Summer 2019

Developing Algorithms to Detect Incidents on Freeways From Loop Detector and Vehicle Re-Identification Data

Biraj Adhikari
Old Dominion University, badhi002@odu.edu

DEVELOPING ALGORITHMS TO DETECT INCIDENTS ON FREEWAYS FROM LOOP DETECTOR AND VEHICLE RE-IDENTIFICATION DATA

by

BirajAdhikari
B.S. December 2014, Tribhuvan University

A Thesis Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

CIVIL ENGINEERING

OLD DOMINION UNIVERSITY
August 2019

Approved by:

Mecit Cetin (Director)

Hong Yang (Member)

SherifIshak (Member)

ABSTRACT

DEVELOPING ALGORITHMS TO DETECT INCIDENTS ON FREEWAYS FROM LOOP DETECTOR AND VEHICLE RE-IDENTIFICATION DATA

BirajAdhikari
Old Dominion University, 2019
Director: Dr. Mecit Cetin

A new approach for testing incident detection algorithms has been developed and is presented in this thesis. Two new algorithms were developed and tested taking California #7, which is the most widely used algorithm to date, and SVM (Support Vector Machine), which is considered one of the best performing classifiers, as the baseline for comparisons. Algorithm #B in this study uses data from Vehicle Re-Identification whereas the other three algorithms (California #7, SVM and Algorithm #A) use data from a double loop detector for detection of an incident. A microscopic traffic simulator is used for modeling three types of incident scenarios and generating the input data. Two incident scenarios are generated by closing either one lane or two lanes of a four-lane highway. The third scenario involves bottleneck blocking two lanes of the freeway with an incident occurring in the upstream of the bottleneck. The highway network is five miles long and simulated in VISSIM. Traffic parameters like occupancy, speed, flow and number of vehicles passing through the loop detector are collected to assess the traffic condition between the sensors or detectors. The proposed performance test inspects whether the algorithms thus tested were able to detect any occurrences and incidences within the first minutes in different scenarios and compares their respective detections to identify the best performing algorithm in all the contingencies. The results indicate that the implementation of this new approach not only reduces the dilemma of selecting thresholds but also checks algorithm performance in different incident scenarios so that the response time for clearing such incidences is as short as possible. Likewise, making use of Re-identification data and travel time makes the incident detection more trivial and self-evident and thus outperformed the algorithms using traditional data like occupancy speed and volume in uncongested traffic conditions. Further different SVM models were trained and tested inspecting the effects of change in location of incident concerning detectors. However, using data from loop detector performed well when the incident happened at the upstream detector while using that from re-identification encountered delays in overall detection time for the same.

This thesis is dedicated to the proposition that “The secret of getting ahead is getting started” by Mark Twain.

ACKNOWLEDGMENTS

I want to extend my sincere appreciation to Dr. Mecit Cetin for his assistance, guidance, and in-depth reviews throughout the study process. He has been instrumental in the development of this thesis. Also I would like to thank Dr. Hong Yang for letting me be involved in his research project related to this thesis and for his reviews as well.

I would also like to thank my co-worker Mr.Olcay Sahin for helping me with coding parallel running the algorithms in R. Finally I would like to thank my father Er. Ratnakar Adhikari and my mother Binu Sapkota. They ensured and appreciated the importance of quality education and provided me a lifetime of support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Surface transportation is one of the most widely used means of transportation around the world and chiefly incorporates movement of a vehicle from one point to another. Roads support the safety and wealth of communities. Their resilience is considered one of the vital aspects of governing and managing cities and societies. Congestion on urban roads has been a serious problem in the US for many decades. The total cost associated with traffic congestion is increasing, which is evident from the following data. In 1982 the total cost of traffic congestion was $41 billion, which soared to $112 billion in 2000 and stood at $153 billion by 2010 [1]. Increased congestion resulted in a subsequent increase in incident delays such that incident delays in the US comprised 61% of all urban freeway delays in 1984 and were estimated to have risen to approximately 70% by 2005[2]. During any incident occurrences, transportation infrastructure is directly or indirectly affected, posing a threat to human safety and also causing a significant impact on social and economic facets of a community or country.

It will be worth looking at the definition of an incident at first. An incident is defined as any occurrence that affects roadway capacity, either by obstructing travel lanes or by causing gawkers to block lanes[3]. This is not limited to accidents; it also includes conditions like flat tires, ticket inspections, breakdowns, abandoned vehicles, spills, etc. Incidents can be classified into seven standard types as follows [4]:

1. Abandoned Vehicles;

2. Accidents and fires;

3. Debris on the highway;

4. Failure in mechanical, electrical, fuel, or cooling system leading to towing away of vehicles;

5.Stalled vehicles, which typically need brief roadside attention only;

6. Tire problems;

7. Other issues, including miscellaneous events such as pedestrians walking along freeways, roadside fires, etc.

Traffic flow is reduced by incidents either directly, by lane closure, or indirectly, by gawkers slowing down to look at an incident. It was found that congestion due to such incidents constitutes

somewhere between one-half and three-fourths of the total congestion on urban freeways[5]. The incidents that have the most adverse effect on travel times are the ones in which there is a lane blockage. Even in freeways where there are shoulders, it was found that about 20% of the incidents lead to lane blockage[3].

Incidents can be a major source of congestion on freeways. Minor incidents, which can be short duration fender-benders, may have minor effects on traffic flow. However, some large to medium scale incidents block several lanes and can even last for hours disrupting normal traffic and forming several mile-long queues. According to research done on urban freeways in the central Puget Sound region of Washington State, lane blocking incidents generally accounts for between 2 and 20 percent of total daily, whereas they are responsible for the delay between 10% and 35% of all non-recurring delay[6]. Consequently, responding to such incident situations, DOT, freeway safety patrol, police, or fire department need to work together to detect, clear and restore the normal traffic flow.

To restore the traffic to normal condition, detecting the incident condition plays a vital role, for which there has been extensive research over time. Numerous techniques have been applied in incident detection model and algorithm, but it is still challenging to understand the correlates of traffic incidents in freeways, and there is plenty of room for improvement in detecting them effectively. The techniques or algorithms used for incident detection need to be robust enough, such as the developed model can identify the incident accurately and precisely. In addition, the time to detect the incident, also called detection time, should be as low as possible. Responding to such incident situations may take less time compared to an algorithm with high detection time.

1.1 INCIDENT DETECTION AND VERIFICATION

Using standard algorithms like the California algorithm, it is possible to detect an incident from point sensor data. Once the incident is detected and confirmed by the algorithms in the traffic control center, the concerned authorities are immediately notified. Then the incident is verified from information obtained from road users or surveillance cameras. An emergency team is deployed to assist those who were involved in the incident. Incident detection based on detector data has been a major research area for many years. The capacity of the freeway drops when the incident occurs, and if the demand is greater than the reduced capacity, the vehicles arriving at the incident location will face delays, and the delays of vehicles arriving will continue to increase as long as the capacity is less than the demand. Many algorithms, both analytic and Artificial Intelligence-based, were developed and tested to get quick and accurate detection of incidents on freeways.

1.1.1 Response

Once the occurrence of the incident has been confirmed, ambulances, tow trucks, freeway safety patrol, police, and fire departments rush to the scene. This is the response phase, and the total time taken, from the time of occurrence of the incident to the time when the emergency response units reach the scene, is known as the response time. The literature indicates that response time on freeways varies from 10 to 25 minutes, and it depends on many factors [4]. In this phase too, the reduced capacity of the freeway continues and causes delay to the vehicles arriving at the scene of the incident.

1.1.2 Clearance

Once emergency personnel reach the scene of an incident, they work on clearing the incident from the freeway and providing assistance to those involved in the incident. Prediction of incident clearance time was discussed by Wang and by ADVANCE project group [7]. Clearance time for incidents was modeled based on the following factors: lane closure, number of cars involved, number of trucks involved, number of personal injuries, number of fatalities, hazmat involvement, fire involvement, the time frame of occurrence, prevailing weather, prevailing temperature, and land use type. Classification and regression tree (CART) analyses were conducted to study the effect of the various hypothesized variables on the clearance of incidents. Based on these results, prediction/decision trees were constructed.

1.1.3 Recovery

Once the incident is cleared, the queue that built up will start dissipating,and traffic flow will eventually reach normal conditions. The time taken for traffic to reach normal circumstances once the incident is cleared,also known as recovery time, is directly dependent on the rate of discharge after the incident. Various researchers have studied the rate at which vehicles are discharged from a bottleneck and after the clearance. It isfound that the average rate at which vehicles are discharged can be 10% lower than the rate measured before the queue's formation based on a study of queues at bottlenecks near Toronto, Canada[8]. In a similar study of a freeway bottleneck in Texas, it was found that the average free- flow rates ranged from 2096 to 2210 vehicles per hour per lane across all lanes; the queue discharge flow rates averaged approximately 2175 vehicles per hour per lane for the study sites[9].

1.2 RESEARCH OBJECTIVES AND METHODOLOGY

This study is based on VISSIMmicrosimulation modeling and simulation run for Incident detection strategy and research. The main objectives of this study can be identified as:

- Creating a simulation of anincident on a freeway from which data can be extracted to apply, evaluate, and study the incident detecting algorithms.

- Calibrating and tuning the algorithm thresholds, tuning parameters, and also the features that are used while training the algorithms so that optimal performance is obtained.
- Developing some new algorithms based on knowledge gained from the analysis of existing algorithms to detect the incident.
- Applying and evaluating the performances of the incident detection algorithms.
- Studying the factors that can affect the algorithms such as bottlenecks and number of lanes blocked and features of the road such as speed, flow, occupancy, and travel time during the incident.

In the study of an incident and its corresponding factors, this thesis makes use of simulation data of different simulated incidents created in simulation software “PTVVISSIM.” For development of the algorithm and analysis of the incident, different scenarios were developed in VISSIM, blocking one to several lanes in different random scenariosand also with different traffic volume.To collect the data, data collection points, which can be double loop detectors in actuality, are placed along the road at every quarter mile. Traffic parameters like occupancy, speed, flow, and number of vehicles passing at the loop detector are collected to assess the traffic condition in between the sensors or detectors; in other words, the traffic conditions are checked for each 0.25-mile segment along the road. Moreover, the study is focused on reviewing the existing algorithms on incident detectionanddeveloping, analyzing, implementing, and experimenting with an algorithm for detecting an incident on a hypothetical road created in VISSIM.

After the simulation, the required data is collected with COM interface using VBA (visual basic application) in Microsoft Excel. This data is later processed and analyzed using incident detection algorithms in R-studio, which is capable of handling big data generated from the simulation software. The details of these processes are explained in later chapters. In Chapter II, a literature review relevant to this study is briefly presented with results and research done by other scholars in this field. Chapter III deals with how the research is done and the approach of the study, describing the detailed processes and assumptions of this thesis. In Chapter IV, different incident scenarios and algorithms used for the study are described in detail. Results are presented in Chapter VI, and finally, conclusions and recommendation for future work are discussed in Chapter VII.

CHAPTER II

LITERATURE REVIEW

2.1 AUTOMATIC INCIDENT DETECTION SYSTEM (AID)

Automatic Incident Detection (AID) systems, which have been in development since the early 1970s, make use of algorithms to analyze traffic data and eventually detect incidents to reduce their adverse effects. A variety of freeway incident algorithms are developed based on traffic flow theory, pattern recognition, statistical techniques and, recently, using artificial intelligence and fuzzy logic. Diverse methods of traffic incident detection are grouped intothesecategories:

- Pattern-based such as California algorithm;
- Statistical based such as standard normal deviate (SND);
- Smoothing/Filtering Algorithms such as double exponential smoothing (DES);
- Image-based;
- Artificial intelligence-based.

2.1.1 Pattern-based

California Algorithm is a comparative algorithm which uses preset thresholds to classify road conditions in real time [10]. The California Algorithm needs only occupancy data from two adjacent detector stations. The algorithm calculates the spatial difference in occupancy, OCCDF, the relative spatial difference of occupancies, OCCRDF, and the relative temporal difference in downstream occupancy, DOCCTD. In the equations below,i indicates thedetector number, i.e. i for upstream detector and i+1 for downstream detector and t for time.

$$OCCDF(i,t) = OCC(i,t) - OCC(i+1,t) \quad (1)$$

$$OCCRDF(i,t) = (OCC(i,t) - OCC(i+1,t)) / OCC(i,t) \quad (2)$$

$$DOCCTD(i,t) = (OCC(i+1,t+2) - OCC(i+1,t)) / OCC(i+1,t) \quad (3)$$

The algorithm predictsin what state the road currently is in after comparisons of thresholds and inputs. Flowcharts for the basic California algorithm and California Algorithm #7 are shown in Figure1 and Figure2 respectively.

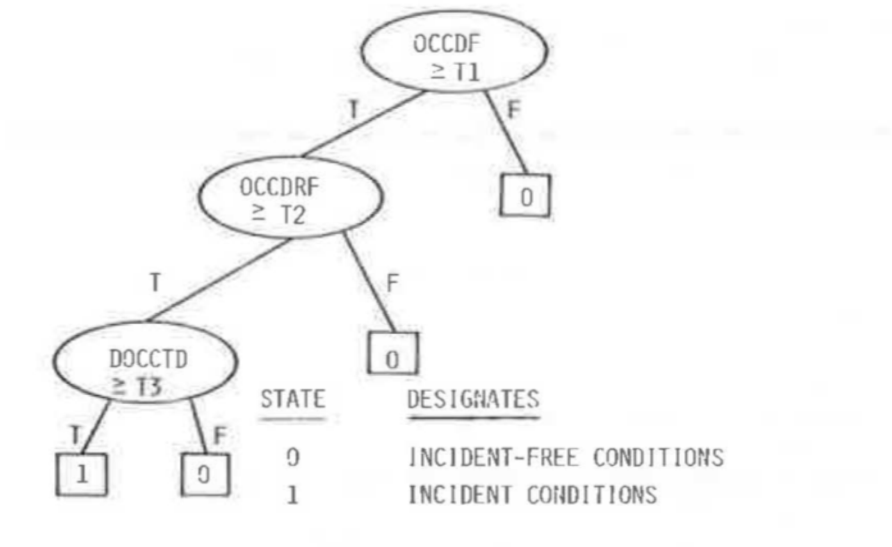


Figure1:Decision tree for basic California algorithm[10]

California Algorithm #7, one of the most widely used algorithms, outputs 4 possible states. The road is in state 0 when there are no incidents; in state 1 when there is a possibility of incident but still there aren't any detected incidents; in state 2 when the incident is detected and state 3 when the incident continues. In California #7 the third parameter was replaced with a measurement of the current downstream occupancy so that compression waves that recur due to false alarms, could be easily acknowledged. A persistence check which required incident conditions to persist for at least two iterations was also added.The logic behind the California Algorithm can be found elsewhere.[10]

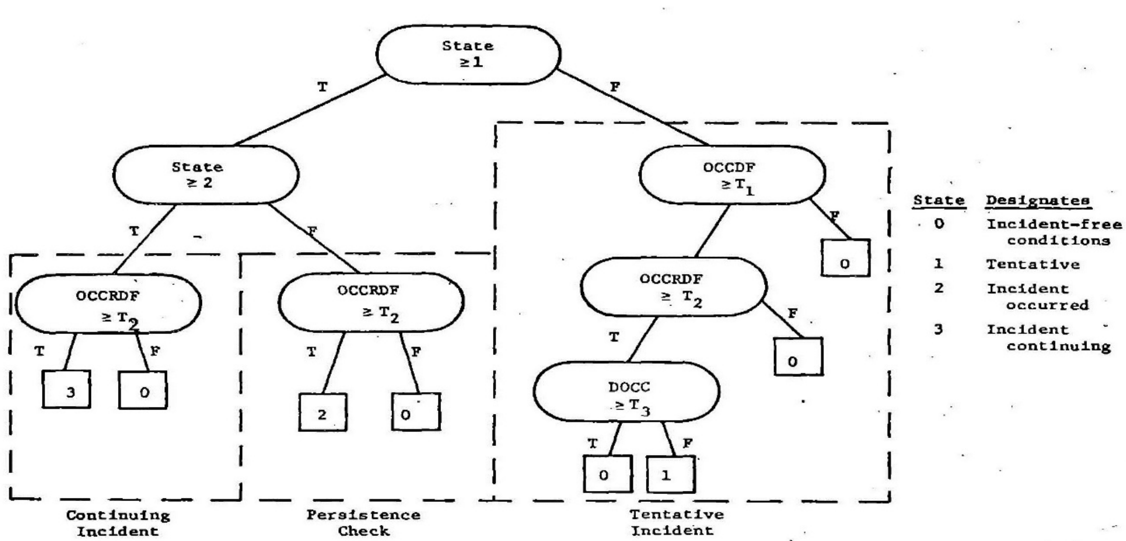


Figure2: Decision tree for California algorithm #7[10]

The ARIMA model, which is another pattern-based algorithm, assumes that differences in a traffic variable are measured in the current time slice t and the same traffic variable in the previous time slice t-1. These variables can be predicted by averaging the errors between the predicted and observed traffic variable from the past three-time slices, which can be further used to detect the incident.[11]

2.1.2 Statistical based

The Standard Normal Deviate (SND) algorithm was developed by the Texas Transportation Institute (TTI) for use in the initial surveillance and control center in Houston, TX in the early 1970s. The SND algorithm calculates the normal standard deviation of the traffic control measure, which is the number of deviations of a particular value of a traffic variable that deviates from the mean of that particular traffic variable. The algorithm’s working principle is based on the principle that a sudden change in a measured traffic variable suggests that an incident has occurred. It compares 1-minute average occupancy measurements to archived occupancy values of the mean and SND defining the thresholds for detecting the incidents. An SND value greater than the critical value indicates the presence of an incident. Two successive intervals are used to make a consistency test.[12]

Similarly, the Bayesian algorithm makes use of the relative difference of the occupancies used in the California algorithms as the traffic measure to compute the likelihood that an incident signal is caused by a lane-blocking incident[13]. However, it calculates the conditional probability using Bayesian statistics. Bayesian theory undertakes the frequency distributions of the upstream and downstream occupancies during the incident and incident-free conditions. Three databases are identified for satisfying the requirement of the Bayesian algorithm: 1) traffic occupancy and volume data during incident conditions; 2) traffic occupancy and volume data during incident-free conditions; and 3) archived data on the type, location, and severity of incidents.

2.1.3 Smoothing/Filtering Algorithms

Smoothing and filtering techniques make use of the mathematical technique for producing a weighted average of a given traffic variable and are designed to remove short-term noise or inhomogeneity from traffic data that cause false alarms and hence permit exact traffic patterns to be more visible to detect true incidents [14] more readily.Filtering algorithms use a linear filter that allows the low-frequency components of the detector data to pass while removing the undesirable high-frequency portions of the detector data. The representative smoothing/filtering algorithms consist of the double exponential smoothing (DES) algorithm [15], low-pass filter (LPF) algorithms [16].

The DES algorithm, expressed mathematically as a double exponential smoothing function, with smoothing constant weights past and present volume, occupancy and speed observations for forecasting short-term traffic conditions that are expected to resemble true traffic conditions closely. Incidents are detected using a tracking signal, which is the algebraic sum (to the present minute) of all the 12 previous errors between the predicted and observed traffic variable. Under incident-free conditions, the tracking signal should dwell around zero since predicted and observed traffic conditions should be similar.

Likewise, theLPF algorithm series, also known as the Minnesota algorithms or the DELOS (detector logic with smoothing) algorithms, is based on a simple comparison of the occupancy levels at two adjacent detector stations. This algorithm removes sharp or high-frequency fluctuations or noise in the data while allowing wide or low-frequency fluctuations normallyrelatedto incident conditions to pass through the pre-set filter. Two filters on two levels were applied to distinguish incident and bottleneck congestion better and hence reduce the false alarm rate. The algorithm employs 3- minute and 5-minute moving average occupancies [16], uses three types of smoothing techniques, i.e., statistical median occupancies [16], or exponential smoothing occupancies to reduce the false alarm rate [16]. A more comprehensive discussion and evaluation of these smoothing algorithms can be found elsewhere [16].

2.1.4 Imaged based

The principle behindthe AIDA algorithm is that it watches for rapid traffic breakdowns, comparing speed and occupancy with the preset thresholds for determining congestion levels. It takes advantage of temporal variations of traffic characteristics in addition to spatial ones. The AIDA was later improved to include ancillary information provided by video detection which uses information about stopped vehicles and shock wave signature recognition. The main advantages of the image processing-based incident detection technique are that a detected incident can be verified visually in a short time. Also, it is capable of monitoring traffic and detecting incidents outside of through lanes, e.g., ramps, shoulders, intersections, orunder any traffic size. Compared with traditional pattern and statistical-based approaches, image-based approaches detect incidents through video captured from CCTV cameras and has been successfully used to detect shoulder incidents accurately[17]. However, the large deployment of CCTV cameras is time-consuming and expensive, and the performance will be suspended once an operator wishes to tilt or zoom in a different view.

2.1.5 AI-based

In the meantime, artificial intelligence algorithms, which regard incident detection as a task to construct a classification model according to traffic metrics such as traffic flow, use diverse approaches such as an artificial neural network (ANN), machine learning (ML), or K-means clustering to offer efficient ways by classifying the traffic state. The classical artificial neural networks (NNs) have been

widely studied to detect freeway incidents and proved able to provide fast and reliable incident detection on freeways [18-20]. Diverse models have been developed such as multilayer feed-forward NN (MLFNN), constructive probabilistic NN (CPNN) [21], and probabilistic neural network (PNN) [22]. Meanwhile, machine learning algorithms have also been applied to detect traffic incidents such as support vector machine (SVM), particle swarm optimization (PSO), random forest (RF).Pattern recognition techniques recognize traffic patterns according to their common characteristics. Some recently introduced algorithms, the Fuzzy ART (Adaptive Resonance Theory) and Fuzzy ARTMAP, work on this principle. A recent study evaluating the algorithms using online techniques [23], showed that Fuzzy ART produces a significantly higher detection rate at the same false alarm rate, compared to California Algorithms #7 and #8. The recently developed Logit - based algorithm attempts to recognize incident patterns by using the incident index. The incident index represents the probability of occurring incidents and is estimated by the multinomial logit model. The model was evaluated on a typical signalized artery in Seoul, and it is reported that most of the incidents could be detected accurately.

2.1.5.1 SVM (Support Vector Machine):

For this study, as a machine learning algorithm, support vector machine (SVM), has also been applied to detect traffic incidents. Introduced by Vapnik in 1995, Support vector machine (SVM) is a comparatively new pattern classifier. Study results have shown that SVM offers a lower misclassification rate, higher correct detection rate, lower false alarm rate and slightly faster detection time than the multi-layer feed-forward neural network (MLF) and probabilistic neural network models in arterial incident detection. Also, on the freeway,SVMs have exhibited incident detection performance as good as the MLF, one of the most promising incident detection models developed to date.[24].

AnSVM, with a decision boundary, developed based on the concept of structural risk minimization (of classification error) using the statistical learning theory classifies an input vector into one of two classes. Its algorithm directly finds a separating hyperplanewhich is optimal by being a maximal margin classifier concerning training data. For non-linearly separable data, the SVM uses a kernel method to transform the original input space, where the data is non-linearly separable, into a higher dimensional feature space where an optimal linear separating hyperplane is constructed.In other words, it treats each data item as a point in n-dimensional space (where n is a number of features you have) with the value of each feature being the value of a particular coordinate that is plotted. Then data are classified by finding the hyperplane that differentiates the two classes very well. Support Vectors are simply the coordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line) as shown in Figure3.Based on its learning approach, the SVM is believed to have areasonably good classification rate for high-dimensional data.

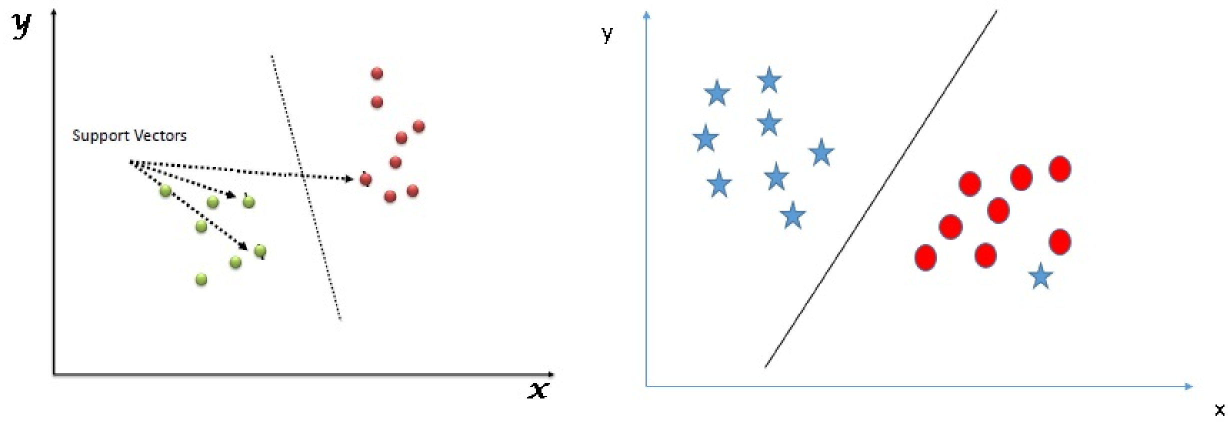


Figure3: Hyper-plane separating two different class or labels.

After that,the hyperplane, which segregates the two classes better, is selected maximizing the distances between the nearest data point (either class). This will help us to decide the right hyper-plane. This distance is called the margin. Another reason for selecting the hyperplane with higher margin is robustness. If we choose a hyper-plane having a low margin, then there is a high chance of miss-classification.

For non-linear hyper-planes, SVM has a technique called the kernel**s**. These are functions which take low dimensional input space and transform it into a higher-dimensional space, i.e. it converts not separable problem to separable problem, and these functions are called kernels. It does some extremely complex data transformations, anddetermines the process for separating the data based on the labels. So, based on this principle, we can easily separate the incident and no – incident state of the road based on the support vectors. For the purpose of incident detectionin this thesis, three SVM models using linear, polynomial, and radial kernels were trained and tested.

It is to be noted that only California algorithm #7, which is one of the most widely used and oftentimes considered as the benchmark for other algorithms, and SVM model, which is regarded as one of the best performing algorithms to date, are used to compare the results obtained from other algorithms developed for this thesis.

2.2 VEHICLE RE-IDENTIFICATION SYSTEM

This section discusses preceding research related to vehicle re-identification systems. First, complementary detector technologies are discussed, and then a basic concept about vehicle re-identification systems and algorithms are presented. There are four emerging detector systems under

Caltrans sponsorship that promise to yield more robust vehicle signatures while being compatible with the re-identification algorithm:

1. Magnetic Vehicle Signatures from Loop Detectors,[25].
2. Vehicle Dimensions and Velocity from Scanning Laser Radar, [26].
3. Vehicle Dimensions and Velocity from Overhead Video Detectors, [27].
4. Visual Vehicle Signatures from Wayside Cameras, [28].

For example, system # 2 above measures vehicle length with an error of 1 inch at free-flow traffic speeds. These emerging technologies use specialized hardware to extract vehicle signatures that are more descriptive than effective length. As a result, the hardware must be deployed before quantifying the benefits. In most cases, the systems have only been installed on small test sites. Other systems use automatic vehicle identification (AVI), e.g., machine-readable "license plates", [29] that make vehicle re-identification trivial, but the systems may compromise privacy. Furthermore, the AVI systems do not measure local conditions at the detectors, and this omission can impact surveillance and control. Although these systems appear promising for free flow and lightly congested conditions, they currently do not perform well under heavy congestion.

2.2.1 Basic Vehicle Re-identification Algorithm and Results:

This section bestows brief details about the vehicle re-identification algorithm. The work builds off of simpler algorithms presented in [30]. This algorithm uses data from two-speed traps,and thenthe concurrent video was collected at each station to serve as ground truth.The section begins by defining the simple vehicle signatures that serve as input parameters and then proceeds through the re-identification algorithm, and finally, the algorithm performance is compared against the ground truth data. Effective vehicle length is the measured by multiplying velocity with time the detector was on. As there are two measurements for each of these parameters, the average of the two measurements is recorded as the effective vehicle.For a given downstream measurement, the algorithm attempts to find the upstream measurement that corresponds to the same vehicle. Unfortunately, it is not possible to find the upstream match directly for several reasons: a vehicle's measured length is not unique, it may be subject to measurement errors, and as noted above it is subject to resolution constraints. But it is possible to eliminate many unlikely matches via this resolution test. Although it is not possible to identify a unique match for an individual vehicle, a sequence of measured lengths rapidly becomes distinct and the sequence can potentially be re-identified at successive detectors. The algorithm looks for short sequences of measured vehicle lengths that exhibit a strong correlation between two stations. Next, a set of reasonable upstream matches is identified for each downstream vehicle; where this set is the last n

successive upstream vehicles in the same lane ending with the last vehicle to pass the upstream speed trap before the downstream observation. Finally, the algorithm identifies the upstream match for each downstream vehicle or each row in the vehicle match matrix. This upstream match corresponds to the column with the longest modified sequence passing through the given row. It is asserted that simple speed trap algorithm found a match for about 90 percent of the vehicles.

Similarly, apart from this basic approach, another researcher, Dr.Cetin the advisor for this thesis, used several modeling approaches to solve the re-identification problem. The research team used Naïve Bayes (NB), Bayesian Models (BM) fitted by mixture models, and the formulation of the re-identification problem as a mathematical assignment problem. The results show that solving the re-identification problem with the mathematical assignment formulation outperforms solving with NB models or Bayesian Models (BM) fitted by mixture models, especially when vehicle-pairs exceeding a high threshold of similarity are matched. [31] This research also concluded that up to 90% matching accuracy is attained when the best combination of re-identification method and similarity measure are implemented, and only those vehicle-pairs exceeding a high threshold of similarity are matched.

2.3 FACTORS AFFECTING ALGORITHM PERFORMANCE

Several factors affect the performance of all types of incident detection algorithms. The key factors that may affect the detection of the incident are shown below:

- Operating Conditions of the Highway

  Some algorithms are good or almost perfectat detecting an incident,but many fail to detect all the operating conditions of the road. The road can be at capacity or well below capacity depending upon the volume of the traffic that can vary from low to medium to high.
- Duration of the Incident

  The incident can be as mundane as a vehicle stopping for a tire check anda flat tire, or it can be as serious as an accident, which can take several minutes to hours to clear. Therefore, the duration of the incident is another factor to be considered for the incident.
- Geometric Factors

  As the geometry of the road may change from location to location, some geometric factors like lane drops, grade, and no ramps can also affect the detection time of the algorithms.
- Environmental

  The road surface is not always the same. Different environmental conditions can change; it may snow, rain, or the road surface might become icy. This may affect the sensors on the road affecting the detection of the incident.

- The severity of the incident.
- Detector spacing
- Location of the incident relative to the detector station.
- Heterogeneity of the vehicle fleet.

2.4 ALGORITHM PERFORMANCE EVALUATION

The most commonly used performance measures are the ability of an algorithm to detect an incident (detection rate) versus its false alarm rate. The detection rate is the number of incidents detected as a percentage of the number of incidents occurred.

$$\beta = \text{detection rate (percent)} = 100 * \frac{N_D}{\mathrm{N_I}} \tag{4}$$

where N_D = number of incidents detected and N_I = number of incidents.

Likewise, the false alarm rate is the number of false alarm signals as a percentage of tests performed by the algorithm.

$$\alpha = \text{False Alarm rate (percent)} = 100\ x \frac{N_{FA}}{N_F} \tag{5}$$

where N_{FA} = total number of false-alarm signals generated by the algorithm and N_F = total number of tests performed by the algorithm

Another performance measurement is average detection time, which is the average time taken by an algorithm to detect the incident or start of the incident.

