A Variable Threshold Incident Detection Model

Hussein DIA
Lecturer
Institute of Transport Studies
Department of Civil Engineering
Monash University
Clayton Victoria 3168
Australia
Fax: +61 3 9905 4944
Email: hussein.dia@eng.monash.edu.au

Geoff ROSE
Senior Lecturer
Institute of Transport Studies
Department of Civil Engineering
Monash University
Clayton Victoria 3168
Australia
Fax: +61 3 9905 4944
Email: geoff.rose@eng.monash.edu.au

abstract: This paper explores an approach for improving the performance of an Artificial Neural Network (ANN) freeway incident detection model. The performance of the model was evaluated using real-world incident and non-incident data. Initial results revealed that a high number of false alarms were caused by congested conditions. To improve operational performance, a model that applies a variable decision threshold (depending on the time of day) was formulated. The results reported here indicate that this model achieves improved false alarm results without sacrificing the detection rate performance. These results also suggest that additional improvements in performance may be obtained by implementing a dynamic decision threshold.

1. INTRODUCTION

Freeway incidents are increasingly contributing to urban congestion, pollution and deteriorated safety conditions. The high cost of congestion caused by incidents such as accidents, disabled vehicles and other events that result in a capacity reduction of the facility, has prompted a growing world-wide interest in developing efficient and effective automated incident detection methods. Such incidents account for a large percentage of the total delays and costs due to congestion on major freeways around the world. The adverse effects of incidents are also expected to increase as freeway facilities in major cities around the world become more congested.

The benefits to be derived from early incident detection and prompt response in terms of providing real-time traveller information and timely dispatch of emergency services can drastically reduce traffic delays, air pollution and improve safety and real-time traffic control. Intelligent Transportation Systems (ITS) technologies are structured to address these needs through Advanced Traffic Management Systems (ATMS) and Advanced Traveller Information Systems (ATIS). For these systems to be effective, it is necessary to develop reliable procedures for detecting incidents in a short time. Few of the algorithms developed or proposed over the last two decades have been implemented in practice due to varying operational levels in terms of detection rate, false alarm rate and mean time-to-detect. Therefore, the need is pressing for more effective real-time incident detection algorithms that maximise detection rate while only generating an operationally acceptable level of false alarms.

This paper first discusses the development of a neural network freeway incident detection model. In contrast to published artificial neural network (ANN) incident detection algorithms which relied on simulated or limited field data for model development and testing, the models described in this paper were trained and tested on a real-world data set of 100 incidents. The incident detection performance of the ANN model is first reported based on a validation-test data set of 40 incidents that were independent of the data used for training. The false alarm rates (FAR) of the model are then evaluated based on non-incident data that were collected from a freeway section which was video-taped for a period of 33 days. The initial FAR results revealed that a high number of false alarms were due to congestion. To improve operational performance, a model that applies a variable decision threshold (depending on the time of day) was formulated. The results for the variable decision threshold model are then reported. These results clearly indicate that it is possible to achieve improved false alarm results without sacrificing the detection rate performance. They also suggest that additional improvements in performance may be obtained by implementing a dynamic decision threshold.

2. AUTOMATIC INCIDENT DETECTION

Automatic incident detection systems involve two main components: a traffic detection system and an incident detection algorithm. The traffic detection system provides the traffic information necessary for detecting an incident while the incident detection algorithm interprets that information and ascertains the presence or absence of incidents or non-recurring congestion. Inductive loop detectors embedded in the freeway pavement are typically used to obtain traffic data, primarily on occupancy (percent of time a detector is occupied) and volume. Dual loop installations also provide speed data. More recently, image-based video detectors have also been used for the same purpose in the AUTOSCOPE (Michalopolous *et al.*, 1993) and IMPACTS (Hoose, 1992) incident detection systems.

