to the SVM model. Random Forest is an ensemble classification method that was first proposed by Breiman [68]. His goal was to find out whether the infinity theory applies to the practical issues of a finite dataset. He determined that Random Forest has a strong tie to the infinity behavior that gives the results one would expect from the infinity analysis. Random Forest constructs many decision trees made up of random subsets of the data. A trained forest can accept variables (speed, occupancy) and report a prediction (incident, non-incident). Liu et al. (2013) implemented Random Forest for incident detection and claims that Random Forest is a superior incident detection model and that it can improve incident detection performance [70].

An advantage of Random Forest (RF) classification , an ensemble learning technique, is that it does not require parameter tuning to obtain unbiased error estimates. For every AI model, some optimization search is required to find optimal parameter values or thresholds which are resolved through the ensemble technique. The ensemble idea helps with finding better results by combining many models. Each tree in the forest depends on the value of random vector sampled independently. RF improves classification performance through a voting process. To classify a new object from an input vector, each tree gives a vote and the forest chooses the class with highest votes.

Random Forest Methodology

The simple Random Forest is a classifier consisting of multiple tree-structured classifiers $\{h(X, \Theta_k), k = 1, \dots\}$ where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree performs a vote for the most popular class at input x [69]. A small group of features split each node. To grow the tree, CART, a binary recursive partitioning methodology, is used in which each parent node only splits into two child nodes and the process repeats. The key elements of CART analysis are to set splitting, deciding, and assigning to class rules. The Gini rule would be applied to look for the best split among all possible splits for all features in the model. For more information on Gini rule, refer to Berzal et. al, 2003 [72]. Then, we repeat the search until further splitting is not possible. In this method, the number of features included in the model is fixed and randomly selected [68].

The aim of the objective function is to maximize the margin. The larger the margin, the more confident is the classification from the forest. Given an ensemble of classifiers $h_1(x), h_2(x), \ldots, h_k(x)$, the margin function is [69]:

$$mr(X,Y) = P_{\Theta}(h(X,\Theta) = Y) - max_{i \neq Y} P_{\Theta}(h(X,\Theta) = j),$$

where X and Y are random vector training data samples, and P_{Θ} is the indicator function. If mr(X,Y) > 0, then the classifiers vote "correct", and if mr(X,Y) < 0, then the set of classifiers vote "incorrect".

The margin measures the extent that the average number of votes at X,Y for the right class exceeds the average vote for any other class. The generalization error is given by [69]:

$$PE^* = P_{X,Y}(mg(X,Y) < 0)$$

where the subscripts X,Y indicate that the probability is over the X,Y space.

Two class models can be simplified as [69]:

$$mr(X,Y) = 2P_{\Theta}(h(X,\Theta) = Y) - 1$$

and the strength (or accuracy) of the classifiers $\{h(X,\Theta)\}$ is [69]:

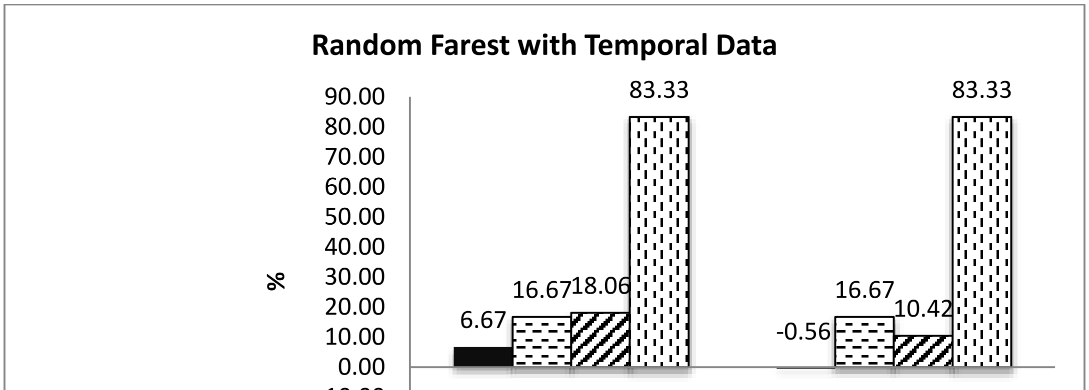
$$s = E_{X,Y}\, mr(X,Y).$$

In order to find the optimum number of trees, different models with various tree numbers should be examined. The best model is the one with steadiest error rate for the lowest number of trees (69).

Prediction Performance and Comparison

In the interest of fair comparison, the data was prepared in the same way as it was for the SVM model. The variables were normalized by their mean and standard deviations. Grid searches were implemented to find the best number of trees and number of variables. First, the incident detection model was developed using only the time slice t in question. Then, 5 and 10 minute temporal data were added to the model. The results from these models are illustrated in Figure 6-5. The result showed that the Random Forest incident detection model on average tends to detect incidents 6 mins after the incident happens. Even though temporal data did not make any changes on the detection rate and false negative rate (number of incidents not detected), better measures of FAR and MTTD were achieved. These results suggest using temporal data improved the accuracy of the model by reducing false alarm rate and produced faster detection time.

As in the previous technique evaluations, the distribution of time to detect was developed and is presented on Figure 6-6. The time to detect distribution showed that there is a time shift between the Random Forest model without temporal data and with temporal data. The use of temporal data on Random Forest helped the model detect incidents around the exact incident start time.



	Random Forest	Random Forest-5,10min temporal data
Mean Time to Detect (MTTD) min	6.67	-0.56
False Negative (Percentage of Incident has not being detected by model) %	16.67	16.67
False Alarm Rate (FAR) %	18.06	10.42
Detection Rate (DR) %	83.33	83.33

Figure 6-5 Performances of RF models with/without temporal data

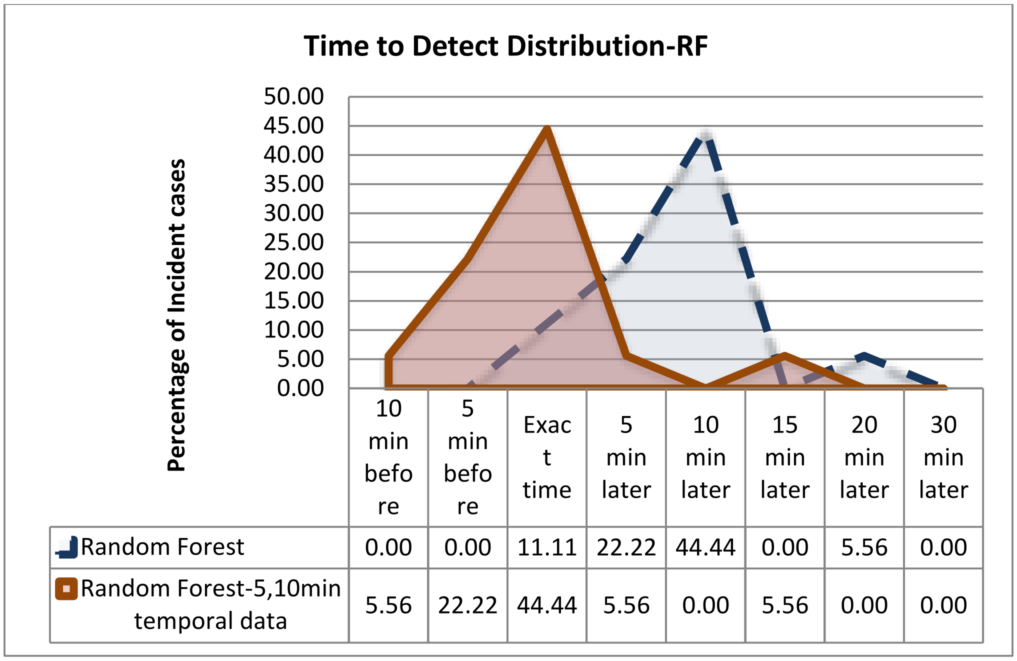


Figure 6-6 RF Time Performance: Time to Detect Distribution

Sensitivity analysis of the RF models

Like the previous technique, the importance of the observed variables was evaluated by implementing sensitivity analysis. The results showed that the upstream speed at the current time step is still the most correlated variable to describe incident occurrence. Although, the order of the variables may be the same as the SVM or NB models with slight changes, but the percentage values were not the same, which could result from the nature of the Random Forest split rule.

Table 6-4 RF Importance of predictors from sensitivity analysis

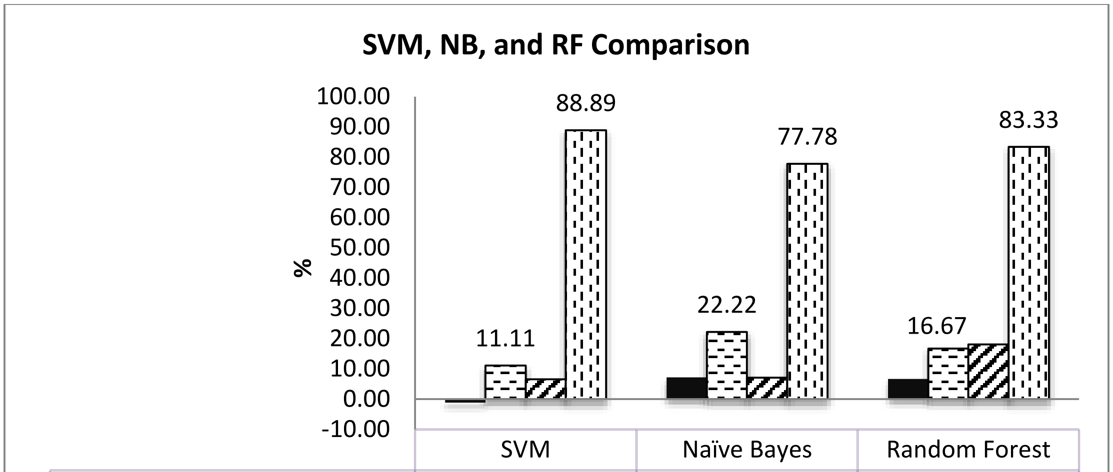
RF Importance (%)		RF with 5,10 min temporal Importance (%)	
S21	100	S21	100
O21	32.14	O21	31.55
V21	25.76	S23	30.74
V11	10.23	V21	24.46
S11	7.82	S22	17.76
O11	0	V11	12.97
		S11	11.76
		V23	8.21
		O11	8.17
		O23	7.85
		O22	7.09
		V13	6.87
		S13	5.72
		V22	5.2
		O13	4.25
		V12	4.04
		S12	2.88
		O12	0

COMPARISON OF SVM, NAÏVE BAYES, AND RANDOM FOREST INCIDENT DETECTION MODELS

The result from SVM, Naïve Bayes, and Random Forest were separately discussed in previous sections. The same measures of performance (MTTD, false negative, FAR, and DR) were used to compare these incident detection models.

The preliminary investigation of the mean time to detect incident (MTTD), indicated the SVM model was able to detect incidents on average 5 minutes before it actually happened, whereas both Naïve Bayes and Random Forest detected the incident after some delay.

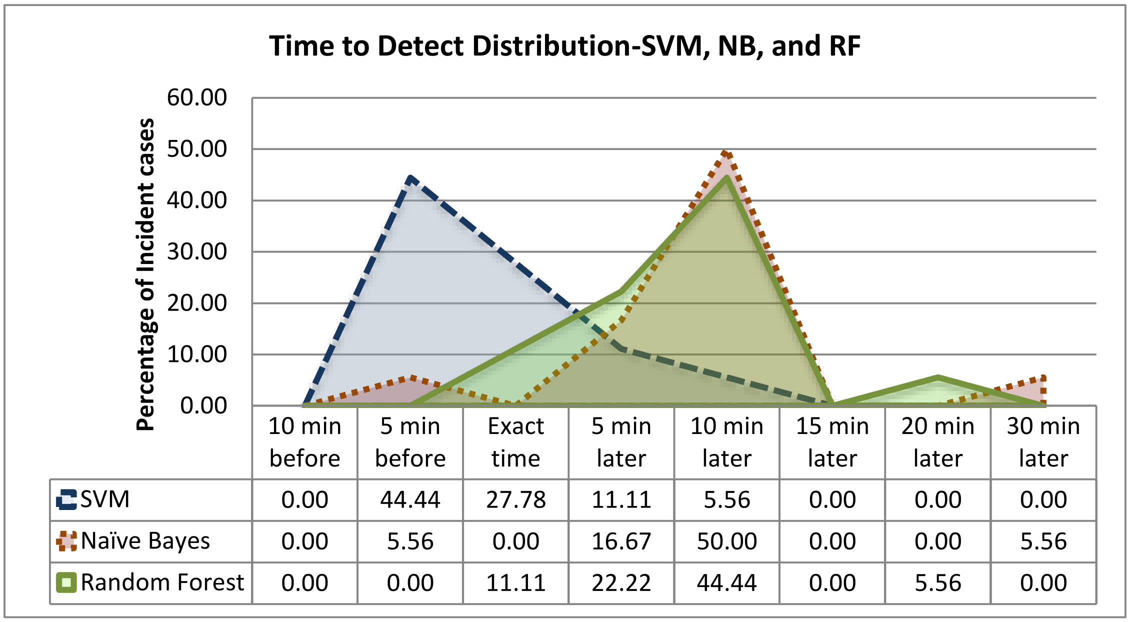
Often times the false negative percentage in incident detection is mistakenly calculated using the confusion matrix (Appendix A) from the result. Nevertheless that matrix contains all the early/ late detection false positive/negative cases. In order to address this, the predicted values must be compared to actual traffic conditions. The case study results showed that SVM has lower false positive/negative than the two other models.



	SVM	Naïve Bayes	Random Forest
Mean Time to Detect (MTTD) min	-1.11	7.22	6.67
False Negative (Percentage of Incident has not being detected by model) %	11.11	22.22	16.67
False Alarm Rate (FAR) %	6.60	7.12	18.06
Detection Rate (DR) %	88.89	77.78	83.33

Figure 6-7 Performances of SVM vs. NB/RF

The distribution of time to detect incident represented in Figure 6-8 showed that none of them were centered around the exact time of incident, however SVM detected nearest to the exact time, but early rather than late.