A summary of the comparative performance of different algorithms is given in table1.

Table 1: Reported Summary of Algorithm Performance

Algorithm		Detection Rate	False Alarm Rate	Average Detection Time
		[%]	[%]	[minutes]
California [10]	Basic	82	1.73	0.85
	California #7	67	0.134	2.91
	California #8	68	0.177	3.04

Table 1: Continued

	APID	86	0.05	2.5
Standard Normal Deviate [12]		92	1.3	1.1
Bayesian [13]		100	0	3.9
Time Series ARIMA [11]		100	1.5	0.4
Exponential Smoothing [14-15]		92	1.87	0.7
Low-Pass Filter [16]		80	0.3	4
Modified McMaster [17]		68	0.0018	2.2
Neural Networks [17-19]	MLF	89	0.01	0.96
	PNN	89	0.012	0.9

Data in the table is taken from different studies, some done in real field tests, whereas others use simulated data. Therefore, the figures indicating the performance may change depending upon the geometry of the road, operating condition, and type of data used. These figures represent the most acceptable values from the literature review.

Table 2: Summarized Table of the Literature Review

Research Paper	Algorithm Used	Type of Data Used	Source Of Data	Detection Rate [%]	False Alarm Rate [%]	Average Detection Time [minutes]	Remarks
(1976)Development And Testing Of Incident Detection Algorithms	California Algorithms	Loop Detectors	Field Data	82	1.73	0.85	
(1993)Smoothing Algorithms For Incident Detection	Double exponential smoothing	Loop Detectors	Field Data(I-35W in Minneapolis)	92	1.87	0.7	
(1995)Automated Detection Of Lane-Blocking Freeway Incidents Using Artificial Neural Network	MLF(Multi-Layer Feedforward)	Loop Detectors	Simulated Data	89	0.01	0.96	
(2001)Classification Of Traffic Incident Patterns	CPNN(Constructive Probabilistic Neural Network)	Loop Detectors	Field Data (I-880 freeway, California.)	91.30%	0.27	7.68 [interval of 30s]	
(2003)Incident Detection Using Support Vector Machines	SVMs	Loop Detectors	Field Data for Freeway (I-880 freeway, California) and Simulation data for arterials	91.3	0.13	4.5 [interval of 30s]	
(2015) Incident Detection Methods Using Probe Vehicles With On-Board GPS Equipment	Algorithm I (From travel time) and Algorithm II (from shockwave)	Probe Data	Field Data (Tokyo Metropolitan Expressway (MEX))	55	4.6	14.8	(In 1% penetration rate- using algorithm 1) Algorithm 2 is very poor in performance

Table 2: Continued

(2018)Fuzzy Deep Learning-Based Urban Traffic Incident Detection	Fuzzy deep neural network (FDNN)	Loop Detectors	Simulated Data	98.23%	0.24%	192.44 [s]	Complex
(1996)Integration Of Probe Vehicle And Induction Loop Data: Estimation Of Travel Times And Automatic Incident Detection	COMETT(Computer Model for Estimation of Travel Times using induction loop detectors)	Both Probe and Loop	Simulation	85%	-	-	Complex(Uses Cumulative flow Diagrams Software)

CHAPTER III

SIMULATION ENVIRONMENT AND DATA PROCESSING

In this chapter, the research approach is explained after some basic explanation of the simulation tool and programming languages used in the study.

3.1 PTV – VISSIM SIMULATION TOOL

In the field of transportation engineering, There are mainly three types of simulation scales for modeling and analysis of traffic: macroscopic, mesoscopic, and microscopic.Simulation models based on the deterministic relationships of the flow, speed, and density of the traffic stream are called macroscopic simulation. The simulation takes place on a section-by-section basis rather than by tracking individual vehicles in a macroscopic model. Macroscopic simulation models were originally developed to model traffic in distinct transportation subnetworks, such as freeways, corridors, surface-street grid networks, and rural highways. Whereas mesoscopic models combine the properties of both microscopic and macroscopic simulation models, these models provide less fidelity than micro simulation tools but are superior to the typical planning analysis techniques. Microscopic simulation models simulate the movement of individual vehicles based on car-following and lane-changing theories. These models are useful in evaluating heavily congested conditions, complex geometric configurations, and system-level impacts of proposed transportation improvements that are beyond the limitations of other tool types.

This research is accomplished in VISSIM (Version 10.1), which is a microscopic simulation tool. PTV VISSIM allows simulating traffic patterns exactly. Motorized private transport, goods transport, rail and road-related public transport, pedestrians and cyclists – as the world's leading software for microscopic traffic simulation.PTV VISSIM displays all road users and their interactions in one model. The software offers flexibility in several respects: the concept of links and connectors allows users to model geometries with any level of complexity. Attributes for driver and vehicle characteristics enable individual parameterization. VISSIM also has other features like realistic lane geometry, accurately representing the position of all network elements, a proven car-following model, a detailed simulation of lane changing and merging which was very useful to model traffic and vehicle behavior during the simulation.

3.2 VISSSIM COM INTERFACE

Although VISSIM offers a user-friendly graphical interface (GUI) through which one can design the geometry of any road networks and set up simulations in a simple way, for several problems the GUI is not satisfying. For example, when the user aims to access and manipulate VISSIM objects during the simulation dynamically. Therefore, an additional interface is offered based on COM, which is a technology to enable inter-process communication between software (Box, 1998). The VISSIM COM interface defines a hierarchical model in which the functions and parameters of the simulator initially provided by the GUI can be manipulated by programming. It can be programmed in any language that can handle COM objects (e.g., C++, Visual Basic, Java, etc.). Through VISSIM COM, the user can manipulate the attributes of most of the internal objects dynamically. In this research, Visual Basic Application (VBA), which is inbuilt in MS-EXCEL, is used as a programming language to handle the COM objects.

3.3 DATA COLLECTION

As this research was completely based on simulation, the data in this study is collected from the simulation models created in VISSIM.There were mainly two types of data collected from the simulation models: detector data and trajectory data. Detector data was collected every 30 seconds; these data are average occupancy, speed and number of vehicles passing through the detector placed along the road in some spacing (a quarter mile in this study). The data collected by the sensor in VISSIM may reflect the data obtained from the double loop detector in real life. The trajectory data is the record of each vehicle, its position, and its speed at every time step of 1sec.These data may reflect the data collected from the GPS equipment in real life.

In total, 62 simulations were run per scenario, 31 for training and 31 for testing. There are three different scenarios which will be explained in later chapters. Apart from these three scenarios, data was also collected for the no-incident condition. Consequently, seven data sets were collected, 6 datasets for training and testing for everythree scenarios, and 1 data set for the no-incident condition.

3.4 R-PROGRAMMING

R is a programming language and free software environment for statistical computing and graphics supported by the R Foundation. Although R has a command-line interface, there are several graphical user interfaces, such as R-Studio, an integrated development environment. In this study, the algorithms used to detect the incidents were written in R- programing language mainly because of its ability to handle big data, such as trajectory data, effectively and efficiently. The data are loaded in R-studio, and algorithms are developed, calibrated, and tested; Also, the results were evaluated and analyzed using this platform. In general, the applied analysis approach can be seen in Figure4.

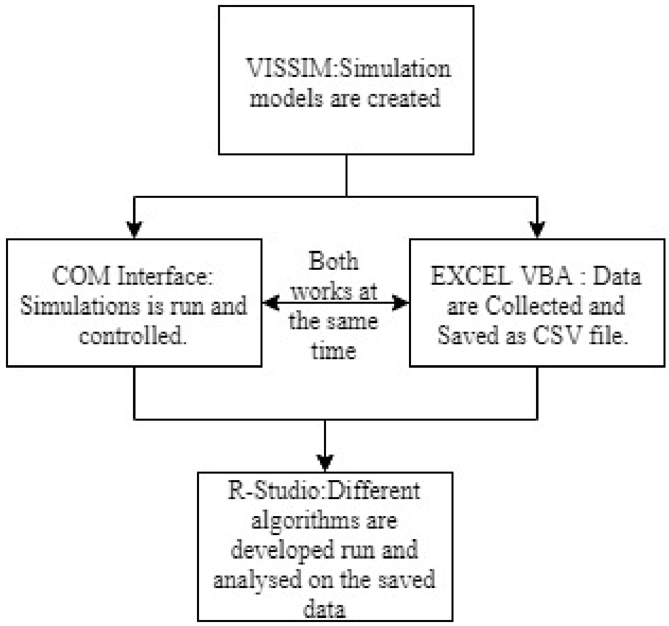


Figure4: Flowchart for Applied Analysis Process.

3.5 ASSUMPTIONS

There are no significant assumptions considered for the analysis in this study. However, the following assumptions are taken into consideration in general in simulation models created:

1. It is considered that change in the roadside environment does not affect driver behavior within the traveled way (e.g., distraction due to stalled vehicles, obstruction of visibility, etc.)
2. In driving behavior, “100 percent safe driving” is assumed, so driving behavior will not be changed throughout the simulation even after the incident.
3. The incident is assumed and modeled as a vehicle stopping in the road as in a parking lot, in a parking lot model, but the combination of signals, lane block option and reduced speed area was used as a final model; the signal is used to stop vehicles assumed to be in an incident.
4. All the vehicle inputs are stochastic, and only passenger carsare used in the final model.

3.6 SIMULATION AND MODEL DEVELOPMENT

This chapter deals with standard input parameters for the simulation model as well as every single detail of each scenario considered for the different simulation models. At first, basic inputs and driving behavior and different layouts used in the simulation will be discussed, and a detailed description of the algorithm and its development and application in detecting incidents are thoroughly described.

3.7 BASIC INPUTS

Many input parameters have been used for the analysis; among them, some basic input parameters used in the simulation are:

- Desired speed distribution for all vehicles = 60-65 mph,
- Incident time = 15 minutes,
- Simulation run time = 1800 seconds,
- Traffic volumes in each scenario = 5000 to 8000 with an increment of 100 (31 different volumes),
- Total number of simulation per scenario = 62 (31 for training + 31 for testing),
- Random Seeds = 369(test) & 738(train),
- Simulation Resolution = 10 time-step(s) per simulation second.

There are 31 simulations for training for each volume from 5000 to 8000 with an increment of 100 vehicles per hour and 31 simulations for testing as well; the only difference in training and testing is a random seed. There are three different scenarios. Namely, an incident closing one lane, an incident closing two lanes, and an incident closing two lanes with a work zone at an upstream location. Again,five different algorithms with different input parameters are used to train these data sets. The chart in Figure5shows the whole picture of the algorithms, data types, input features used by each algorithm, and scenarios considered in this thesis.

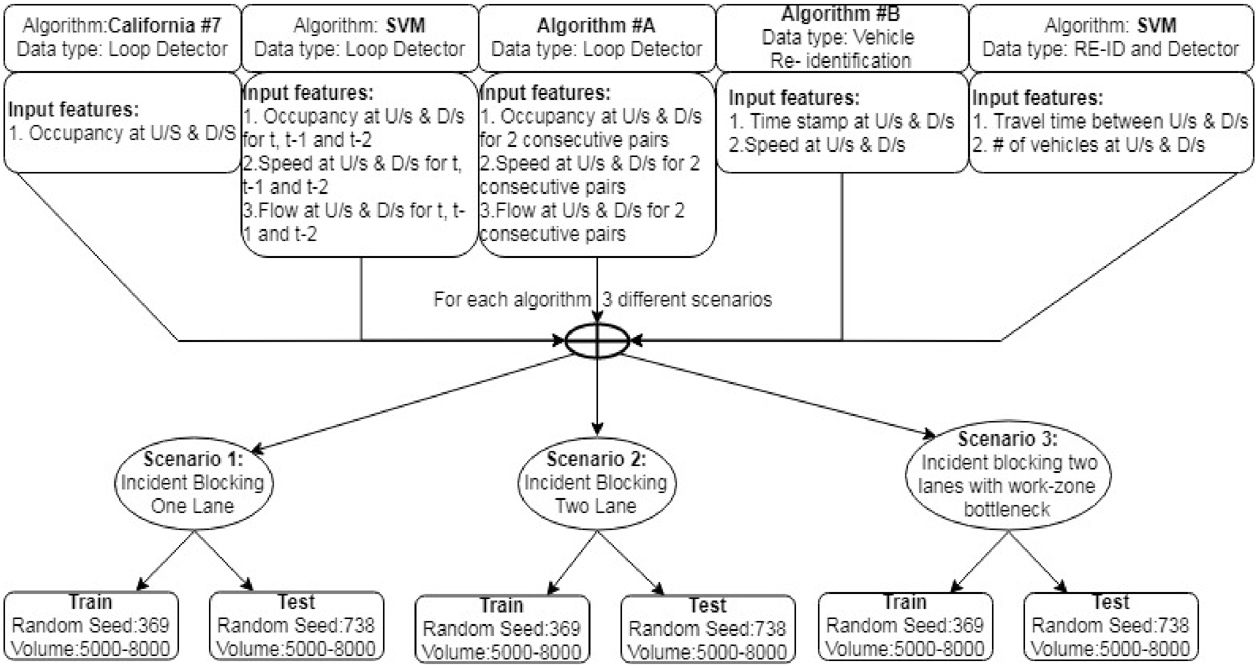


Figure5: Chart Showing Algorithm and Scenarios Considered.

3.8 DRIVING BEHAVIOR

Driving behavior is an important input parameter. It deals with the micro-level driving characteristics,which are followed by the vehicles in the simulation. Figure6 below shows the input window for driving behavior parameter sets passed in VISSIM.

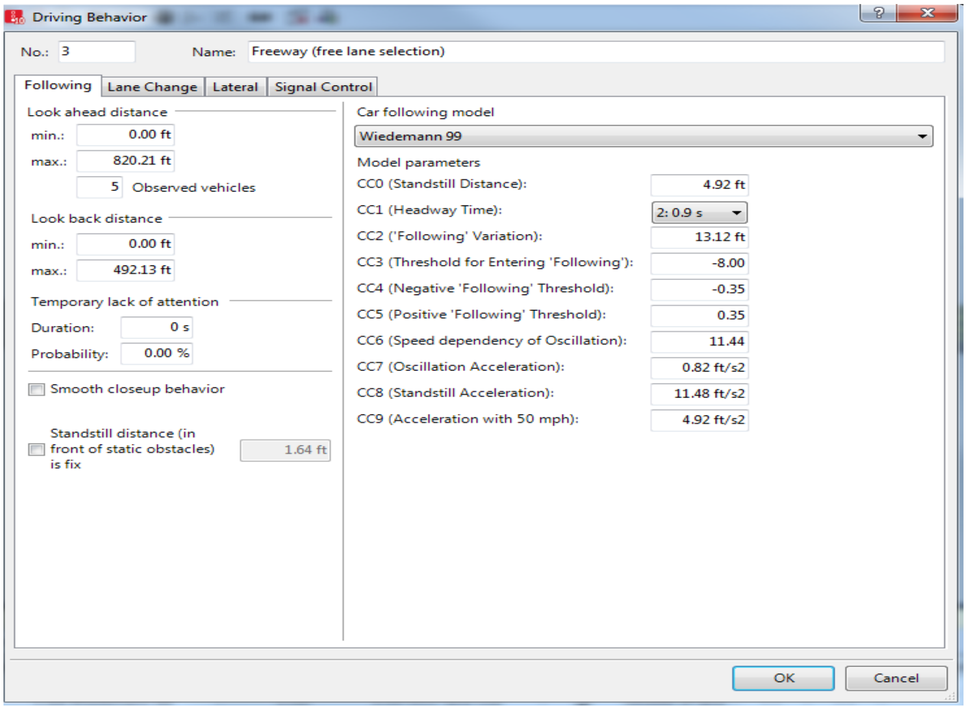


Figure6: Input Window for driving behavior Model Parameters

For this study, Freeway (free lane selection) driving characteristics are considered as we are modeling a 4 lane highway. VISSIM is a state-of-the-art micro-simulation model that deals with behavioral model analysis like Wiedemann theory. The Wiedemann model assumes that a driver can be in one of four driving modes:

Free driving: No influence of preceding vehicles can be observed. In this state, the driver seeks to reach and maintain his desired speed. In reality, the speed in free driving will vary due to imperfect throttle control. It will always oscillate around the desired speed.

Approaching:The driver adapts his speed to the slower speed of a preceding vehicle. While approaching, the driver decelerates, so that there is no difference in speed once he reaches the desired safety distance.

Following: The driver follows the preceding car without consciously decelerating or accelerating. He keeps a safe distance more or less constant. However, again due to imperfect throttle control, the difference in speed oscillates around zero.

Braking: Driver applies medium to high deceleration rates if the distance to the preceding vehicle falls below the desired safety distance. This can happen if the driver of the preceding vehicle abruptly changes his speed or the driver of a third vehicle changes lanes to squeeze in between two vehicles.

In this paper,Wiedemann 99 is selected as the car-following model with the parameters, as shown in Figure5.These parameters can be further described as follows:

CC0: The average desired standstill distance between two vehicles measured in meters.

CC1: Time distribution of the speed-dependent part of desired safety distance (Headway time).This is the distance in seconds a driver wants to maintain at a certain speed. The higher the value, the more cautious the driver. The safety distance is defined in the car following model as the minimum distance a driver will maintain while following another vehicle.

CC2: Restricts the distance difference (longitudinal oscillation) or how much more distance than the desired safety distance a driver allows before he intentionally moves closer to the car in front.

CC3: It controls the start of the deceleration process, i.e., the number of seconds before reaching the safety distance. At this stage, the driver recognizes a preceding slower vehicle.

CC4: It defines negative speed difference during the following process. Low values result in a more sensitive driver reaction to the acceleration or deceleration of the preceding vehicle.

CC5: Defines positive speed difference during the following process. Enter a positive value for **CC5,**which corresponds to the negative cost of **CC4**. Low costs result in a more sensitive driver reaction to the acceleration or deceleration of the preceding vehicle.

CC6: Influence of distance on speed oscillation while in the following process.Value 0: The speed oscillation is independent of the distance larger values: Leads to a higher speed oscillation with increasing distance.

CC7: Oscillation during acceleration

Apart from these, lane change behavior is another essential behavior to be considered; the window in Figure7 shows the lane change behavior parameters window in VISSIM.

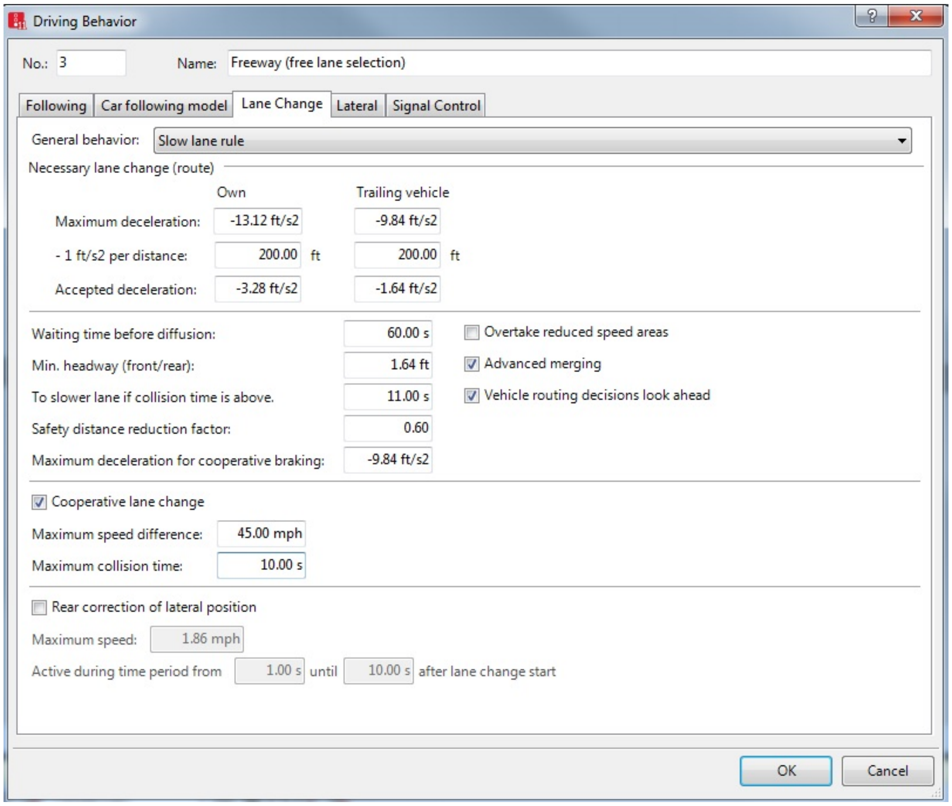


Figure7: Lane changing driving behavior parameters.

There are two important topics to mention: advanced merging and cooperative lane change behavior which is selected (checked) for this study.

Advanced merging: This option is selected in the driving behavior parameter sets by newly created networks. The option is considered for any necessary lane change towards the next connector along the route. If this option is selected, more vehicles can change lanes earlier. Thus, the capacity increases, and the probability that vehicles come to a stop to wait for a gap is reduced. Select the option accordingly to achieve the desired lane change behavior: If vehicle **A** has to change lanes and recognizes that the neighboring vehicle in front **B** on the target lane has approximately the same speed or is only slightly faster (-1.0 m/s < dv < 0.1 m/s), **A** slows down slightly (by 0.5 m/s^2) to move into the gap behind **B**, if the option is selected.

Cooperative lane change: If vehicle **A** observes that a leading vehicle **B** on the adjacent lane wants to change to his lane **A**, then vehicle **A** will try to change lanes itself to the next lane in order to facilitate lane changing for vehicle **B**. For example, vehicle **A** would switch from the right to the left lane when vehicle **B** would like to switch to the left from a merging lane to the right lane. Vehicle **A** behaves during this lane change as if it would have to change lanes due to a connector at a long distance. It accepts

its **Maximum deceleration** and the deceleration of the trailing vehicle **C** on the new lane, by the parameters for the necessary lane change.

Vehicle **A** does not make a cooperative lane change when the following conditions are true:

If the new lane is less appropriate for continuing its route if vehicle **B** is faster than the maximum speed difference (in the example 10.80 km/h (=3 m/s);

If the collision time exceeded the maximum collision time (in the example 10 seconds); and

The speed of vehicle **A**increased by the maximum speed difference (in the example 10.80 km/h).

3.9 VEHICLE INPUT

Vehicle Input is considered for medium to high volume traffic. Vehicle composition includes car and HGV within which are different 2D and 3D models of cars. A different type of vehicle that only uses the unblock lanes before a quarter-mile of incident location was input at 480 sec of stimulation such that this platoon of vehicles reaches the incident area at 600 simulation seconds when the incident happens.Vehicle volume is increased starting from 5000 vehicles per hour to 8000 vehicles per hour in an increment of 100 vehicles per hour.This results in 62 simulations per scenario, 31 for training and 31 for testing.

3.10 INCIDENT SCENARIOS AND LAYOUTS

There are three different scenarios considered and two different layouts of the road used in this study. The scenarios considered for the study are one lane closed, two lanes closed and two lanes closed with bottle-neck in downstream. The road layout is the same for both two-lane and one lane closed conditions whereas one lane is reduced in downstream of the incident area to create a bottleneck in the two-lane closed bottle-neck scenarios. These scenarios are described in detail later in this section.

3.10.1 Basic Layout of the Model

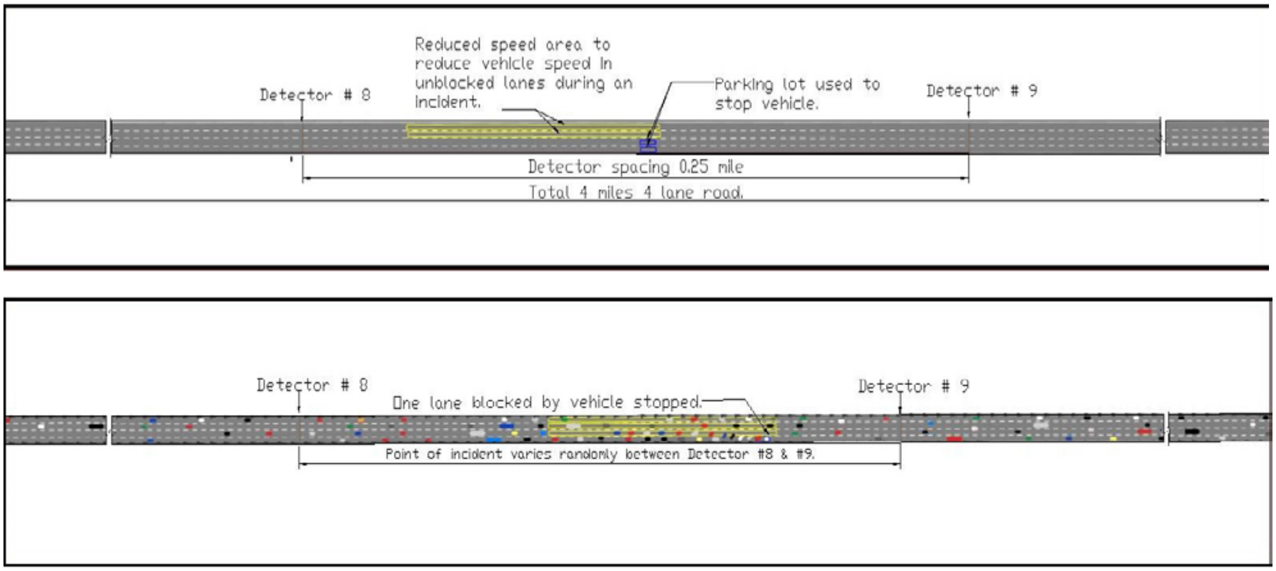


Figure8: Layout of the road in simulation.

As shown inFigure8, a hypothetical four-lane freeway is created in VISSSIM. The total length of the road is 4 miles, and the width of one lane is 12 ft. There are a total of 16 detectors placed along the road with the spacing of a quarter-mile, so there are 16 different data collection points. The incident is always simulated to occur between detector #8 (in upstream) and detector #9 (in downstream). The parking lot is used to create an incident where one vehicle gets stuckfor the total incident duration creating traffic congestion. Generally, in the simulation, the vehicles in the free lane do not reduce their speed near the incident location, whereas reduced speed is usually observed in real-life situations during an incident. Therefore, a 500 ft. long reduced speed area is placed on the lanes that are not blocked by the vehicle. Also, this reduced speed area is activated only for the incident duration to make vehicles reduce their speed from 50-65 mph to 25-45 mph.

After analysis of the data produced from this parking lot model and inspecting the travel time and occupancy collected between detector #8 and detector #9, it was found that the graph produced by this model is somewhat non-realistic and especially the vehicles in the rightmost lane are not affected by the incident which is clearly seen on the travel time plotted in Figure9.

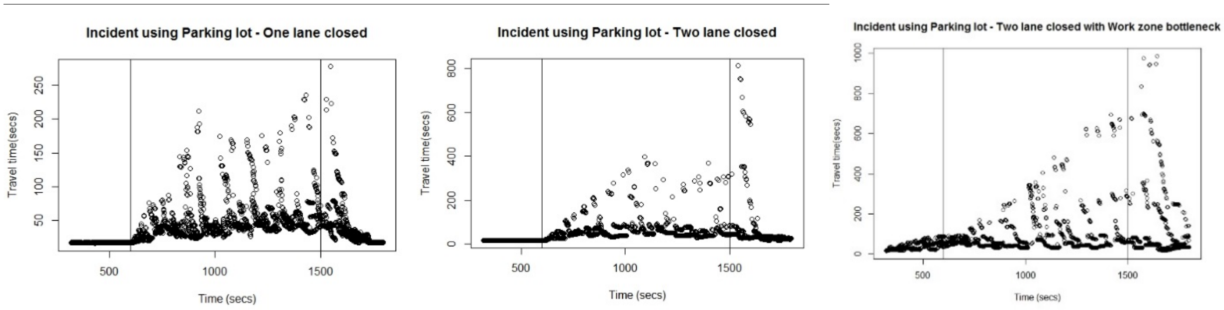


Figure9: Plot of travel time vs. time between detector #8 and detector #9.

Another issue using a parking lot is abnormal parking behavior of the vehicles when the vehicle parks in the corner of the parking space blocking two lanes where it was supposed to block only one lane which can be seen in the clip of the event as shown in Figure10. Apart from this, the incident time can’t be assigned exactly in 600secs, but 10 to 15 secs after 600secs causing a delay in detecting an incident.So to solve these issues, a new, improved model was developed, which is described in the forthcoming sections.

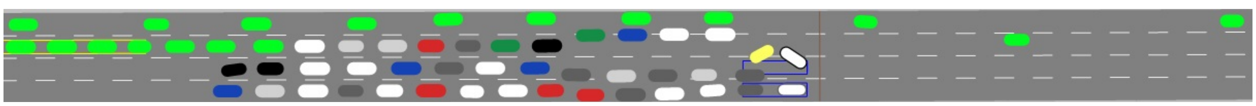


Figure10: Abnormal parking that happens randomly in simulation.

3.10.2 Basic Layout of the Improved Model

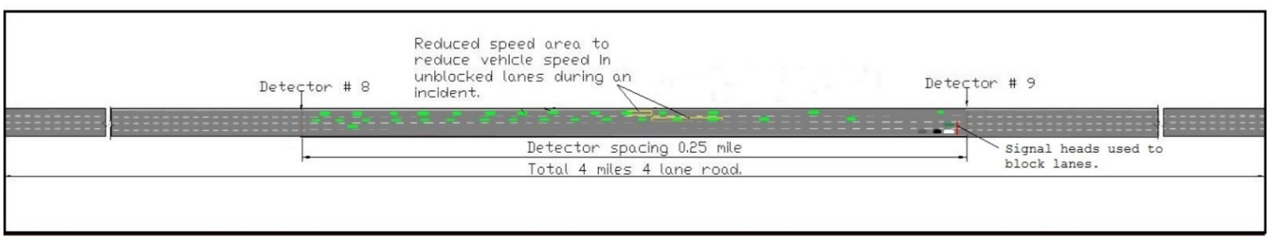


Figure11: Layout of the improved model in the simulation.

This new, improved model, as shown in Figure11, is created to eliminate those problems encountered using the previous model that uses the parking lot. Rather than using the parking lot, signals are used to block the lane. Vehicles in the blocked lanes are also blocked in a link such that most of the vehicles are forced to use the unblocked lanes after entering the section between detector #8 and detector #9. A Reduced speed area is placed in the unblocked lanes and activated for the incident duration.Also, this reduced speed area is activated only for the incident duration to make vehicles reduce their speed

from 50-65 mph to 10-25 mph. The only drawback of this model is that the incident position is fixed at the end of the link; otherwise, the vehicles would not get enough space to change the lane. The same problem occurs that occurred using the parking lot, prompting a long queue and cars stuck in the blocked lane for more time than it should be taking while not affecting the vehicles in the unblocked lanes. To solve this, the detector positions are changed instead of the incident position such that the incident occurs near upstream detector #8, in the middle and near downstream detector # 9.

Travel time plot from this model is more realistic, and the vehicles on the unblocked rightmost lanes are also affected by the incident. The plot of travel time is shown in upcoming sections where the details of the scenarios are taken into account foranalysis of the incident detection algorithms.