2.1 Incident Detection Algorithms

A number of AID algorithms have been developed from a variety of theoretical foundations. Their structure varies in the degree of sophistication, complexity, data requirements and the type of surveillance technology used for data collection. Some of the most widely used algorithms include the comparative or pattern comparison algorithms, eg. the California-type algorithms (Levin and Krause, 1979); time series algorithms, eg. the Auto-Regressive Integrated Moving Average (ARIMA) algorithm (Ahmed and Cook, 1982); and the McMaster algorithm (Persaud and Hall, 1989).

2.2 Artificial Neural Networks

Few of the previously developed algorithms have been implemented in practice due to various limitations and varying operational levels. Therefore, the need is pressing for more effective real-time incident detection algorithms. Furthermore, desired new-generation algorithms should also lend themselves to implementation on new platforms such as parallel computers and must have the required flexibility for the smooth integration with emerging ITS technologies.

One promising approach to address these objectives involves the application of artificial neural networks. Ritchie and Cheu (1993) demonstrated successfully the feasibility of using ANNs for freeway incident detection. They tested a multi-layer feed-forward (MLF) ANN on a freeway section using simulated traffic detector data. The results confirmed their hypothesis that spatial and temporal traffic patterns could be recognised and classified by ANNs. However, their results were limited in the sense that they trained and tested the ANN models on simulated data. Only a small set of field data with several lane-blocking incidents were used to evaluate the trained ANN models (Cheu and Ritchie, 1995). More recently, Stephanedes and Liu (1995) developed an ANN model that was based on real-world incident data collected from a freeway in the Twin Cities Metropolitan area. The results of their work, however, were also limited in that the model was trained and tested on the same data set. In addition, the models developed in both studies used only volume and occupancy data and did not address operational issues such as the impact of detector malfunction and quality of input data on model performance. The work reported here is part of a research program that addresses these unresolved issues.

2.3 Performance Measures for Incident Detection Algorithms

The performance of an incident detection algorithm is measured by three criteria: detection rate (DR), false alarm rate (FAR) and time-to-detect (TTD). The DR is defined as the number of incidents detected by the algorithm divided by the total number of incidents known to have occurred during the recorded time. The FAR is defined as the number of incident-free intervals which gave false alarms divided by the total number of incident free intervals. Finally, the TTD is the difference between the time of occurrence and the time at which the incident was declared or an alarm was raised by the algorithm. When an algorithm is being evaluated, however, it is customary to seek the mean time-to-detect (MTTD) a set of (n) incidents. The occurrence time of an incident is usually not known precisely and an estimate has to be deduced from loop detector data or records kept by police, traffic control centres or towing companies.

The above definitions clearly show that both the DR and FAR measure the effectiveness of the algorithm while the MTTD reflects its efficiency. The DR and FAR are, however, positively correlated. In order to detect more incidents, the algorithm thresholds are relaxed which causes some incident-free intervals to be interpreted as alarms. Since many false alarms are caused by random fluctuations in traffic flow, a persistence test is usually performed by testing warnings in a few consecutive intervals before declaring an alarm. This method, in conjunction with increased duration of the persistence test, has been shown to reduce the FAR. However, this was also found to reduce the efficiency of the algorithm since it increased the MTTD considerably. Clearly the three performance measures are all inter-related. The relative importance of the measures, however, is typically DR, FAR and MTTD.

3. DATA FOR THE DEVELOPMENT OF THE ANN MODELS

In order to train a neural network to perform incident detection, the network must be presented with examples of input detector data (speed, flow and occupancy) and output states for both incident and incident-free conditions. Therefore, the data required should at least have a description of the state of traffic along the freeway in addition to detector data comprising traffic flow measurements at regular time intervals for each detector station.

3.1 Data Collection

The data required for model development were assembled from two data sources held at the VicRoads Traffic Control and Communications Centre (TCCC) in Melbourne, Australia. The first data source comprised information logged by the operators at the TCCC regarding the incidents that occurred on the freeways. This information is received by the operators from a variety of sources including motorists using either private mobile phones or the emergency phones located near the freeway. The second data source comprised the loop-detector data which consisted of speed, flow and occupancy measurements in 20-second cycles. Each of the collected incidents was then examined individually. The log entry for the incident was compared against the detector data to find out if the incident could be detected from the loop data. This was accomplished by monitoring the immediate upstream and downstream stations from the incident using a graphical computer program that was developed to plot the detector data. A total of sixty incidents were clearly detectable from the detector data and could be confirmed from the operator's log. These incidents were used for model development in this study.