	10 min before	5 min before	Exact time	5 min later	10 min later	15 min later	20 min later	30 min later
SVM	0.00	44.44	27.78	11.11	5.56	0.00	0.00	0.00
Naïve Bayes	0.00	5.56	0.00	16.67	50.00	0.00	0.00	5.56
Random Forest	0.00	0.00	11.11	22.22	44.44	0.00	5.56	0.00

Figure 6-8 SVM, NB, and RF Time Performance: Time to Detect Distribution

ENSEMBLE MODEL

Many machine-learning techniques have been successfully applied in the field of automated incident detection. However, these techniques are intended for optimization problems and are often sensitive to the choice of optimal parameter values. To avoid the burden of finding the optimal parameter value and threshold and in order to improve the detection rate, the ensemble technique has been introduced. The idea of using an ensemble comes from the idea that combining many models can help to find better results. One of the earliest ensemble methods was proposed by Dasarathy et al. in which the study partitioned feature space using multi-classifiers [75]. In 1990, Hansen et al. showed that using an ensemble method improved the performance of neural networks [76]. This technique has been successfully applied to a wide variety of fields since then. Specifically, Chen et al. (2007) used a neural network ensemble model in the field of transportation to detect traffic incidents [77]. They found that using an ensemble model leads to better results than using a single model. A few years later, Chen et al. (2009) [33]

applied the SVM ensemble technique to create an improved incident detection model. They were able to achieved high detection rates with low false alarm rates, 88.7 and 1.57 respectively. However, they had only implemented the SVM ensemble technique on the I-880 dataset from 1993.

Ensemble Methodology

When creating an ensemble classifier, the training subsets must be selected. Each classifier requires a multi-subset in which each training subset should have the least possible common data points with the other subsets. There are quite a few methods to accomplish this goal, however, a resampling technique is often used. This study adopted bagging to obtain the training subsets. Bagging uses bootstrapping, which randomly draws a sample with replacement, to generate training subsets. If the training dataset is denoted by *S* and test dataset by *D*, then the training subset can be denoted as:

$$\{S_1, S_2, \dots, S_n\},$$

where S_i is training subset and n is the number of subsets or individual classifiers.

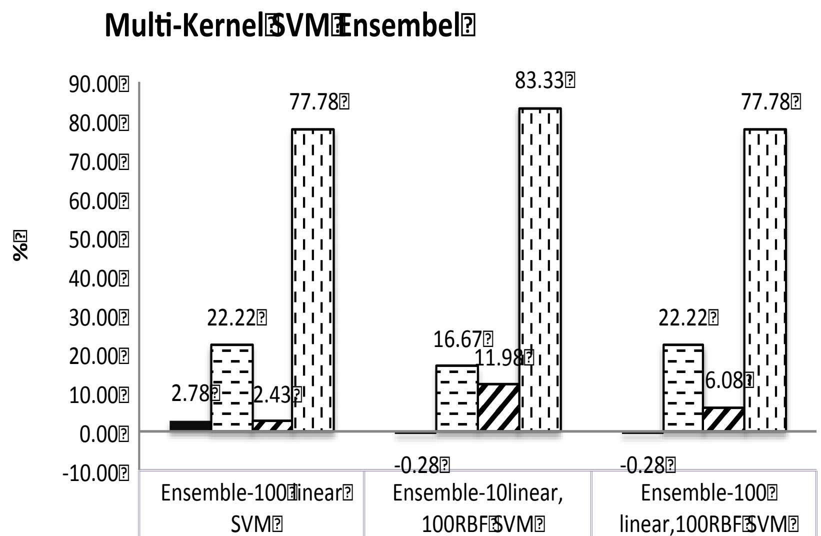
The output of the n individual SVM classifiers can be denoted as $\{y_1(x), y_2(x), \dots, y_n(x)\}$. To obtain the final decision of the ensemble model, this study applied an un-weighted voting method. If the $y_f(x)$ is the final decision of the ensemble model, then final output of the model will be:

$$y_f(x) = sgn\left(\sum_{i=1}^{n} y_i\,(x)\right)$$

This function indicates the sum of the total vote, negative means there is no accident and positive means incident alarm. The same performance measures as previous sections were implemented here as well.

Prediction Performance and Comparison

Multiple scenarios with different combinations of kernel functions were defined and tested for this part of study. The three best scenarios based on observations are presented in Figure 6-9. Comparison among the scenarios indicates that the SVM ensemble models tend to detect incidents around the exact time of incident. Evaluation of FAR indicated that the rate is highly sensitive to the type and number of kernel functions implemented. However, the detection rate and false negative rate were steady.



	Ensemble-100 linear SVM	Ensemble-10linear, 100RBF SVM	Ensemble-100 linear,100RBF SVM
Mean Time to Detect (MTTD) min	2.78	-0.28	-0.28
False Negative (Percentage of incident has not being detected by model) %	22.22	16.67	22.22
False Alarm Rate (FAR) %	2.43	11.98	6.08
Detection Rate (DR) %	77.78	83.33	77.78

Figure 6-9 Performances of SVM Ensemble Models

The time to detect distribution of ensemble models indicates the fact that implementing an ensemble method does in fact improve the detection time. From Figure 6-10, all scenarios were found to have almost the same distribution with the mean close to the exact time of incident and almost the same variance.

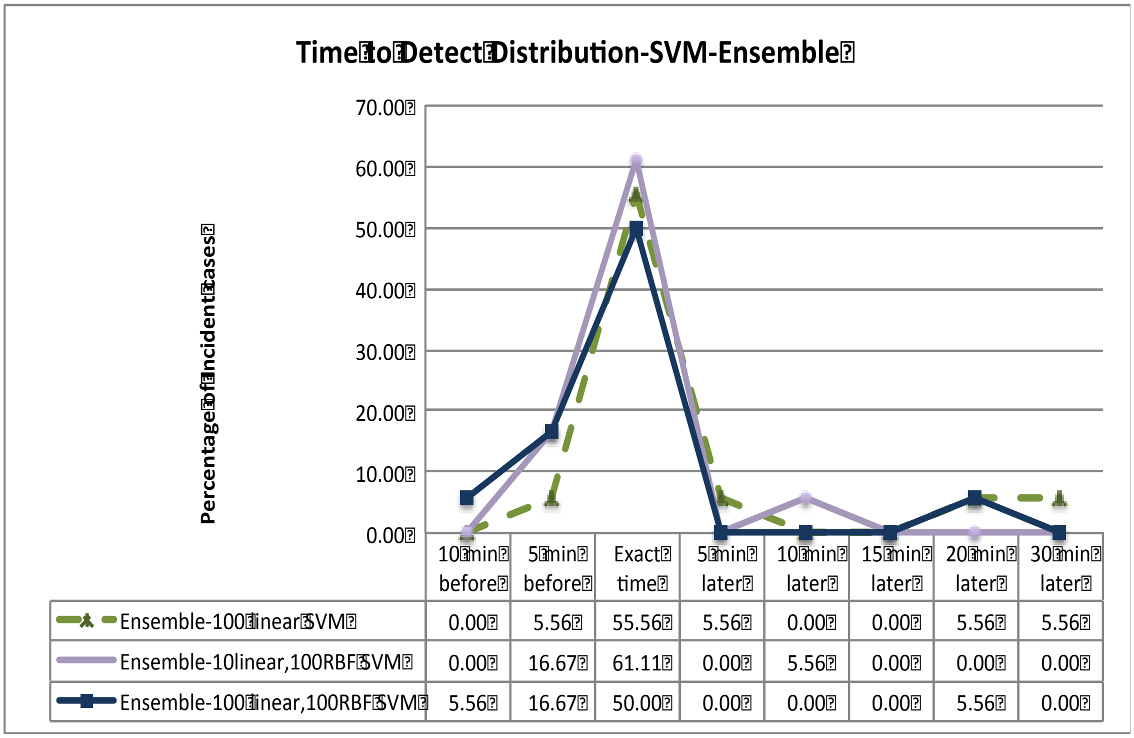
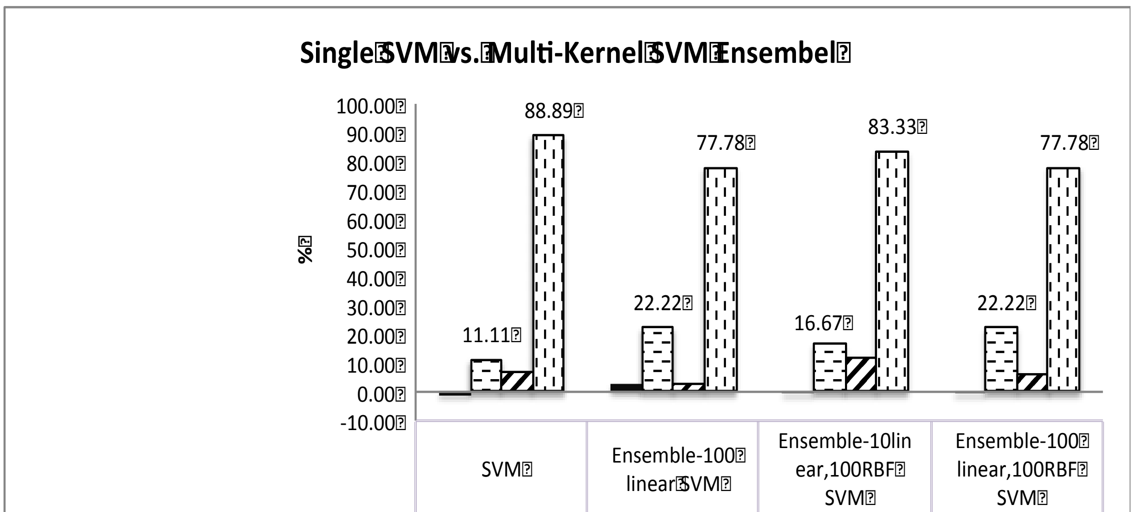


Figure 6-10 SVM Ensemble Time Performance: Time to Detect Distribution

SVM, and SVM Ensemble Incident Detection Models Comparison

An intuitive description of the relationship between the single SVM model and SVM ensemble model is represented in graphically in Figures 6-11 and 6-12.

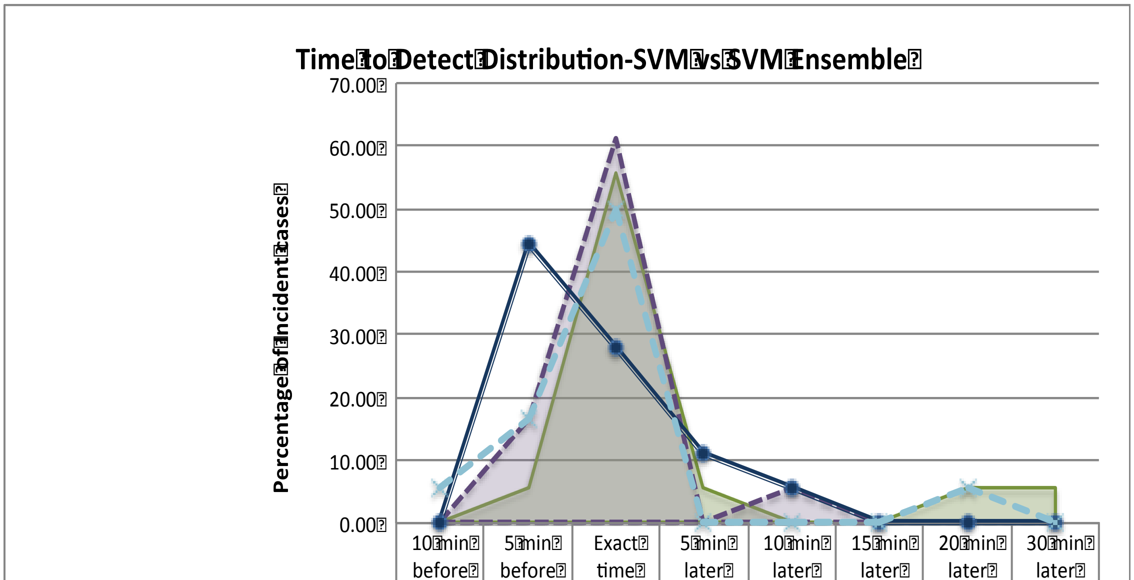
The results from Figure 6-11 indicate the ensemble model using only a linear kernel function was able to lower the FAR. However, note that the detection rate decreased. Another finding from this graph is the higher false negative, which means implementing the ensemble model increased the chances of an incident not being detected by the model.



	SVM	Ensemble-100 linear SVM	Ensemble-10linear,100RBF SVM	Ensemble-100 linear,100RBF SVM
Mean Time to Detect (MTTD) min	-1.11	2.78	-0.28	-0.28
False Negative (Percentage of Incident has not being detected by model) %	11.11	22.22	16.67	22.22
False Alarm Rate (FAR) %	6.60	2.43	11.98	6.08
Detection Rate (DR) %	88.89	77.78	83.33	77.78

Figure 6-11 Performances of SVM ensemble models vs. single SVM model

However, that leads to the question: why have these models become more popular? This could be because of the fewer false alarms and the fact that the ensemble models tend to detect an incident closer to the exact time it occurs. Figure 6-12 illustrates the time to detect distributions for both the simple SVM model and the ensemble models. Generally, detecting an incident at the exact time it occurs has been commonly accepted as a better alternative to early detection, since detecting an incident that has not yet occurred falls into the false alarm category. If we alter our definition of a successful detection model being the one that detects the most incidents, then we come to a different conclusion. Adapting this definition leads to the conclusion that the simple SVM model (despite early detection resulting in a higher FAR) is the best model.



	10 min before	5 min before	Exact time	5 min later	10 min later	15 min later	20 min later	30 min later
Ensemble-100 linear SVM	0.00	5.56	55.56	5.56	0.00	0.00	5.56	5.56
Ensemble-10linear,100RBF SVM	0.00	16.67	61.11	0.00	5.56	0.00	0.00	0.00
SVM	0.00	44.44	27.78	11.11	5.56	0.00	0.00	0.00
Ensemble-100 linear,100RBF SVM	5.56	16.67	50.00	0.00	0.00	0.00	5.56	0.00

Figure 6-12 SVM vs. SVM ensemble Time Performance: Time to Detect Distribution

Comparison Of Models Developed In This Study

A comparison of the relationship between the models is presented in graphical form in Figure 6-13 through Figure 6-17. Each graph depicts all models specific performance criteria. By examining these graphs, one can determine strengths and weaknesses of each type of model. It can be seen, generally that Naïve Bayes took longer to detect an incident, where as the Ensemble SVM detection time was mostly closer to the exact time of the real incident.