3.9.3 Scenario 1: Incident Blocking One Lane

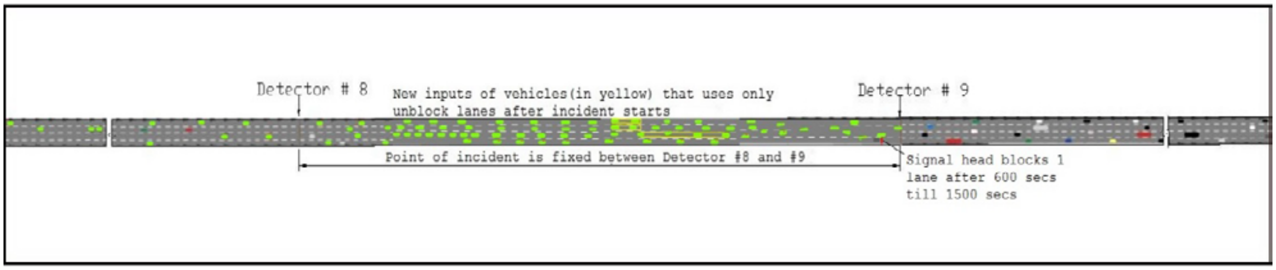


Figure12: One-lane-closed scenario in VISSIM

In the one lane closed scenario, only one lane, i.e.,the rightmost lane is blocked by the vehicle between detector #8 and detector #9. The incident starts at 600 simulation seconds and ends after 15 mins at 1500 simulation seconds. The reduced speed area is placed on the unblocked lanes and activated for the incident duration.Also, this reduced speed area is activated only for the incident duration to make vehicles reduce their speed from 50-65 mph to 10-25 mph.The new volume of incident type vehicles is assigned after the incident such that they use only the unblocked lanes, which are shown in yellow in Figure12. The total length of the road is 4 miles, and the width of one lane is 12 ft. There are a total of 48 detectors placed along the road with the spacing of nearly 1/12th mile and shifted, to take into account the effect of position of the incident on the algorithms. The incident is mainly categorized as near to the upstream detector, in between upstream and downstream (or incident in the middle), and near to the downstream detector.In the travel time plotted in Figure13, one lane scenarios indicate the delay caused by an incident is evenly distributed, and almost all vehicles are affected by the incident which ismore satisfactory and more realistic than the previous plot.

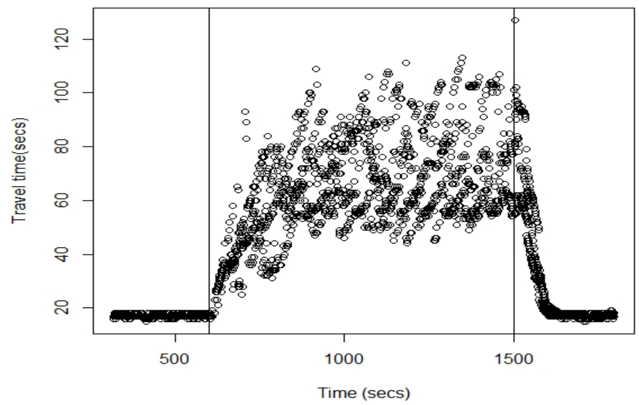


Figure13: Plot of travel time vs time in one lane closed scenario between detector 8 and 9.

3.9.4 Scenario 2: Incident Blocking Two Lanes

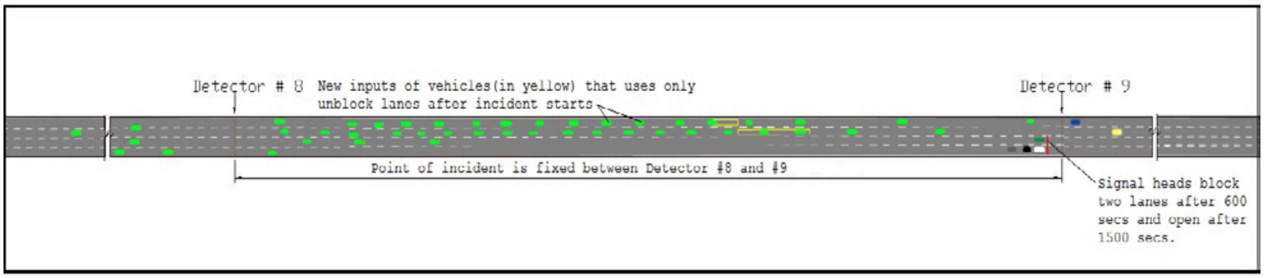
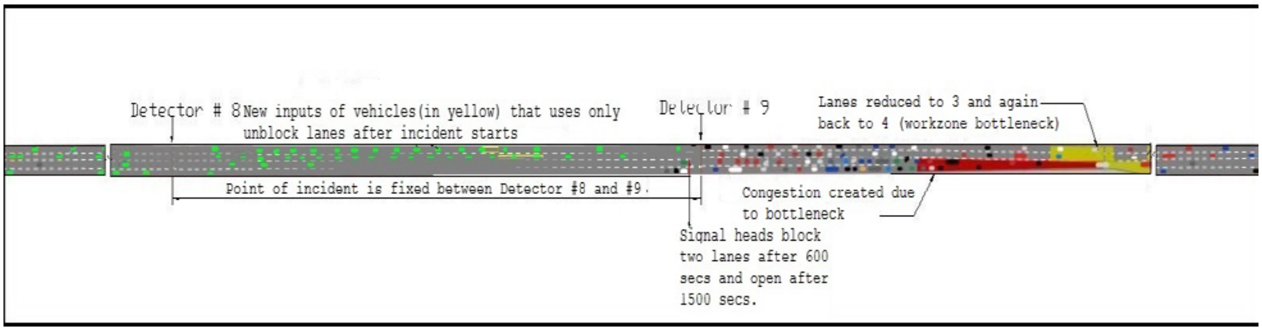


Figure14: Two-lanes-closed scenario in VISSIM

Similarly, for the two-lane closed scenario, only two lanes, i.e., the two right lanes, are blocked by the vehicle between detector #8 and detector #9. The incident starts at 600 simulation seconds and ends after 15 mins at 1500 simulation seconds. The reduced speed area is placed in the unblocked lanes and activated for the incident duration. The total length of the road is 4 miles, and the width of one lane is 12 ft. There are a total of 48 detectors placed along the road with the spacing of nearly 1/12th mile and shifted to take into account the effect of position of the incident on the algorithms. The incident is mainly categorized as near the upstream detector, in between upstream and downstream (or incident in the middle), and near the downstream detector.Similarly, for the travel time plotted in Figure15, two-lane closed scenarios indicate the delay caused by an incident is evenly distributed and almost all vehicles are affected by the incident which is more satisfactory and more realistic than the previous plot.



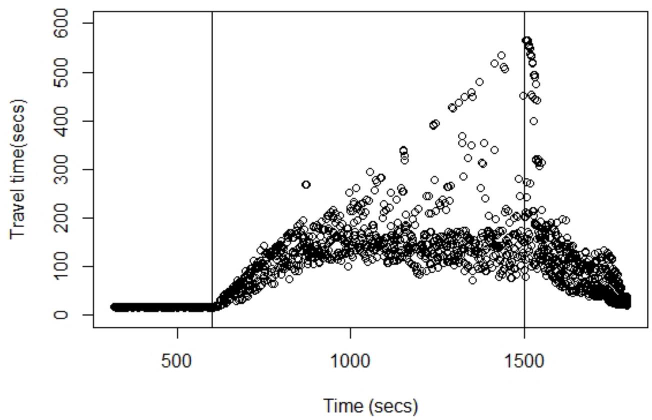


Figure15: Travel time Vs. Time in a two-lane closed scenario between detector 8 and 9.

3.9.5 Scenario 2WZ: Incident blocking two lanes in the upstream of a work-zone bottleneck

Figure16: VISSIM Screenshot for Scenario 2WZ

In this scenario, a bottleneck is created on the downstream of the incident position between detector #9 and detector #10.This bottleneck is a lane reduction from 4 lanes to three lanes and again back to 4 lanes before detector #10. The reason behind the introduction of this bottleneck is to check whether the algorithm candetect an incident in congested traffic conditions. The congestion created by the bottleneck at the downstream usually reaches detector # 9 at volume levels above 6,000 vehicles per hour. In this way, it is possible to apprehend the effect of the bottlenecks on the incident detection algorithms using the detector and travel time data. Therefore, for two-lane closed with the bottle-neck scenario, only two lanes, i.e., the two right lanes, are blocked by the vehicle between detector #8 and detector #9. The incident starts at 600 simulation seconds and ends after 15 mins at 1500 simulation seconds. The Reduced speed area is placed on the unblocked lanes and activated for the incident duration. A Total of 48 detectorsare placed along the road with the spacing of nearly 1/12th mile and shifted, to take into account the effect of position of the incident on the algorithms. The incident is mainly categorized as near the

upstream detector, in between upstream and downstream (or incident in the middle), and near the downstream detector.

Also for this congested scenario, the travel time plotted in Figure15 clearly indicates the delay caused by an incident is evenly distributed and almost all vehicles are affected by the incident which is more satisfactory and more realistic than the previous plot. Apart from the delay caused by the incident, it could also be seen that there is some delay caused due to a bottleneck before 600 seconds; this effect is accessed in this scenario on incident algorithms.

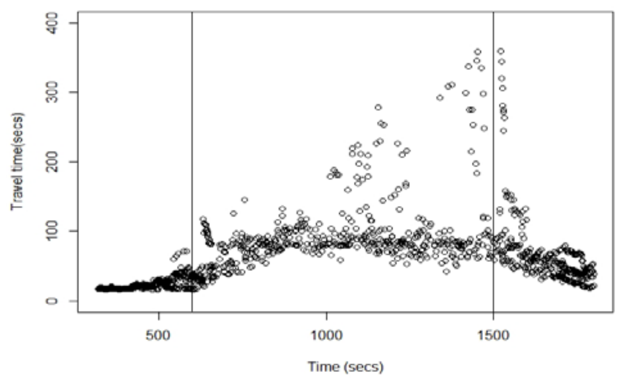


Figure17: Travel time Vs Time plot in 2 lanes closed work-zone bottleneck condition.

CHAPTER IV

INCIDENT DETECTION ALGORITHMS AND PARAMETER OPTIMIZATION

As described in Chapter II, there are lots of algorithms developed and used in the past; most of them are checked and analyzed too. In this chapter, the new algorithms developed and used in this study are described thoroughly. For the survey, we used four different algorithms: California #7, Algorithm using SVM, Research Algorithm #A, which uses detector data, and Research Algorithm #B that uses trajectory data. As California #7 and SVM is already explained in Chapter II, the other algorithms are described further in this subchapter.

4.1 CALIBRATING CA#7 PARAMETERS

For the calibration of thresholds in the California algorithm, heuristic searches on sets of thresholds were done. Threshold T1 was investigated from 5 to 55 in increments of step 5, i.e., 5, 10, 20…55. Similarly, T2 was investigated from 0.10 to 0.8 in increments of step 0.05 and T3 from 5 to 80 in increments of step 5.The best threshold sets (T1, T2, T3)that result in giving minimum false alarm were selected as the calibrated thresholds for the particular algorithms. Then these best performing threshold sets were further investigated in the No-incident scenario to select the better thresholds set among the best.

4.2 NO- INCIDENT SCENARIO FOR THRESHOLDS VALIDATION

There is No-incident simulated in this scenario, and the road is in free-flow condition. This scenario might be thought to be redundant and less significant as the same condition is likely to happen in other parts of the road where there is no incident even in the incident scenarios. However, when algorithms are run in this scenario, the results validate the calibrated thresholds for the given algorithm to make certain that there is no false-alarm generated while the road is in the incident-free condition. All the algorithms studied in this thesis were checked under no incident. No-incident data was collected from 31 different simulations with volumes increasing from 5000 vph (vehicles per hour) to 8000 vph at the increment of 100 vph per simulation with different random seeds. Apart from California Algorithm #7, no other algorithms produced a false alarm in the No-incident dataset. This data-set was used to select the best-calibrated parameter for the studied algorithms. For example, the calibrated thresholds T1, T2, T3 of California Algorithm #7 had many calibrated values having the same performance in 2 lanes closed incident scenario as shown in table 25. when tested under the No-incident scenario, the performance varies for different calibrated thresholds and the dilemma of selecting the best thresholds for incident scenarios can be solved testing this threshold in the No-incident scenario. We can select the thresholds

that produce minimum false alarm in No-Incident scenarios. All sets of thresholds for CA#7 for all scenarios was checked on No-incident and the best one was selected according to their performance in No-incident.

Table 3: Calibrated thresholds in No-Incident and two lanes closed Incident scenarios.

Calibration of CA#7 in Incident Condition					Application in No-Incident Condition	
T1	T2	T3	Detection Rate	False Alarm Rate	Detection Rate	False Alarm Rate
5	0.2	35	96.77%	3.23%	74.19%	25.81%
5	0.22	35	96.77%	3.23%	83.87%	16.13%
5	0.24	35	96.77%	3.23%	90.32%	9.68%
5	0.28	35	96.77%	3.23%	100.00%	0.00%

In table25, different thresholds sets performance in No-Incident, and two lanes closed Incident scenarios are presented. The detection rate at Incident condition was the same for all the given sets of calibrated thresholds, whereas the detection rate in the No-Incident condition when checking over all the detectors varies and performed best after T1=5, T2=0.28, and T3= 35 with no false alarm. Hence, this set of thresholds was selected for two lanes closed scenario for CA #7.

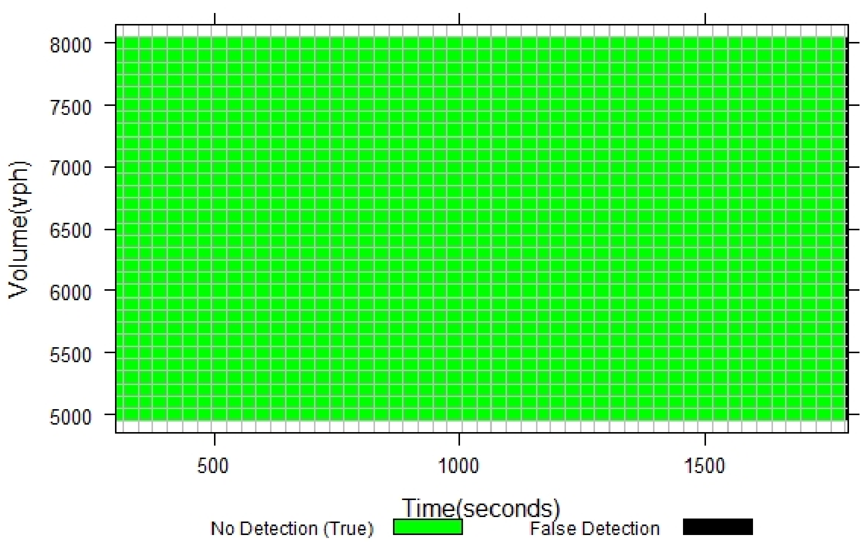


Figure18:: Result of California #7 in No-Incident Scenario using 2- lane closed thresholds

The result for the No-Incident scenario is presented figuratively. The result of California #7 in a no-incident scenario using 1-lane closed thresholds (T1=5, T2=0.44, T3=35) is shown in Figure47. In this

case, the performance measures did not change when all the thresholds were checked in the No- incident scenario. Only in 3 of 31 simulations, the algorithm didn't detect the incident.

Similarly, the result of California #7 in a no-incident scenario using 2-lane closed thresholds (T1=5, T2=0.28, T3=35) is shown in Figure46. In this case, there was no false alarm, and the algorithm didn't detect any false incident.

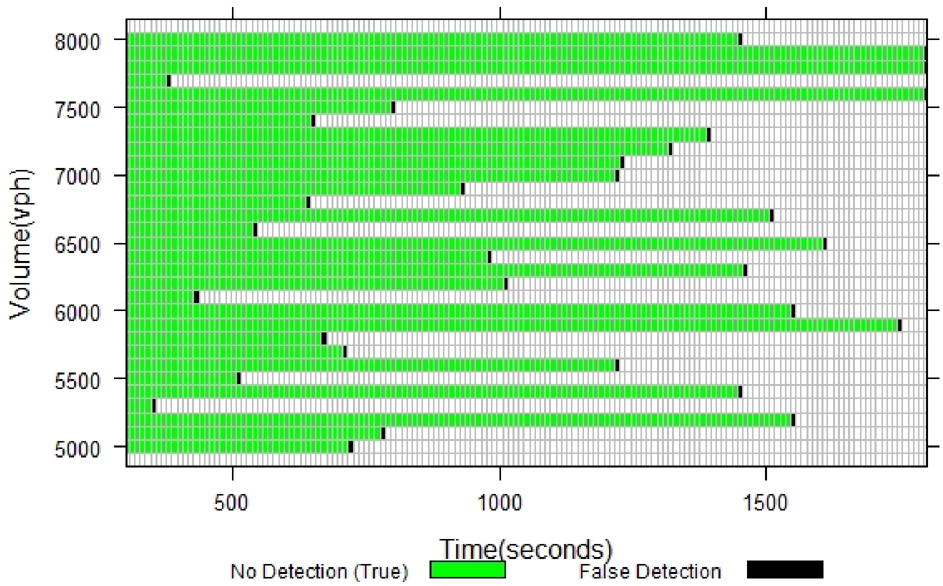


Figure19: Result of California #7 in No-Incident Scenario using 1- lane closed thresholds.

Table 4: Calibrated thresholds in No-Incident and one lane closed Incident scenarios.

Best thresholds for one lane closed CA#7 In Incident Scenario					Performance in No-Incident Scenario	
T1	T2	T3	Detection Rate	False Alarm Rate	Detection Rate	False Alarm Rate
5	0.44	40	65.00%	35.00%	9.68%	90.32%
5	0.44	40	65.00%	35.00%	9.68%	90.32%
5	0.44	45	65.00%	35.00%	9.68%	90.32%
5	0.44	50	65.00%	35.00%	9.68%	90.32%
5	0.44	55	65.00%	35.00%	9.68%	90.32%
5	0.44	60	65.00%	35.00%	9.68%	90.32%
5	0.44	65	65.00%	35.00%	9.68%	90.32%
5	0.44	70	65.00%	35.00%	9.68%	90.32%
5	0.44	75	65.00%	35.00%	9.68%	90.32%
5	0.44	80	65.00%	35.00%	9.68%	90.32%

In table26, different threshold sets' performance in the No-Incident and 1 lane closed Incident scenarios are presented. The detection rate at Incident and No-incident conditions were the same for all the calibrated thresholds.

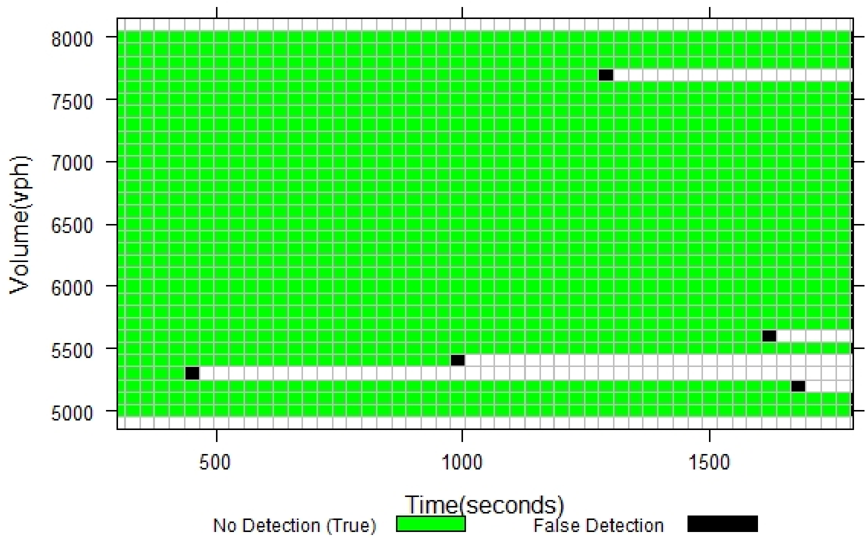


Figure20: Result of California #7 in No-Incident Scenario using 2- lane closed work-zone bottleneck thresholds.

The result of California #7 in a no-incident scenario using 2-lane closed work zone thresholds (T1=5, T2=0.22, T3=40) is shown in Figure48. In this case, the performance measures did not change that drastically when all the thresholds were checked in the No- incident scenario. Only in 5 out of 31 simulations, the algorithm detects the incident when there was no incident.

Apart from California #7, no other algorithms detected any incident when tested under the no-Incident scenario.

4.3 ALGORITHM USING SVM ON DETECTOR DATA:

A brief introduction about Support Vector Machine (SVM) and how it classifies the sets of data is already explained in Chapter II. However, to use SVM in predicting incident, several parameters used to train SVM need to be analyzed. In addition, SVM needs to be tuned properly to be used as the predictor. Therefore, different sets of parameters were passed to explore the best parameters and checked to train the SVMmodel. Apart from these parameters,SVM also uses different types of kernels to classify the data. Also, it is recommended to check which kernelsare best suited for our data sets and can easily be used for the detection of the incidents. For the exploratory analysis Detection rate and false alarm,the rate is

calculated for every time step, the Detection rate is calculated as (True positive + True negative) / Total # of the test) and False alarm Rate as (False Positive/Total). The summary of the results on SVM model training for exploring the best parameters is given in table3.

Note: All calibration or training of the model on SVM is done using “e1071’ and “caret” packages in R.

Table 5: Summary of SVM Model Trained and Tested for Incident Detection.

SVM Models		# of Parameters	Linear Kernel			Polynomial Kernel			Radial Kernel		
Set	Parameters Passed		FAR	DR	MTD(sec)	FAR	DR	MTD(sec)	FAR	DR	MTD(sec)
1	V, Q, O,DTO,UTO,DSO,DSNoV,DSV	8	3.32 %	92.68 %	672.58	3.33 %	92.92 %	672.58	3.43 %	92.37 %	672.58
2	V, Q, O,DTO,UTO,DSO,CumV,DSCumV	8	3.06 %	93.00 %	674.52	3.24 %	92.53 %	674.52	3.13 %	93.08 %	674.52
3	V,Q ,O,UTO, UUTO,DSO,CumV,DSCumV,USO	9	2.83 %	93.63 %	675.48	2.89 %	93.71 %	675.48	2.87 %	93.55 %	675.48
4	V,Qand O of U/s and D/s for t-4 time steps	30	2.92 %	93.31 %	674.52	2.66 %	95.16 %	674.52	2.95 %	93.15 %	674.52
5	V,Qand O of U/s and D/s for t-3 time steps	24	2.96 %	93.15 %	683.23	2.58 %	95.48 %	683.23	2.95 %	93.06 %	683.23
6	V,Qand O of U/s and D/s for t-2 time steps	18	3.18 %	92.34 %	678.39	2.70 %	95.08 %	678.39	3.03 %	92.98 %	678.39
7	V,Qand O of U/s and D/s for t-1 time steps	12	3.24 %	92.42 %	678.39	3.23 %	95.16 %	678.39	3.27 %	92.10 %	678.39

*For Linear Kernel, tuning range used : epsilon = 0 to 1 step 0.1 &cost=(0.01,0.15,0.25,0.5,1,5,10,50,100,250)

*For Polynomial Kernel tuning range used : degree=(2,3,4,5), coef0=(0.1,0.5,1,2,3,5,10)

* Only Trained for Detector # 7, 8 and 9

Abbreviations in table:

V= Speed
Q= Flow
O= Occupancy
USO= Upstream Occupancy
DTO= Occupancy at t-1
UTO=Occupancy at t+1
UUTO= Occupancy at t+2
DSO= Occupancy at Downstream(D/S)
DSNOV= Number of the vehicle at D/S
DSV= Downstream Speed
CumV = Cumulative # of vehicles passing at t;
DSCUMV= D/S Cumulative # of vehicles passing at t

There is a change in the performance of classification as the tuning parameters change. Taking set 6 as a sample case, the results are presented in Table4below. The table shows the linear kernel tuning table for the set 6 SVM model. The model is trained with different values epsilon and cost, and the best tune is obtained at minimum error and dispersion. In the sample case,as shown in table4,the best tuning parameter was epsilon =0 and cost = 50 highlighted in yellow, which has the lowest error as well as dispersion. This highlighted value of epsilon and cost is used in testing the data-set for the application, and results were presented intable4.

Table 6: Sample for selecting best tuning parameter in SVM.

	epsilon	cost	error	dispersion
1	0	0.01	0.009723502	0.001730413
2	0.1	0.01	0.009723502	0.001730413
3	0.2	0.01	0.009723502	0.001730413
.	.	.	.	.
.	.	.	.	.
12	0.3	0.15	0.005207373	0.00128611
23	0	0.25	0.003963134	0.001427828
34	0	0.5	0.003317972	0.001281515
45	0	1	0.003041475	0.001288859
56	0	5	0.002488479	0.001193816
67	0	10	0.002396313	0.001103428
78	0	50	0.002304147	0.001018932
89	0	100	0.002442396	0.001129842
100	0	250	0.00281106	0.001311544

In table3, the best performing sets of parameters are shown. We can see that the performance of all these models with different sets of parameters hasa similar result. Since these results are pretty close to each other, it is better to look at whether these differences were statistically significant. The confidence interval for the difference between two means contains all the values of (μ_1 - μ_2) (the difference between the two population means) which would not be rejected in the two-sided hypothesis test of H_0: $\mu_1 = \mu_2$ against H_a: $\mu_1 \neq \mu_2$, i.e.,H_0: $\mu_1 - \mu_2 = 0$ against H_a: $\mu_1 - \mu_2 \neq 0$.

If the confidence interval includes 0, we can say that there is no significant difference between the means of the two populations, at a given level of confidence. The confidence interval for the difference in means $\mu_1 - \mu_2$ is given by:

$$(\overline{x_1} - \overline{x_2}) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where, t^* is the upper (1-*C*)/2 critical value for the *t* distribution with *k* degrees of freedom (with *k* equal to either the smaller of n_1-*1* and n_1-*2* or the calculated degrees of freedom).So if 0 falls between the lower and upper limit of the confidence interval, the means are the same and we can say that there is no significant difference between the two data sets.

It was found that all the means are the same for all of these models. Table 4 shows the result of the last three sets compared with each other to see if there is any significant difference between their means. So after this analysis, any of the sets as mentioned earlier,could be used for our analysis. However, for this study in the SVM model, Set 6, which uses speed flow and occupancy of upstream and downstream Detector for t-2 time steps, was used. As it was too time-consuming to train Radial SVM, Only linear and Kernels were used and tuned to train the SVM model.

In Table 4,

Assumption: As our sample size is greater than 30. According to the central limit theorem, we can assume sample to be normally distributed.

x1 and x2 = mean of sample 1and 2, respectively.

n1 and n2 = number of samples 1 and 2, respectively.

s1 and s2 = standard deviation of samples 1 and two respectively.

LowerCI = Lower value of confidence interval.

UpperCI= Upper value of confidence interval.

Level of confidence interval = 95% i.e. $\alpha = 0.05$.

Result = the same mean if 0 falls between the lower and upper value of confidence interval.

Result = Different mean if 0 doesn’t fall between the range of confidence interval.

Table 7: Comparing the means of performance measures on a different set of SVM Model.

Models	Check On	x1	x2	n1	n2	s1	s2	Lower CI	Upper CI	Result
SVM Set 4 and SVM set 5	False alarm rate from linear kernel	2.919	2.965	31	31	4.446	4.529	-2.374	2.282	Same mean
	Detection rate from linear kernel	93.306	93.14	31	31	14.48	14.53	-7.365	7.687	Same mean
	False alarm rate from polynomial kernel	2.657	2.581	31	31	3.016	2.639	-1.393	1.547	Same mean
	Detection rate from polynomial kernel	95.161	95.48	31	31	4.130	3.441	-2.294	1.649	Same mean
	False alarm rate from radial kernel	2.949	2.949	31	31	4.130	4.281	-2.182	2.182	Same mean
	Detection rate from radial kernel	93.145	93.06	31	31	14.53	14.53	-7.458	7.619	Same mean
SVM Set 5 and SVM set 6	False alarm rate from linear kernel	2.965	3.180	31	31	4.529	4.417	-2.535	2.105	Same mean
	Detection rate from linear kernel	93.145	92.33	31	31	14.53	14.44	-6.710	8.323	Same mean
	False alarm rate from polynomial kernel	2.581	2.704	31	31	2.639	2.871	-1.553	1.308	Same mean
	Detection rate from polynomial kernel	95.484	95.08	31	31	3.441	3.679	-1.445	2.251	Same mean
	False alarm rate from radial kernel	2.949	3.026	31	31	4.281	4.439	-2.339	2.185	Same mean
	Detection rate from radial kernel	93.065	92.984	31	31	14.530	14.555	-7.463	7.624	Same mean
SVM Set 6 and SVM set 7	False alarm rate from linear kernel	3.180	3.241	31	31	4.417	4.644	-2.412	2.289	Same mean
	Detection rate from linear kernel	92.339	92.419	31	31	14.447	14.979	-7.714	7.553	Same mean
	False alarm rate from polynomial kernel	2.704	3.226	31	31	2.871	5.051	-2.653	1.609	Same mean
	Detection rate from polynomial kernel	95.081	95.161	31	31	3.679	4.279	-2.150	1.989	Same mean
	False alarm rate from radial kernel	3.026	3.272	31	31	4.439	4.549	-2.577	2.086	Same mean
	Detection rate from radial kernel	92.984	92.097	31	31	14.555	14.876	-6.747	8.521	Same mean

So for the study in the SVM model, Set 6, which uses speed flow and occupancy of upstream and downstream Detector for t, t-1, and t-2 time steps were used. As it was time-consuming to train Radial SVM, only linear Kernels were used and tuned to train the SVM model. The input matrix for training the SVM is shown in table6below, where t indicates time step and each row represents a pair of consecutive sensors. Ground truth is 1 for sensor pair 8-9 between 600 to 1500 secs and 0 for all.

Table 8: Matrix input for the training and testing the SVM model.

	From U/S Detector									From D/S Detector									Ground Truth
	Occupancy			Speed			Flow			Occupancy			Speed			Flow			
Sensor pair	t	t-1	t-2	t	t-1	t-2	t	t-1	t-2	t	t-1	t-2	t	t-1	t-2	t	t-1	t-2	
1-2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2-3	0.12	0.13	0.15	5	6	7	.	.	.	.	.	.	.	.	.	.	.	.	0
8-9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

4.3.1 Algorithm #A:

This algorithm was developed to observe the change in occupancy difference and number of vehicles between two sensors at a particular section (i.e., section between the detectors) as experienced by the moving vehicles. The basic principle behind this algorithm is that there is an increase in the parameters like the difference in occupancy and number of vehicles in that section of road where there is an incident relative to the section of road without an incident. So, as the vehicles reach the part of the road where there is an incident, there is an abrupt change in occupancy difference and also in the number of vehicles (i.e., vehicle density). Algorithm #A detects this change and predicts the incident. The flow diagram of Algorithm #A is shown in Figure18.