3.2 Assignment of Desired Output States

As was mentioned previously, training a neural network to perform incident detection involves presenting the network with examples of input detector data (speed, flow and occupancy) for the upstream and/or downstream stations in addition to providing the desired output states or correct responses for each input vector. Two output states are used to describe the traffic conditions within the section under consideration: State 1, $\{0\}$, representing incident-free conditions and State 2, $\{1\}$, representing incident conditions.

3.3 Creation of Training and Training-Test Data Sets

The next activity involved compiling the training and training-test data sets that will be used for training the ANN incident detection models. The training data set will be used for determining the network parameters while the training-test data set will be used to prevent the network from learning the idiosyncrasies in the training data set and thereby enables the model to generalise better (Masters, 1993). Therefore, the two data sets are essentially used for training the ANN model and are thus referred to as the 'training data sets'. The ANN models should be trained on a set of incidents that are representative of the population to which the network will ultimately be applied. Training an ANN model with a wide range of incidents that include different patterns under a variety of flow conditions and traffic periods helps improve the robustness of the model in detecting incidents under varying conditions. Therefore, the data was stratified according to incident severity (in terms of the number of lanes blocked due to the incident), prevailing flow conditions prior to the occurrence of incidents (heavy, moderate and light), traffic period of the day (peak or off-peak) and incident duration. Two data sets, each comprising 30 incidents, were then selected randomly from the sixty incidents to form the master training and training-test data sets (Dia and Rose, 1995).

4. DATA FOR THE VALIDATION OF ANN INCIDENT DETECTION MODELS

In addition to the training data sets required for model development, a third data set is also needed for validating the performance of the trained incident detection models. This data set should be independent of the data sets used for model training. The validation data set compiled in this study comprised 40 incidents that were detectable from the detector data. Of these, 25 occurred on the Tullamarine Freeway and 15 occurred on the South Eastern Freeway in Melbourne, Australia.

The 100 incidents collected for this study (60 for training and 40 for validation) are believed to be the largest set of 'real world' lane-blocking incidents available for the development of freeway incident detection models using field data. The 60 incidents in the training set comprised a total of 25,333 observations while the 40 incidents in the validation-test data set comprised a total of 14,149 observations. In total, the 100 incidents comprised 22,186 incident-free observations and 17,296 incident-conditions observations.

5. A FRAMEWORK FOR AUTOMATED INCIDENT DETECTION USING ANNS

The general framework for automated incident detection using artificial neural networks is shown in *Figure 1* (Rose and Dia, 1995). Consider the section of freeway shown in *Figure 1(a)* which is defined by upstream and downstream detector locations. A corresponding ANN incident detection model structure is shown in *Figure 1(b).* The detector station data form the input to the ANN. The output is a {0,1} variable indicating the absence or presence of an incident in the freeway section, respectively. In order to train the ANN (in a supervised mode) to perform incident detection, the network must be presented with input detector data and output states for both incident and incident-free conditions. Therefore, the input to the ANN model comprises real-time speed, flow and occupancy measurements in 20-second intervals from each of the upstream and downstream stations. The output of the ANN model is the traffic state within the section. Output State 1 {0} represents incident-free conditions and output State 2 {1} represents incident-conditions.

Cheu and Ritchie (1995) tested three ANN architectures suitable for incident detection and real-time classification problems. These included the multi-layer feed-forward (MLF) neural network, the self-organising feature map (SOFM) and the adaptive resonance theory (ART). The MLF, implemented with the back-propagation (BP) training algorithm, proved to be superior to the other architectures tested. The MLF was chosen for implementation in this study based on its demonstrated superior incident detection performance over the other ANN architectures. In particular, the standard three-layer feed-forward neural network, shown in *Figure 1(b)*, has been chosen for this study (Dia and Rose, 1995).