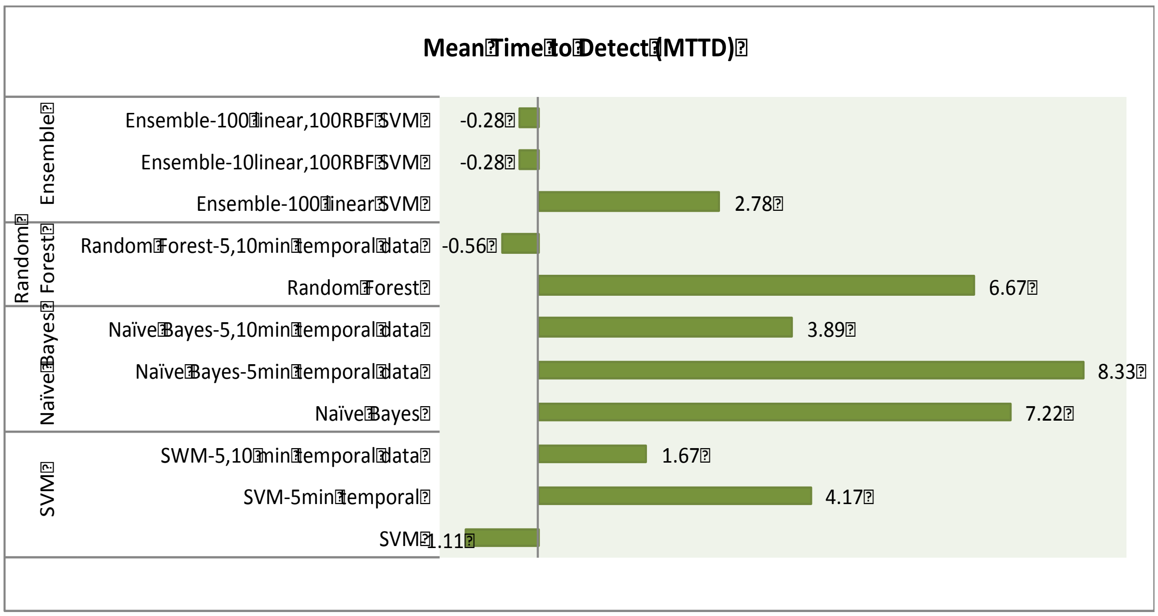


Figure 6-13 Mean time to detect (MTTD) performance measure for all scenarios

Figure 6-14 graphs results of FAR for all scenarios. The SVM ensemble model with the 100 linear kernel function had the lowest alarm rate. Random forest had the highest FAR, but was reduced by applying temporal data to the model.

Looking at the detection rate (Figure 6-15) indicates Naïve Bayes with temporal data has the best performance. Each technique has some strengths and also some weaknesses. In Figure 6-16, Naïve Bayes with temporal data had the lowest false negative rates. One may chose the most practical model based on their preference to one or more specific performance measures.

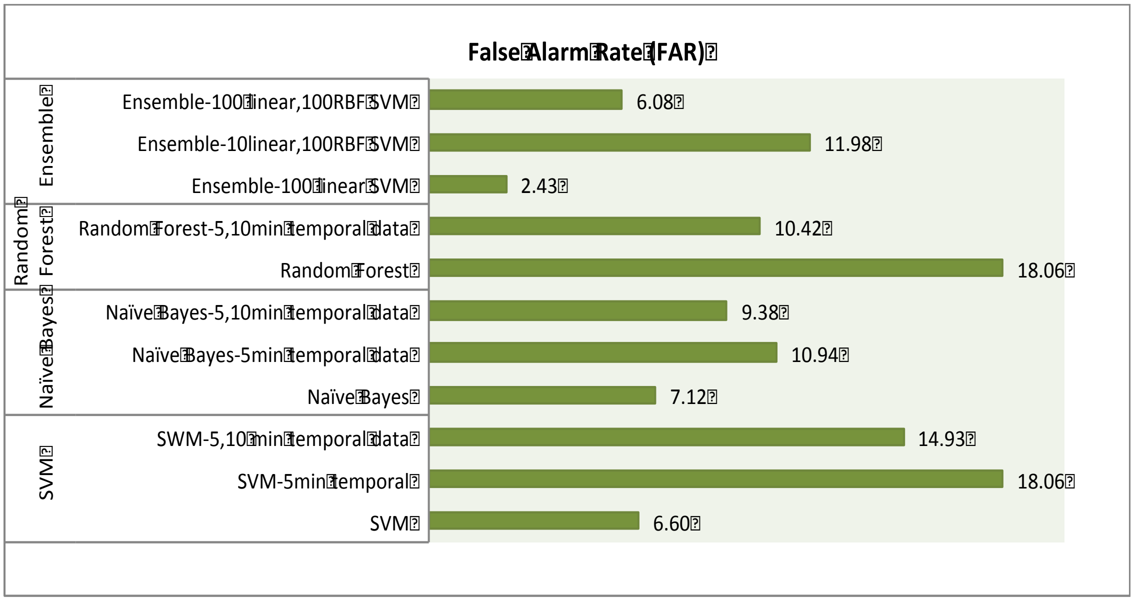


Figure 6-14 False alarm rate (FAR) performance measure for all scenarios

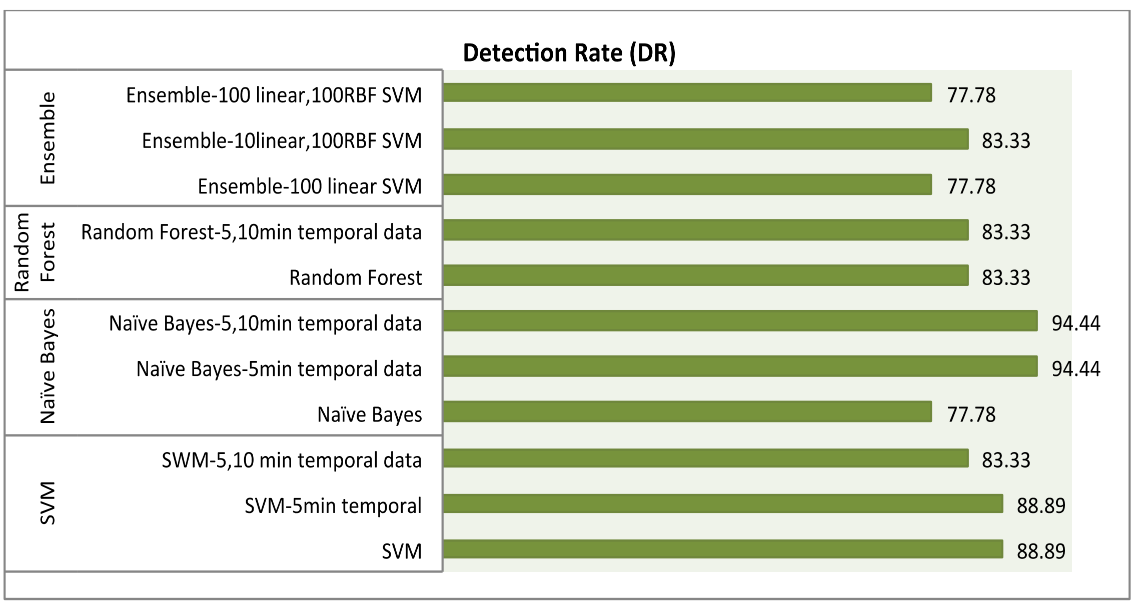


Figure 6-15 Detection rate (DR) performance measure for all scenarios

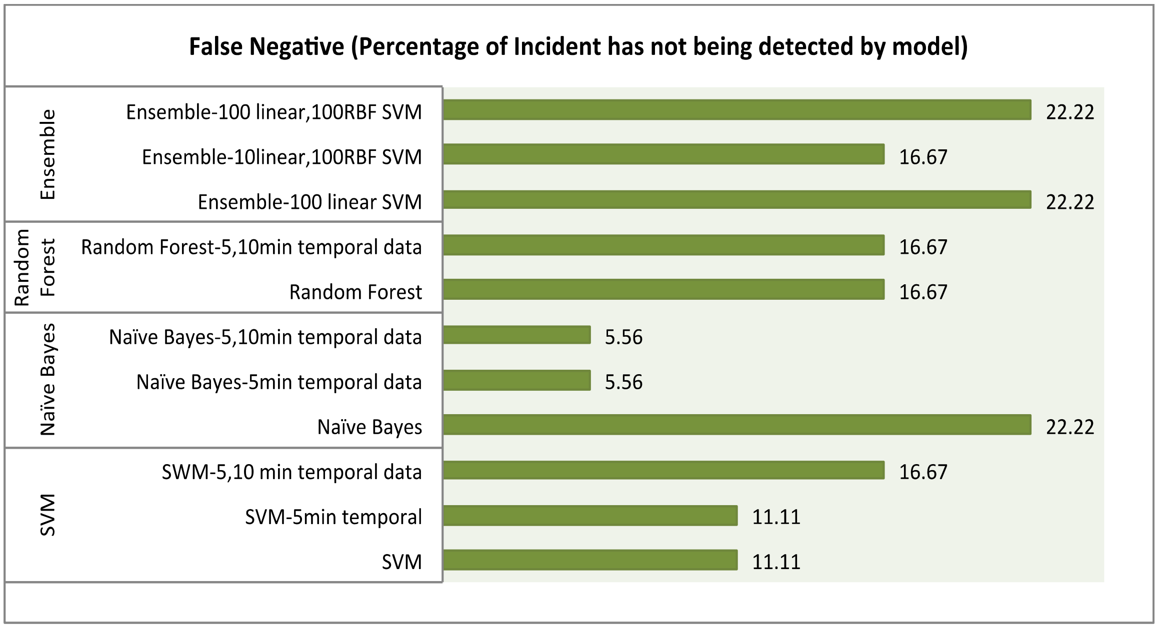


Figure 6-16 False Negative (percentage of incidents not detected) performance measure for all scenarios

Figures 6-17, 6-18 were prepared to present a better image of all the developed models in this section. The Naïve Bayes model had the lowest performance measure, which was improved through adapting temporal data. Naïve Bayes with temporal data holds the lowest false negative rate (percentage of incidents not detected), however it tends to have longer mean time to detect. Random Forest had trustworthy performance and was also improved with the adaptation of temporal data. However, it does not fall into the best performing incident detection model group. The single SVM has reasonably good performance regardless of its sensitivity to the appropriate kernel function or parameters. The parameters for best performing single SVM model was achieved through trial and error.

Overall, the time to detect distribution of each incident detection model added valuable insight as a performance measure. Figure 6-18 illustrates the wide range of distributions for the incident detection techniques that have been studied in this chapter. It is well shown that the SVM ensemble incident detection model has the best performance with respect to detection time, reaching over 61%.

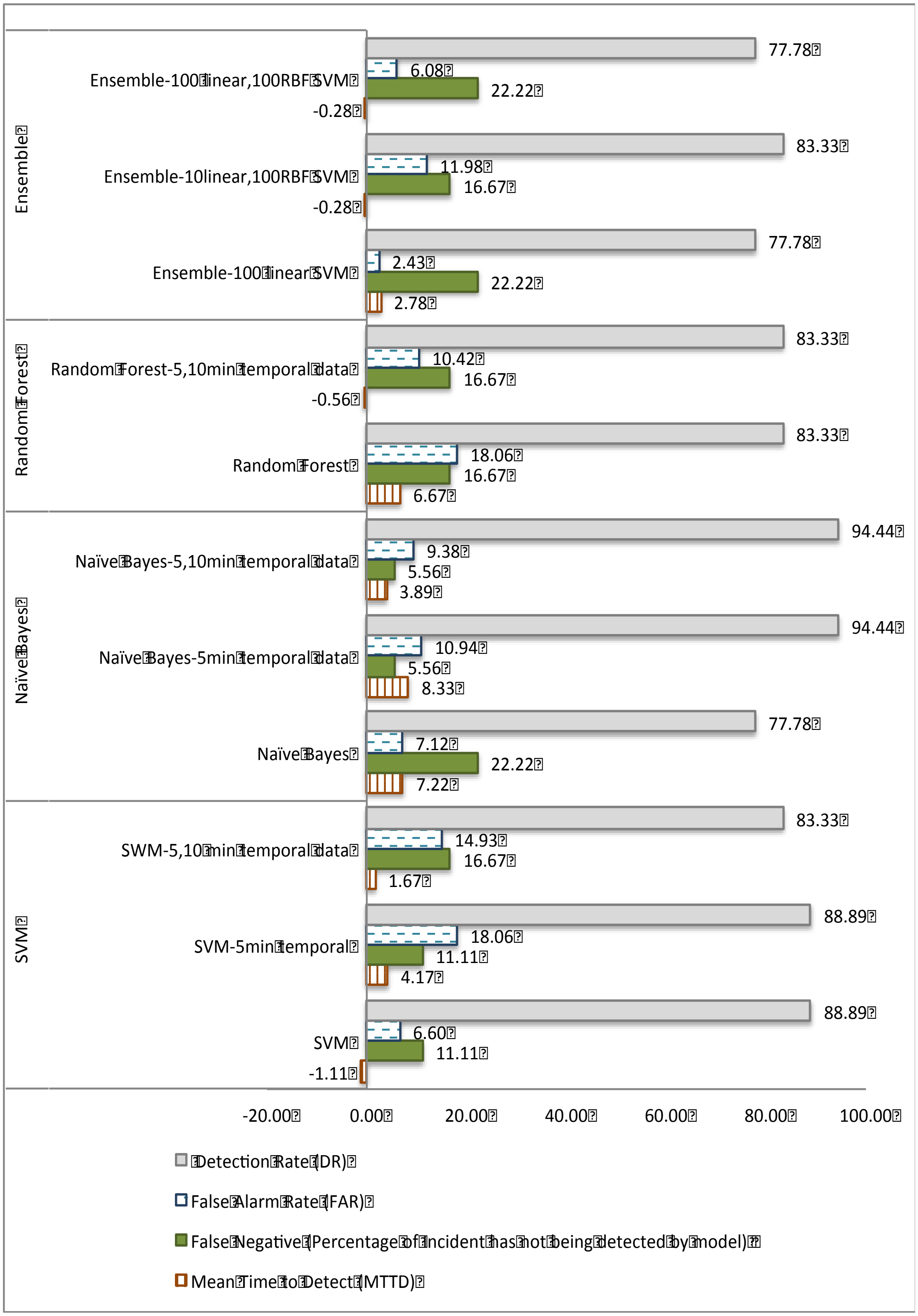
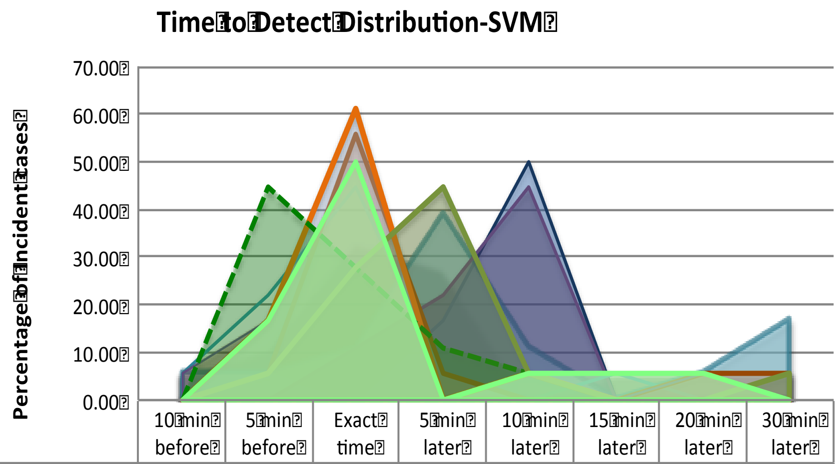