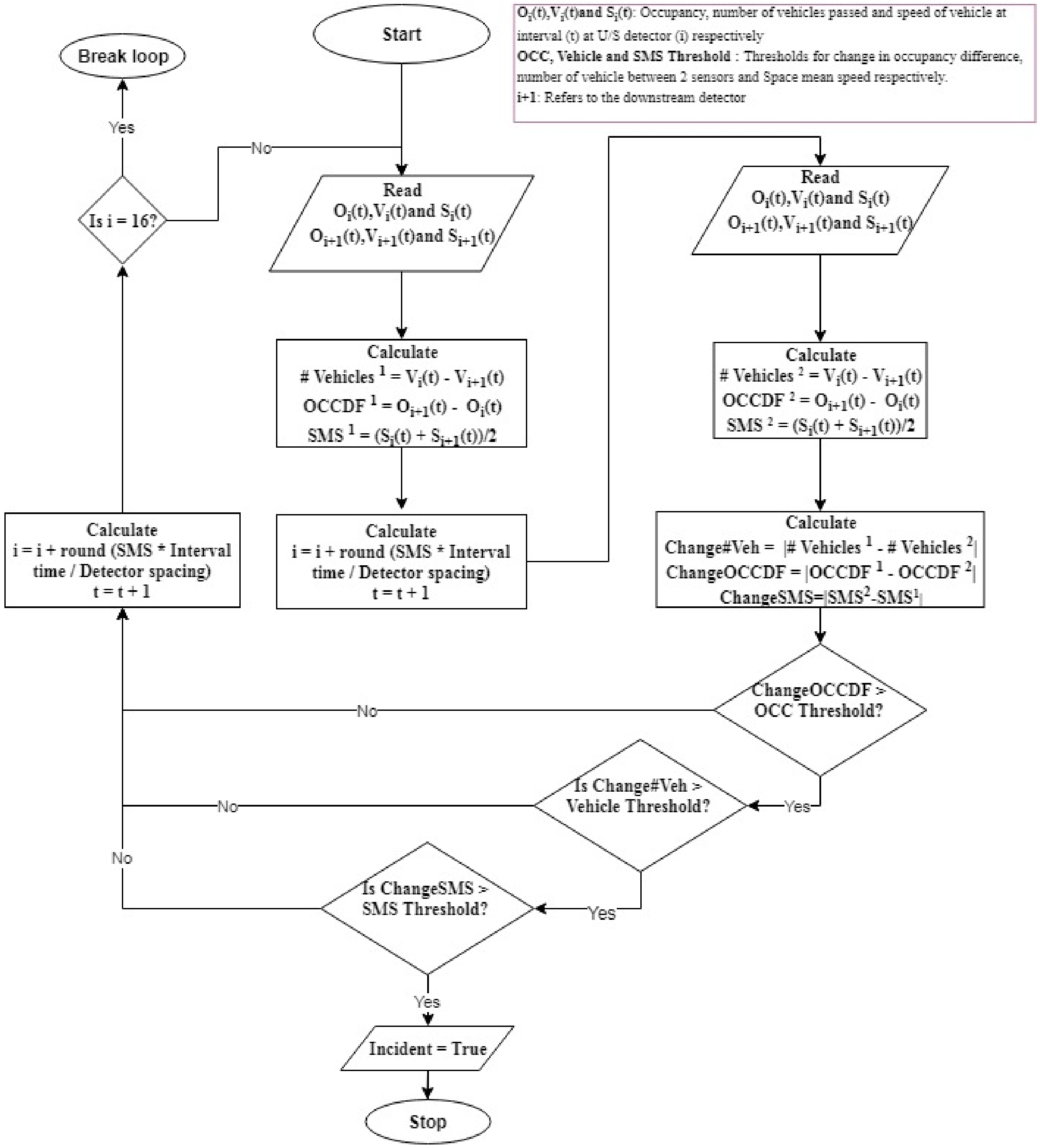


Figure21: Flowchart for Algorithm #A

In Figure18, the flowchart of algorithm #A is presented, where $O_i(t)$, V_i (t) and S_i (t) is the occupancy at interval t, a number of vehicles passed during interval t and Speed of vehicle at interval (t) at the detector (i) respectively. Similarly, i+1 refers to the downstream detector. The process shown in the flow diagram is carried out for each interval (t) the data is collected starting from(t=1). These parameters can be directly read from the detectors. After reading O(t), V (t) and S (t) for bothi and i+1 detectors, the

number of vehicles and occupancy difference between these detectors are also calculated. In addition to this space mean speed is also calculated to get the next i, where the same process is carried out once again to calculate a number of vehicles and occupancy difference between these detectors. After this,the change in the number of vehicles, occupancy difference and space mean speed (SMS) between these two portions of the road are calculated. Then these changes are checked on thresholds that are calibrated from the training dataset. If the change is found to be higher than both of the thresholds, it is confirmed that there is an incident between these 2 segments of the road. This process gets repeated until the incident is detected or the end of the road is reached.

4.3.2 Algorithm #B:

Algorithm #B, unlike any other algorithms used in this study, is based on trajectory data instead of detector data. In Algorithm #B, the incident is detected by comparing the travel time of the vehicles with the historical thresholds. The principle behind algorithm # B it that the vehicle’s travel time increases in incident condition, so if the travel time of the vehicles increases it can be asserted that there is an incident on the road.

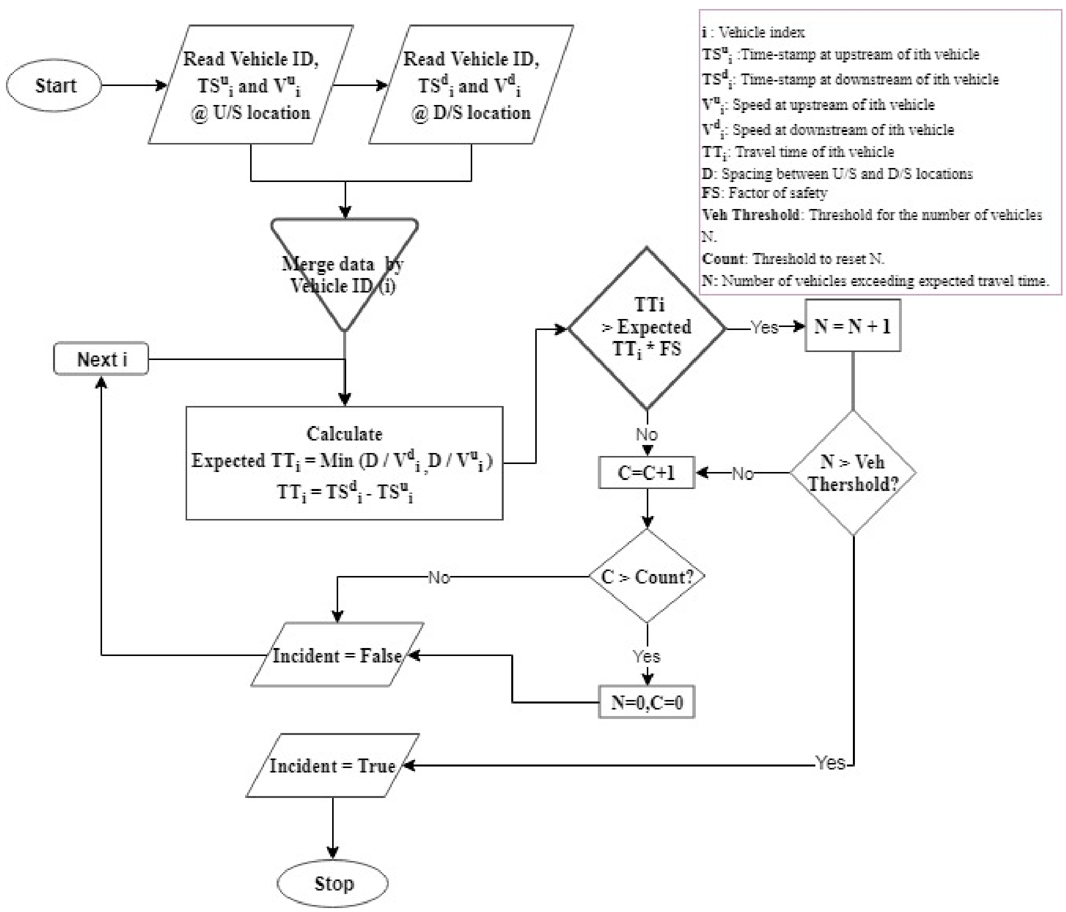


Figure22: Flowchart for Algorithm #B.

The flow diagram of algorithm #B is presented in Figure12. In Figure19, i represents vehicle index, TS^{u}_{i} represents time-stamp in the upstream. Similarly,TS^{d}_{i} represents time-stamp at downstream of i^{th} vehicle. TT represent travel time, and D is the spacing between U/S and D/S locations.FS indicates the factor of safety and Veh threshold is the threshold of the number of vehicles exceeding the travel time, which are selected or calibrated according to the historical data.

At first, the data is read from trajectories or maybe from some vehicle identification system at the upstream position and again ata downstream position. This data are then merged by vehicle identification number (i) and sorted according to upstream time-stamps. This is prepared so that the same vehicle is identified at the downstream location. Now anticipated travel time (Expected TT) and actual travel time (TT) is calculated and checked whether the Actual travel time is greater than expected travel time * factor of safety, and if so, how many vehicles had an actual travel time greater than expected travel time? The N parameter, which is the number of vehicles exceeding the expected travel time, is reset whenever the total number of vehicles checked exceeds a certain limit, i.e. Count. If the number of vehicles having greater travel time than anticipated exceeds acceptable thresholds, or in other words if N out of some observed number of vehicles (Count) exceeds the expected travel time then the incident is detected; otherwise, the process is repeated for the next vehicle. The algorithms run until the detection of an incident and after that, the algorithm stops and must be re-run.

Although this algorithm works in normal conditions where there is no bottleneck, in bottle-neck conditions, there may be an error in the estimate of the travel time from point speed at upstream and downstream. As the vehicles are in a congested state, the speed measured at these points can lead to greater expected travel time than the vehicles traveled, which in turn is the reason for such a high factor of safety. This not only makes an algorithm detect the incident late on time but also is the main cause ofthe wrong detection. Hence, to eliminate this problem, we can use any other detector at the downstream of this detector so that the estimate of the expected travel time is taken under the un-congested state. However, it is possible to track the same vehicle in 3 different sensors and to get the traveled time from these three detectors; this travel time only represents the free-flow travel time and cannot be used in this bottle-neck condition. To solve this problem, a modified algorithm #B is developed where the incident is verified using the third detector loop. The flowchart to this algorithm is presented in Figure20below.

In Figure 14, algorithm #B is slightly modified by adding the loop for the 3rd detector. The incident is first detected at the 3rd detector, for which the travel time of each vehicle is tested with expected travel time. After detection of the incident at this detector where there is no congestion, the algorithm goes back to the logic in algorithm #B, where it detects the incident based on how many vehicles (N) are late for the given number of vehicles observed (Count). In other words, if (N) late

vehicles out of some observed number of vehicles (Count) exceeds the expected threshold (Veh threshold) then the incident is detected; otherwise, the process is repeated for the next vehicle.

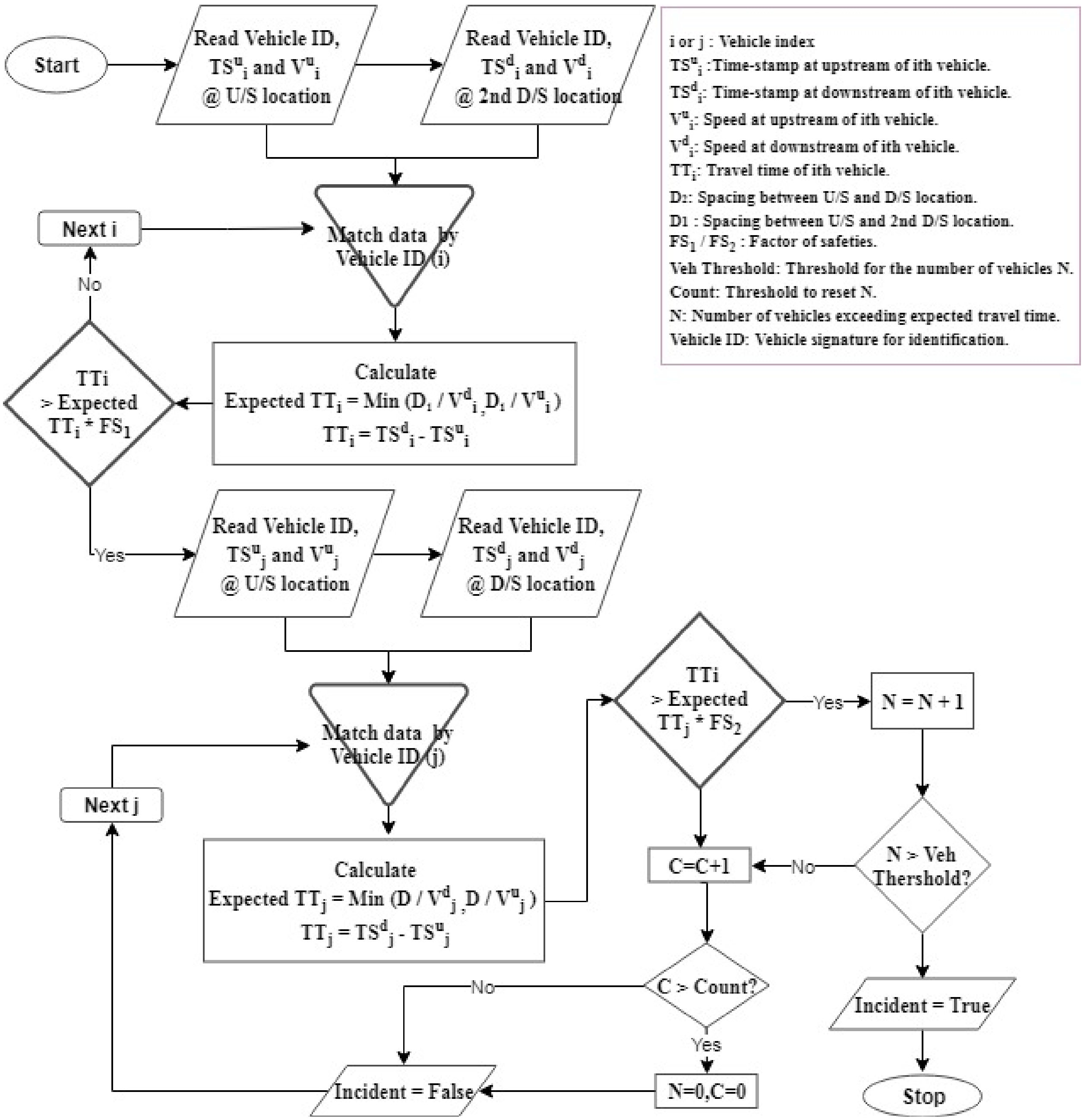


Figure23: Flow Diagram for Modified Algorithm #B

The results from modified algorithm B are not presented in this thesis; it is a simple idea that works in the case of bottleneck conditions. It was tested and proved using data from a parking lot model but not tested on the new model due to the time constraint to develop other SVM models using travel time and volume using similar data using only two detectors.

4.4 ALGORITHM USING SVM ON RE-IDENTIFICATION DATA:

Similar to SVM using detector data in this case, we make use of re-identification data such as travel time based on the timestamp and number of vehicles passing through upstream and downstream points.The principle behind it is that the vehicle’s travel time increases in an incident condition, so if the travel time of the vehicles increases it can be asserted that there is an incident on the road which was already apprehended in previous algorithms and was found to be performing good in an uncongested condition. So for the congested condition,two more features were added to the SVM model, i.e. number of vehicles at upstream and number of vehicles at down-stream. The input matrix passed to SVM model training is presented in Table7. Here we use travel time aggregated over the 30-second interval, the number of vehicles at upstream is shown as NOvehU/s in the table, and a number of vehicles at downstream is shown as NovehD/s in the table. Ground truth is 1 for the time between 600 – 1500 and the rest are 0, where 1 points to an incident condition and 0 to a No-incident condition.

Table 9: Matrix input for training SVM using Re-Id data.

#	Time30secInt	Volume	Travel Time	NOvehU/s	NOvehD/s	Ground truth
1	330	5000	16.67	44	37	0
2	360	5000	17	42	53	0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
11	630	5000	20.68	58	21	1
12	660	5000	34	44	24	1
.	.	.	.	.	.	.
.	.	.	.	.	.	.

The matrix shown in Table 7 is used to train the SVMmodel which is the data from random seed 369 and after the training, the same three features are used to predict (or detect) the incident.

4.5 NEW PERFORMANCE TESTING:

As we know, with the means of standard algorithms, it is possible to detect the incident from the point data, and once the incident is detected and confirmed by the algorithms in the traffic control center,

the concerned authorities are immediately notified for a response. Before that, the truth about the incident needs to be verified from the information obtained from road users, or from surveillance cameras. What if the information is not obtained? Also there are lots of false alarms generated by these algorithms which can confuse responders or operators working in the control center. Even if the detection rate is good, the use of this algorithm might be complicated because of the false alarms generated and the inability to verify the truth. Therefore, to deal with this problem, a new performance test is performed in this thesis study, which can at least compare different algorithm's practical performance.

Consequently,a new performance test is carried out, the incident detection algorithms are tested such that algorithms can detect the incident only after the incident occurs but within 5 minutes of the incident. If the algorithm detects the incident before the incident occurs, it's a false alarm. If it detects an incident after 5 minutes, it's also not useful and also considered a false alarm. The algorithm should detect the incident only after the incident happens so that the response team is less confused by the results of the algorithm. There are four different conditions that we get after running the algorithm, as shown in the figures below. In all cases, the test is carried out for 10 mins, 5 minutes before the incident and 5 minutes after the incident for incident data and for also 10 minutes for the non-incident data.

The vertical axis in the figures represents the only 0 and one where 0 is no incident, and 1 is an incident detected by the algorithm. The horizontal axis represents the time, which starts 5 minutes before the incident and ends 5 minutes after the incident.

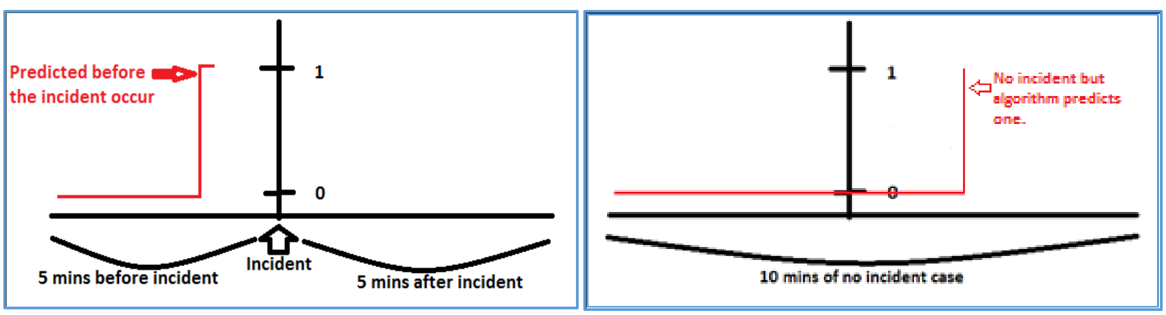


Figure24: False alarm (Type 1)

The first condition is when the algorithm detects (or predicts) the incident before the incident happens. This should not happen because it may confuse the responders as they might get to the location before the incident and realize it is a false alarm. Also, in the no incident scenario, if an algorithm predicts that there is an incident any time during a 10 minute time frame, then it is categorized as a false alarm. These types of false alarms are indicated as alarm type 1 in this study.

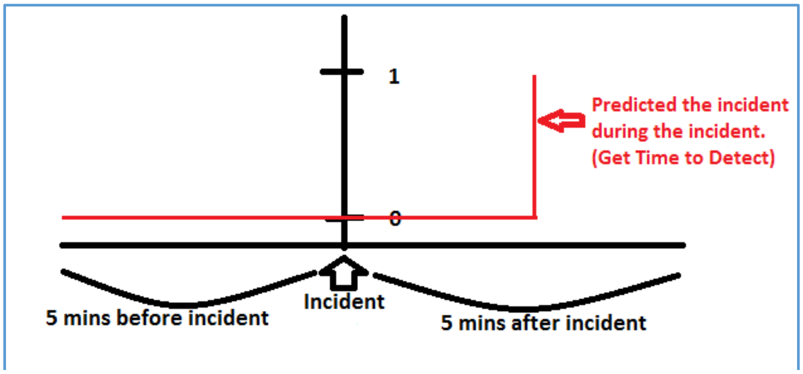


Figure25: True alarm (Type 2)

The second type of outcome is that the incident is detected after an incident happens and within 5 minutes of the incident. This is a true alarm, so we get the time to detect the incident from this type of alarm. These types of alarm are said to be alarm type 2 for this study. The nexttwo outcomes are of no alarm outcomes, where the algorithms don't detect any incident until 10 minutes after the incident.

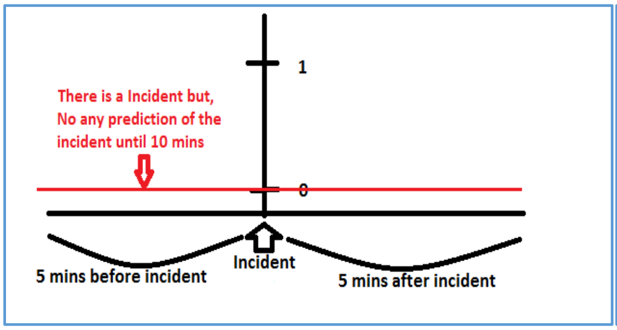


Figure26: No alarm False (Type 3)

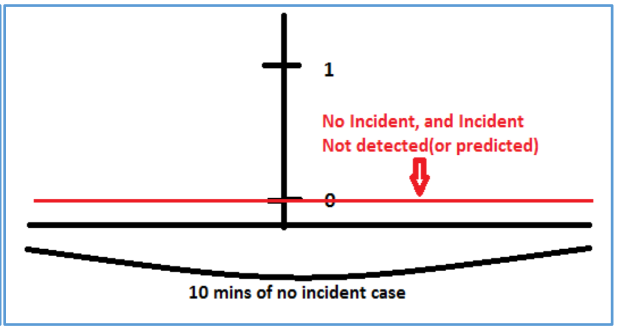


Figure27: No alarm True ((Type 4)

In these outcomes,the incident is not detected, but there is an incident in reality in the third case (type 3), whereas the fourth case (Type 4) has no incident. Therefore, the 3rd case is no alarm but with the false result, it doesn't detect incident, but there is an incident. And the 4th case is true as there is no incident and the algorithm doesn't detect any incident. In summary, we can assert that Type 1 and Type 3 are false positive and false negative, respectively. However, type 2 and type 4 represent true positive and true negative, respectively.

Now, a new detection rate is calculated as a percent of the sum of true positive and true negative over the total numberof outcomes.

$$\text{Detection rate (percent)} = 100 * \frac{\# \, of \, Type \, 2 \; + \; \# \, of \, Type \, 4}{\text{Total \# of outcomes}} = 100 * \frac{TP + TN}{\text{Total}} \ldots \ldots \ldots \ldots \ldots \ldots \ldots (6)$$

Similarly, the false alarm rate is calculated asa percent of the sum of false positives and false negatives over the total number of outcomes.

$$\text{False Alarm Rate (percent)} = 100 * \frac{\# \, of \, Type \, 1 \; + \; \# \, of \, Type \, 3}{\text{Total \# of outcomes}} = 100 * \frac{FP + FN}{\text{Total}} \ldots \ldots \ldots \ldots \ldots \ldots (7)$$

Mean detection time, which is the average time taken by an algorithm to detect the incident or start of the incident, is taken only from type 2 outcomes. In this way, we can say if the algorithm is capable of detecting the incident in a practical way, instead of looking at the detection rate and false alarm rate based just on a number of incidents and a false alarm generated on each time step the algorithm is run. In this way, as the detection rate + false alarm rate is equal to 100, we can consider only one of these measures to calibrate the thresholds. Previously, to calibrate the threshold minimum false alarm was selected for a given threshold generally greater than 80% and there arises the dilemma forchoosing the thresholds when detection rate is high and false alarm is also high and detection rate is low (but higher than 80) and the false alarm is too small. This dilemma of selecting the thresholds is solved when using this performance measure.

CHAPTER V

RESULTS

This chapter includes all the relevant results obtained during and after implementation of the algorithms in detecting the incident. As stated in the previous chapter, there are three different scenarios for our case study and four different algorithms used in this thesis. The detailed result of these cases is further described in this section part by part according to the scenarios considered for the study.

5.1 ONE LANE CLOSED SCENARIO:

In this scenario,only one rightmost lane was blocked by vehicles after 600 simulation seconds for 15 minutes. The location of the incident varies between sensors #8 and #9. A total of 62 numbers of the simulation were run varying volume from 5000 vehicles per hour to 8000 vehicles per hour in an increment of 100 vehicles per hour with two different seeds. The total duration of the simulation was 30 minutes, where 5 mins of the warm-up time were considered, and data was collected only after 300 simulation seconds. Calibration of the detection models (algorithms) was performed on a training dataset of a 1st random seed, and testing was carried out for the 2nd random seed of the simulation. Five different algorithms mentioned above, as described in previous chapters, were used. Results obtained from these algorithms are described one by one and summarized to demonstrate the difference in the performance of these algorithms on this specific scenario. The other difference for the one lane closed scenario is that 10-sec interval data from the detectors are used instead of 30-sec interval except for SVM using Re-Id where 30 sec is used as it was found that even using 10 secs the algorithm only detects after 30 secs of the incident in this algorithm.

California # 7:

At first, the calibration of the thresholds is carried out to find the best thresholds that are used in the California # 7 algorithm. The dataset with 1st random seed, i.e. 369, is used for calibration. Different sets of T1, T2, and T3 are tested, and the threshold set with minimum false alarm rate is selected as the best threshold. In this case, the best threshold found was T1 = 5, T2 = 0.44 and T3 = 35 with a detection rate of 38.71% and false alarm rate of 61.29% when calculated from equation 5 and 6 in chapter IV. The results are shown in Table8 below.

Table 10: Result of California #7 In One Lane Closed Scenario

Thresholds: T1= 5, T2= 0.44, T3= 35					
Volume	Alarm type	Detect time (sec)	Time to detect (sec)	Position (miles)	Remarks
5000	1	470	-130	2	Correct position
5100	1	400	-200	1	Detected at wrong position
5200	3	1110	510	0.25	Detected at wrong position
5300	1	420	-180	1.5	Detected at wrong position
5400	3	1120	520	0.75	Detected at wrong position
5500	1	410	-190	1.5	Detected at wrong position
5600	3	1240	640	1.25	Detected at wrong position
5700	3	1650	1050	1	Detected at wrong position
5800	3	1690	1090	1.75	Detected at wrong position
5900	3	1700	1100	2	
6000	1	370	-230	1.25	Detected at wrong position
6100	1	410	-190	1	Detected at wrong position
6200	3	1650	1050	1.5	Detected at wrong position
6300	3	1610	1010	1	Detected at wrong position
6400	1	380	-220	0.75	Detected at wrong position
6500	3	970	370	2	Correct position
6600	3	990	390	2	Correct position
6700	3	1140	540	1.25	Detected at wrong position
6800	2	870	270	2	Correct position
6900	2	830	230	2	Correct position
7000	2	780	180	2	Correct position
7100	2	620	20	2	Correct position
7200	2	800	200	2	Correct position
7300	3	1210	610	0.75	Detected at wrong position
7400	2	810	210	2	Correct position
7500	2	770	170	2	Correct position

Table 8: Continued

7600	2	790	190	2	Correct position
7700	2	780	180	2	Correct position
7800	2	830	230	2	Correct position
7900	2	890	290	2	Correct position
8000	2	740	140	2	Correct position

For easy visualization of the results, a plot is created as shown in Figure25.The Y-axis represents volumes used in each simulation, and the X-axis represents time in simulation seconds, the firstdark black vertical line at 600seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. According to the performance test, the algorithm should detect the incident in between these two lines. The algorithm was run every 10 secs in this case, and the green box, red box, blue box, and black box indicate no detection (True), i.e. there is no incident and no detection, no detection (False), i.e. there is an incident but no detection, True detection, and False detection respectively. Each gray grid is 10 sec apart, as in this case 10 secs data is collected, and the algorithm is run in each time interval of 10 seconds. Looking at the figure, the California algorithm at this scenario doesn't perform well, among 31 volumes checked. 19 times the incident is not correctly detected until the end of the simulation; 12 of them are detected late and 7 of them are detected before the incident actually happened and 15 of them are detected in the wrong position.Out of incidents occurring in 31 different volumes, only 12 of them were correctly and timely detected.

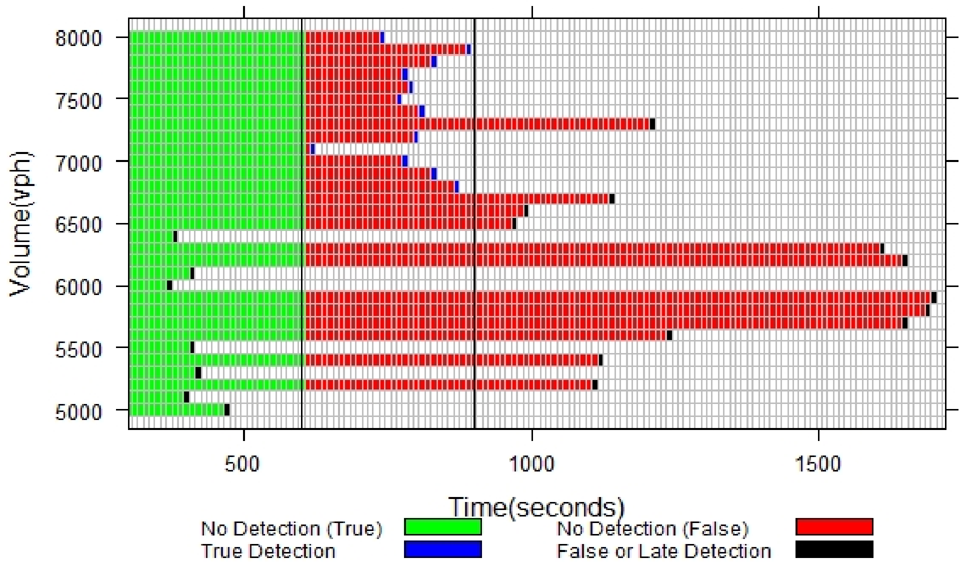


Figure28: Result of California #7 In One Lane Closed Scenario

Algorithm using SVM on detector data:

The calibration of the SVM is carried out to find the best SVM model parameters that are used in this algorithm. The dataset with 1st random seed, i.e. 369, is used for calibration. The parameters used in SVM are speed, occupancy and number of the vehicle for t, t-1 and t-2 time steps for both upstream and downstream detectors, so in total there are 18 parameters. The SVM model is then tuned under the range of epsilon from 0 to 1 with an increment of 0.01 and different cost from 1 to 100.

Optimum Parameters after tuning SVM:

- SVM-Type: C-classification
- SVM-Kernel: radial
- Cost: 10
- Epsilon (gamma): 0.05555556
- Number of Support Vectors: 1353

Table 11: Result of SVM in One Lane Closed Scenario

Volume	Alarm type	Detect time (sec)	Time to detect (sec)
5000	2	630	30
5100	2	630	30
5200	2	630	30

Table 9: Continued

5300	2	640	40
5400	2	630	30
5500	2	630	30
5600	2	630	30
5700	2	630	30
5800	2	630	30
5900	2	630	30
6000	2	630	30
6100	2	630	30
6200	2	630	30
6300	2	630	30
6400	2	630	30
6500	2	630	30
6600	2	630	30
6700	2	630	30
6800	2	630	30
6900	2	630	30
7000	2	630	30
7100	2	630	30
7200	2	630	30
7300	2	630	30
7400	2	630	30
7500	2	630	30
7600	2	630	30
7700	2	630	30
7800	2	630	30
7900	2	630	30
8000	2	640	40

For easy visualization of the results, a plot is created as shown in Figure26, in this figure the first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. According to the performance test, the algorithm should detect the incident in between these two lines. The algorithm was run every 30 secs; the green box, red box, blue box, and black box indicate No detection (True), i.e. there is no incident and no detection, No detection (False), i.e. there is an incident but no detection, True detection, and False detection respectively. This

algorithm at this scenario detected the incident with a detection rate of 100% and a false alarm rate of 0% when calculated from equations 6 and 7 in Chapter IV.