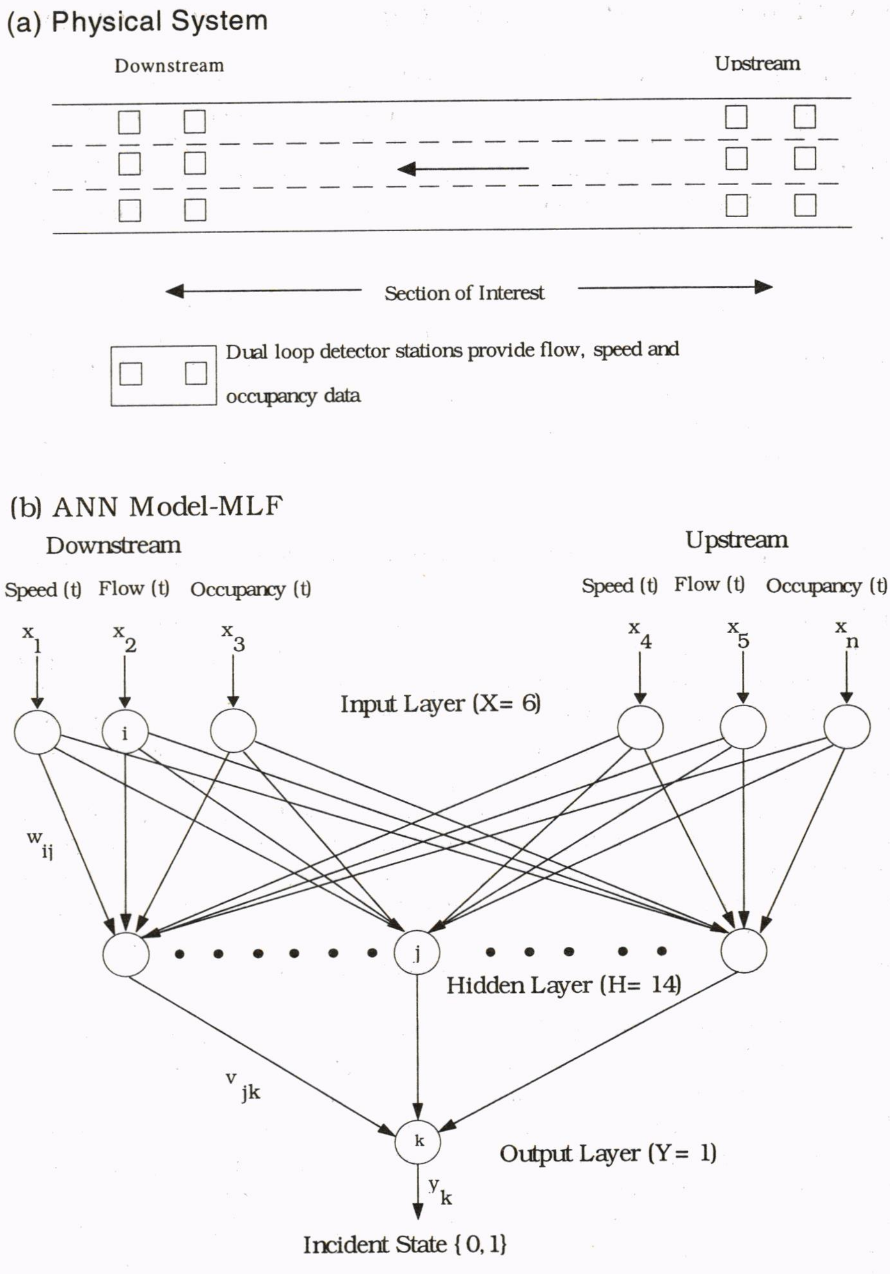

Figure 1- ANN modelling framework

5.1 Incident Detection Parameters

One of the main issues in incident detection modelling is the selection of an appropriate set of input features. The choice of traffic flow variables, detection logic and other related parameters is a function of the desired complexity of the model and the surveillance technology used. The issues related to the input features that were investigated in this study included the selection of the number of stations required to identify incidents within a section (upstream, downstream and both stations were investigated), the number of preceding time intervals needed for each decision regarding the presence or absence of incidents at any time interval *t* (intervals *t* to *t-4* were investigated), and the station input data (data provided on a lane-by-lane basis, from the fast lane and averaged across all lanes were investigated). The detailed results of this investigation are provided in Dia and Rose (1997). These results clearly indicate that models using dual stations perform better than models using a single station. The model's fault-tolerance under conditions of corrupt or missing data and the impact of loop detector failure/malfunction on the performance of the model are also provided in Dia and Rose (1997).

5.2 Model Evaluation Using Performance Envelopes

A graph that helps to show the relationship between DR and FAR can be obtained by evaluating the DRs and FARs for many possible decision thresholds (DTs). Typically, a decision threshold of half-activation (a value of 0.5) is chosen for making the decision regarding the presence or absence of incident conditions. If a vector of input data is presented to the ANN model which results in the output PE being activated to at least a value of 0.5 (half-activation), it is concluded that incident conditions are present for that time period. Otherwise, the