Figure 6-17 Performance for all scenarios



	10 min before	5 min before	Exact time	5 min later	10 min later	15 min later	20 min later	30 min later
Naïve Bayes	0.00	5.56	0.00	16.67	50.00	0.00	0.00	5.56
Naïve Bayes-5min temporal data	5.56	5.56	11.11	38.89	11.11	0.00	5.56	16.67
Naïve Bayes-5,10min temporal data	5.56	11.11	33.33	27.78	0.00	5.56	5.56	5.56
Random Forest	0.00	0.00	11.11	22.22	44.44	0.00	5.56	0.00
Random Forest-5,10min temporal data	5.56	22.22	44.44	5.56	0.00	5.56	0.00	0.00
Ensemble-100 linear SVM	0.00	5.56	55.56	5.56	0.00	0.00	5.56	5.56
Ensemble-10linear,100RBF SVM	0.00	16.67	61.11	0.00	5.56	0.00	0.00	0.00
Ensemble-100 linear,100RBF SVM	5.56	16.67	50.00	0.00	0.00	0.00	5.56	0.00
SVM	0.00	44.44	27.78	11.11	5.56	0.00	0.00	0.00
SVM-5min temporal	0.00	5.56	27.78	44.44	5.56	0.00	0.00	5.56
SVM-5,10 min temporal data	0.00	16.67	50.00	0.00	5.56	5.56	5.56	0.00

Figure 6-18 All Scenarios Time Performance: Time to Detect Distributions

COMPARISON OF MODELS DEVELOPED IN THIS STUDY TO PREVIOUS STUDIES

To assess the competitiveness of the algorithms developed in this study, several advanced incident detection models were evaluated, including the SVM ensemble method, Dynamic Time Warping, Multi-Kernel SVM, Bayesian Network, Naïve Bayes, Decision Tree, and Dynamic Bayesian Network. One of main criteria in case collection comparison is whether the model was tested on real field data and not the simulation data because all the above advanced machine learning models are data driven and sensitive to field site characteristics.

The literature review of the above mentioned models (listed in Table 6-5) indicated that a majority of recent studies were still testing models on the I-880 dataset from 1993. However, it is difficult to evaluate performance of two algorithms with different theoretical foundations on the same dataset. It is critical to implement the model on new sites to find out the extent of applicability for these models. There were only a few studies that tested their models on sites other than I-880. Model 5, 6, and 10 (from Table 6-5) have higher FAR. The resolution of the data is another possible factor on accuracy of the model as well. Considering the lack of field data available, there is no case study on the sensitivity of the data collection resolution on incident detection performance. Nevertheless, the current sample of recent studies shows that longer duration data collection intervals lead to higher false alarm rates. Another criterion for incident detection performance is mean time to detect (MTTD). Table 6-5 shows that the incident detection models developed using 5-minute data resolution did not calculate the MTTD. Historically dynamic time warping holds the highest detection rate (DR), but also has the longest mean time to detect (MTTD). Table 6-5 provides a good sense of what to expect from incident detection models described in the literature. However, the models developed in this study were tested on different sites from different states to evaluate the transferability of the models. Therefore, the performance measures might be less comparable to the results from the literature review. Another key element to be noticed, is these models were developed for normal traffic condition.

This study was mainly focused on resolving incident detection (high FAR/low DR) when the subject freeway carries high traffic volume (congested stop-and-go conditions). From Figure 6-17, the Naïve Bayes algorithm with temporal data produced the best detection rate (DR) 94% with an acceptable FAR of 9.3%. Table 6-5 shows that the second model with DTW has the same detection rate (94%) but longer detection time. In this study, the SVM ensemble model has the lowest FAR, 2.4%, with 78% DR and 2.7-minute MTTD. On the other hand, from Table 6-5, as mentioned earlier, the models were only tested on I-880 therefore they exhibited lower FAR. If one is only looking at the models tested on freeways other than I-880, one finds that Model 7 has the lowest FAR of 4.98% with 79% detection rate. However, in that study, the resolution of data collection was much higher in this case (30 sec) and the effort was not designed for incident detection for high volume traffic conditions.

Table 6-5 Recently developed incident detection models summary

Model No	Author	Year	Dataset	Data resolution	Technique	MTTD	FAR	DR	CR
1	Chen, S., Wang, W., Zuylen, Z.	2009	I-880 Freeway 1993	30 sec	SVM Ensemble (bagging-boosting)	3.94	1.6	88.7	-
2	Hi-ri-o-tappa, K. ; et. al.	2011	Daokanong-Suksawat freeway, Bangkok, Thailand 2010	1 min	Dynamic Time Warping	5.67	5	94	-
3	Jianli Xiao, Yuncai Liu	2012	I-880 Freeway 1993	30 sec	Proposed SVM Ensemble	1.84	3.6	86.3	95.6
4	Jianli Xiao and Yuncai Liu	2012	I-880 Freeway 1993, PeMS Noisy Data Set 1993	30 sec	MKL-SVM	0.61	14.3	63.6	81.6
5	Moinul, H., Muromachi, Y.	2012	Shibuya 3 and Shinjuku 4 expressways, Tokyo, Japan 2008-2009	5 min	Bayesian Network	-	13	58	82
6	Qu, X., Wang, W., Wenfu Wang, Liu, P.	2013	The Milwaukee area I-894, 9.3 mile, 2011	5min	Sideswipe crash prediction, SVM (RB)	-	15.4	88.5	86.5
7	Liu, Q., Lu, J., Chen, S., Zhao, K.	2014	I-880 Freeway 1993, Ayer Rajah Expressway Singapore (5.8 km)	30 sec	Naïve Bayes	1.46	4.98	79.6	87.8
8	LU, J., LIU, Q., YUAN, L., CHEN, S.	2014	I-880 Freeway 1993	30 sec	Decision Tree formed by C4.5	0.83	0.98	87.62	98.5
9	Xiao, J., Gao, X., Qing-Jie Kong, Liu, Y.	2014	I-880 Freeway 1993	30 sec	Multiple Kernel SVM Ensemble	1.86	3.56	85.96	95.57
10	Jie Sun, Jian Sun	2015	Yan-an expressway, North-South expressway, Shanghai, China 2010	5 min	Dynamic Bayesian Network model	-	23.7	76.4	76.3

One may notice the incident detection models developed implementing only I-880 dataset (Model 1, 3, 8, and 9) has higher that 85% detection rate and very low FAR (less than 3.6%). There aere few models developed using field data other than I-880 dataset. However they did not address peak hour incident detection issue. If one look at the models developed implementing more recent field data (model 2, 5, 10), one can find that the either they suffer from long detection time (MTTD, model 2), low DR (model 5), or very high FAR (model 10). This could be because of the nature of the data.

This section provides a good picture of what current incident detection model's performances are. Comparing this with the results obtained form previous section shows the model developed in this study were able to increase the performance of incident detection over all.

CONCLUDING REMARKS

Different techniques were adopted to develop and test incident detection models on datasets from different sites. The temporary data concept was introduced and enhancements of the model performance were evaluated. Adoption of 5 and 10 minute temporal data improved the SVM algorithm time to detect, by orienting the detection toward exact incident time. However, the gain of more detection accuracy caused an increase in false positives and negatives. The effect of temporal data on improvement of Naïve Bayes (NB) and Random Forest (RF) was more significant. Comparison of single SVM with NB and RF demonstrated an expected better performance of the single SVM algorithm. But implementation of temporal data could improve the NB and RF. To be able to pick the right model for specific traffic control center, one might need to define the priority of the performance measures first and pick the better fit model based on their case.

The ensemble models were also constructed and compared to the single SVM model. The performance measure comparison indicated that the ensemble model was able to shift the detection time toward the exact time of incident.

In the last part of this chapter, the competitiveness of this study with regard to several advanced models was assessed. The performance comparison between models developed in this study and other advanced freeway AID algorithms also showed the strong competitiveness of SVM algorithm.

Chapter 7: Effects of Traffic State on Freeway Incident Duration

Estimating freeway incident duration is a significant incident management challenge for traffic operations centers. This part of the study examined the effect of V/C (volume/capacity) ratio and level of service (LOS) on the time duration of traffic incidents. The Dallas dataset was used for this part of the study because it had more detail available. Companion field incident data included incident location, affected lane(s), time of incident detection, time cleared, speed, number of vehicles involved, peak/off peak, and type of incident. This chapter used LOS, as described by the Highway Capacity Manual, as an alternative to using V/C ratio. Weather conditions before incident start and incident detection mode effects were also evaluated. The difference and possible advantage of using LOS is that LOS offers discrete classes of conditions as opposed to the continuous nature of the V/C ratio. The discrete class variable tends to reduce random variability in the data, thus yielding a better fitting model. The results from this Chapter can benefit traffic control centers by providing techniques to improve the accuracy of estimated incident duration thereby providing more reliable traveler information guidance.

IMPORTANCE OF INCIDENT DURATION ESTIMATION

Traffic incidents including vehicle crashes, disabled vehicles, and lost cargo have become a typical part of urban freeway travel. As a result, traffic incident management has become an important component of urban freeway traffic management. Reliable incident duration prediction in real-time is vital for advanced traffic incident management (ATIM) so that travelers can have accurate and current travel-time forecasts. Under normal traffic conditions, road users choose routes based on their plan or experience whereas under incident conditions they hope to acquire accurate traffic information from a traffic control center or other traffic data source to avoid travel delays.

Despite the number of past studies in this field, predicting incident duration is still a challenge. Developing an accurate prediction model requires extensive amounts of detailed traffic data. While technology growth has improved traffic data collection techniques, not all of the specific traffic data is always available. Studies in this field have shown that the use of different incident data sources, related variables, and prediction techniques can have significant negative effects on prediction results. Therefore, using new robust data sources could lead to an improved prediction model.

The objective of this study is to explore the use of a different perspective to solve the incident duration prediction problem. Instead of observing specific incident related parameters to predict the incident duration, we evaluate using observed traffic characteristics and representing them as volume to capacity (V/C) ratios and as levels of service (LOS). We also evaluate the effect of speed and location on incident duration prediction.

HISTORY OF INCIDENT DURATION PREDICTION MODELS

In the past decades, many studies have investigated incident duration prediction models. Various regression and Bayesian classification models have been traditionally used. Usually, a Bayesian classifier has shown better prediction performance (64). More recently, classification and survival analysis has become more popular in the incident detection field (58,59,60). These models have high prediction accuracy for short duration incidents, but commonly have large errors in long duration incidents (<60 min).

Most incident duration models relate incident duration to incident type and severity, lane closure(s), and vehicle type [57,58]. Incident duration is also affected by incident location, average speed at the time of incident, traffic flow condition, and level of service (LOS). Recent studies have applied generalized traffic conditions, such as peak hour and congestion presence as binary variables [59,62,63]. Kang et al. (2011) found that including a binary peak hour variable enhances incident duration prediction. Chang et al. (2013) applied a decision tree classification with twelve variables including a peak hour variable. They achieved overall accuracy of 75.1% and 96% for short duration incidents (5-41 min) prediction [59]. Hojati et al. (2014) considered more traffic factors into the incident duration prediction model including average speed at the time of incident, and V/C before and after the time of incident. They found that all traffic measures included in their model had significant effects on incident duration prediction [61]. They also found morning peak period incident durations are 40% shorter than other times.

MODEL BUILDING

This section highlights the two most important considerations for creating a multiple regression model followed by a description of data preparation and model building.

Let us assume there are n-1 predictor variables, therefore there are n beta coefficients (including the intercept). The multinomial regression model is:

$$Y_i = B_0 + B_1 X_{i1} + \cdots + B_{n-1} X_{i,n-1} + \varepsilon_i,$$

so that

$$E\{Y\} = B_0 + B_1 X_{i1} + \cdots + B_{n-1} X_{i,n-1}$$

where the outcome variable is Y_i, the covariate variables are X, and the random error is ε_i. E{Y} is a linear function of the covariate variables.

The parameter β_k indicates the change in E{Y} per unit increase in X_k (continuous) when all the other predictor variables are held constant. This model assumes there is no interaction among variables. That means the effect of any predictor variable on E{Y} is the same regardless of the levels of the other variables. One common mistake is to mix up the concept of interaction and correlation. Correlation happens when a predictor is redundant whereas interaction happens when the effect one variable depends on the value of another variable. Interaction between two predictor variables can be entered into a model as the cross-product of those two variables, as such

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i1} * X_{i2} + e_i.$$

However, including interaction in the model requires a different interpretation of the beta coefficients. More details are provided in the result section.

Another consideration for multiple regression models is multicollinearity which occurs when there are variables that overlap with other variables. In this case, redundant predictor variables must be removed from the model.

To select a set of predictors, the forward stepwise procedure is most commonly used. Each step in the procedure adds or deletes an x variable. The criterion for adding or deleting an x variable can be stated in terms of error sums of squares reduction, coefficient of partial correlation, the t statistic, the F statistic, or the p-value.

Data Structure

Data from Dallas, Texas is used for this investigation (the same data used for incident prediction). For purposes of this effort, these two datasets were merged. First, the incident data was cleaned by removing highly unlikely values and missing data elements. Volume and speed data were extracted one hour before each incident started upstream of each incident location. Volume to capacity (V/C) ratios and levels of service (LOS) were calculated using Highway Capacity Manual procedures. Then, these variables were added to the incident dataset. Traffic conditions were extracted for 107 incident cases on 4 different freeways (Table 7-1).