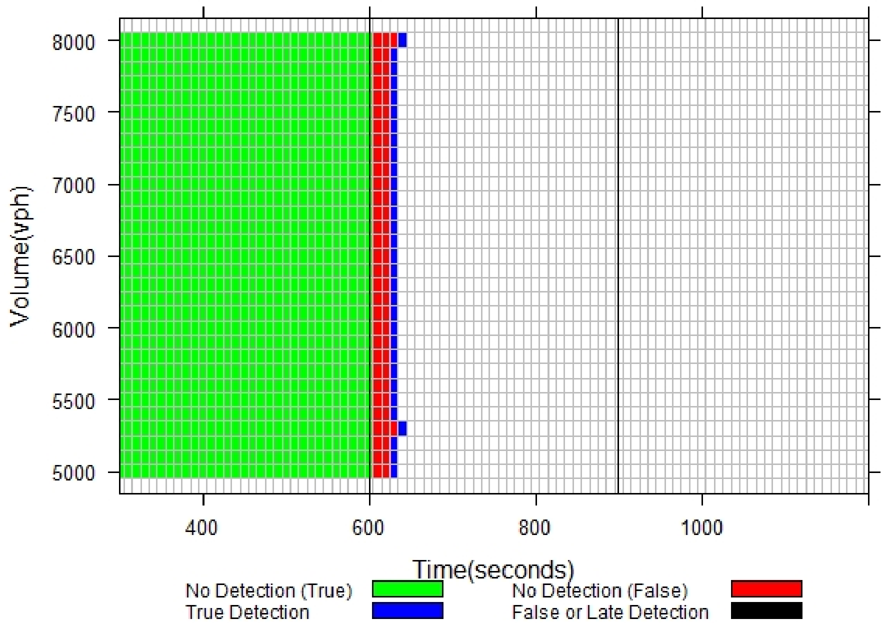


Figure29: Result of SVM in One Lane Closed Scenario

Algorithm #A:

Similar to the California algorithm at first, the calibration of the thresholds needs to be carried out to find the best thresholds used for algorithm #A. The dataset with a 1st random seed, i.e. 369, is used. Different sets of OCC Threshold, Vehicle Threshold, and SMS threshold are tested, and the threshold set with minimum false alarm rate is selected as the best threshold. In this case, the best threshold found was found to be OCC Threshold (T1) = 0.05, Vehicle Threshold (T2) = 10 and SMS threshold (T3) = 2.The results are shown in table10and Figure27below.

OCC Threshold (T1) = 0.8, Vehicle Threshold (T2) = 10 and SMS threshold (T3) =0

Table 12: Result of Algorithm #A in One Lane Closed Scenario

Thresholds: (T1) = 0.05, (T2) = 10 and (T3) = 2					
Volume	Alarm type	Detect time (sec)	Time to detect (sec)	Position (miles)	Remarks
5000	2	620	20	1.9	
5100	2	620	20	1.9	
5200	2	640	40	1.9	
5300	2	620	20	2.1	
5400	2	620	20	1.9	
5500	2	620	20	1.9	
5600	2	620	20	1.9	
5700	2	640	40	1.9	
5800	2	640	40	1.9	
5900	2	640	40	1.9	
6000	2	650	50	2.1	
6100	2	660	60	1.9	
6200	2	630	30	1.9	
6300	2	620	20	1.9	
6400	2	620	20	1.9	
6500	2	620	20	2.1	
6600	2	620	20	1.9	
6700	2	620	20	1.9	
6800	2	630	30	1.9	
6900	2	620	20	1.9	
7000	2	620	20	1.9	
7100	2	620	20	1.9	
7200	2	640	40	1.9	
7300	2	620	20	1.9	
7400	2	620	20	1.9	
7500	2	620	20	1.9	
7600	2	630	30	2.1	
7700	2	630	30	1.9	
7800	2	620	20	1.9	
7900	2	630	30	1.9	
8000	2	640	40	1.9	

For easy visualization of the results, a plot is created as shown in Figure27, in this figure the first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. According to the performance test, the algorithm should detect the incident in between these two lines. The algorithm was run every 10 secs in this case, and the green box,

red box, blue box, and black box indicate no detection (True), i.e. there is no incident and no detection, no detection (False), i.e. there is an incident but no detection, True detection, and False detection respectively. Each gray grid is 10 secs apart, as in this case, 10 secs of data is collected and analyzed in each time interval of 10 seconds.This algorithm at this scenario detected the incident with a detection rate of 100% and a false alarm rate of 0% when calculated from equations 6 and 7 in Chapter IV.

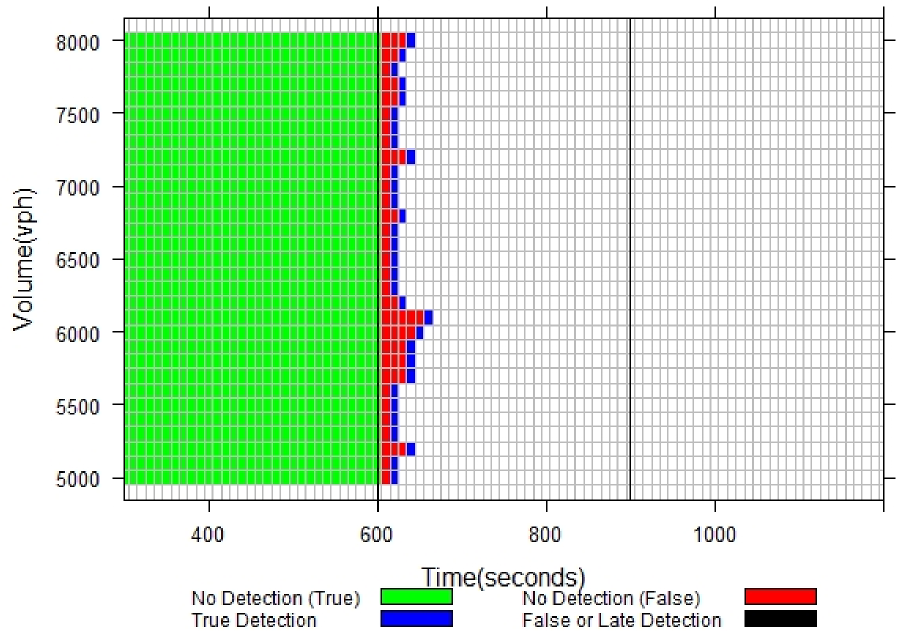


Figure30: Result of Algorithm #A in One Lane Closed Scenario

Algorithm #B:

This algorithm is based on AVI (Automatic Vehicle Identification), data and is collected only at two stations, detectors #8 and #9, within which the incident occurs. Only speed and timestamp of vehicles at upstream and downstream are used. However, calibration is done similarly as in the California algorithm; the calibration of the thresholds is carried out to find the best thresholds that give the minimum false alarm rate. the dataset with a 1st random seed, i.e. 369, is used for calibration. Different sets of thresholds, the threshold for a number of vehicles and threshold for the factor of safety are tested, and the threshold set with minimum false alarm rate is selected as the best threshold. In this case, the best threshold found was the threshold for a number of vehicles = 1, the threshold for the factor of safety = 1.25 with a detection rate of 100% and with no false alarm rate when calculated from equations 6 and 7 in Chapter IV. The results are shown in table11 and Figure28 below.

Table 13: Result of Algorithm #B in One Lane Closed Scenario

Thresholds: Number of vehicles = 1, Factor of safety = 1.25, Count=5			
Volume	Alarm type	Detect time (sec)	Time to detect (sec)
5000	2	617	17
5100	2	621	21
5200	2	620	20
5300	2	619	19
5400	2	618	18
5500	2	614	14
5600	2	617	17
5700	2	622	22
5800	2	613	13
5900	2	618	18
6000	2	622	22
6100	2	616	16
6200	2	618	18
6300	2	616	16
6400	2	615	15
6500	2	619	19
6600	2	618	18
6700	2	617	17
6800	2	618	18
6900	2	615	15
7000	2	619	19
7100	2	620	20
7200	2	614	14
7300	2	615	15
7400	2	617	17
7500	2	618	18
7600	2	615	15
7700	2	617	17
7800	2	618	18
7900	2	616	16
8000	2	615	15

For easy visualization of the results, in Figure28, the first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. According to the performance test, the algorithm should detect the incident in between these two lines. The algorithm was run every 10 secs in this case, and the green box, red box, blue box, and black box indicate no detection (True), i.e. there is no incident and no detection, no detection (False), i.e. there is an

incident but no detection, True detection, and False detection respectively. Each gray grid is 10 secs apart, as in this case, 10 secs of data is collected and analyzed in each time interval of 10 seconds.

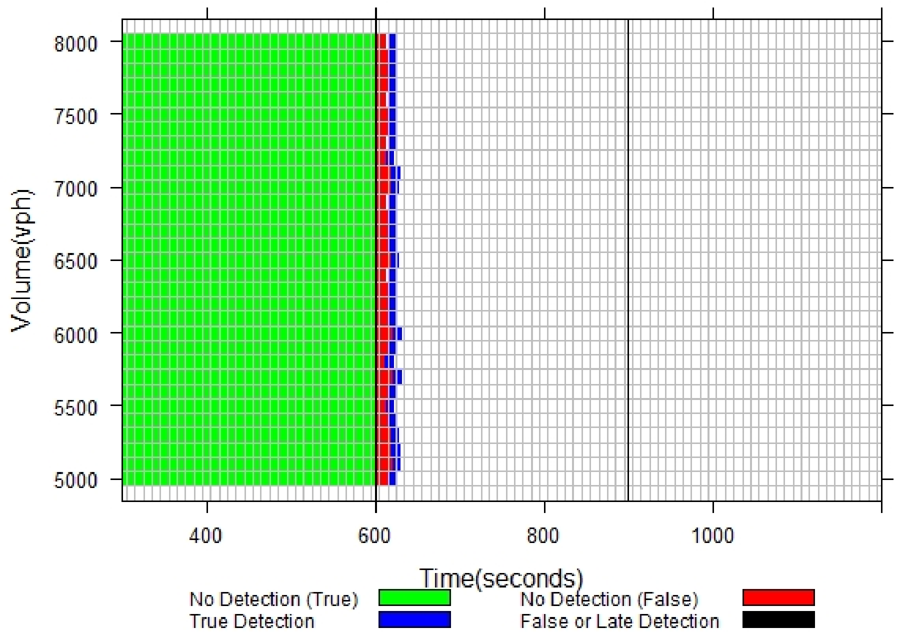


Figure31: Result of Algorithm #B in One Lane Closed Scenario

It is to be noted that algorithm #B runs on Re-identification data, so the percent of vehicles that are correctly re-identified needs to be accessed. Figure29shows the performance of the algorithm in different percentages of appropriately re-identified vehicles.

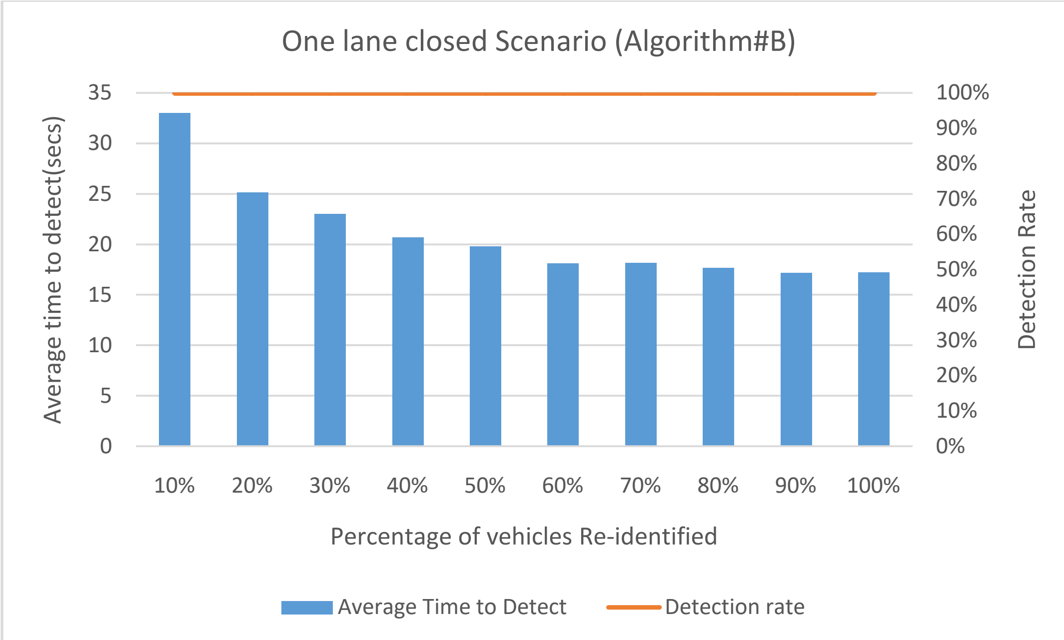


Figure32: Detection Rate and Average Time to Detect under Different Percent of Re-identified Vehicles.

In Figure29, the primary vertical axis at left indicates the average time to detect the incident, whereas the second axis represents the detection rate. It can be asserted that as the percentage of vehicles re-identified decreases the average time to detect increases. The detection rate is constant at 100%, so the incident is correctly identified and does not change upon the percentage of vehicles re-identified even if only 10% of vehicles are re-identified. Most of the literature asserts that 90% of the vehicles can be correctly identified, and the result shows that the performance of the algorithm at 90% of vehicles re-identified has no significant change in detection rate or average time to detect compared to 100% re-identified vehicles.

SVM using Re-Identification:

This algorithm is based on AVI (Automatic Vehicle Identification); data is collected only at two stations, detectors #8 and #9, within which the incident occurs. Only the number of vehicles and the timestamp of vehicles at upstream and downstream are used. However, calibration is done similarly as in the SVM example using detector data. The dataset with a 1st random seed, i.e. 369, is used for calibration. The parameters used in this SVM are travel time and number of the vehicle for ttime steps of the 30-sec time interval for both upstream and downstream detectors, so in total there are three parameters. The SVM model is then tuned under the range of epsilon from 0 to 1 with an increment of 0.01 and different cost from 1 to 100. The results are shown in Table 12.

Table 14: Result of SVMRe-ID in One Lane Closed Scenario

Volume	Alarm type	Detect time	Time to detect
5000	2	630	30
5100	2	630	30
5200	2	630	30
5300	2	630	30
5400	2	630	30
5500	2	630	30
5600	2	630	30
5700	2	630	30
5800	2	630	30
5900	2	630	30
6000	2	630	30
6100	2	630	30
6200	2	630	30
6300	2	630	30
6400	2	630	30
6500	2	630	30
6600	2	630	30
6700	2	630	30
6800	2	630	30
6900	2	630	30
7000	2	630	30
7100	2	630	30
7200	2	630	30
7300	2	630	30
7400	2	630	30
7500	2	630	30
7600	2	630	30
7700	2	630	30
7800	2	630	30
7900	2	630	30
8000	2	630	30

For easy visualization of the results, a plot is created as shown in Figure30. In this figure the first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. According to the performance test, the algorithm should detect the incident in between these two lines. The algorithm was run every 10 secs in this case, and the green box, red box, blue box, and black box indicate no detection (True), i.e., there is no incident and no detection,

no detection (False), i.e. there is an incident but no detection, True detection, and False detection respectively. Each gray grid is 30 secs apart, as in this case 30 secs of data is collected and analyzed in each time interval of 30 seconds.

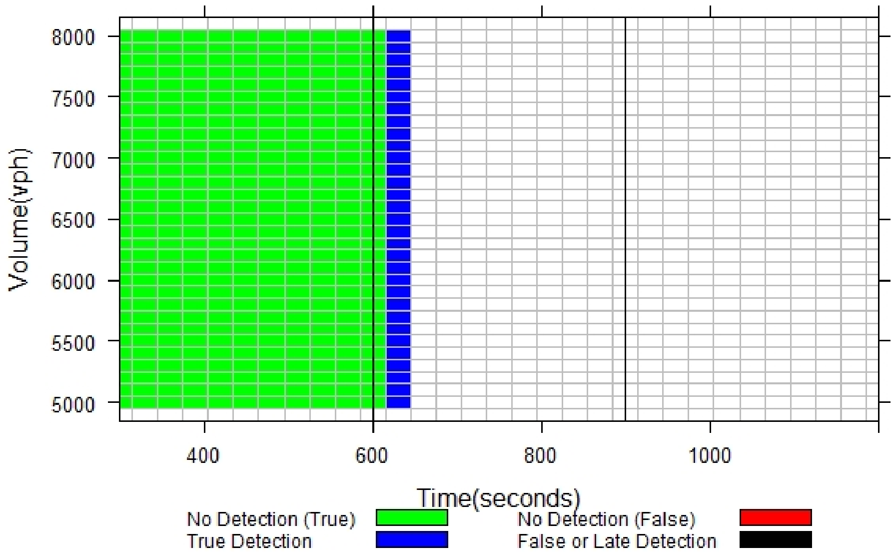


Figure33: Result of SVM Re-ID in One Lane Closed Scenario

Also, this algorithm runs on Re-identification data, so the percent of vehicles that are correctly re-identified needs to be accessed. Figure31 and Figure29 show the performance of the algorithm in different percentages of appropriately re-identified vehicles.

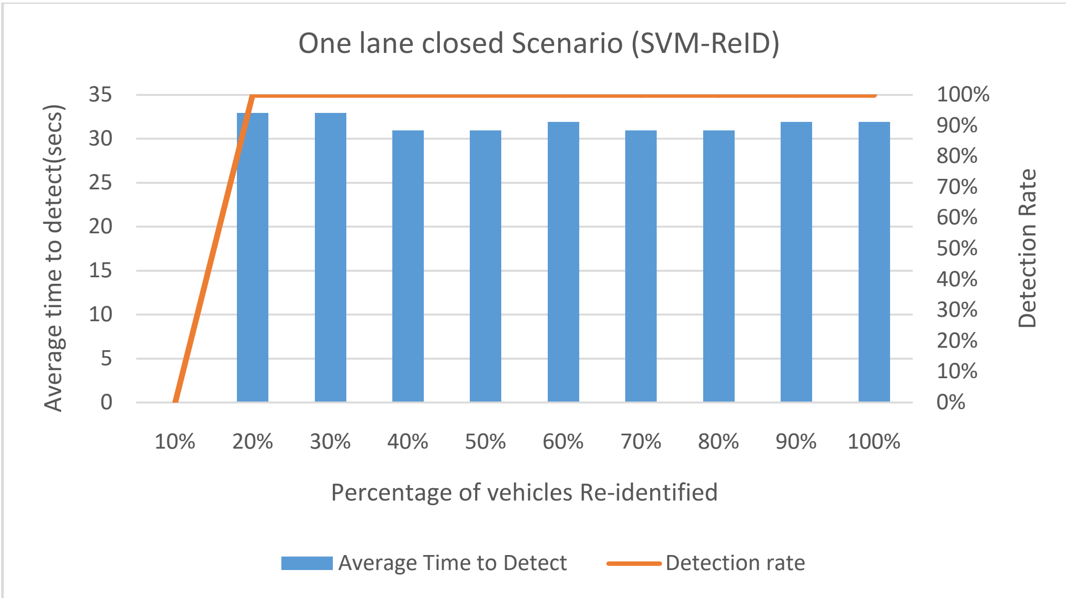


Figure34: Detection Rate and Average Time to Detect under Different Percent of Re-identified Vehicles (SVM).

The detection rate is constant at 100%. That means the incident is correctly detected until 20 % and doesn't change upon the percentage of vehicles re-identified;only even if 10% of vehicles are re-identified does the detection rate drop to 0. Most of the literature asserts that 90% of the vehicles can be correctly identified, and the result shows that the performance of the algorithm at 90% of vehicles re-identified has no significant change in detection rate or average time to detect compared to 100% re-identified vehicles.

Summary of Results in one lane closed scenario:

Looking at the figures and tables of results in the 1 lane closed scenario, it is clear that the SVM models were the best performing models. The SVM model has no false alarm rate and the average time to detect the incident of about 30.65secs when tested at all volumes from 5000 to 8000, as shown in the figure. The California model is comparatively wrong in detecting an incident in this scenario. California #7 has the worst minimum detection time. However, Algorithm # B has excellent performance with 100% detection rate and average minimum time to detect just 17.32secs. Hence, in the 1 lane closed scenario, the best results were achieved from the SVM model in detector data and Algorithm #B in overall results. The summary is presented in table13below.

Table 15: Summary of Result in One Lane Closed Scenario

#	Performance measures	California #7	SVM(Detector)	Algorithm #A	Algorithm #B	SVM(REID)
1	**Alarm Type 1**	7	0	0	0	0
2	**Alarm Type 2**	12	31	31	31	31
3	**Alarm Type 3**	12	0	0	0	0
4	**Average Detection Time (secs)**	1066.25	630.65	627.74	617.32	630.00
5	**Detection Rate**	38.71%	100.00%	100.00%	100.00%	100.00%
6	**False Alarm Rate**	61.29%	0.00%	0.00%	0.00%	0.00%

5.2 TWO LANE CLOSED SCENARIO:

In this scenario, the two rightmost lanes were blocked by vehicles after 600 simulation seconds for 15 minutes. The location of the incident varies between sensors #8 and #9. A total of 62 numbers of the simulation was run varying volume from 5000 vehiclesper hour to 8000 vehiclesper hour in an increment of 100 vehiclesper hour with two different seeds. The total duration of the simulation was 30 minutes, where 5 minsof the warm-up time were considered, and data was collected only after 300 simulation seconds. Calibration of the detection models (algorithms) was performed on a training dataset of a 1st random seed, and testing was carried out for 2nd random seed of the simulation. Five different algorithms above, as described in previous chapters, were used. Results obtained from these algorithms are described one by one and further summarized to see the differences in the performance of these algorithms on this specific scenario.

California # 7:

At first, the calibration of the thresholds is carried out to find the best thresholds that are used in the California # 7 algorithm. The dataset with a 1st random seed, i.e. 369, is used for calibration. Different sets of T1, T2, and T3 are tested, and the threshold set with minimum false alarm rate is selected as the best threshold. In this case, the best threshold found was T1 = 5, T2 = 0.28 and T3 = 35 with a detection rate of 96.77% and false alarm rate of 3.23% when calculated from equations 6 and 7 in Chapter IV, which is just one false alarm out of 31; however, when tested it did not produce any false alarms. The results are shown in table14and Figure32 below.

Table 16: Result of California #7 In Two Lane Closed Scenario

Thresholds: T1= 5, T2= 0.28, T3= 35					
Volume	Alarm type	Detection time (sec)	Time to detect (sec)	Position (miles)	Remark
5000	2	720	120	2	
5100	2	720	120	2	
5200	2	720	120	2	
5300	2	720	120	2	
5400	2	660	60	2	
5500	2	750	150	2	
5600	2	720	120	2	
5700	2	720	120	2	
5800	2	660	60	2	
5900	2	720	120	2	
6000	2	660	60	2	
6100	2	660	60	2	
6200	2	660	60	2	
6300	2	660	60	2	
6400	2	780	180	2	
6500	2	720	120	2	
6600	2	720	120	2	
6700	2	660	60	2	
6800	2	660	60	2	
6900	2	720	120	2	
7000	2	660	60	2	
7100	2	660	60	2	
7200	2	660	60	2	
7300	2	690	90	2	
7400	2	720	120	2	
7500	2	660	60	2	
7600	2	660	60	2	
7700	2	660	60	2	
7800	2	660	60	2	
7900	2	660	60	2	
8000	2	660	60	2	

For easy visualization of the results, a plot was created and shown in Figure32. The first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. According to the performance test, the algorithm should detect the incident in between these two lines. The algorithm was run every 30 secs, and the green box, red box, blue box, and

black box indicate No-detection (True), i.e., there is no incident and no detection, No-detection (False), i.e. there is an incident but no detection, True detection, and False detection, respectively.

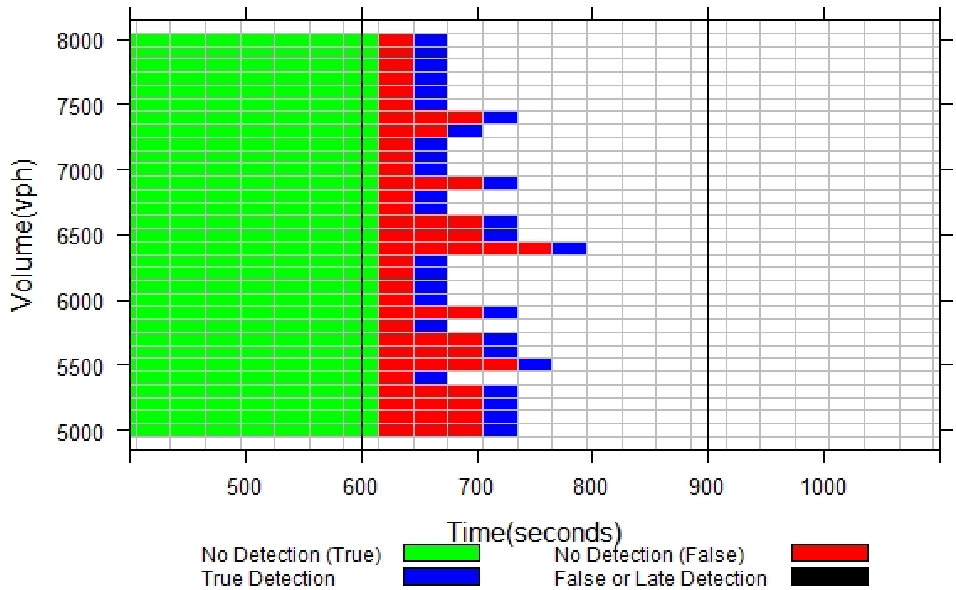


Figure35: Result of California #7in Two-Lane Closed Scenario

Algorithm using SVM:

Similarly, the calibration of the SVM is carried out to find the best SVM model parameters that are used in this algorithm. The dataset with a 1st random seed, i.e., 369, is used for calibration. The parameters used in SVM are speed, occupancy, and number of the vehicle for t, t-1 and t-2 time steps, so in total, there are 18 parameters. The SVM is tuned under the range of epsilon from 0 to 1with an increment of 0.01 and different cost from 1 to 100.

Optimum Parameters found inthe two-lane closed case areas follows:

- SVM-Type: C-classification
- SVM-Kernel: radial
- Cost: 50
- Epsilon (gamma): 0.05555556
- Number of Support Vectors: 217

Table 17: Result of SVM in Two-Lane Closed Scenario.

Volume	Alarm type	Detect time (sec)	Time to detect (sec)
5000	2	630	30
5100	2	660	60
5200	2	660	60
5300	2	660	60
5400	2	630	30
5500	2	660	60
5600	2	630	30
5700	2	630	30
5800	2	630	30
5900	2	630	30
6000	2	630	30
6100	2	630	30
6200	2	630	30
6300	2	630	30
6400	2	630	30
6500	2	630	30
6600	2	630	30
6700	2	630	30
6800	2	630	30
6900	2	630	30
7000	2	630	30
7100	2	630	30
7200	2	630	30
7300	2	630	30
7400	2	630	30
7500	2	630	30
7600	2	630	30
7700	2	630	30
7800	2	630	30
7900	2	630	30
8000	2	630	30

As before for the visualization of the results of the SVM model, a plot is created as shown in Figure33. In this figure, the first dark black vertical line at 600 seconds is the point when the incident starts and the second black vertical line is 5 minutes after the incident. The green box, red box, blue box, and black box indicates No detection (True), i.e. there is no incident and no detection, No detection (False), i.e. there is an incident but no detection, True detection, and False detection, respectively.

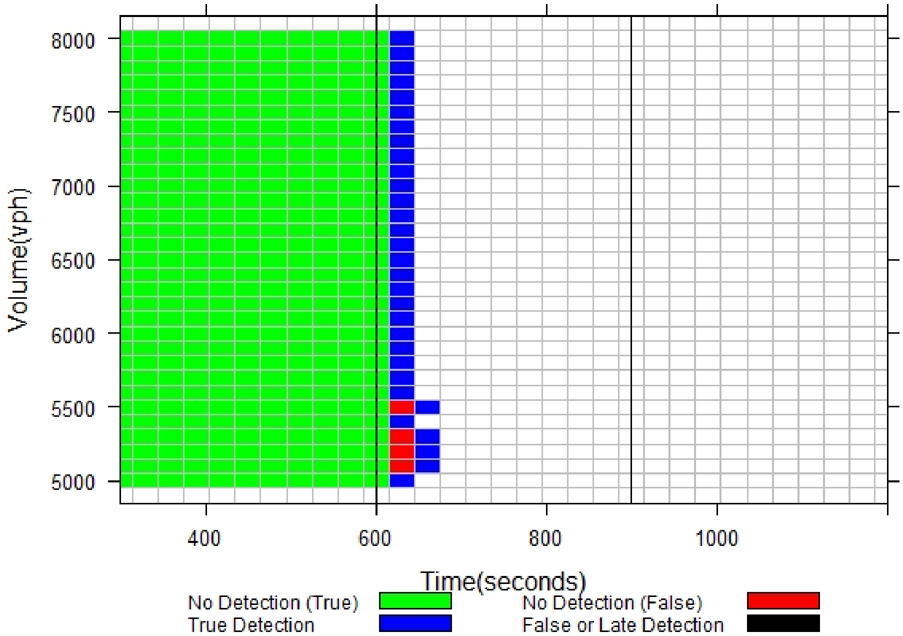


Figure36: Result of SVM inTwo-Lane Closed Scenario.

Algorithm #A:

Similar to the California algorithm at first, the calibration of the thresholds needs to be carried out to find the best thresholds used for algorithm #A. The dataset with a 1st random seed, i.e. 369, is used for calibration. Different sets of OCC Threshold, Vehicle Threshold, and SMS threshold are tested, and the threshold set with minimum false alarm rate is selected as the best threshold. In this case, the best threshold found was found to be OCC Threshold (T1) = 0.8, Vehicle Threshold (T2) = 10 and SMS threshold (T3) =0. The results are shown in table16 and Figure34 below.

Table 18: Result of Algorithm #A in Two-Lane Closed Scenario.

Thresholds: (T1) = 0.8, (T2) = 10 and (T3) = 0			
Volume	Alarm type	Detection time (sec)	Time to detect (sec)
5000	2	750	150
5100	2	750	150
5200	2	750	150
5300	2	780	180
5400	2	750	150
5500	2	750	150
5600	2	750	150
5700	2	750	150

Table 16: Continued

5800	2	750	150
5900	2	750	150
6000	2	750	150
6100	2	750	150
6200	2	750	150
6300	2	750	150
6400	2	750	150
6500	2	720	120
6600	2	720	120
6700	2	720	120
6800	2	720	120
6900	2	720	120
7000	2	720	120
7100	2	720	120
7200	2	720	120
7300	2	720	120
7400	2	720	120
7500	2	720	120
7600	2	720	120
7700	2	720	120
7800	2	720	120
7900	2	720	120
8000	2	720	120

As before for the visualization of the results of Algorithm #A, a plot is created as shown in Figure34. In this figure the first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. The green box, red box, blue box, and black box indicate No detection (True), i.e. there is no incident and no detection, No detection (False), i.e. there is an incident but no detection, True detection, and False detection, respectively.

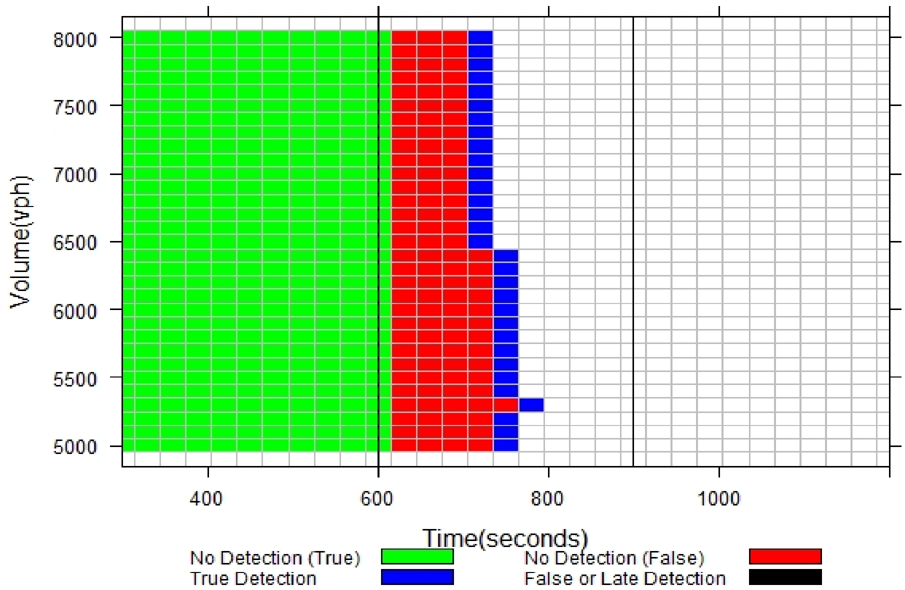


Figure37: Result of Algorithm #A in Two-Lane Closed Scenario.

Algorithm #B:

Apart from other algorithms used, this algorithm is based on AVI (Automatic Vehicle Identification) technology, and data is collected only at two stations, detectors #8 and #9, within which the incident occurs. Only speed and timestamp of vehicles at upstream and downstream are used. However, calibration is done similarly as in the California algorithm; the calibration of the thresholds is carried out to find the best thresholds that give a minimum false alarm rate. The dataset with 1st random seed, i.e. 369, is used for calibration. Different sets of thresholds, the threshold fora number of vehicles and threshold for the factor of safety, are tested, and the threshold set with the minimum false alarm rate is selected as the best threshold. In this case, the best threshold found was the threshold for a number of vehicles = 1, the threshold for the factor of safety = 1.25 with a detection rate of 100% and with no false alarm rate when calculated from equations 1 and 2 in Chapter IV. The results are shown in table17andFigure35below.

Table 19: Result of Algorithm #B in Two-Lane Closed Scenario.

Thresholds: Number of vehicles = 1, Factor of safety = 1.25, Count = 5			
5000	2	617	17
5100	2	621	21
5200	2	620	20
5300	2	619	19

Table 17: Continued

5400	2	618	18
5500	2	614	14
5600	2	617	17
5700	2	622	22
5800	2	613	13
5900	2	618	18
6000	2	622	22
6100	2	616	16
6200	2	618	18
6300	2	616	16
6400	2	615	15
6500	2	619	19
6600	2	618	18
6700	2	617	17
6800	2	618	18
6900	2	615	15
7000	2	619	19
7100	2	620	20
7200	2	614	14
7300	2	615	15
7400	2	617	17
7500	2	618	18
7600	2	615	15
7700	2	617	17
7800	2	618	18
7900	2	616	16
8000	2	615	15
5000	2	617	17

Againfor the visualization of the results of Algorithm #A, a plot is created as shown in Figure35. In thefigure, the first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. The green box, red box, blue box, and black box indicate No detection (True), i.e., there is no incident and no detection, No detection (False), i.e., there is an incident but no detection, True detection, and False detection, respectively.

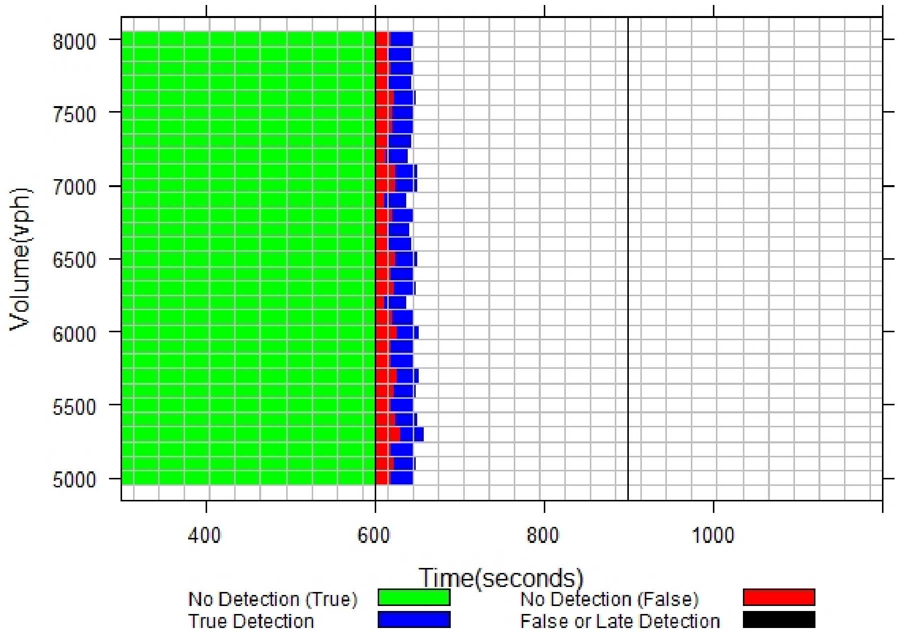


Figure38: Result of Algorithm #B in Two-Lane Closed Scenario.

It is to be noted that algorithm #B runs on Re-identification data, so the percentage of vehicles that are correctly re-identified needs to be accessed. The figure shows the performance of the algorithm in a different percentage of appropriately re-identified vehicles.

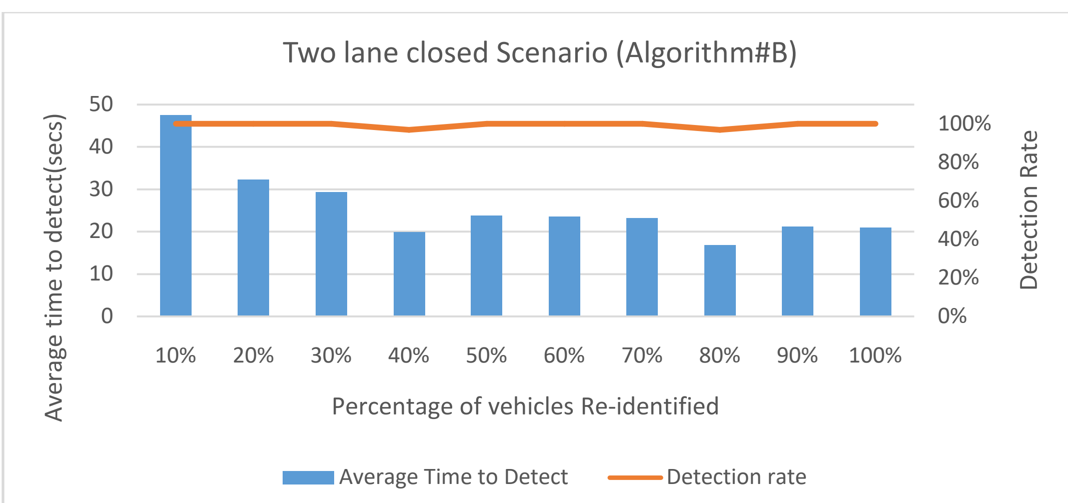


Figure39: Detection Rate and Average Time to Detect under Different Percent of Re-identified Vehicles.

InFigure36, the primary vertical axis at left indicates the average time to detect the incident, whereas the second axis represents the detection rate. It can be asserted that as the percentage of vehicles re-identified decreases the average time to detect increases. The detection rate is varied at 80%,which means the incident is not-correctly detected but again comes back at 70% till 40% change upon the percentage of vehicles re-identified. Most of the literature asserts that 90% of the vehicles can be correctly identified, and the result shows that the performance of the algorithm at 90% of vehicles re-identified has no significant change in detection rate and a very slight change in average time to detect compared to 100% re-identified vehicles.

SVM using Re-Identification:

This algorithm is based on AVI (Automatic Vehicle Identification); data is collected only at two stations, detectors #8 and #9, within which the incident occurs. Only the number of vehicles and timestamp of vehicles upstream and downstream are used. However, calibration is done similarly asSVM using detector data. The dataset with 1st random seed, i.e. 369, is used for calibration. The parameters used in this SVM are travel time and number of the vehicle for ttime steps of the 30sec time interval for both upstream and downstream detectors, so in total there are three parameters. The SVM model is then tuned under the range of epsilon from 0 to 1 with an increment of 0.01 and different cost from 1 to 100. The results are shown in Table 12.

Table 20: Result of SVM Re-ID in Two-Lane Closed Scenario

Volume	Alarm type	Detect time	Time to detect
5000	2	660	60
5100	2	630	30
5200	2	630	30
5300	2	630	30
5400	2	630	30
5500	2	630	30
5600	2	630	30
5700	2	630	30
5800	2	630	30
5900	2	630	30
6000	2	630	30
6100	2	630	30
6200	2	630	30
6300	2	630	30
6400	2	630	30
6500	2	630	30

Table 18: Continued

6600	2	630	30
6700	2	630	30
6800	2	630	30
6900	2	630	30
7000	2	630	30
7100	2	630	30
7200	2	630	30
7300	2	630	30
7400	2	630	30
7500	2	630	30
7600	2	630	30
7700	2	630	30
7800	2	630	30
7900	2	630	30
8000	2	630	30

For easy visualization of the results, a plot is created as shown in Figure30. In this figure the first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. According to the performance test, the algorithm should detect the incident in between these two lines. The algorithm was run every 10 secs in this case, and the green box, red box, blue box, and black box indicates; no detection (True), i.e. there is no incident and no detection, no detection (False), i.e. there is an incident but no detection, True detection, and False detection respectively. Each gray grid is 30 secs apart, as in this case, 30 secs of data is collected and analyzed in each time interval of 30 seconds.

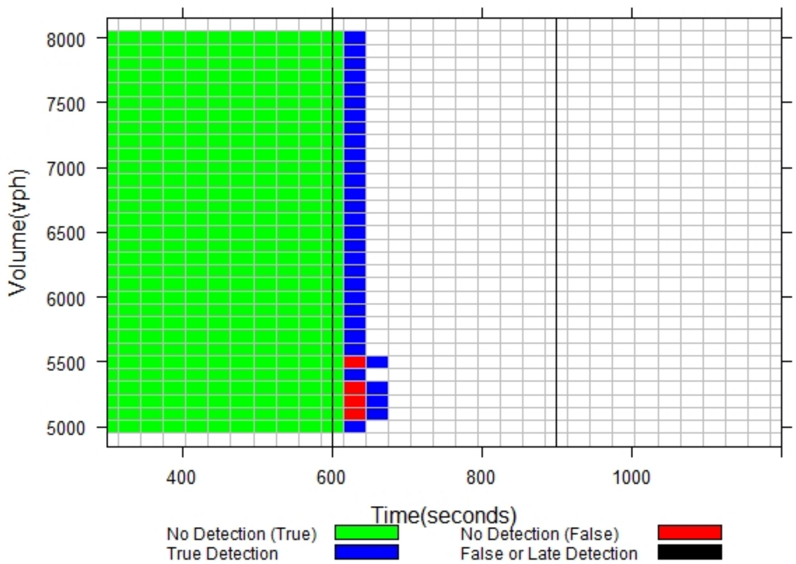


Figure40: Result of SVM Re-ID in One Lane Closed Scenario

Also, this algorithm runs on Re-identification data, so the percentage of vehicles that are correctly re-identified needs to be accessed. Figure31 and Figure29 show the performance of the algorithm in different percentages of appropriately re-identified vehicles.

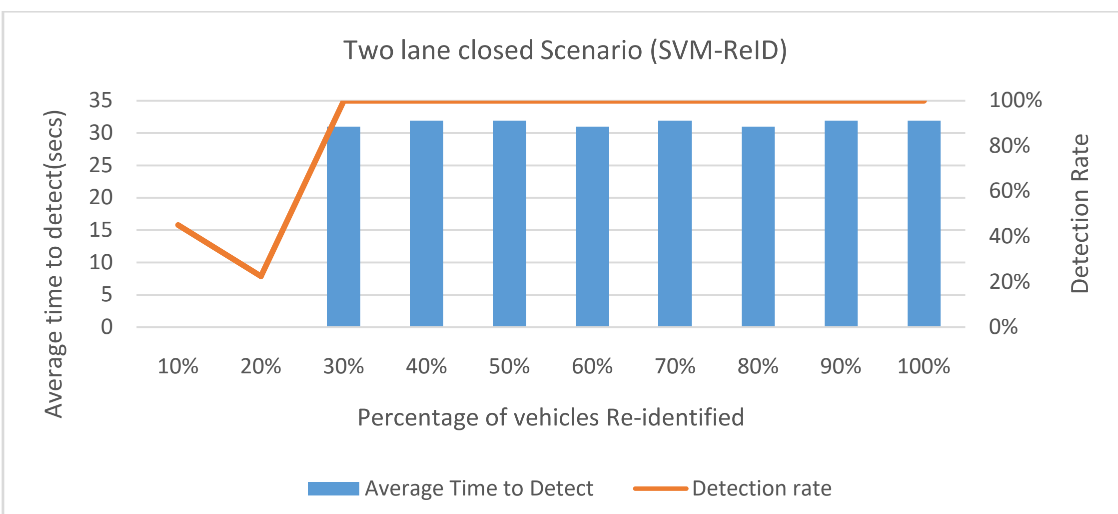


Figure41: Detection Rate and Average Time to Detect under Different Percent of Re-identified Vehicles (SVM).

In this scenario, the detection rate is constant at 100%. That means the incident is correctly detected until 30 % and does not change with the percentage of vehicle re-identified. Only if 20% of vehicles are re-identified does the detection rate drop. Most of the literature asserts that 90% of the vehicles can be correctly identified, and the result shows that the performance of the algorithm at 90% of vehicles re-identified has no significant change in detection rate or average time to detect compared to 100% re-identified vehicles.

Summary of Results in the two-lane closed scenario:

Looking at the figures and tables of results in the 2 lanes closed scenario, it is clear that the SVM model was the best performing using the detector data. The SVM model has no false alarm rate and an average time to detect the incident of about 33.87secs when tested at all volumes from 5000 to 8000, as shown in the figure. The California model is comparatively better than Algorithm #A in detecting an incident in this scenario, although it has detection time better than algorithm # B. Algorithm #A has the worst minimum detection time. However, Algorithm # B has perfectperformance with 100% detection rate and average minimum time to detect just 17.32secs. Hence, in the 2 lanes closed scenario, the best results were achieved from the SVM model in detector data and Algorithm #B in overall results.The summary is presented in table19 below

.

Table 21: Summary of Result in Two-Lane Closed Scenario

#	Performance measures	California #7	SVM(Detector)	Algorithm #A	Algorithm #B	SVM(REID)
1	**Alarm Type 1**	0	0	0	0	0
2	**Alarm Type 2**	31	31	31	31	31
3	**Alarm Type 3**	0	0	0	0	0
4	**Average Detection Time (secs)**	689.03	633.87	735.48	617.32	630.97
5	**Detection Rate**	100.00%	100.00%	100.00%	100.00%	100.00%
6	**False Alarm Rate**	0.00%	0.00%	0.00%	0.00%	0.00%

5.3 TWO LANE CLOSED WITH WORK-ZONE BOTTLENECK SCENARIO:

In this scenario the two rightmost lanes were blocked by vehicles after 600 simulation seconds for 15 minutes, and apart from that there is a work-zone-bottleneck in D/S of the incident location between detectors #9 and #10, where the lane drops from 4 to 3 lanes and again back to 4 lanes before detector #10.This scenario is considered to analyze the algorithm performance under the congested road condition. A total of 62 numbers of the simulation were run varying volume from 5000 vehicles per hour to 8000 vehicles per hour in an increment of 100 vehicles per hour with two different seeds. The total duration of

the simulation was 30 minutes, where 5 mins of the warm-up time were considered, and data was collected only after 300 simulation seconds. Calibration of the detection models (algorithms) was performed on a training dataset of a 1st random seed, and testing was carried out for 2nd random seed of the simulation. Five different algorithms above, as described in previous chapters, were used. Results obtained from these algorithms are described one by one and further summarized to see the differences in the performance of these algorithms on this specific scenario.

California #7:

As this case is similar to the 2 lanes closed situation, but congestion is reaching our incident area; after the calibration of the thresholds, the calibrated parameters are also found to be similar. The dataset with a 1st random seed, i.e. 369, is used for calibration. Different sets of T1, T2, and T3 are tested, and the threshold set with minimum false alarm rate is selected as the best threshold. In this case, the best threshold found was T1 = 5, T2 = 0.22 and T3 = 40 with a detection rate of 100.00% and false alarm rate of 0.00% when calculated from equations 6 and 7 in Chapter IV. The results are shown in table20 and Figure39 below.

Table 22: Result showing California #7 in Two-Lane Closed Work-zone Scenario.

Thresholds: T1= 5, T2= 0.22, T3= 40					
Volume	Alarm type	Detection time (sec)	Time to detect (sec)	Position (miles)	Remarks
5000	2	720	120	2	
5100	2	660	60	2	
5200	3	1800	1200	1.75	Wrong position
5300	2	720	120	2	
5400	2	720	120	2	
5500	2	750	150	2	
5600	2	720	120	2	
5700	2	690	90	2	
5800	2	690	90	2	
5900	2	720	120	2	
6000	2	720	120	2	
6100	2	720	120	2	
6200	2	720	120	2	
6300	2	660	60	2	
6400	2	690	90	2	
6500	2	660	60	2	
6600	2	660	60	2	
6700	2	690	90	2	
6800	2	660	60	2	

Table 20: Continued

6900	2	660	60	2	
7000	2	690	90	2	
7100	2	660	60	2	
7200	2	690	90	2	
7300	2	690	90	2	
7400	2	660	60	2	
7500	2	660	60	2	
7600	2	660	60	2	
7700	2	690	90	2	
7800	2	690	90	2	
7900	2	660	60	2	
8000	1	570	-30	2	Early Detection

Same as before for the visualization of the results, a plot is created as shown in Figure39. The first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. The green box, red box, blue box, and black box indicate No detection (True), i.e. there is no incident and no detection, No detection (False), i.e. there is an incident but no detection, True detection, and False detection, respectively.

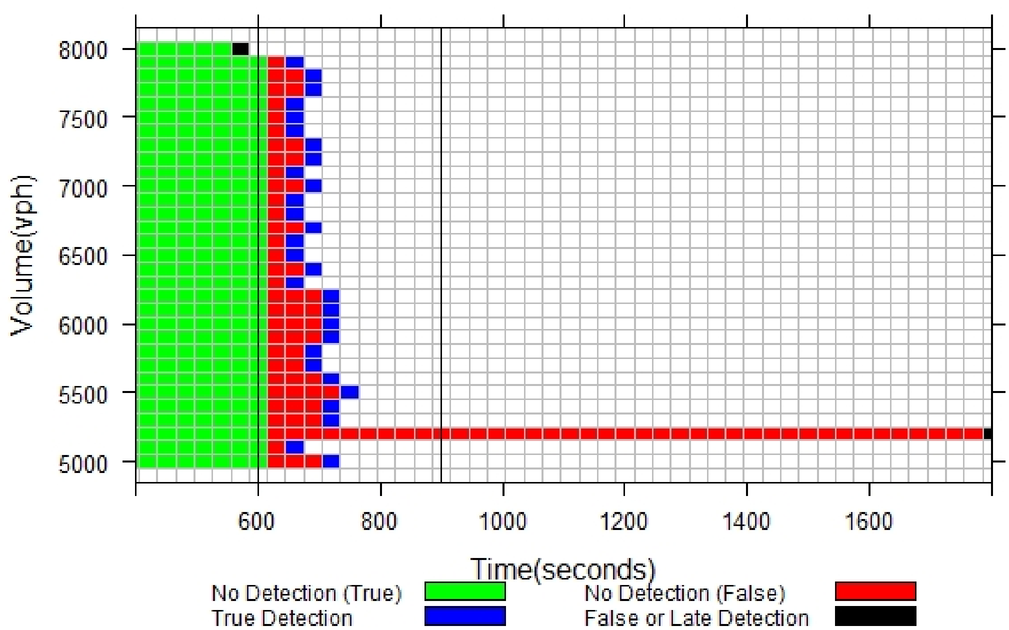


Figure42: Result of California #7 in Two-Lane Closed Work-zone Scenario.

Algorithm SVM using detector data:

Similarly, the calibration of the SVM is carried out to find the best SVM model parameters that are used in this algorithm. The dataset with a 1st random seed, i.e. 369, is used for calibration. The parameters used in SVM are speed, occupancy, and the number of the vehicle for t, t-1, and t-2 time steps, so in total there are 18 parameters. The SVM is tuned under the range of epsilon from0 to 1 with an increment of 0.01 and different cost from 1 to 100.

Optimum Parameters after tuning SVM:

- SVM-Type: C-classification
- SVM-Kernel: radial
- Cost: 10
- Epsilon (gamma): 0.05555556
- Number of Support Vectors: 430

The results of SVM prediction are shown in table21 and Figure40 below:

Table 23: Result showing the performance ofSVM in Two-Lane Closed Work-zone Scenario.

Volume	Alarm type	Detect time (sec)	Time to detect (sec)
5000	2	660	60
5100	2	660	60
5200	2	660	60
5300	2	660	60
5400	2	660	60
5500	2	660	60
5600	2	660	60
5700	2	630	30
5800	2	660	60
5900	2	660	60
6000	2	660	60
6100	2	660	60
6200	2	660	60
6300	2	660	60
6400	2	660	60
6500	2	660	60
6600	2	630	30
6700	2	660	60
6800	2	660	60
6900	2	660	60
7000	2	630	30

Table 21: Continued

7100	2	660	60
7200	2	660	60
7300	2	660	60
7400	2	660	60
7500	2	660	60
7600	2	690	90
7700	2	660	60
7800	2	660	60
7900	2	690	90
8000	2	660	60

For visualization of the results, a plot is created as shown in Figure40. the first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. The green box, red box, blue box, and black box indicate No detection (True), i.e., there is no incident and no detection, No detection (False), i.e., there is an incident but no detection, True detection, and False detection, respectively.

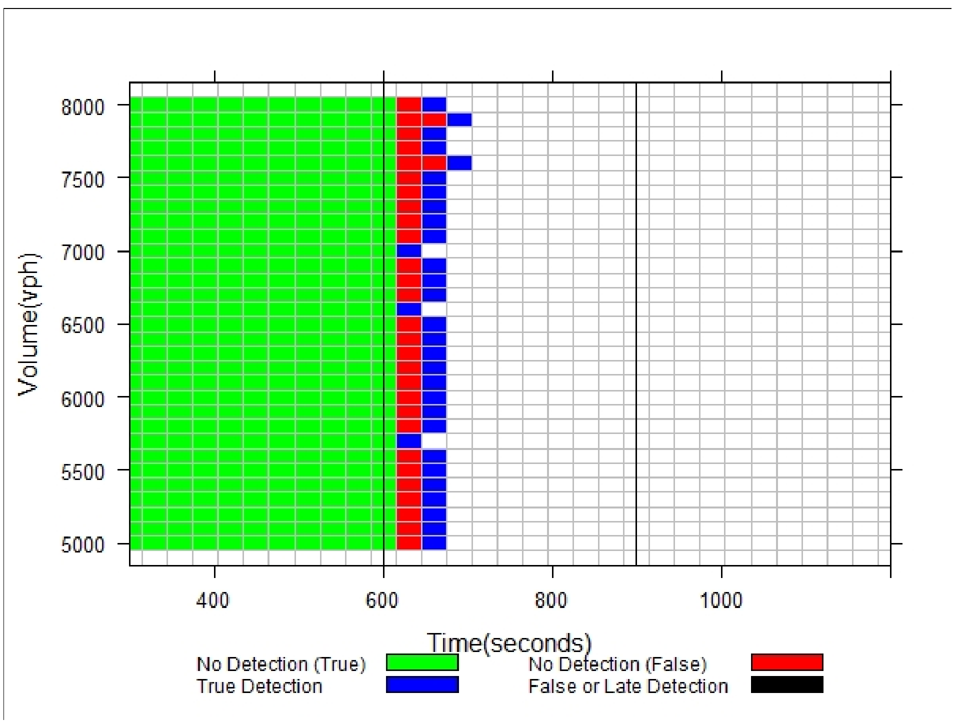


Figure43: Result of SVM in Two-Lane Closed Work-zone Scenario.

Algorithm #A:

At first, the calibration of the thresholds is carried out to find the best thresholds used for algorithm #A. The dataset with a 1st random seed, i.e. 369, is used for calibration. Different sets of OCC Threshold, Vehicle Threshold, and SMS threshold are tested, and the threshold set with minimum false alarm rate is selected as the best threshold. In this case, the best threshold found was found to be OCC Threshold (T1) = 0.75, Vehicle Threshold (T2) = 5 and SMS threshold (T3) = 1.The results are shown in the table and figure below.

Table 24: Result of Algorithm #A in Two-Lane Closed Work-zone Scenario.

Thresholds: (T1) = 0.75, (T2) = 5 and (T3) = 1				Position	Remark
Volume	Alarm type	Detect time (sec)	Time to detect (sec)		
5000	2	750	150	1.9	
5100	2	840	240	1.9	
5200	2	750	150	1.9	
5300	2	780	180	1.9	
5400	2	750	150	1.9	
5500	1	780	180	1.5	Wrong position
5600	2	780	180	1.9	
5700	2	720	120	1.9	
5800	2	660	60	1.9	
5900	1	750	150	1.5	Wrong position
6000	2	750	150	1.9	
6100	2	750	150	1.9	
6200	2	750	150	1.9	
6300	2	720	120	1.9	
6400	2	690	90	1.9	
6500	2	720	120	1.9	
6600	2	720	120	1.9	
6700	2	750	150	1.9	
6800	2	630	30	1.9	
6900	2	780	180	1.9	
7000	2	720	120	1.9	
7100	2	720	120	1.9	
7200	2	720	120	1.9	
7300	2	690	90	1.9	
7400	2	720	120	1.9	
7500	2	600	0	1.9	

Table 22: Continued

7600	2	780	180	1.9	
7700	2	720	120	1.9	
7800	2	750	150	1.9	
7900	3	1170	570	0.9	Wrong position
8000	2	720	120	1.9	

For visualization of the results, a plot is created as shown in Figure41. the first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. The green box, red box, blue box, and black box indicate No detection (True), i.e., there is no incident and no detection, No detection (False), i.e., there is an incident but no detection, True detection, and False detection, respectively.

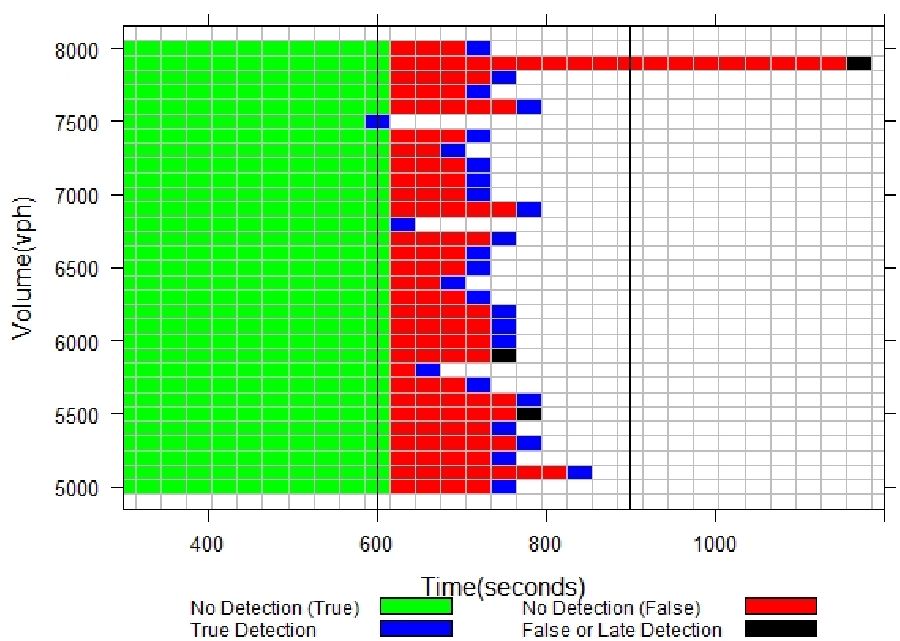


Figure44: Results showing performance of Algorithm #A in Two-Lane Closed with Work-zone Scenario.

Algorithm #B:

This algorithm is based on Re-ID (Re-Identification) technology; data is collected only at two stations, detectors #8 and #9, between which the incident occurs. Only speed and timestamp of vehicles at upstream and downstream are used. However, calibration is done similarly as the California algorithm; the calibration of the thresholds is carried out to find the best thresholds that give the minimum false

alarm rate. The dataset with a 1st random seed, i.e. 369, is used for calibration. Different sets of thresholds, the threshold for a number of vehicles and threshold for the factor of safety are tested, and the threshold set with minimum false alarm rate is selected as the best threshold. In this case, the best threshold found was the threshold for a number of vehicles = 0, the threshold for the factor of safety = 8.5 with a detection rate of 100% and with no false alarm rate when calculated from equations 6 and 7 in Chapter IV. The results are shown in table23 and Figure42 below.

Table 25: Result showing performance of Algorithm #B in Two-Lane Closedwith Work-zone Scenario.

Thresholds: Number of vehicles = 9, Factor of safety = 2 , Count=10			
Volume	Alarm type	Detect time	Time to detect
5000	2	661	61
5100	2	697	97
5200	2	642	42
5300	2	695	95
5400	2	654	54
5500	2	629	29
5600	2	650	50
5700	2	672	72
5800	2	617	17
5900	2	624	24
6000	2	678	78
6100	2	652	52
6200	2	637	37
6300	2	648	48
6400	2	632	32
6500	2	645	45
6600	2	648	48
6700	2	639	39
6800	2	642	42
6900	2	649	49
7000	2	648	48
7100	2	659	59
7200	2	644	44
7300	2	636	36
7400	2	649	49
7500	2	669	69
7600	2	668	68
7700	2	644	44

Table 23: Continued

7800	2	629	29
7900	1	377	-223
8000	2	647	47

For visualization of the results, a plot is created as shown in Figure42. The first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. The green box, red box, blue box, and black box indicate No detection (True), i.e. there is no incident and no detection, No detection (False), i.e. there is an incident but no detection, True detection, and False detection, respectively.

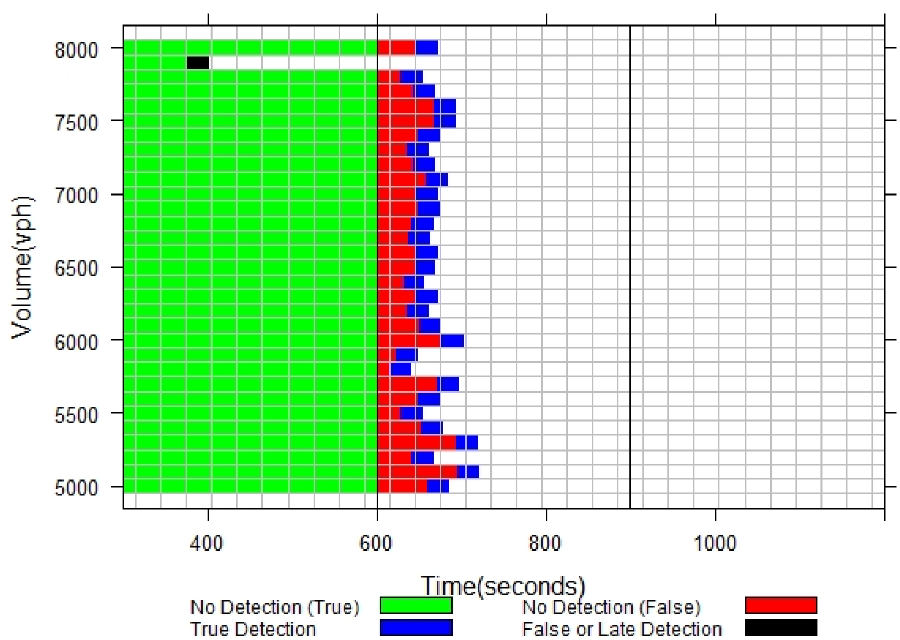


Figure45: Result of Algorithm #B in Two-Lane Closed Work-zone Scenario.