Table 7- 1 Part of the structured dataset for incident duration prediction

Location ID	IncidentDuration	ffectedLanesTe	TypeText	NumberOfInvolvedVehicles	faxAffectedLanesTe	Volume	Status	No.lane	Volume/ln	V/C	LOS
147	0:50	Lane1	DisabledVehicle	-1	Lane1	3846	peak	4	962	0.48	C
138	1:43	None	Other	2	Lane1, Lane2, Lane3	4546.667	Peak	4	1137	0.57	C
102	1:42	e1, Lane2, Lane3, I	Accident	3	Lane1, Lane2, Lane3, Hov	1797	peak	4	449	0.22	C
147	0:44	None	Other	-1		4181.333	peak	4	1045	0.52	D
148	0:30	Lane1, Lane2	Accident	2	Lane1, Lane2	4405.333	peak	4	1101	0.55	E
138	1:36	None	Accident	-1	Lane3, Lane4	4790.667	peak	4	1198	0.60	D
118	1:09	Lane1, Lane2	Accident	2	Lane1, Lane2	4088	off-peak	4	1022	0.51	C
109	0:52	Lane3, Lane4	Accident	2	Lane3, Lane4	4186.667	peak	4	1047	0.52	E
102	0:37	Lane1, Lane2	Accident	3	Lane1, Lane2	5160	peak	4	1290	0.65	E
140	2:17	RightShoulder	Accident	-1	ightShoulder, Lane2, Lane	6221	peak	4	1555	0.78	E
140	0:23	RightShoulder	Accident	1	RightShoulder, Lane3	6830	peak	4	1708	0.85	E
109	0:24	Lane3	DisabledVehicle	1	Lane3	4626	peak	4	1157	0.58	F

First Model Results

The first model building effort used ordinary least squares regression as the estimation tool. Regression allows the analyst to easily compare the usefulness of predictor variables by assessing whether or not the predictor variables are responsible for the variability in the dependent variable. Regression is a simple estimation tool that has been widely used and is well understood. Using ordinary least squares regression brings transparency to the estimation process. There are several presumed data requirements for regression analysis. One of the primary assumptions is a normally distributed error pdf. Satisfying this assumption has been checked at each step. The VIF test were impelemened to measured how much the variance of an estimated regression coefficient is increased because of collinearity.

Different scenarios were defined to explore the significance of candidate predictor variables for the incident duration model (Table 7-2).

Table 7- 2 First Regression Model Scenarios (Dependent variable: Incident duration)

Scenario	Predictor Details	R^2	Adj. R^2	Sig.
1	V/C ratio	0.003	-0.007	0.587
2	LOS (A-F)	0.047	0	0.427
3	V/C ratio, Volume, Status, No of Vehicle, Max Affected Lanes, Type	0.222	0.15	0.003
4	LOS, Volume, Status, No of Vehicle, Max Affected Lanes, Type	0.338	0.154	0.024
5	Volume, Status, No of Vehicle, Max Affected Lanes, Type	0.22	0.156	0.002

First Regression Model Discussion

The results show that both the LOS and V/C ratio do not appear to be good predictors of incident duration. The R^2 in scenarios 1 and 2 indicate that LOS is a better predictor than V/C ratio, but regardless neither suffices. The results could be telling us several different facts. One might be the LOS was almost the same across most incidents, in which case the predictor variable did not vary with the dependent variable. One might think that if the V/C ratio is high (i.e., close to one) it should cause longer incident duration, nevertheless the result from this dataset do not support this hypothesis. Based on these results, one can claim there is no linear association between incident duration and V/C ratio or LOS. Besides LOS and V/C, volume per lane does not appear to be a significant predictor, possibly because of limited volume variation across the data. The nature of the data is slightly problematic because it does not cover a significant range for each parameter.

As a next step we consider the fact that according to the empirically based speed-volume relationship of the HCM, speeds are essentially constant as volumes change from LOS A through C or even D. Therefore, we classify LOS A and B as condition one, LOS C and D as condition two, and LOS E and F as condition three.

Second Model Results

The overhead time duration is the time required for emergency service vehicles to reach a particular incident site. Overhead time duration is influenced by the physical characteristics of the freeway section, such as proximity to ramps or elevated or depressed vertical profiles. Overhead time duration tends to have a huge effect on incident duration prediction. In order to account for these physical characteristics, variable Freeway_Loc was introduced to the incident duration prediction model (Table 7-3).

Table 7- 3 Descriptive Statistics

Freeway_Loc	Mean (min)	Std. Deviation (min)	n (incidents)
1.00	17.7500	5.91031	12
2.00	26.8333	2.40580	12
3.00	36.6316	4.17945	19
4.00	69.3281	25.29449	64
Total	52.9720	28.54175	107

In order to examine each potential predictor variable, a stepwise regression search algorithm in terms of the t statistic and its associated p-value was implemented. And for each scenario normal distribution of the error term and residual were examined. Weather conditions before the incident were also considered as a possible influencing feature on incident duration. Weather data were collected separately from the Weather Channel and classified as wet or dry conditions. Table 7-4 summarizes the model scenarios identifying predictor variable(s) found to be insignificant.

Table 7- 4 Second Model Scenario Summery

Variables	Significance Test Result
V/C ratio	Not significant
LOS / Grouped LOS	Not significant
Speed	Not significant
Location/Freeway	Not significant
Traffic Status (peak/off-peak)	Not significant
Volume	Not significant
Number of involved vehicle	Not significant
Max affected main lane	**Significant**
Weather condition	**Significant**
Incident Type	Not significant

Second Regression Model Discussion

The results showed the incident location is not a significant explanatory variable for incident duration. Grouping LOS improved the predictor, but not enough to make it a significant predictor of incident duration. Similar to the first model, the procedure to find the "best" set of predictors is a sequence of regression models, at each step adding or deleting an X variable. The criterion for adding or deleting an X variable is partial correlation, the t statistic, normal distribution of the error term, and the p-value. If the corresponding p-value is less than the pre-determined level for entry, then the predictor is retained, otherwise it is deleted.

Examination of speed showed that speed has a linear association with incident duration. However based on this specific sample, the models did not find speed a significant predictor of incident duration. The result from this part of study showed there is no linear association between incident duration and V/C ratio, or number of involved vehicles.

The interactions of different variables have been tested, however, no significant interaction effects were found among these variables.

For model comparisons, R^2 and adjusted R^2 were used. Since R^2 usually can be inflated by including more predictor variables, the adjusted R^2 was used. It can be calculated as:

$$R_a^2 = 1 - \frac{\frac{SSE}{n-p}}{\frac{SSTO}{n-1}} = 1 - \left(\frac{n-1}{n-p}\right)\frac{SSE}{SSTO}$$

Where p is the number of variables in the model. The result from best scenario chosen as the best-fit model is presented in Table 7-5.

Table 7- 5 Best Scenario Parameter Estimates

Parameter	B	Std. Error	t	Sig.
Intercept	33.95	5.12	6.62	1.55E-09
Max Affected Lane	12.05	3.049	3.953	0.000141
weather.ID2 (wet)	29.68	13.43	2.209	0.029345

The results of Table 7-5 show as expected that there is a positive linear relationship (B coefficient) between incident duration and maximum number of affected lane. The more lanes involved in an incident, we expect to have heavier congestion that leads to longer incident durations. The weather condition during and before the incident happens shows that incident durations increase significantly when the weather condition is wet.

To assess the goodness of fit of the model, we use the residuals defined by

$$e_i = y_i - \hat{y}_i \quad \text{where} \quad \hat{y}_i = b_0 + b_1 X_{i1} + \cdots b_{p-1} X_{i,p-1}$$

This shows that:

Residual standard error: 25.93 on 104 degrees of freedom
R-squared: 0.1903,
Adjusted R-squared: 0.1748

Third Model Results

The evaluations were extended to include additional test sites to strengthen confidence in the performance of the generalized incident detection model. Here, we expand our evaluation to include an additional 1500 incident cases from Dallas dataset (US 75 North-South and I 635 East-West). This data includes every incident that lasts longer than five minutes from July 2012 to September 2012. The recorded features on this dataset are listed as:

- Incident type (IT)
- Maximum number of affected lanes (MaxLane)
- Detection mode (i.e. camera) (DM)
- Incident location (IL)
- Number of involved vehicles (No.Veh)

Weather conditions before the incident were considered as a possible influencing feature on incident duration and were collected separately from the Weather Channel.

To define the significance of these features, many scenarios were defined. The stepwise regression search algorithm in terms of the t statistic and its associated p-value was implemented. To find the correlation between nominal/ordinal variables, Pearson's Chi-Squared test was used (Table 7-6). The Pearson Test showed strong correlation among several predictor variables, which suggests that they should not be included at the same time in the model. For example, number of involved vehicles is a function of road geometry (location) and weather conditions. A correlation between weather type, number of involved vehicles, and type of incident can be expected. The type of incident is highly correlated with number of involved vehicles and maximum number of affected lanes.

Table 7- 6 Pearson Correlation Test

		RoadId	Weather ID	Number Of Involved Vehicles	Max Number Of Affected Main Lanes	Type
RoadId	Pearson Correlation	1	-.065*	.008	.026	-.012
	Sig. (1-tailed)		.011	.381	.160	.326
	N	1508	1223	1508	1508	1508
Weather ID	Pearson Correlation	-.065*	1	-.009	-.044	-.033
	Sig. (1-tailed)	.011		.383	.064	.125
	N	1223	1223	1223	1223	1223
Number Of Involved Vehicles	Pearson Correlation	.008	-.009	1	.037	-.161**
	Sig. (1-tailed)	.381	.383		.075	.000
	N	1508	1223	1508	1508	1508
Max Number Of Affected Main Lanes	Pearson Correlation	.026	-.044	.037	1	-.243**
	Sig. (1-tailed)	.160	.064	.075		.000
	N	1508	1223	1508	1508	1508
Type	Pearson Correlation	-.012	-.033	-.161**	-.243**	1
	Sig. (1-tailed)	.326	.125	.000	.000	
	N	1508	1223	1508	1508	1508

A simple linear regression model was developed for each of the 6 variables. Each variable separately was significant. Therefore, these variables are significant when they are the first added in the model, but given the other correlated variables, each variable is no longer significant (Table 7-10, Scenario 1 & 4). This indicates that the predictors are correlated and there is multicollinearity. The simplest measure of multicollinearity is the correlation matrix of the predictor variables (Table 7-6). However, linear relationships between more than two predictor variables do not always lead to large bivariate correlations and so these are hard to detect from the correlation matrix. To select a set of predictor variables, the forward stepwise procedure was used. The F statistic, and the p-value were used for adding or deleting an x variable. At each step, variables with smallest p-value were candidates for the next step.

Some of these models are presented in Appendix B. A summary of selected scenarios is represented in Table 7-10. Based on our evaluation in the previous section, incident location was recognized as an important influencing feature. Therefore, in this section with a new large dataset, the first scenario reevaluated the effect of incident location on incident duration prediction. The result showed that incident location is a significant feature, however in a large scale it is not able to describe the large variability of incident duration. That could be as a result of volatility in the larger dataset. Scenario 2 represents the fact that incident detection mode has a significant effect on incident duration. Table 7-7 shows all of the collected detection modes present in our dataset. According to the fact that detection by Camera and Courtesy Patrol had the highest rank, to reduce the prediction variability, we introduced a Modified detection mode (Table 7-8).

Table 7- 7 Incident detection mode

Detection Mode	n (incidents)
531 radio	27
Camera	575
Courtesy Patrol	718
DART	18
DPD	19
Garland PD	3
LBJ Express	5
Media	47
Mesquite PD	4
Police	3
Public	95

The statistics in Table 7-8 show that detection by camera and public information requires more time than detection by Courtesy Patrol. The comparison between Scenario 2 and 13 shows that using modified category versus original detection mode categories did not change the adjusted R^2, however, it doubled the F-test value (Appendix B).

Table 7- 8 Modified incident detection mode

Descriptive Statistics

Dependent Variable: TotalDuration

Modified Detection Mode	Mean (minutes)	Std. Deviation (minutes)	n (incidents)
Camera	40.08	27.547	575
Courtesy Patrol	19.18	15.619	718
Other	47.98	28.540	120
Public	30.05	20.288	95
Total	30.12	24.806	1508

Observing the weather condition categories (Table 7-9), the two dominant groups are "Mostly Cloudy" and "Partly Cloudy". The intuitive reason for considering weather conditions on incident duration prediction was that rain/heavy rain/ thunderstorms seem like they could be responsible for causing delays in clearing an incident. Very little effect on incident duration is expected from "Clear Sky", "Partly Cloudy", or "Mostly Cloudy" conditions.

Table 7- 9 Weather condition categories

Weather	n (incidents)
Clear	20
Haze	1
Heavy Rain	1
Heavy Thunderstorms and Rain	1
Light Rain	2
Light Thunderstorms and Rain	3
Mostly Cloudy	307
Overcast	23
Partly Cloudy	578
Thunderstorm	3

Therefore, modified weather conditions were introduced with two categories: 1) dry conditions and 2) wet conditions.

When there is no interaction, the effect of any predictor variable on E{Y} is the same no matter the values/levels (numerical or categorical) of other variables. Scenario 11 showed that the modified weather condition is a significant predictor for incident duration prediction. However, it describes a small portion of incident duration variability (R^2 = 0.073). Scenario 16 indicates that the modification did not remove the correlation between weather and number of vehicles involved in the incident. In Scenario 8, we introduce a lanes involved ratio (LaneInv.Ratio):

LaneInv.Ratio = (MaxLane/Total No of lane)

The purpose behind this was to normalize the variable to take the effective number of lanes into account. This variable now represents the fraction of capacity lost on the freeway. The results showed that this variable itself could explain 21% of incident duration variability.

Table 7- 10 Summery of all incident duration scenarios

Scenario	Predictor Details	R^2	Adjusted R^2	Sig.	Insignificant Variable(s)
1	IL (US75N, US75S,I635E,I635W)	0.022	0.019	0	-
2	DM	0.197	0.194	0	-
3	IT,DM, Interaction IT*DM	0.297	0.28	0	-
4	IL, IT, DM, Interaction IT*DM, Interaction IT*IL, Interaction DM*IL	0.348	0.296	0	IL, Interaction IT*IL
5	MaxLane, DM	0.262	0.26	0	-
6	MaxLane, No.Veh, DM, IT, Interaction IT*DM	0.323	0.306	0	-
7	MaxLane, No.Veh, DM, IT, Weather, Interaction DM*weather, Interaction IT*weather,	0.362	0.327	0	Interaction DM*weather, Interaction IT*weather
8	LaneInv.Ratio	0.214	0.214	0	-
9	LaneInv.Ratio, DM, IT	0.307	0.299	0	Interaction IT*DM
10	LaneInv.Ratio, No. Veh, Weather, IT, DM, Interaction (Weather*DM, DM*IT, Weather*DM*IT)	0.362	0.327	0	Interaction (Weather*DM, DM*IT, Weather*DM*IT)
11	Modified weather	0.073	0.072	0	-
12	Modified weather, IL	0.10	0.095	0	IL,Interaction IL*weather
13	Modified DM	0.195	0.194	0	-
14	LaneInv.Ratio, Modified DM	0.262	0.262	0	-
15	LaneInv.Ratio, Modified DM, IT, Interaction Modified DM*IT	0.307	0.299	0	Interaction Modified DM *IT
16	LaneInv.Ratio, No.Veh, Modified weather, Modified DM, Interaction Modified DM*Modified weather	0.26	0.255	0	Modified weather, Interaction Modified DM*Modified weather
17	Modified weather, Modified DM, LaneInv.Ratio, IT	0.281	0.276	0	-

Both incident location (RoadId) and modified weather conditions before an incident are significant predictors, and only describe respectively 0.02 and 0.07 variability on incident duration. The correlation table 7-6 shows strong correlation between location and weather in this dataset. They also tend to be highly correlated with other predictors (IT, No.Veh).

Parametes	***B***	***Std. Error***	***t***	***Sig.***
Intercept	36.859	2.113	17.44	0.00
Weather.ID2 - wet	32.807	6.59	4.97	7.72E-07
ModifiedDetectionMode-Courtesy Patrol	-8.102	1.66	-4.88	1.23E-06
ModifiedDetectionMode-Other	4.158	2.045	2.034	0.04224
Lane Involvement Ratio	20.347	4.37	4.647	3.80E-06
Incident Type 2 (Disabled Vehicle)	-10.476	2.179	- 4.808	1.75E-06
Incident Type 3 (Debris)	-18.705	4.77	- 3.915	9.63E-05
Incident Type 4 (Others)	10.95	3.353	3.266	0.00113

Table 7- 11 Best Scenario Parameter Estimates

The results of Table 7-11 show that the incident duration is expected to have 32 minutes longer duration when the weather condition is wet. The base detection mode in this model is camera. The results show that if the detection mode is courtesy patrols one can expect 8 minutes less incident duration on average compared to the base detection mode. There is a positive linear relationship (B coefficient) between incident duration and maximum number of affected lanes. The more lanes involved in an incident, we expect to have heavier congestion that leads to a longer incident. The base incident type in this model is accident. The results show that less incident duration is expected if the incident type is disabled vehicle/ debris by 10.47/18.7 minutes.

The summary of the best model is:

Residual standard error: 19.67 on 1041 degrees of freedom
R-squared: 0.281, Adjusted R-squared: 0.2762

CONCLUSION

Incident duration has certain randomness, and apart from recorded variables, is influenced by unrecorded factors such as response time, weather conditions, and the approach used by incident management personnel. Different methods of incident management for similar incident types can cause large model prediction errors. We investigated LOS and V/C ratio rather extensively, but both were poor predictors of incident duration. These data seem to indicate the traffic conditions before incidents were pretty similar across most incident occurrences.

By grouping LOS and the number of affected lanes, a lot of data scatter was removed and LOS change was not large enough to make it significant.

After including a location variable to the model we were able to improve the model significantly, indicating that incident duration is influenced by the physical characteristic of the freeway. We also found that speed and number of affected lanes has a linear relationship with incident duration.

Chapter 8: Conclusions

Most AID algorithms fail to deliver high detection rates during heavy traffic conditions. When the problem is too complex, data driven techniques appear to be more successful in finding a pattern. A majority of recent advanced studies have proven the success of these data driven techniques in the field of AID (Table 6-5). However because of the low performance during heavy traffic conditions of the currently available incident detection methods, traffic control centers around the nation do not implement AID in practice. The objective of this dissertation was to enhance the current capabilities of AID techniques for heavy traffic conditions. Additionally, this dissertation evaluated the transferability of the developed model. The second part of this dissertation focused on developing an incident duration prediction model by introducing and evaluating new predictors.

SUMMARY OF COMPARISON TECHNIQUES

Relatively diverse and recent field datasets (in comparison to previous studies) were extensively tested for this study. The first stage of study used a Dallas dataset (containing 2 directions of 2 different freeways) to develop a real-time incident detection model with a focus on high volume traffic conditions. All data used in this proposed model are readily available to traffic control centers across the country. The results from the first stage indicated that the SVM incident detection technique complies with the transferability requisite, however the model needed to be tested further with even more diverse and extensive datasets. Although transferability was the driving motivation behind testing in a new site, validating the model was another important motive because of the data driven nature of the algorithm. The data driven algorithms are highly sensitive to the configuration of data. The more diverse the data is for training, the more one can rely on transferability of the model. Therefore in the second stage of study, data from several different states (Florida, Georgia, and Maryland) was requested to further test model transferability. Testing the model on new datasets could suggest possible modifications that could improve the general incident detection model. Florida DOT was kind enough to share their data for our model evaluation. With their help, the model developed with Dallas data and previously tested with Miami data was again tested with Florida data. Although it is very common to receive high false alarm rates when testing the model on a new site, the developed model was able to achieve very satisfying results with the Florida data (DR: 88.9, FAR: 6.8). Next, the SVM algorithm was evaluated against other competing techniques such as Naïve Bayes (NB), Random Forest (RF), and SVM ensemble for building and testing incident detection models. Furthermore, this study introduced the new concept of adding temporal data to improve incident detection. A summary of the results was presented in Figure 6-17 and 6-18. The results show that SVM is more robust than NB and RF in tackling the incident detection problem. Further, implementation of temporal data added more accuracy to NB and RF models, but it did not improve the SVM model, which could be because of the different theoretical foundations of these techniques. To choose the best model, between SVM and the NB and RF with temporal data, the traffic control centers must prioritize the incident detection performance measure and make their decision based on that, because each model could have its own advantages. Incident detection time is a critical criterion for practical implementation; mean time to detect (MTTD) is commonly used as a performance measure for AID algorithms. However this performance measure is not sufficiently intuitive. This study introduced time to detect distribution (TTDD), which provided a very intuitive understanding of the performance of each algorithm in regards to detection time. Generally, NB and RF tend to detect incidents a few minutes after the incident occurs. The ensemble model and the addition of temporal data were able to shift the detection time toward the exact time of event.

Incident duration prediction which is a key element of any incident management plan was the second part of this study. Many studies have tried to improve the duration prediction accuracy, but have not yet produced very satisfying results. This could be because the random nature of incident duration makes improving duration prediction difficult. Studying the literature revealed that a few variables- LOS, v/c ratio, and weather condition- have not been tested for incident duration prediction. Despite the expectation, no significant relation between LOS or v/c ratio on incident duration was found. However, this could also be as a result of the sample dataset for this study. The traffic data before incident were similar across most incident cases because we focused on detection during heavy traffic conditions. This study established the significance of speed and involved number of lanes in describing incident duration variability. This study was also able to show the significant effect of weather conditions, incident type, and detection mode on incident duration prediction.

CONCLUDING RECOMMENDATIONS

The models developed in this study have the potential to be effectively implemented in traffic control centers. However, one might need to consider that field data are often subject to collection bias. While working with field data, it was apparent that incidents are logged with delay. In other words, the start time of incidents are recorded later than they occur. Singliar and Hauskrecht (2010) used dynamic bayesian network (DBN) to address this problem. However, DBN requires information on actual time of incident for training the model. Unfortunately, this information was not collected/provided. If concerned with the validity of this model, the model can be tested on new data. The technique is data-driven, therefore a more diverse dataset can enhance the performance of the incident detection model.

While the incident has influence on traffic flow fluctuation upstream and downstream of incident; it is necessary to note this influence is a function of the distance of the data collection points from the incident location. The first issue to address is the distance between detectors. Additional detectors may be needed to cover the area fully, depending on the distance. The model cannot currently resolve the issue of having enough detectors by itself. However, including the distance between detectors in the model might improve it to some extent.

Incident duration has certain randomness apart from the examined recorded variables. It is likely influenced by unrecorded factors. For example, the geometry of the freeway, response time, weather conditions, the distance from previous entrance to the incident location, and management plans are all key varying variables that have not been addressed in the modeling. Detection mode and incident type appear to be a surrogate for and explain some proportion of the unrecorded factors. A significant amount of the incident duration variation appears to be explained by the incident detection mode variable. Thus suggesting that the incident detection mode and also the number of involved lanes influence the incident management plan. Evaluation of weather conditions showed the significance of this variable on incident duration. However, because our dataset was limited to the summer in Texas, this variable was not able to explain a large fraction of incident duration variability. We suggest further investigation into duration variability with a broader dataset.

APPENDIX A

SVM Model Results:

Confusion Matrix and Statistics

	Reference	
Prediction	Non	Yes
Non	538	45
Yes	38	84

Accuracy : 0.8823
95% CI : (0.8562, 0.9051)
No Information Rate : 0.817
P-Value [Acc > NIR] : 1.548e-06

Kappa : 0.5978
Mcnemar's Test P-Value : 0.5102

Sensitivity : 0.9340
Specificity : 0.6512
Pos Pred Value : 0.9228
Neg Pred Value : 0.6885
Prevalence : 0.8170
Detection Rate : 0.7631
Detection Prevalence : 0.8270
Balanced Accuracy : 0.7926
ROC curve variable importance

SVM Model Results with temporal data: 5 min

Confusion Matrix and Statistics

Prediction	Reference Non	Yes	
Non	470	42	
Yes		104	87

Accuracy : 0.7923
95% CI : (0.7604, 0.8217)
No Information Rate : 0.8165
P-Value [Acc > NIR] : 0.9542

Kappa : 0.4158
Mcnemar's Test P-Value : 4.455e-07

Sensitivity : 0.8188
Specificity : 0.6744
Pos Pred Value : 0.9180
Neg Pred Value : 0.4555
Prevalence : 0.8165
Detection Rate : 0.6686
Detection Prevalence : 0.7283
Balanced Accuracy : 0.7466
ROC curve variable importance

SVM Model Results with temporal data: 5,10 min

Confusion Matrix and Statistics

	Reference		
Prediction	Non	Yes	
Non	488	44	
Yes		86	85

Accuracy : 0.8151
95% CI : (0.7844, 0.8431)
No Information Rate : 0.8165
P-Value [Acc > NIR] : 0.5620608

Kappa : 0.452
Mcnemar's Test P-Value : 0.0003232

Sensitivity : 0.8502
Specificity : 0.6589
Pos Pred Value : 0.9173
Neg Pred Value : 0.4971
Prevalence : 0.8165
Detection Rate : 0.6942
Detection Prevalence : 0.7568
Balanced Accuracy : 0.7545

NAÏVE BAYES MODEL RESULTS WITHOUT TEMPORAL DATA

Confusion Matrix and Statistics

	Reference	
Prediction	Non	Yes
Non	533	61
Yes	41	68

Accuracy : 0.8549
95% CI : (0.8267, 0.8801)
No Information Rate : 0.8165
P-Value [Acc > NIR] : 0.004059

Kappa : 0.4848
Mcnemar's Test P-Value : 0.059934

Sensitivity : 0.9286
Specificity : 0.5271
Pos Pred Value : 0.8973
Neg Pred Value : 0.6239
Prevalence : 0.8165
Detection Rate : 0.7582
Detection Prevalence : 0.8450
Balanced Accuracy : 0.7279
ROC curve variable importance

NAÏVE BAYES CLASSIFIER WITH TEMPORAL DATA 5MIN

Confusion Matrix and Statistics

	Reference	
Prediction	Non	Yes
Non	536	53
Yes	38	76

Accuracy : 0.8706
95% CI : (0.8435, 0.8945)
No Information Rate : 0.8165
P-Value [Acc > NIR] : 7.169e-05

Kappa : 0.5476
Mcnemar's Test P-Value : 0.1422

Sensitivity : 0.9338
Specificity : 0.5891
Pos Pred Value : 0.9100
Neg Pred Value : 0.6667
Prevalence : 0.8165
Detection Rate : 0.7624
Detection Prevalence : 0.8378
Balanced Accuracy : 0.7615

NAÏVE BAYES CLASSIFIER WITH TEMPORAL DATA: 5MIN, 10MIN

Confusion Matrix and Statistics

	Reference		
Prediction	Non	Yes	
Non	492	31	
Yes		82	98

Accuracy : 0.8393
95% CI : (0.81, 0.8657)
No Information Rate : 0.8165
P-Value [Acc > NIR] : 0.06369

Kappa : 0.5349
Mcnemar's Test P-Value : 2.556e-06

Sensitivity : 0.8571
Specificity : 0.7597
Pos Pred Value : 0.9407
Neg Pred Value : 0.5444
Prevalence : 0.8165
Detection Rate : 0.6999
Detection Prevalence : 0.7440
Balanced Accuracy : 0.8084

RANDOM FOREST WITHOUT TEMPORAL DATA

Confusion Matrix and Statistics

	Reference	
Prediction	Non	Yes
Non	482	55
Yes	92	74

Accuracy : 0.7909
95% CI : (0.7589, 0.8204)
No Information Rate : 0.8165
P-Value [Acc > NIR] : 0.962559

Kappa : 0.372
Mcnemar's Test P-Value : 0.002985

Sensitivity : 0.8397
Specificity : 0.5736
Pos Pred Value : 0.8976
Neg Pred Value : 0.4458
Prevalence : 0.8165
Detection Rate : 0.6856
Detection Prevalence : 0.7639
Balanced Accuracy : 0.7067

RANDOM FOREST WITH TEMPORAL DATA: 5MIN, 10MIN

Confusion Matrix and Statistics

	Reference	
Prediction	Non	Yes
Non	514	36
Yes	60	93

Accuracy : 0.8634
95% CI : (0.8358, 0.888)
No Information Rate : 0.8165
P-Value [Acc > NIR] : 0.0005323

Kappa : 0.5749
Mcnemar's Test P-Value : 0.0189035

Sensitivity : 0.8955
Specificity : 0.7209
Pos Pred Value : 0.9345
Neg Pred Value : 0.6078
Prevalence : 0.8165
Detection Rate : 0.7312
Detection Prevalence : 0.7824
Balanced Accuracy : 0.8082

APPENDIX B

Models for Chapter 7

INCIDENT DURATION – ROAD ID

Descriptive Statistics

Dependent Variable:TotalDuration

RoadId	Mean	Std. Deviation	N
7	24.43	20.277	142
8	33.97	28.358	238
9	33.17	26.639	280
10	21.93	19.359	143
13	31.23	24.135	375
14	29.82	24.059	336
Total	30.19	24.824	1514