It is to be noted that algorithm #B runs on Re-identification data, so the percentage of vehicles that are correctly re-identified needs to be assessed. The figure shows the performance of the algorithm in different percentages of appropriately re-identified vehicles.

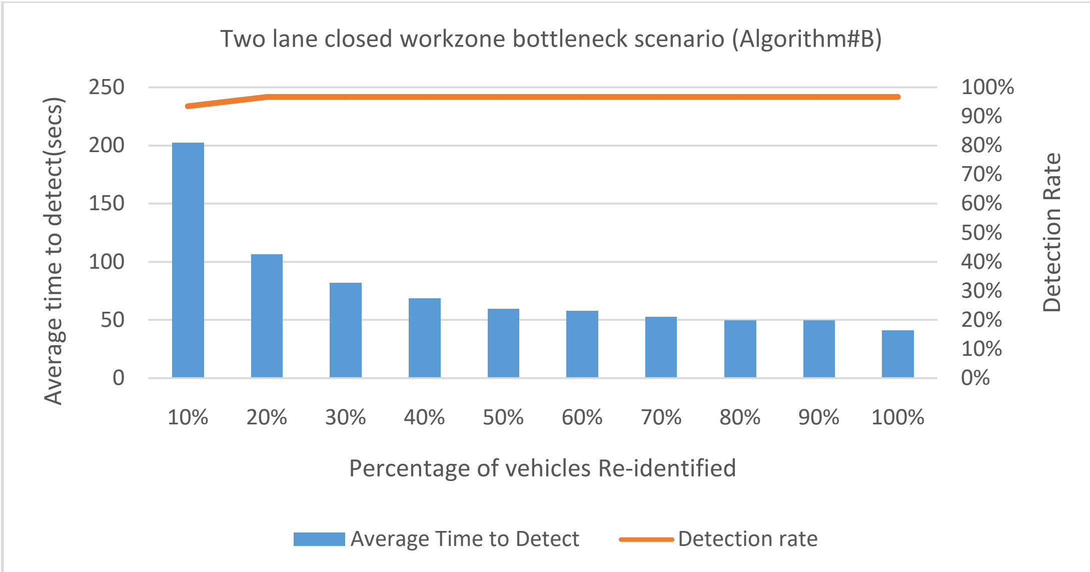


Figure46: Detection Rate and Average Time to Detect under Different Percent of Re-identified Vehicles.

In Figure43, the primary vertical axis at left indicates the average time to detect the incident whereas the second axis represents the detection rate. It can be asserted that as the percentage of vehicles re-identified decreases the average time to detect increases. The detection rate is decreasing as the percentage of vehicles re-identified decreases. Thus, the incident is more correctly detected at a higher percent of re-identified vehicles andfails to detect correctly below 20%; from 30% to 100% the detection rate is almost between 97% percent. Most of the literature asserts that 90% of the vehicles can be correctly identified, and the result shows that the performance of the algorithm at 90% of vehicles re-identified has no significant change in detection rate and a very slight change in average time to detect compared to 100% re-identified vehicles.

SVM using Re-Identification data:

In this bottleneck condition, there may be an error in the estimated incident from the travel time only. As the vehicles are in a congested state, the travel time measured at these points can lead to greater travel time than the vehicles actually travel when compared to the no bottleneck condition. This, in turn, is the reason for such a high number of vehicles checked, i.e. 9 out of 10 calibrated on algorithm # B. This not only makes an algorithm to detect the incident late on the time but also is the main causefor incorrect detection. To eliminate this problem, the number of vehicles downstream and upstream is used in this SVM model so that the incident is not merely based on the travel time but also on the number of vehicles counted at these two points. Theresult of modified SVM using Re-Identification is presented below in Figure44.

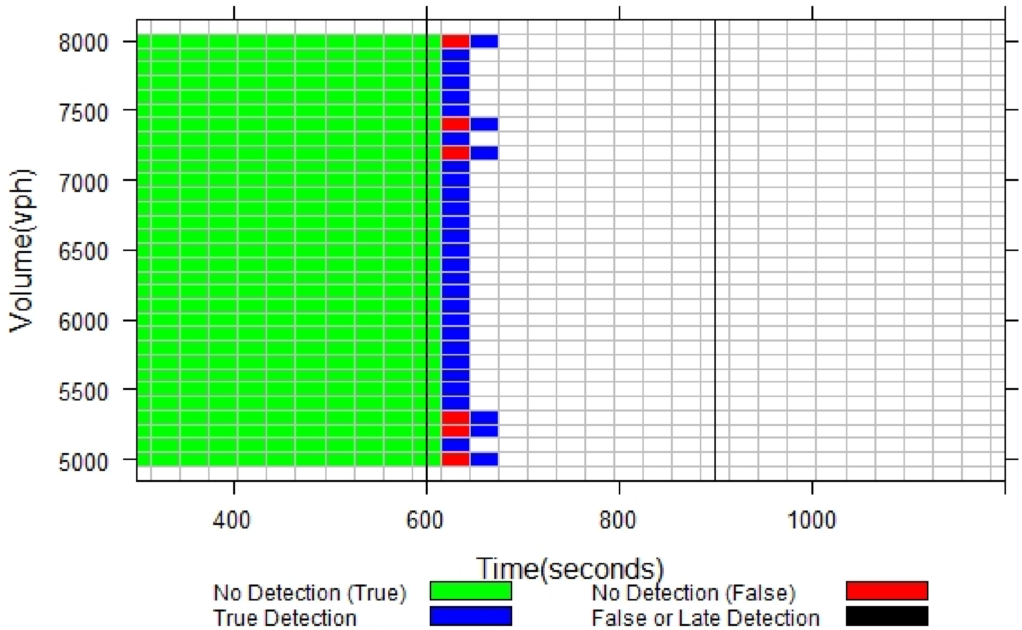


Figure47: Result of SVM using Re-Id in Two-Lane Closed Work-zone Scenario.

The first dark black vertical line at 600 seconds is the point when the incident starts, and the second black vertical line is 5 minutes after the incident. The green box, red box, blue box, and black box indicate No detection (True), i.e. there is no incident and no detection, No detection (False), i.e. there is an incident but no detection, True detection, and False detection respectively.The result indicates that the average detection time is reduced to about 630 secs from 650. Not only is the time taken to detect the incident reduced,but the detectionrate is better – 100% with no false detection.

Again, to access the performance under the different percent of vehicles that are correctly re-identified the algorithm is run under the different percentages of vehicles re-identified and presented below inFigure45.

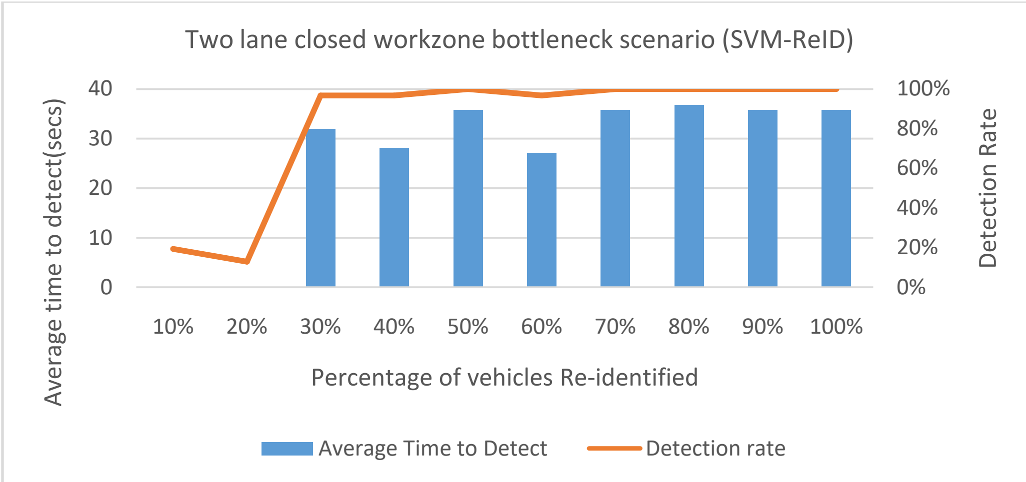


Figure48: Detection Rate and Average Time to Detect under Different Percent of Re-identified Vehicles.

In Figure45, the primary vertical axis at left indicates the average time to detect the incident, whereas the second axis represents the detection rate. There is nosignificant change in the incident's detection rate or average time to detect from 70% to 100% of vehicles re-identified, whereas the detection rate decreases at 60% of vehicles re-identified but also indicated the drop in average time to detect. The algorithm fails to detect any incident at 20% and 10% with nearly 0% detection rate. This figure demonstrates that the algorithm performs as good in 30% to 100% of re-identified vehicles.

Summary of Results in two-lane closed work-zone bottleneck scenario:

Looking at the figures and tables of results in this scenario, it is clear that the SVM model was the best performing using the detector data. The SVM (Re-ID) model has no false alarm rate with an average time to detect the incident of about 35.81 secs when tested at all volumes from 5000 to 8000, as shown in the figure. The California model and Algorithm #Awere comparatively bad in detecting an incident in this scenario, although the California algorithm has a detection time better than algorithm #B. Algorithm #A has the worst minimum detection time; however, Algorithm #B has an excellent performance measure with an approximately 96% detection rate and average minimum time to detectat just 50secs. Hence, in the 2 lanes closed bottleneck scenario, the best results were achieved from the SVM model in detector data and SVM (in Re-ID) in overall results. The summary is presented in table24 below.

Table 26: Summary of Result in Two-Lane Closed Bottleneck Scenario

#	Performance measures	California #7	SVM(Detector)	Algorithm #A	Algorithm #B	SVM(REID)
1	**Alarm Type 1**	1	0	2	1	0
2	**Alarm Type 2**	29	31	28	30	31
3	**Alarm Type 3**	1	0	1	0	0
4	**Average Detection Time (secs)**	726.00	659.03	751.00	650.13	635.81
5	**Detection Rate**	93.55%	100.00%	90.32%	96.77%	100.00%
6	**False Alarm Rate**	6.45%	0.00%	9.68%	3.23%	0.00%

5.4 SUMMARY OF OVERALL RESULTS:

Heretofore, the results were presented case by case. Now it is noteworthy to look at the overall performance of the algorithms tested in different road conditions. The table below shows the overall result of the algorithms used in this study. The result indicates that the overall performance of SVM is the best among the algorithms that make use of detector data. Algorithm #A, developed in this study, also performed reasonably well with a detection rateof 98.92%and false alarm rate of 1.08%; that is as good as the SVM model, but the SVM model had zero false alarms in both cases and is thus considered the superior model among all these algorithms. California #7 was good at two-lane closed scenarios, but it performed very poorly in the one lane closed scenario.Hence, we can say that the algorithms developed in this study werebetter than or as good as the algorithms that have been used in the field to date.

Table 27: Summary of Overall Performance.

<table>
<tr><td>Algorithms</td><td rowspan="2">#</td><td rowspan="2">Performance measures</td><td>California #7</td><td>SVM(Detector)</td><td>Algorithm #A</td><td>Algorithm #B</td><td>SVM(REID)</td></tr>
<tr><td>Data Type</td><td>Loop Detector</td><td>Loop Detector</td><td>Loop Detector</td><td>Re-Identification</td><td>Re-Identification</td></tr>
<tr><td rowspan="6">One Lane closed</td><td>1</td><td>Alarm Type 1</td><td>7</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>2</td><td>Alarm Type 2</td><td>12</td><td>31</td><td>31</td><td>31</td><td>31</td></tr>
<tr><td>3</td><td>Alarm Type 3</td><td>12</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>4</td><td>Average Detection Time (secs)</td><td>1066.25</td><td>630.65</td><td>627.74</td><td>617.32</td><td>630</td></tr>
<tr><td>5</td><td>Detection Rate</td><td>38.71%</td><td>100.00%</td><td>100.00%</td><td>100.00%</td><td>100%</td></tr>
<tr><td>6</td><td>False Alarm Rate</td><td>61.29%</td><td>0.00%</td><td>0.00%</td><td>0.00%</td><td>0%</td></tr>
<tr><td rowspan="6">Two Lane closed</td><td>1</td><td>Alarm Type 1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>2</td><td>Alarm Type 2</td><td>31</td><td>31</td><td>31</td><td>31</td><td>31</td></tr>
<tr><td>3</td><td>Alarm Type 3</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>4</td><td>Average Detection Time (secs)</td><td>689.03</td><td>633.87</td><td>735.48</td><td>617.32</td><td>630.97</td></tr>
<tr><td>5</td><td>Detection Rate</td><td>100.00%</td><td>100.00%</td><td>100.00%</td><td>100.00%</td><td>100%</td></tr>
<tr><td>6</td><td>False Alarm Rate</td><td>0.00%</td><td>0.00%</td><td>0.00%</td><td>0.00%</td><td>0%</td></tr>
</table>

Table 27: Continued

Two Lane closed - Bottleneck	1	**Alarm Type 1**	1	0	2	1	0
	2	**Alarm Type 2**	29	31	28	30	31
	3	**Alarm Type 3**	1	0	1	0	0
	4	**Average Detection Time (secs)**	726.00	659.03	751.00	650.13	635.81
	5	**Detection Rate**	93.55%	100.00%	90.32%	96.77%	100%
	6	**False Alarm Rate**	6.45%	0.00%	9.68%	3.23%	0%
Overall Performance	1	**Alarm Type 1**	8	0	2	1	0
	2	**Alarm Type 2**	72	93	90	92	93
	3	**Alarm Type 3**	13	0	1	0	0
	4	**Average Detection Time (secs)**	827.09	641.18	704.74	628.26	632.26
	5	**Detection Rate**	77.42%	100.00%	96.77%	98.92%	100.00%
	6	**False Alarm Rate**	22.58%	0.00%	3.23%	1.08%	0.00%

5.5 RESULTS AFTER VARYING INCIDENT POSITION:

These results were only based on the one location of the incident which is an incident occurring nearthe downstream detector. Thus, it is necessary to determine whether the best performing algorithm using SVM is able to detect an incident when the position of the incident changes. As discussed in Chapter IV, there are a total of 48 detectors placed along the road with the spacing of nearly 1/12th mile and shifted to take into account the effect of position of an incident on the algorithms. The detectors are moved such that the incident occursnear an upstream sensor, in between the upstream and downstream (or an incident in the middle), and near the downstream sensor.This way we can analyze the effect of change in position of the incident on these best performing algorithms. The incident is mainly categorized as near the upstream detector, in between upstream and downstream (or an incident in the middle), and near the downstream detector.

Only SVM models, found to be the best performing in the previous result, were further tested. These tests are also categorized in 3 different scenarios as before which is discussed in subsequent sections.

5.5.1 One lane closed scenario:

The SVM models, one using detector data and the other using Re-Id(especially travel time and number of vehicles) were trained on one random seed and tested on a different random seed similar to the previous test. The only difference in this test is the incident position; instead of training and testing on one particular location of the incident, the location is changed by changing the position of a detector such that the incident happens at U/s, at middle and D/s between 2 detectors and trained combining all three incident locations. The result is shown in Table 28 and in Table 29

Table 28: Performance of SVM models in one lane closed scenario.

	Performance measures	Incident at downstream	Incident at middle	Incident at upstream	Incident at downstream	Incident at middle	Incident at upstream
One lane closed scenario		SVM using detectors 5,6,7,8,9,10and 11			SVM using all detectors		
1	**Alarm Type 1**	5	3	5	0	1	0
2	**Alarm Type 2**	26	28	26	31	30	31
3	**Alarm Type 3**	0	0	0	0	0	0
4	**Average Detection Time (secs)**	634.62	632.14	633.46	639.68	633.00	632.90
5	**Detection Rate**	83.87%	90.32%	83.87%	100.00%	96.77%	100.00%
6	**False Alarm Rate**	16.13%	9.68%	16.13%	0.00%	3.23%	0.00%
One lane closed scenario		SVM using all detectors			SVM using Re-ID & Flow		
1	**Alarm Type 1**	0	1	0	0	0	0
2	**Alarm Type 2**	31	30	31	31	31	31
3	**Alarm Type 3**	0	0	0	0	0	0
4	**Average Detection Time (secs)**	639.68	633.00	632.90	630.00	631.94	660.00
5	**Detection Rate**	100.00%	96.77%	100.00 %	100.00%	100.00 %	100.00%
6	**False Alarm Rate**	0.00%	3.23%	0.00%	0.00%	0.00%	0.00%
One lane closed scenario		SVM using only Travel time			SVM using Re-ID & Flow		
1	**Alarm Type 1**	0	0	0	0	0	0
2	**Alarm Type 2**	31	31	31	31	31	31
3	**Alarm Type 3**	0	0	0	0	0	0
4	**Average Detection Time (secs)**	630.00	634.84	660.00	630.00	631.94	660.00
5	**Detection Rate**	100.00%	100.00%	100.00 %	100.00%	100.00 %	100.00%
6	**False Alarm Rate**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

In this scenario, the results are further categorized according to the position of the incident. The performance measures are also color-coded such that red means bad, green means good, and yellow (or orange) means comparatively satisfactory. There are three different comparisons made within this scenario using four different SVM models which were trained separately according to a data type as indicated on the table.

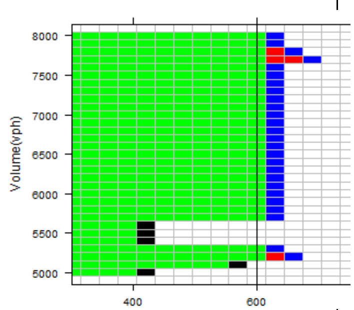
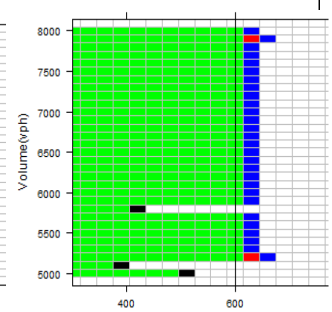
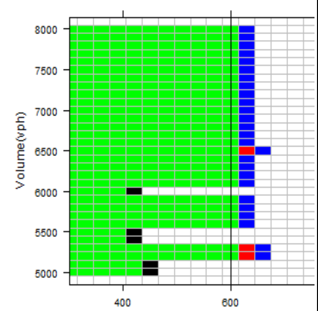
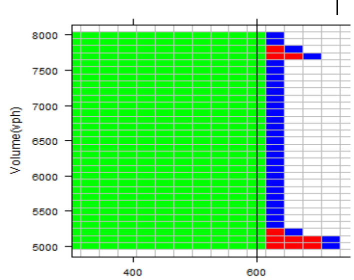
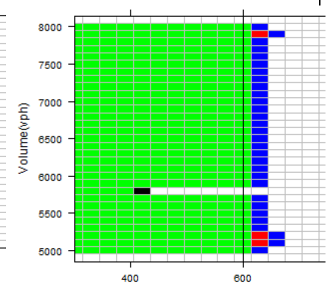
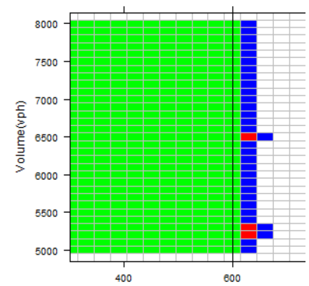
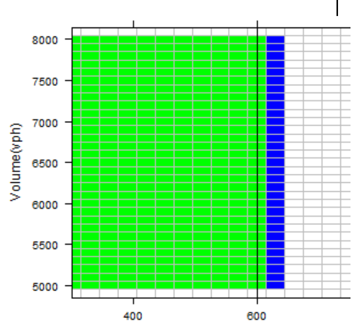
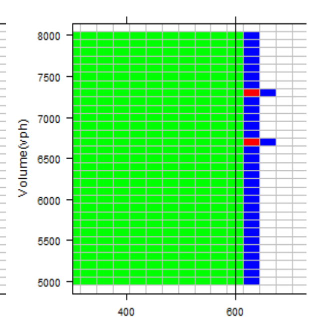
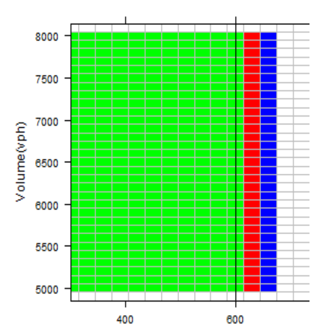
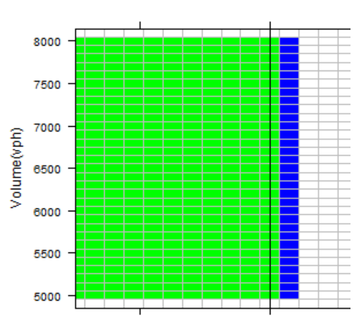
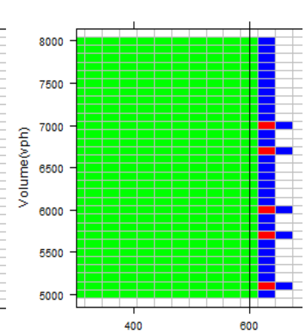
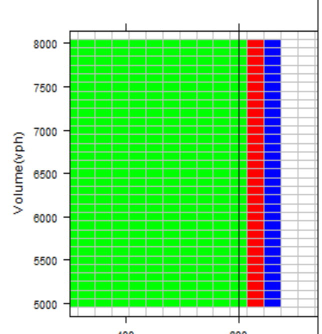
The first comparison is between the SVM models on detector type data; one uses only data from detector 5,6,7,8,9,10 and 11, whereas the other uses data from all 16 detectors. In the first comparison, it is clear that training data sets using all detectors is better as the misclassification error depends greatly upon the number of samples trained as they will affect the weights that the SVM gives to each feature for

predicting the incident. Looking at the figures and colors the model uses, only 5,6,7,8,9,10, and 11 detectors performed better only when the incidents occurred in the middle section. When the SVM is trained based on data from all sensors the detection rate improved from 83 % to 100 % in detection rate whereas the detection time was not changed; instead, it detected a little late for an incident downstream because the incident was detected in the next time step rather than earlier when it was falsely detected.

The next comparison is between the best of SVM using detector type and Vehicle Re-Identification data. In this comparison,it is seen that SVM using travel time and flow performed better than the SVM using detector data. Not only the time to detect is better using travel time and flow but also the detection rate is better when an incident is at U/s and in the middle of the two observation points. While the time to detect incident occurring upstream uses travel time more than detector data because travel time of vehicles when observed between the detectors may not significantly increase immediately when the incident occurs near the U/s observation point.

The last comparison iswithinSVM using Re-identification data; however, one only uses travel time, and the other uses travel time and the number of vehicles observed in between 2 points.The results for both of the models are almost similar, and the only difference is when an incident occurs in the middle, where the model uses travel time and vehicle flow sometimes detects the incident slightly before then using only travel time when tested on the same samples. This can be easily visualized in Table 29.In Table 29, the figures show at what time the incident was detected in blue and black, where the first dark black vertical line at 600 seconds is the point when the incident starts.The green box, red box, blue box, and black box indicate No detection (True) i.e. there is no incident and no detection, No detection (False) i.e. there is an incident but no detection, True detection, and False detection, respectively.

Table 29: Table of Figures showing performance of SVM models in one lane closed scenario.

Data and features used	Legends: No Detection (True) · True Detection · No Detection (False) · False or Late Detection		
	Incident At Downstream	Incident at Middle	Incident at Upstream
SVM using Detector data (trained using only detectors 5,6,7,8,9,10,11) Flow at U/s and D/s for time t, t-1 and t-2 Occupancy at U/s and D/s for time t, t-1 and t-2 Speed at U/s and D/s for time t, t-1 and t-2			
SVM using Detector data (trained using all detectors) Flow at U/s and D/s for time t, t-1 and t-2 Occupancy at U/s and D/s for time t, t-1 and t-2 Speed at U/s and D/s for time t, t-1 and t-2			
SVM using REID and number of vehicles. Travel Time Number of vehicles passing through U/s and D/s			
One Lane closed SVM using REID Only Travel Time			

It is to be noted that SVM using Re-ID runs on Re-identification data so that the percentage of vehicles that are correctly re-identified needs to be assessed. The figureshows the performance of the algorithm in different percentages of appropriately re-identified vehicles.The primary vertical axis at left indicates the average time to detect the incident whereas the second axis represents the detection rate. It can be asserted that as the percentage of vehicles re-identified decreases the average time to detect increases. The detection rate is decreasing as the percentage of vehicles re-identified decreases, so the incident is more correctly detected at a higher percentage of re-identified vehicles and fails to detect correctly below 20%. Most of the literature asserts that 90% of the vehicles can be correctly identified, and the result shows that the performance of the algorithm at 90% of vehicles re-identified has no significant change in detection rate and a very slight change in average time to detect compared to 100% re-identified vehicles.

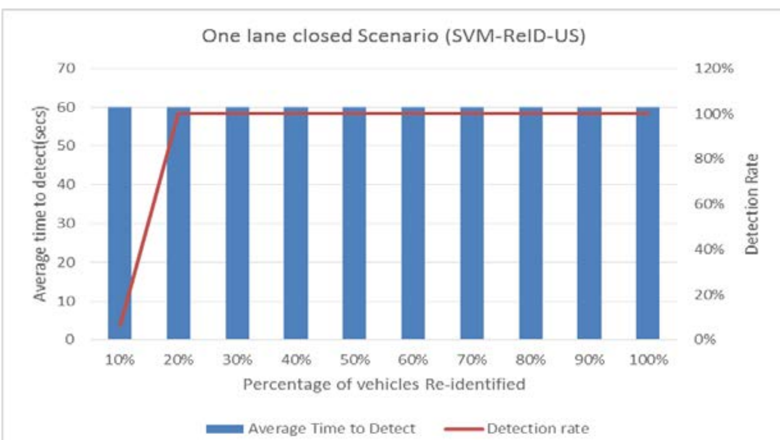


Figure49: Average time to detect and detection rate according to the percentage of vehicles re-identified incident at D/s.

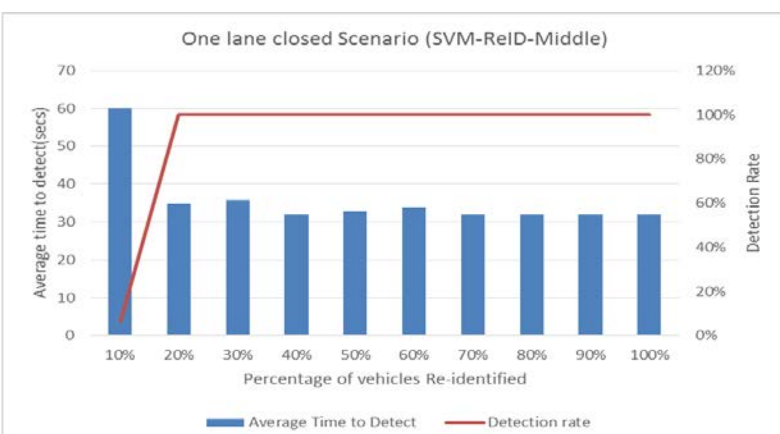


Figure50: Average time to detect and detection rate according to the percentage of vehicles re-identified incident at the middle.

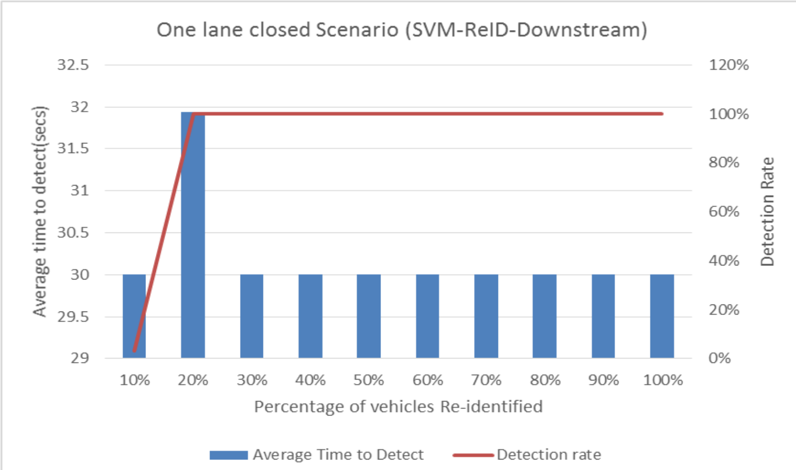


Figure51: Average time to detect and detection rate according to the percentage of vehicles re-identified incident at U/s.

5.5.2 Two-lane closed scenario:

Like the one lane scenario, the SVM models in this scenario also use detector data and Re-Id (especially travel time and number of vehicles) and were trained on one random seed and tested on a different random seed. The difference in this test is the incident position; instead of training and testing on one particular location of the incident, the location is changed by changing the position of the detector such that the incident happens U/s, in the middle and D/s between 2 detectors and trained combining all three incident locations. The results are presented in Table 30 and Table 31.

Table 30: Performance of SVM models in the two-lane closed scenario.

Two-lane closed scenario		SVM using Detector Data			SVM using Re-ID & Flow		
#	Performance measures	Incident at downstream	Incident at middle	Incident at upstream	Incident at downstream	Incident at middle	Incident at upstream
1	**Alarm Type 1**	1	0	0	3	3	5
2	**Alarm Type 2**	30	31	31	28	28	26
3	**Alarm Type 3**	0	0	0	0	0	0
4	**Average Detection Time (secs)**	640.00	634.84	636.77	630.00	647.14	654.23
5	**Detection Rate**	96.77%	100.00%	100.00%	90.32%	90.32%	83.87%
6	**False Alarm Rate**	3.23%	0.00%	0.00%	9.68%	9.68%	16.13%
Two-lane closed scenario		SVM using Detector Data			SVM using only Travel time		
#	Performance measures	Incident at downstream	Incident at middle	Incident at upstream	Incident at downstream	Incident at middle	Incident at upstream
1	**Alarm Type 1**	1	0	0	0	0	0
2	**Alarm Type 2**	30	31	31	31	31	31

Table 30: Continued

3	**Alarm Type 3**	0	0	0	0	0	0
4	**Average Detection Time (secs)**	640.00	634.84	636.77	660.00	657.10	660.00
5	**Detection Rate**	96.77%	100.00%	100.00%	100.00%	100.00%	100.00%
6	**False Alarm Rate**	3.23%	0.00%	0.00%	0.00%	0.00%	0.00%

Also in this scenario, the results are further categorized according to the position of incident. The performance measures are also color-coded such that red means bad, green means good, and yellow (or orange) means comparatively satisfactory. There are three different comparisons made within this scenario using three different SVM models which were trained separately according to a data type as indicated on the table.