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	20814.241[a]	5	4162.848	6.887	.000	.022	34.434	.999
Intercept	1109889.421	1	1109889.421	1836.144	.000	.549	1836.144	1.000
RoadId	20814.241	5	4162.848	6.887	.000	.022	34.434	.999
Error	911536.975	1508	604.467					
Total	2312286.000	1514						
Corrected Total	932351.215	1513						

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	20814.241[a]	5	4162.848	6.887	.000	.022	34.434	.999
Intercept	1109889.421	1	1109889.421	1836.144	.000	.549	1836.144	1.000
RoadId	20814.241	5	4162.848	6.887	.000	.022	34.434	.999
Error	911536.975	1508	604.467					
Total	2312286.000	1514						
Corrected Total	932351.215	1513						

a. R Squared = .022 (Adjusted R Squared = .019)

b. Computed using alpha = .05

INCIDENT DURATION – ROAD ID, TYPE OF INCIDENT, DETECTION MODE

Between-Subjects Factors

		N
Type Text	Accident	414
	Debris	32
	Disabled Vehicle	994
	Other	74

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	324502.403[a]	112	2897.343	6.678	.000	.348	747.929	1.000
Intercept	133553.876	1	133553.876	307.822	.000	.180	307.822	1.000
TypeText	14815.220	3	4938.407	11.382	.000	.024	34.147	.999
DetectionMode	12657.769	10	1265.777	2.917	.001	.020	29.174	.980
RoadId	1948.182	5	389.636	.898	.482	.003	4.490	.325
TypeText * DetectionMode	19973.134	19	1051.218	2.423	.001	.032	46.035	.997
TypeText * RoadId	9792.411	15	652.827	1.505	.095	.016	22.570	.880
DetectionMode * RoadId	17796.728	27	659.138	1.519	.043	.028	41.019	.981
TypeText * DetectionMode * RoadId	24708.073	32	772.127	1.780	.005	.039	56.948	.998
Error	607848.813	1401	433.868					
Total	2312286.000	1514						
Corrected Total	932351.215	1513						

a. R Squared = .348 (Adjusted R Squared = .296)

b. Computed using alpha = .05

INCIDENT DURATION – TYPE OF INCIDENT, DETECTION MODE

Between-Subjects Factors

		N
Detection Mode	531 radio	27
	Camera	575
	Courtesy Patrol	718
	DART	18
	DPD	19
	Garland PD	3
	LBJ Express	5
	Media	47
	Mesquite PD	4
	Police	3
	Public	95

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	276713.360[a]	33	8385.253	18.928	.000	.297	624.637	1.000
Intercept	116329.302	1	116329.302	262.595	.000	.151	262.595	1.000
TypeText	11190.829	3	3730.276	8.421	.000	.017	25.262	.994
DetectionMode	15963.368	10	1596.337	3.603	.000	.024	36.035	.995
TypeText * DetectionMode	20769.365	20	1038.468	2.344	.001	.031	46.884	.997
Error	655637.856	1480	442.999					
Total	2312286.000	1514						
Corrected Total	932351.215	1513						

a. R Squared = .297 (Adjusted R Squared = .281)

b. Computed using alpha = .05

INCIDENT DURATION – MAX NO. OF AFFECTED LANE, NO. OF INVOLVED VEHICLE, TYPE OF INCIDENT, DETECTION MODE

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	300717.397[a]	35	8591.926	20.105	.000	.323	703.668	1.000
Intercept	67400.155	1	67400.155	157.714	.000	.096	157.714	1.000
MaxNumberOfAffectedMainLanes	19576.291	1	19576.291	45.808	.000	.030	45.808	1.000
NumberOfInvolvedVehicles	3839.863	1	3839.863	8.985	.003	.006	8.985	.850
DetectionMode	10165.222	10	1016.522	2.379	.009	.016	23.786	.943
TypeText	9403.072	3	3134.357	7.334	.000	.015	22.003	.985
DetectionMode * TypeText	16733.713	20	836.686	1.958	.007	.026	39.156	.987
Error	631633.818	1478	427.357					
Total	2312286.000	1514						
Corrected Total	932351.215	1513						

a. R Squared = .323 (Adjusted R Squared = .306)

b. Computed using alpha = .05

INCIDENT DURATION – LANE INVOLVEMENT RATIO

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	198524.827[a]	1	198524.827	410.222	.000	.214	410.222	1.000
Intercept	536269.215	1	536269.215	1108.119	.000	.424	1108.119	1.000
LaneInvolvmentRatio	198524.827	1	198524.827	410.222	.000	.214	410.222	1.000
Error	728821.736	1506	483.945					
Total	2295850.000	1508						
Corrected Total	927346.562	1507						

a. R Squared = .214 (Adjusted R Squared = .214)

b. Computed using alpha = .05

INCIDENT DURATION – INCIDENT DETECTION MODE

Descriptive Statistics

Dependent Variable:TotalDuration

DetectionMode	Mean	Std. Deviation	N
531 radio	43.93	29.226	27
Camera	40.08	27.547	575
Courtesy Patrol	19.18	15.619	718
DART	48.00	25.864	18
Media	52.57	30.501	47
Police Department	44.14	26.320	28
Public	30.05	20.288	95
Total	30.12	24.806	1508

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	183105.555[a]	6	30517.592	61.548	.000	.197	369.291	1.000
Intercept	473276.884	1	473276.884	954.514	.000	.389	954.514	1.000
DetectionMode	183105.555	6	30517.592	61.548	.000	.197	369.291	1.000
Error	744241.008	1501	495.830					
Total	2295850.000	1508						
Corrected Total	927346.562	1507						

a. R Squared = .197 (Adjusted R Squared = .194)

b. Computed using alpha = .05

INCIDENT DURATION – LANE INVOLVEMENT RATIO, NO. OF INVOLVED VEHICLE, WEATHER, TYPE OF INCIDENT, DETECTION MODE

Between-Subjects Factors

Weather		N
	Clear	20
	Haze	1
	Heavy Rain	1
	Heavy Thunderstorms and Rain	1
	Light Rain	2
	Light Thunderstorms and Rain	3
	Mostly Cloudy	307
	Overcast	23
	Partly Cloudy	578
	Thunderstorm	3

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	335418.108[a]	78	4300.232	10.381	.000	.362	809.747	1.000
Intercept	62033.370	1	62033.370	149.757	.000	.095	149.757	1.000
LaneInvolvmentRatio	24404.982	1	24404.982	58.917	.000	.040	58.917	1.000
NumberOfInvolvedVehicles	4081.092	1	4081.092	9.852	.002	.007	9.852	.880
Weather	10815.575	10	1081.558	2.611	.004	.018	26.110	.963
DetectionMode	7208.785	6	1201.464	2.901	.008	.012	17.403	.898
TypeText	7387.862	3	2462.621	5.945	.000	.012	17.835	.957
Weather * DetectionMode	9812.950	20	490.648	1.184	.258	.016	23.690	.855
Weather * TypeText	10801.279	10	1080.128	2.608	.004	.018	26.076	.963
DetectionMode * TypeText	8002.512	15	533.501	1.288	.201	.013	19.319	.809
Weather * DetectionMode * TypeText	4762.419	11	432.947	1.045	.403	.008	11.497	.593

Error	591928.454	1429	414.226						
Total	2295850.000	1508							
Corrected Total	927346.562	1507							

a. R Squared = .362 (Adjusted R Squared = .327)

b. Computed using alpha = .05

INCIDENT DURATION – DETECTION MODE

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	183105.555[a]	6	30517.592	61.548	.000	.197	369.291	1.000
Intercept	473276.884	1	473276.884	954.514	.000	.389	954.514	1.000
DetectionMode	183105.555	6	30517.592	61.548	.000	.197	369.291	1.000
Error	744241.008	1501	495.830					
Total	2295850.000	1508						
Corrected Total	927346.562	1507						

a. R Squared = .197 (Adjusted R Squared = .194)

b. Computed using alpha = .05

INCIDENT DURATION – MODIFIED DETECTION MODE

Descriptive Statistics

Dependent Variable:TotalDuration

ModifiedDetectionMode	Mean	Std. Deviation	N
Camera	40.08	27.547	575
Courtesy Patrol	19.18	15.619	718
Other	47.98	28.540	120
Public	30.05	20.288	95
Total	30.12	24.806	1508

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	181257.399[a]	3	60419.133	121.796	.000	.195	365.387	1.000
Intercept	857030.566	1	857030.566	1727.641	.000	.535	1727.641	1.000
ModifiedDetectionMode	181257.399	3	60419.133	121.796	.000	.195	365.387	1.000
Error	746089.163	1504	496.070					
Total	2295850.000	1508						
Corrected Total	927346.562	1507						

a. R Squared = .195 (Adjusted R Squared = .194)

b. Computed using alpha = .05

INCIDENT DURATION – LANE INVOLVEMENT RATIO, MODIFIED DETECTION MODE

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	242946.538[a]	4	60736.635	133.383	.000	.262	533.531	1.000
Intercept	411148.561	1	411148.561	902.917	.000	.375	902.917	1.000
LaneInvolvmentRatio	61689.139	1	61689.139	135.475	.000	.083	135.475	1.000
ModifiedDetectionMode	44421.712	3	14807.237	32.518	.000	.061	97.554	1.000
Error	684400.024	1503	455.356					
Total	2295850.000	1508						
Corrected Total	927346.562	1507						

a. R Squared = .262 (Adjusted R Squared = .260)

b. Computed using alpha = .05

INCIDENT DURATION – LANE INVOLVEMENT RATIO, TYPE OF INCIDENT, MODIFIED DETECTION MODE

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	284467.072[a]	16	17779.192	41.234	.000	.307	659.751	1.000
Intercept	78668.323	1	78668.323	182.452	.000	.109	182.452	1.000
LaneInvolvmentRatio	22923.959	1	22923.959	53.166	.000	.034	53.166	1.000
TypeText	16636.672	3	5545.557	12.862	.000	.025	38.585	1.000
ModifiedDetectionMode	5222.478	3	1740.826	4.037	.007	.008	12.112	.843
TypeText * ModifiedDetectionMode	6621.138	9	735.682	1.706	.083	.010	15.356	.788
Error	642879.490	1491	431.173					
Total	2295850.000	1508						
Corrected Total	927346.562	1507						

a. R Squared = .307 (Adjusted R Squared = .299)

b. Computed using alpha = .05

INCIDENT DURATION – MODIFIED WEATHER

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	67527.972[a]	1	67527.972	118.277	.000	.073
Intercept	1181854.961	1	1181854.961	2070.057	.000	.579
ModifiedWeather	67527.972	1	67527.972	118.277	.000	.073
Error	859818.591	1506	570.929			
Total	2295850.000	1508				
Corrected Total	927346.562	1507				

a. R Squared = .073 (Adjusted R Squared = .072)

Parameter Estimates

Dependent Variable:TotalDuration

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Partial Eta Squared
					Lower Bound	Upper Bound	
Intercept	43.666	1.389	31.441	.000	40.941	46.390	.39
[ModifiedWeather=1.00]	-16.848	1.549	-10.876	.000	-19.887	-13.809	.07
[ModifiedWeather=2.00]	0[a]	.	.	.	.	.	

a. This parameter is set to zero because it is redundant.

INCIDENT DURATION – MODIFIED WEATHER, ROAD ID

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	92399.683[a]	7	13199.955	23.714	.000	.100
Intercept	152647.927	1	152647.927	274.235	.000	.155
ModifiedWeather	5970.039	1	5970.039	10.725	.001	.007
Road ID	1004.230	3	334.743	.601	.614	.001
ModifiedWeather * Road ID	3756.203	3	1252.068	2.249	.081	.004
Error	834946.879	1500	556.631			
Total	2295850.000	1508				
Corrected Total	927346.562	1507				

a. R Squared = .100 (Adjusted R Squared = .095)

INCIDENT DURATION – LANE INVOLVEMENT RATIO, NO. OF INVOLVED VEHICLE, MODIFIED WEATHER, MODIFIED DETECTION MODE

Tests of Between-Subjects Effects

Dependent Variable:TotalDuration

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	240955.369[a]	9	26772.819	58.430	.000	.260
Intercept	241841.647	1	241841.647	527.802	.000	.261
NumberOfInvolvedVehicles	8369.803	1	8369.803	18.267	.000	.012
LaneInvolvmentRatio	39309.056	1	39309.056	85.789	.000	.054
ModifiedWeather	27.077	1	27.077	.059	.808	.000
ModifiedDetectionMode	19812.044	3	6604.015	14.413	.000	.028
ModifiedWeather * ModifiedDetectionMode	244.431	3	81.477	.178	.911	.000
Error	686391.194	1498	458.205			
Total	2295850.000	1508				
Corrected Total	927346.562	1507				

a. R Squared = .260 (Adjusted R Squared = .255)

References

1. Abuelela, Mahmoud, Stephan Olariu, and Gongjun Yan. "Enhancing automatic incident detection techniques through vehicle to infrastructure communication." In Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on, pp. 447-452. IEEE, 2008.
2. Payne H, Tignor S. “Freeway incident-detection algorithms based on decision trees with states.” Transportation Research Record (682), 1978.
3. Levin M, Krause G. “Incident detection: a Bayesian approach.” Transportation Research Record (682), 1978.
4. Persaud B, Hall F. “Catastrophe theory and patterns in 30-second freeway traffic data-implications for incident detection.” Transportation Research Part A: General 1989, 23(2), pp. 103–113.
5. Ahmed, S. A., and Cook, A. R. “Application of Time Series Analysis Techniques to Freeway Incident Detection.” Transportation Research Record 682, TRB, Washington, D.C., 1980, pp. 19-21.
6. TAN Z., LU X., “A Combination Algorithm of Freeway Traffic Automatic Incident Detection,” ASCE First International Conference on Transportation Information and Safety (ICTIS), Volume I: Highway Transportation, pp. 1106-1112, 2011.
7. Trivedi M, Mikic I, Kogut G. “Distributed video networks for incident detection and management.” Proceedings of Intelligent Transportation Systems. IEEE, 2000, pp. 155–160.
8. Parkany, E. and C. Xie, “A Complete Review of Incident Detection Algorithms and Their Deployment: What Works and What Doesn’t,” The New England Transportation Consortium, 2005.
9. Karim, Asim, and Hojjat Adeli. "Incident detection algorithm using wavelet energy representation of traffic patterns." Journal of Transportation Engineering128, no. 3, 2002, pp. 232-242.

10. Persaud, Bhagwant N., and Fred L. Hall. "Catastrophe theory and patterns in 30-second freeway traffic data—Implications for incident detection."Transportation Research Part A: General 23, no. 2, 1989, pp. 103-113.
11. Cook, A. R., and Cleveland, D. E. ‘‘Detection of freeway capacity-reducing incidents by traffic stream measurements.’’ Transp. Research Record, 495, Transportation Research Board, Washington, D.C., 1974, pp. 1–11.
12. Ma Y. "A Real-Time Traffic Condition Assessment and Prediction Framework Using Vehicle-Infrastructure Integration (Vii) with Computational Intelligence." Ph.D. Dissertation, Clemson University, Clemson, SC, USA. Advisor(s) Mashrur A. Chowdhury, 2008.
13. FHWA: ‘VII architecture and functional requirements version 1.1’. ITS Joint Program Office, US DOT, 2005.
14. PATH: ‘VII California bay area test bed development plan’, 4 April 2006.
15. ITS America: ‘Primer on VII’, http://www.itsa.org/itsa/files/pdf/VIIPrimer.pdf, accessed 11 July 2007.
16. Hi-ri-o-tappa, K., Likitkhajorn, C., Poolsawat, A., Thajchayapong, S., "Traffic incident detection system using series of point detectors", Intelligent Transportation Systems (ITS),15th International IEEE Conference, 2012, pp. 182-187.
17. Xiao, Jianli, Xiang Gao, Qing-Jie Kong, and Yuncai Liu. "More robust and better: a multiple kernel support vector machine ensemble approach for traffic incident detection." Journal of Advanced Transportation, 2013, pp. 858-875.
18. Samant A, Adeli H. "Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis." Computer-Aided Civil and Infrastructure Engineering 2000, 15(4), pp. 241–250.
19. Cristianini N. and J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge, U.K.: Cambridge Univ. Press, 2000.

20. Yuan, Fang, and Ruey Long Cheu. "Incident detection using support vector machines." Transportation Research Part C: Emerging Technologies 11, no. 3, 2003, pp. 309-328.
21. Vapnik, Vladimir. "The support vector method of function estimation." InNonlinear Modeling, Springer US, 1998, pp. 55-85.
22. Tigor, S.C. and Payne, H.J. (1977). "*Improved freeway incident detection algorithms.*" Public Roads, Vol. 41, No. 1, FHWA, pp. 32-40.
23. Levin, M. and Krause, G.M. (1978). "*Incident detection: a Bayesian approach.*" Transportation Research Record, No. 682, TRB, National Research Council, pp. 52-58.
24. Ahmed, M.S. and Cook, A.R. (1980). "*Time series models for freeway incident detection.*" Journal of Transportation Engineering, Vol. 106, No. 6, ASCE, pp. 731-745.
25. Ahmed, M.S. and Cook, A.R. (1982). "*Application of time-series analysis techniques to freeway incident detection.*" Transportation Research Board, No. 841, TRB, National Research Council, pp. 19-21.
26. Hall, F.L., Shi, Y. and Atala, G. (1993). "*On-line testing of the McMaster incident detection algorithm under recurrent congestion.*" Transportation Research Record, No. 1394, TRB, National Research Council, pp. 1-7.
27. Adeli, H. and Samant, A. (2000). "*An adaptive conjugate gradient neural network-wavelet model for traffic incident detection.*" Computer-Aided Civil and Infrastructure Engineering, Vol. 15, No. 4, pp. 251.260.
28. Yuan, Fang, and Ruey Long Cheu. "Incident Detection Using Support Vector Machines." Transportation Research Part C: Emerging Technologies 11, no. 3, 2003, pp. 309-328.
29. Chowdhury, M., A. Sadek, Y. Ma, N. Kanhere and P. Bhavsar. "Applications of artificial intelligence paradigms to decision support to real-time traffic management." Transportation Research Record, 1968, Transportation Research Board, Washington D.C., 2006, pp. 92-98

30. Bhavsar, P., Chowdhury, M. A., Sadek, A., Sarasua, W., and Ogle, J. “Decision support system for predicting traffic diversion impacts across transportation networks using support vector regression.” Transportation Research Board Annual Meeting (CD-ROM), Washington D.C., 2007.
31. Chen, Shuyan, Wei Wang, and Henk Van Zuylen. "Construct support vector machine ensemble to detect traffic incident." Expert systems with applications36, no. 8, 2009, pp. 10976-10986.
32. Li, Y., Cai, Y.-z., Yin, R.-p., & Xu, X.-m. “Fault diagnosis based on support vector machine ensemble.” In Proceedings of the fourth international conference on machine learning and cybernetics, Guangzhou, August , 2005, pp. 3309–3314.
33. Xiao, Jianli, and Yuncai Liu. "Traffic Incident Detection Using Multiple-Kernel Support Vector Machine." Transportation Research Record: Journal of the Transportation Research Board 2324, no. 1, 2012, pp. 44-52.
34. Hsu, Chih-Wei, Chih-Chung Chang, and Chih-Jen Lin. "A practical guide to support vector classification." National Taiwan University, Taipei. (2010). Available at: https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf. Accessed 7/30/2014.
35. Xiao, Jianli, and Yuncai Liu. "Traffic incident detection by multiple kernel support vector machine ensemble." In Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference, 2012, pp. 1669-1673.
36. Ma, Yongchang, Mashrur Chowdhury, Mansoureh Jeihani, and Ryan Fries. "Accelerated incident detection across transportation networks using vehicle kinetics and support vector machine in cooperation with infrastructure agents."*Intelligent Transport Systems*, IET 4, no. 4, 2010, pp 328-337.
37. Qu, Xu, Wei Wang, Wenfu Wang, and Pan Liu. "Real-time freeway sideswipe crash prediction by support vector machine." IET Intelligent Transport Systems7, 2013, no. 4, pp. 445-453.

38. Motamed, M., Machemehl, R. B., “Freeway Real-time Incident Detection Using SVM During the Peak Hour", Submitted to Transportation Research Part C, 2014.
39. Keerthi, S. Sathiya, and Chih-Jen Lin. "Asymptotic behaviors of support vector machines with Gaussian kernel." Neural computation 15, no. 7, 2003, pp. 1667-1689.
40. Jin, X., D. Srinivasan, and R.L. Cheu. , “Classification of freeway traffic patterns for incident detection using constructive probabilistic neural networks,” IEEE Transactions on Neural Networks, vol. 12, no. 5, 2001, pp. 1173–1187. Accessible at: http://www.hindawi.com/journals/mpe/2014/383671/
41. Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." ACM Transactions on Intelligent Systems and Technology (TIST)2, no. 3, 2011,pp. 1- 27.
42. Singh A., “Practical Issues in Machine Learning Overfitting and Model selection and Model selection”, Carnegie Mellon, School of computer science, Feb 3, 2010. Available at: http://www.cs.cmu.edu/~epxing/Class/10701-10s/Lecture/lecture8.pdf
43. MARTIN P.T., PERRIN J., HANSEN B.: ‘Incident detection algorithm evaluation’, Prepared for Utah Department of Transportation, March 2001.
44. Cheu, Ruey Long, Dipti Srinivasan, and Eng Tian Teh. "Support vector machine models for freeway incident detection." In Intelligent Transportation Systems Proceedings.IEEE, vol. 1,2003, pp. 238-243.
45. Hi-ri-o-tappa, K., et.al, “A Novel Approach of Dynamic Time Warping for Short-Term Traffic Congestion Prediction,” Transportation Research Board of the National Academies, Washington, D.C., 2011.
46. Oh, J.-S., C. Oh, S. G. Ritchie, and M. Chang, “Real-Time Estimation of Accident Likelihood for Safety Enhancement,” ASCE Journal of Transportation Engineering, pp. 358–363, 2005.
47. Vapnik, Vladimir. The nature of statistical learning theory. springer, Berlin, 1995.

48. Lin, Chih Jen, C. W. Hsu, and C. C. Chang. "A practical guide to support vector classification." National Taiwan University, 2003. Accessible at: www. csie. ntu. edu. tw/cjlin/papers/guide/guide. pdf (2003).
49. Chang, Chih-chung, and Chih-jen Lin. “LIBSVM: A library for support vector machines.” 2001. Software available at: http://www. csie. ntu. edu. tw/cjlin/libsvm.
50. Hi-ri-o-tappa, K., C. Likitkhajorn, A. Poolsawat, and S. Thajchayapong. "Traffic incident detection system using series of point detectors." In Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on, pp. 182-187. IEEE, 2012.
51. Ruey Long Cheu, Srinivasan, D. ; Eng Tian Teh, “Support vector machine models for freeway incident detection.” Intelligent Transportation Systems, Proceedings. 2003 IEEE (Volume:1), 2003, pp. 238-243.
52. Huang B, Kinsner W. “ECG Frame Classification Using Dynamic Time Warping." In W Kinsner, A Sebak, K Ferens (eds.), Proceedings of the Canadian Conference on Electrical and Computer Engineering- IEEE CCECE 2002, volume 2, pp. 1105-1110, 2002.
53. Syeda-Mahmood T, Beymer D, Wang F. “Shape-Based Matching of ECG Recordings." In A Dittmar, J Clark, E McAdams, N Lovell (eds.), Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE, pp. 2012-2018, 2007.
54. Wei L, Keogh E, Xi X. “SAXually Explicit Images: Finding Unusual Shapes." In CW Clifton, N Zhong, J Liu, BW Wah, X Wu (eds.), Sixth International Conference on Data Mining 2006-ICDM '06, pp. 711-720. IEEE Computer Society, Los Alamitos, CA, USA 2006.
55. Chandrasekaran, Gayathri, Tam Vu, Alexander Varshavsky, Marco Gruteser, Richard P. Martin, Jie Yang, and Yingying Chen. "Tracking vehicular speed variations by warping mobile phone signal strengths." In Pervasive Computing and Communications (PerCom), 2011 IEEE International Conference on, pp. 213-221. IEEE, 2011.
56. Zhang, Ning, Yi Shi, and Wei Huang. "Traffic Event Automatic Detection Based on OGS-DTW Algorithm." Journal of Highway and Transportation Research and Development (English Edition) 6, no. 1 (2012): 54-60.
57. Golob, T.F., Recker, L. (1986). “An analysis of the severity and incident duration of truck-involved freeway accident.” Institute of Transportation Studies, Univ. of California, Irvine. UCI-ITS-WP-86-9.

58. Wu, Wei-wei, Shu-yan Chen, and Chang-jiang Zheng. "Traffic incident duration prediction based on support vector regression." Proceedings of the ICCTP 2011 (2011): 2412-2421.
59. Chang, Hsin-Li, and Tse-Pin Chang. "Prediction of Freeway Incident Duration based on Classification Tree Analysis." Journal of the Eastern Asia Society for Transportation Studies 10 (2013): 1964-1977.
60. Li, Ruimin, and Pan Shang. "Incident duration modeling using flexible parametric hazard-based models." Computational intelligence and neuroscience 2014 (2014): 723427-723427.
61. Hojati, Ahmad Tavassoli, Luis Ferreira, Simon Washington, Phil Charles, and Ameneh Shobeirinejad. "Modelling total duration of traffic incidents including incident detection and recovery time." Accident Analysis & Prevention 71 (2014): 296-305.
62. Li, Ruimin, Francisco C. Pereira, and Moshe E. Ben-Akiva. "Competing risks mixture model for traffic incident duration prediction." Accident Analysis & Prevention 75 (2015): 192-201.
63. Kang, Guoxiang, and Shou-en Fang. "Applying survival analysis approach to traffic incident duration prediction." In Reston, VA: ASCEProceedings of the First International Conference on Transportation Information and Safety, June 30. July 2, 2011, Wuhan, China| d 20110000. American Society of Civil Engineers, 2011.
64. Demiroluk, Sami, and Kaan Ozbay. "Structure learning for the estimation of non-parametric incident duration prediction models." In Transportation Research Board 90th Annual Meeting, no. 11-3143. 2011.
65. Zhang, Kun, and Michael AP Taylor. "Effective arterial road incident detection: a Bayesian network based algorithm." Transportation Research Part C: Emerging Technologies 14, no. 6 (2006): 403-417.
66. Šingliar, Tomáš, and Miloš Hauskrecht. "Learning to detect incidents from noisily labeled data." *Machine learning* 79, no. 3 (2010): 335-354.
67. Pande, A., C. Nuworsoo, and C. Shew. "Proactive Assessment of Accident Risk to Improve Safety on a System of Freeways." California Polytechnic State University, San Luis Obispo, 2012.
68. Breiman, Leo. "Some infinity theory for predictor ensembles." Technical Report 579, Statistics Dept. UCB, 2000.

69. Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.
70. Liu, Qingchao, Jian Lu, and Shuyan Chen. "Traffic Incident Detection Using Random Forest." In Transportation Research Board 92nd Annual Meeting, no. 13-1610. 2013.
71. Liu, Qingchao, Jian Lu, Shuyan Chen, and Kangjia Zhao. "Multiple Naïve Bayes Classifiers Ensemble for Traffic Incident Detection." Mathematical Problems in Engineering 2014 (2014).
72. Berzal, Fernando, Juan-Carlos Cubero, Fernando Cuenca, and María J. Martín-Bautista. "On the quest for easy-to-understand splitting rules." Data & Knowledge Engineering 44, no. 1 (2003): 31-48.
73. Sun, Jie, and Jian Sun. "A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data." Transportation Research Part C: Emerging Technologies 54 (2015): 176-186.
74. Hossain, Moinul, and Yasunori Muromachi. "A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways." Accident Analysis & Prevention 45 (2012): 373-381.
75. Opitz, David, and Richard Maclin. "Popular ensemble methods: An empirical study." Journal of Artificial Intelligence Research (1999): 169-198.
76. B.V. Dasarathy and B.V. Sheela, “Composite classifier system design: Concepts and methodology,” Proceedings of the IEEE, vol. 67, no. 5, pp. 708–713, 1979.
77. L.K. Hansen and P. Salamon, “Neural network ensembles,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 10, pp. 993–1001, 1990.
78. Chen, Shuyan, Wei Wang, Gaofeng Qu, and Jian Lu. "Application of neural network ensembles to incident detection." In Integration Technology, 2007. ICIT'07. IEEE International Conference on, pp. 388-393. IEEE, 2007.
79. Asgari, H., and X., Jin (2015). 'Towards a Comprehensive Telecommuting Analysis Framework; Setting the Conceptual Outline", Transportation Research Record: Journal of the Transportation Research Board, No. 2496, pp.1-9.
80. Asgari, H., Jin, X. & A. Mohseni (2014). ‘Choice, Frequency, and Engagement – A Framework for Telecommuting Behavior Analysis and Modeling’. Transportation Research Record: Journal of the Transportation Research Board, No. 2413, pp. 101-109.