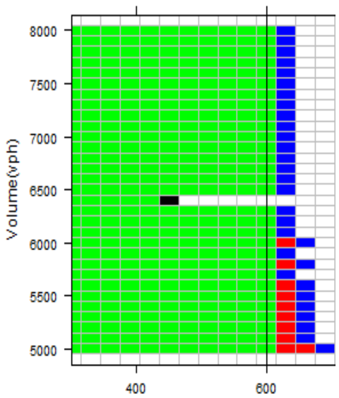
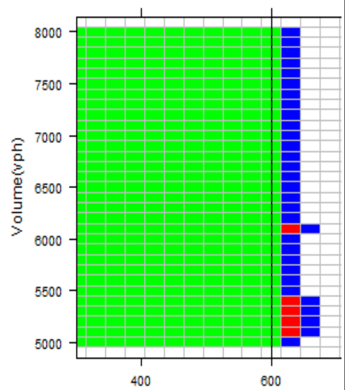
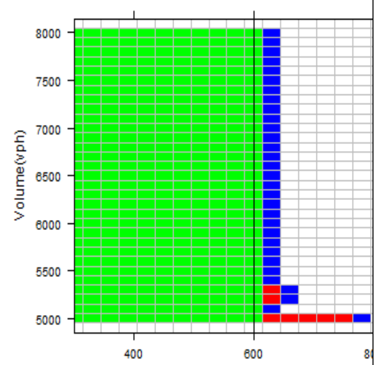
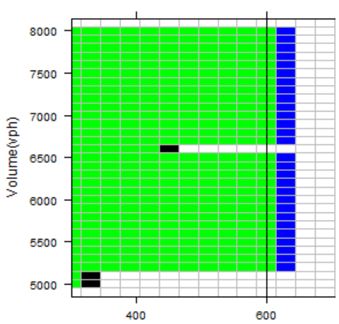
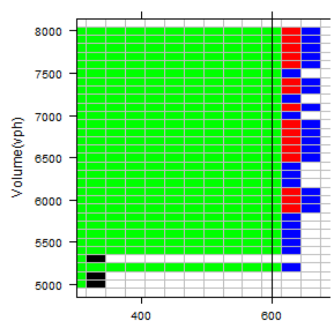
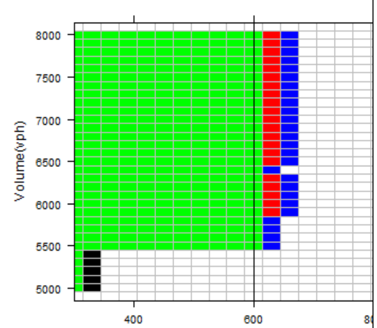
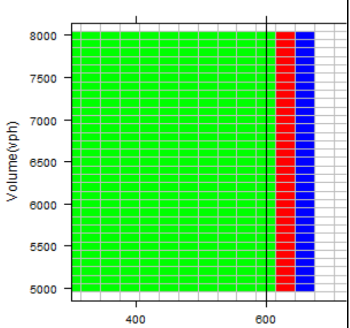
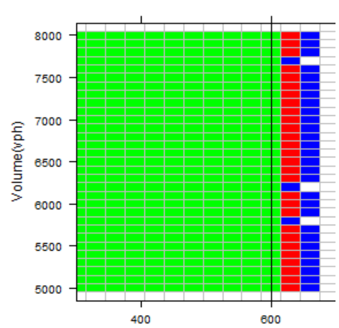
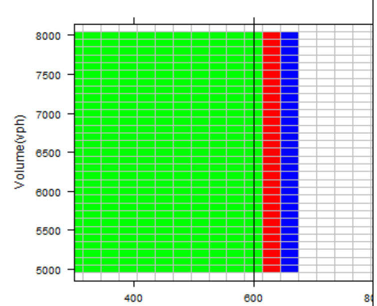
The first comparison is between the SVM using detector data and Vehicle Re-Identification data. In this comparison, it is seen that SVM using detector data performed better than SVM using travel time and flow.Even though the average time to detect is better, the detection rate is always better when detector data is used rather than travel time and flow in all three incident cases.

The SVM model with flow and travel time performed worse than even the SVM trained only using travel time in this scenario. It is because of the weights it gives to the flow; the SVM gives equal weights to travel time as well as flow, and looking at the data on error samples, it was found that the flow difference between 2 observations was large perhaps due to platoon movement of vehicles which in turn classified this situation as an incident. Probably due to this fact and maybe that enough of this type of sample were not trained so that the weight is shifted towards travel timemore than the flow difference. Subsequently, there are more misclassification errors in SVM using travel time and flow than in SVM using only travel time.

The last comparison is between the better SVM model, which is SVM using detector data, and SVMusing Re-identification data; however,this one only uses travel time. Thedetection results for both of the models are similar, and the only difference is for an incident occurring downstream, where the model using travel time detects the incident at a better rate. Even though the detection rate is better, the SVM using detector data predicts the incident around 20 secs faster than the SVM using travel time.

In Table 31, the figures show at what time the incident was detected in blue and black, where the first dark black vertical line at 600 seconds is the point when the incident starts.The green box, red box, blue box, and black box indicate No detection (True), i.e. there is no incident and no detection, No detection (False), i.e. there is an incident but no detection, True detection, and False detection, respectively.

Table 31: Table of Figures showing the performance of SVM models in two-lane closed scenario.

Data and features used	Legends: No Detection (True) — green; True Detection — blue No Detection (False) — red; False or Late Detection — black Incident At Downstream	Incident at Middle	Incident at Upstream
SVM using Detector data (trained using only detectors 5,6,7,8,9,10,11) Flow at U/s and D/s for time t, t-1 and t-2 Occupancy at U/s and D/s for time t, t-1 and t-2 Speed at U/s and D/s for time t, t-1 and t-2			
SVM using REID and number of vehicles. Travel Time Number of vehicles passing through U/s and D/s			
SVM using REID Only Travel Time			

It is to be noted that SVM using Re-ID runs on Re-identification data, so the percentage of vehicles that are correctly re-identified needs to be assessed. The figure shows the performance of the algorithm in a different percentage of appropriately re-identified vehicles.The primary vertical axis at left indicates the average time to detect the incident, whereas the second axis represents the detection rate. It

can be asserted that as the percentage of vehicles re-identified decreases detection rates decrease whereas the average time to detect increases. The detection rate is decreasing as the percentage of vehicles re-identified decreases, so the incident is more correctly detected at a higher percent of re-identified vehicles and fails to detect correctly below 30%. Most of the literature asserts that 90% of vehicles can be correctly identified, and the result shows that the performance of the algorithm at 90% of vehicles re-identified has no significant change in detection rate and a very slight change in average time to detect compared to 100% re-identified vehicles. Average time to detect in the middle and upstream scenario at 20% is blank because when an incident is not detected it gives infinity as the detecting time.

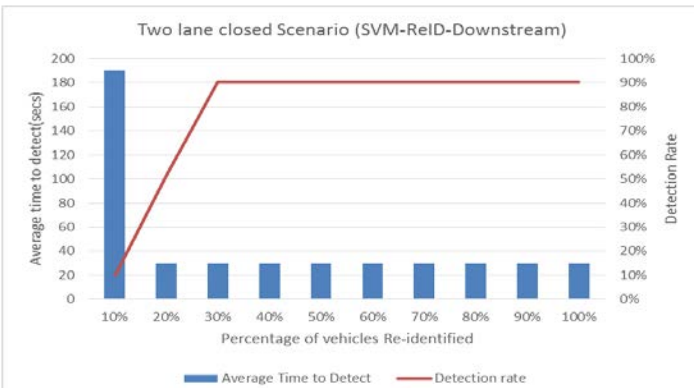


Figure52: Average time to detect and detection rate according to the percentage of vehicles re-identified incident at D/s.

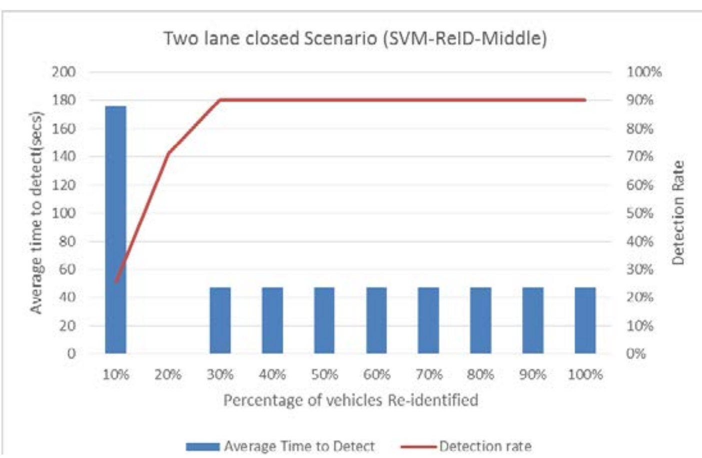


Figure53: Average time to detect and detection rate according to the percentage of vehicles re-identified incident at the middle.

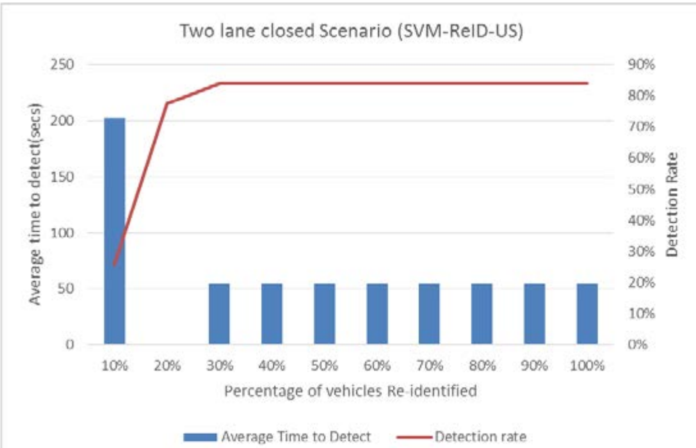


Figure54:Average time to detect and detection rate according to the percentage of vehicles re-identified incident at U/s

5.5.3 Two lanes closed with work-zone bottleneck scenario:

Like the two-lane scenario, the SVM models in this scenario were trained on one random seed and tested on a different random seed, one using detector data and the other using Re-Id (especially travel time and number of vehicles). Instead of training and testing on one particular location of the incident, the location is changed by changing the position of the detector such that the incident happens U/s, in the middle and D/s between 2 detectors, and the model is trained combining all three incident locations. The results are presented inTable 32inTable 33.

Table 32: Performance of SVM models in two lanes closed with a bottleneck scenario.

Two lane closed with work zone bottleneck		SVM using only Travel time			SVM using Re-ID & Flow		
#	Performance measures	Incident at downstream	Incident at middle	Incident at upstream	Incident at downstream	Incident at middle	Incident at upstream
1	**Alarm Type 1**	20	9	7	3	3	4
2	**Alarm Type 2**	11	22	18	28	28	27
3	**Alarm Type 3**	0	0	6	0	0	0
4	**Average Detection Time (secs)**	653.33	658.64	856.25	630.00	656.79	645.56
5	**Detection Rate**	35.48%	70.97%	58.06%	90.32%	90.32%	87.10%
6	**False Alarm Rate**	64.52%	29.03%	41.94%	9.68%	9.68%	12.90%
Two lane closed with work zone bottleneck		SVM using Detector Data			SVM using Re-ID & Flow		
#	Performance measures	Incident at downstream	Incident at middle	Incident at upstream	Incident at downstream	Incident at middle	Incident at upstream
1	**Alarm Type 1**	1	0	0	3	3	4
2	**Alarm Type 2**	30	31	31	28	28	27

Table 32: Continued

3	**Alarm Type 3**	0	0	0	0	0	0
4	**Average Detection Time (secs)**	635.17	649.35	630.00	630.00	656.79	645.56
5	**Detection Rate**	96.77%	100.00%	100.00%	90.32%	90.32%	87.10%
6	**False Alarm Rate**	3.23%	0.00%	0.00%	9.68%	9.68%	12.90%

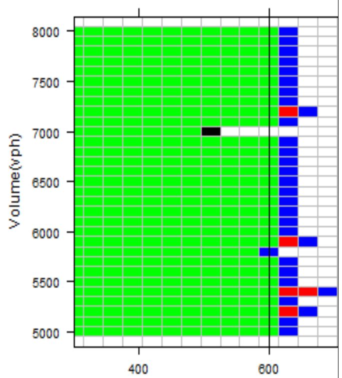
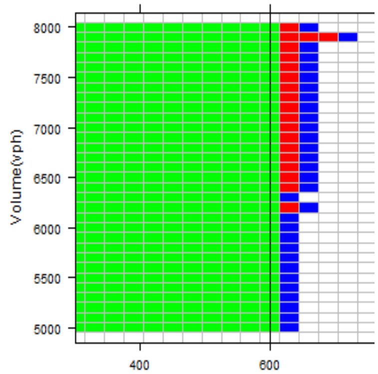
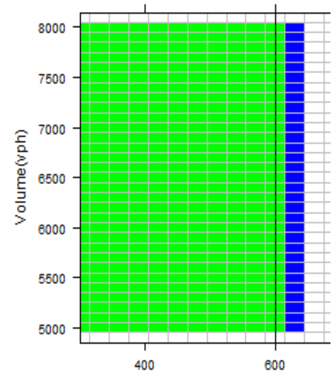
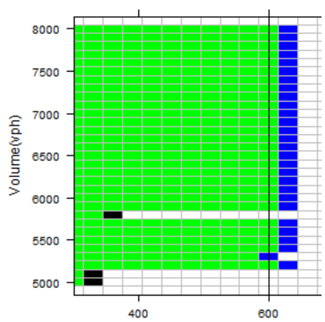
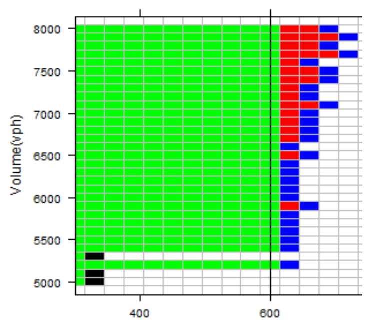
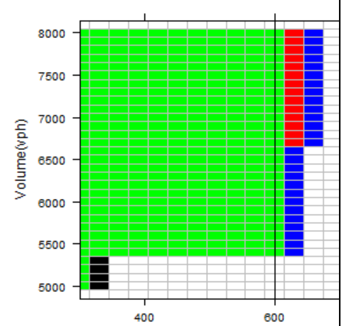
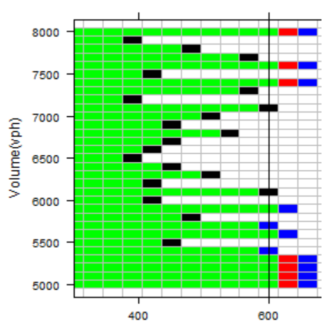
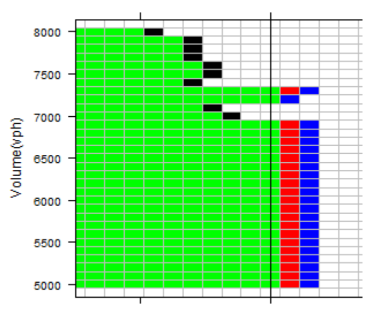
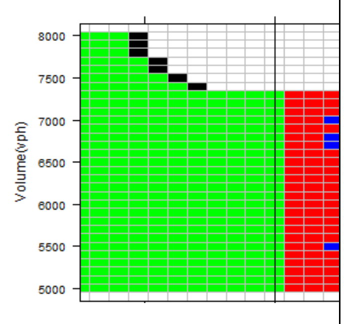
The performance measures are color-coded such that red means bad, green means good, and yellow (or orange) means comparatively satisfactory. There are three different comparisons made within this scenario using three different SVM models which were trained separately according to a data type as indicated in Table 32.

The road is in the congested condition in this scenario. The SVM model with flow and travel time performed better than the SVM trained only using travel time in this scenario. It is because of the weights it gives to the flow; the SVM gives equal weights to travel time as well as flow. Therefore, incorporating flow as well as travel time in the SVM model improved the performance of the model. However, looking at the data on error samples especially in low volumes for SVM using Re-id and flow, as seen in Table 33, it was found that the flow difference between 2 observations was large perhaps due to platoon movement of vehicles which in turn classified this situation as an incident in Incident using flow and travel time. Subsequently, there are more misclassification errors in SVM using travel time and flow than SVM using detector data.

The last comparison is between the SVM using detector data and Vehicle Re-Identification data. In this comparison, it is seen that the SVM using detector data performed better than the SVM using travel time and flow. Even though the average time to detect is better, the detection rate is always better when detector data is used rather than travel time and flow in all three incident cases. Therefore, in this scenario, the SVM using detector data outperformed both the SVM models using re-identification data whether it be only in travel time or in both travel time and flow.

For easy visualization of the tested sample and results,figures are presented in Table 33; the figures show at what time the incident was detected in blue and black, where the first dark black vertical line at 600 seconds is the point when the incident starts.The green box, red box, blue box, and black box indicate No detection (True), i.e. there is no incident and no detection, No detection (False), i.e. there is an incident but no detection, True detection, and False detection, respectively.

Table 33: Table of Figures showing performance of SVM models in two lanes closed with bottleneck scenario.

Data and features used	Legends: No Detection (True) — True Detection — No Detection (False) — False or Late Detection		
	Incident At Downstream	Incident at Middle	Incident at Upstream
SVM using Detector data (trained using only detectors 5,6,7,8,9,10,11) Flow at U/s and D/s for time t, t-1 and t-2 Occupancy at U/s and D/s for time t, t-1 and t-2 Speed at U/s and D/s for time t, t-1 and t-2			
SVM using REID and number of vehicles. Travel Time Number of vehicles passing through U/s and D/s			
SVM using REID Only Travel Time			

Again,SVM using Re-ID runs on Re-identification data, so the percentage of vehicles that are correctly re-identified needs to be assessed. The figure shows the performance of the algorithm in the different percentages of appropriately re-identified vehicles.The primary vertical axis at left indicates the

average time to detect the incident, whereas the second axis represents the detection rate. Most of the literature asserts that 90% of the vehicles can be correctly identified, and the result shows that the performance of the algorithm at 90% of vehicles re-identified has no significant change in detection rate and a very slight change in average time to detect compared to 100% re-identified vehicles. The blank in time to detect in the middle case at 20% and 30% vehicles re-identified is because when an incident is not detected it gives infinity as the detecting time. This is impossible to show in the graph and is not considered a number.

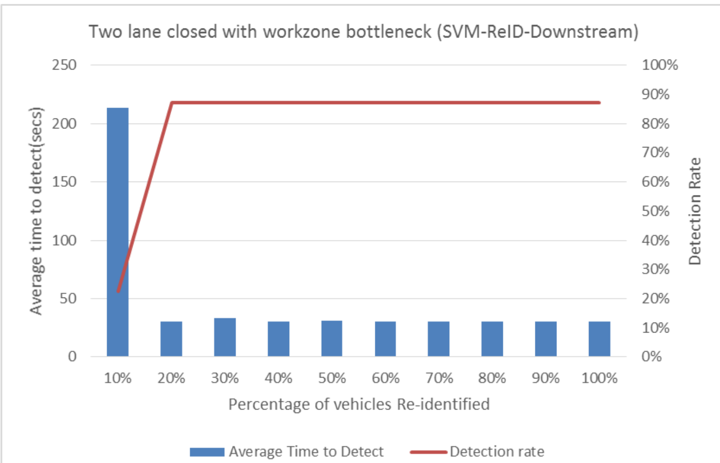


Figure55: Average time to detect and detection rate according to percentage of vehicles re-identified incident at D/s.

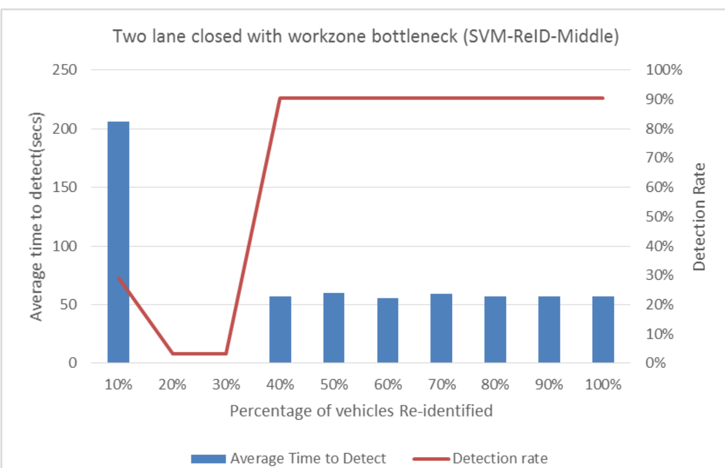


Figure56: Average time to detect and detection rate according to percentage of vehicles re-identified incident at middle.

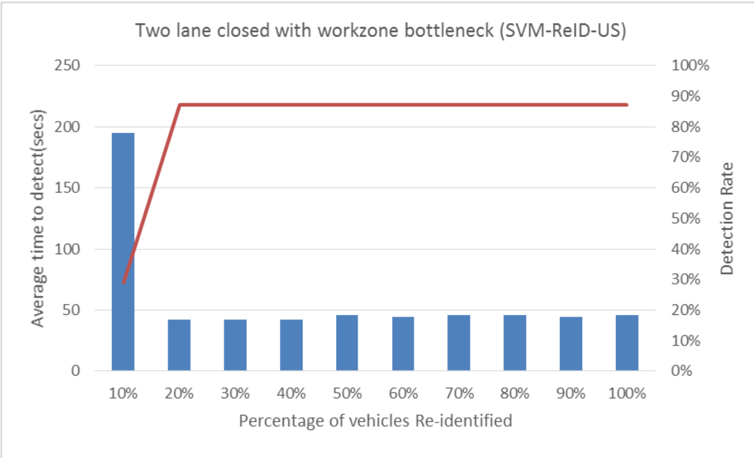


Figure57: Average time to detect and detection rate according to the percentage of vehicles re-identified incident at U/s.

CHAPTER 6

CONCLUSION

In this study of incident detection on a freeway, this thesis makes use of simulation data of different simulated incidents created in simulation software “PTVVISSIM”. Different scenarios were developed in VISSIM, blocking several lanes at different random traffic seeds and also with different traffic volume. With the purpose of collecting the data, data collection points, which can be double loop detectors in actuality, are placed along the road every quarter mile. Traffic parameters like occupancy, speed, flow, and the number of vehicles passing at the loop detector are collected to assess the traffic condition in between the sensors or detectors. In other words, traffic conditions are checked for each 0.25-mile segment along the road. Moreover, the study is focused on reviewing the existing algorithm’s effectiveness at incident detection and on developing, analyzing, implementing and experimenting with the algorithm for detecting incidents on a hypothetical road created in VISSIM. After the simulation, the required data is collected with COM interface using VBA (visual basic application) in Microsoft Excel. This data is later processed and analyzed using incident detection algorithms in R-studio, which is capable of handling big data such as trajectory data generated from the simulation software.

In conclusion, a simulation of an incident on a freeway is created from which data was extracted to apply, evaluate and study the incident detecting algorithms. Then calibrating and tuning the algorithms thresholds, tuning parameters and also the features that are used while training the algorithms was carried out so that optimal performance is obtained. After developing some new algorithms based on knowledge gained from analysis of existing algorithms to detect the incident, these new algorithms were applied and the performances of these algorithms were also evaluated. The factors that can affect the algorithms such as bottlenecks (or congested road conditions), the number of lanes blocked during an incident and the position of an incident relative to detector positions were also evaluated and studied by incorporating these conditions into the simulation itself.

After data collection and modeling, the algorithms were tested in different scenarios with a new performance measure that measured the overall performance of the algorithms in the detection of the incident. The overall performance of the algorithms was tested in different road conditions. The result shows that the overall performance of SVM is the best among the algorithms that make use of detector data. Algorithm #B, developed in this study, also performed fairly well with a detection rate of 100% which is as good as the SVM model in uncongested road conditions but was not useful in congested road conditions as it produced some false alarms. California #7 was good in two-lane closed scenarios, but it

performed very poorly in theone lane closed scenario.Thus, the best algorithms developed in this study were further investigated to see the effects of a change in incident location on these algorithms. SVM using only Re-identification performed better only under free-flow conditions. Even though incorporating flow and travel time in SVM performed better in congested conditions, it still fails to get 100 percent resultsincongested conditions whereas the SVM with detector data is better at congested traffic conditions.

As this study was completely based on simulation data, this study can be performed on real-world field data in order to get results of the algorithm's performance measures and to verify its practicality. apart from this, the incidents are not calibrated according to the incident from field data like speed and travel time of the vehicles involved in different incident scenarios. Further investigation and research on how to improve and give more weight to significant features used in SVM for predicting an incident without increasing the data size is needed. This would eliminate misclassification errors even on small sample sizes.

REFERENCES

1. Schrank, D., B. Eisele, and T. Lomax, *TTI's 2012 urban mobility report.* Texas A&M Transportation Institute. The Texas A&M University System, 2012. **4**.
2. Lindley, J.A., *Urban freeway congestion: quantification of the problem and effectiveness of potential solutions.* ITE Journal, 1987. **57**(1): p. 27-32.
3. Giuliano, G., *Incident characteristics, frequency, and duration on a high volume urban freeway.* Transportation Research Part A: General, 1989. **23**(5): p. 387-396.
4. Sullivan, E.C., *New model for predicting freeway incidents and incident delays.* Journal of Transportation Engineering, 1997. **123**(4): p. 267-275.
5. Kolenko, S. and J. Albergo, *Study of disabled vehicles and accidents on the Congress Expressway*. Chicago Surveillance Project, 1962.
6. Hallenbeck, M.E., et al., *Measurement of Recurring versus Non-Recurring Congestion: Technical Report*. 2003: Washington State Department of Transportation.
7. Sethi, V., *Duration and travel time impacts of incidents*. 1994: Transportation Center, Northwestern University.
8. Cassidy, M.J. and R.L. Bertini, *Some traffic features at freeway bottlenecks.* Transportation Research Part B: Methodological, 1999. **33**(1): p. 25-42.
9. Ringert, J. and I. Urbanik, *Study of freeway bottlenecks in Texas.* Transportation Research Record, 1993(1398).
10. Payne, H.J., E. Helfenbein, and H. Knobel, *Development and testing of incident detection algorithms, volume 2: Research methodology and detailed results*. 1976.
11. Ahmed, S. and A.R. Cook, *Application of time-series analysis techniques to freeway incident detection.* Transportation Research Record, 1982. **841**: p. 19-21.
12. Dudek, C.L., C.J. Messer, and N.B. Nuckles, *Incident detection on urban freeways.* Transportation Research Record, 1974. **495**: p. 12-24.
13. Tsai, J. and E. Case, *Development of freeway incident-detection algorithms by using pattern-recognition techniques.* Transportation Research Record, 1979. **722**: p. 113-116.
14. Balke, K., C.L. Dudek, and C.E. Mountain, *Using probe-measured travel times to detect major freeway incidents in Houston, Texas.* Transportation Research Record, 1996. **1554**(1): p. 213-220.
15. Cook, A.R. and D.E. Cleveland, *Detection of freeway capacity-reducing incidents by traffic-stream measurements.* Transportation Research Record, 1974. **495**: p. 1-11.
16. Chassiakos, A.P. and Y.J. Stephanedes, *Smoothing algorithms for incident detection.* Transportation Research Record, 1993. **1394**: p. 8-16.
17. Martin, P.T., et al., *Incident detection algorithm evaluation.* Prepared for Utah Department of Transportation, 2001.
18. Cheu, R.L. and S.G. Ritchie, *Automated detection of lane-blocking freeway incidents using artificial neural networks.* Transportation Research Part C: Emerging Technologies, 1995. **3**(6): p. 371-388.
19. Lee, D.-H., et al., *Taxi dispatch system based on current demands and real-time traffic conditions.* Transportation Research Record: Journal of the Transportation Research Board, 2004(1882): p. 193-200.
20. Dia, H. and G. Rose, *Development and evaluation of neural network freeway incident detection models using field data.* Transportation Research Part C: Emerging Technologies, 1997. **5**(5): p. 313-331.
21. 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, 2001. **12**(5): p. 1173-1187.
22. Wen, H., et al. *A new algorithm of incident detection on freeways*. in *Vehicle Electronics Conference, 2001. IVEC 2001. Proceedings of the IEEE International*. 2001. IEEE.
23. Al-Deek, H.M. and S. Ishak, *Implementation of incident detection algorithms*. 1999.

24. Yuan, F. and R.L. Cheu, *Incident detection using support vector machines.* Transportation Research Part C: Emerging Technologies, 2003. **11**(3-4): p. 309-328.
25. Kühne, R., et al. *Section-related measures of traffic system performance*. in *76th annual TRB meeting, Transportation Research Board*. 1997.
26. Cheng, H.H., et al., *A real-time laser-based detection system for measurement of delineations of moving vehicles.* IEEE/ASME Transactions on mechatronics, 2001. **6**(2): p. 170-187.
27. MacCarley, A.C., *Video-Based Vehicle Signature Analysis and Tracking Phase 1: Verification of Concept Preliminary Testing.* 1998.
28. Coifman, B., et al., *A real-time computer vision system for vehicle tracking and traffic surveillance.* Transportation Research Part C: Emerging Technologies, 1998. **6**(4): p. 271-288.
29. McCasland, W., *Monitoring freeway traffic conditions with automatic vehicle identification systems.* ITE Journal, 1994. **64**(3).
30. Coifman, B. and M. Cassidy, *Vehicle Reidentification and Travel Time Measurement in Real-Time on Freeways Using the Existing Loop Detector Infrastructure, Transportation Research Record 1643, Transportation Research Board.* 1998.
31. Basar, G., M. Cetin, and A.P. Nichols, *Comparison of vehicle re-identification models for trucks based on axle spacing measurements.* Journal of Intelligent Transportation Systems, 2018. **22**(6): p. 517-529.

VITA

Norfolk, VA 23508| (571) 123-4567
www.linkedin.com/in/biraj-adhikari-ba0b3a173

EDUCATION
Old Dominion University, Norfolk, VA May 2019
Master of Science in Engineering
Major**:** Transportation Engineering

Institute Of Engineering, Tribhuwan University, Pulchowk, Nepal December 2014
Bachelor of Science in Civil Engineering
Major: Civil Engineering
Related Coursework: Transportation, Structural, Geo-Technical, Water Resource, and Environmental.

RELATED EXPERIENCE
Old Dominion University, Norfolk, VA
Research Assistant (January 2018—Present)
Evaluation of Strategies to Reduce Truck Turnarounds at the Hampton Roads Bridge-Tunnel (HRBT)
Assisted Professor, Dr. Cetin, in developing strategies and evaluating these strategies to reduce the number of over-height trucks (turnarounds) entering the HRBT tunnel.
Report preparation, data collection and coding the related codes in R-programming to process the raw data from LIDAR and surveillance camera videos.
Developing a simulation model for the road section, analyzing and presenting the results, etc.

MMM Group (Canada) In Association with ITECO Nepal (P) Ltd. -TMS- Materials Test, Kathmandu, Nepal.
Civil Engineer (August 2016—July 2017)
Preparing of survey, design, rate analysis and Preparation of Engineer Estimate as per engineering technical specifications for each work item. Of Roadsand 87 nos. of RCC Bridges and Number of Box Culvertsunder SASEC Roads Improvement Project.
Involved in the Feasibility study, detailed survey, detailed design, preparation of tender documents, IEE study, and social study for land acquisition, and composition of road projects are Total Feasibility Study 527 km & Detail Design of 525KM 2 Lanes Road 320KM and 4 Lanes Road.
SKILLS

Computer Skills: Proficient in Microsoft Office, Word, Excel, and PowerPoint, Proficient in Auto CAD 3D/2D, Adobe Acrobat, Sketch Up 3D, PTV-VISSIM, Synchro.
Programming languages: SAS, C+, R-programming, Python, Excel VBA, Java-script.