data for that time period is classified as non-incident conditions. Therefore, the value selected for the decision threshold plays an important role in the classification of the input data and consequently in determining the incident detection performance of the model. The plot of DR against FAR (shown in *Figure 2*) is called the Performance Envelope Curve (PEC) of the network.

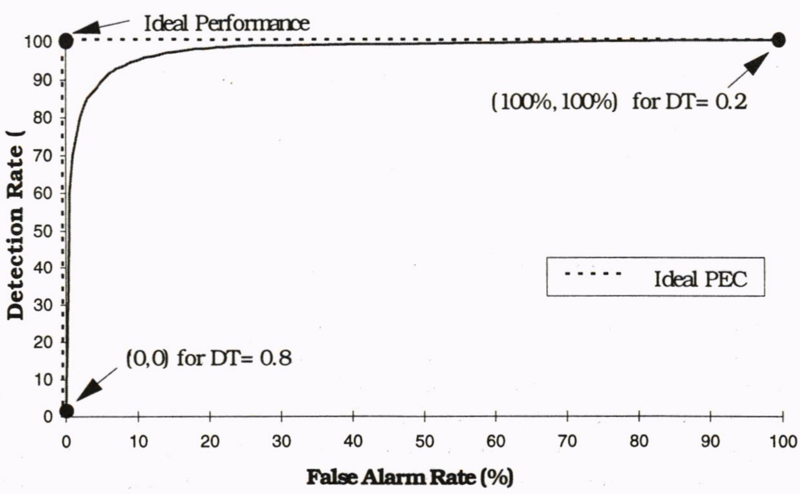

Figure 2- A performance envelope curve

In this study, the ANN model was trained to produce an output value between 0.2 and 0.8. The lower-left corner (0,0) of the PEC will always be one endpoint of this curve. It corresponds to a DT slightly larger than 0.8. At this DT, an observation is classified as an incident condition if it activates the output PE of the network to a value slightly greater than 0.8. Since this is not possible, all observations will be classified as non-incident conditions. Therefore, both the DR rate and FAR will be zero. The upper-right corner (100%,100%) will be the other endpoint of the PEC. This point corresponds to a DT slight smaller than 0.2. At this DT, an observation will be classified as a non-incident condition if it activates the output PE to a value less than 0.2. Again, since this is not possible, all observations will be classified as incident conditions. Therefore, both the DR and FAR will be 100%.

A network that is able to correctly classify all observations would have a PEC that is at a right angle (as shown by the dashed lines in *Figure 2*). At the ideal performance DT, the DR would be 100% and the FAR would be zero. For all lower thresholds, the DR would remain at 100%, while the FAR would increase to 100%. For greater thresholds, the FAR would remain at zero, while the DR would drop to zero. The quality of performance of the network is demonstrated by the degree to which the PEC pushes upward and to the left. This can be quantified by the area under the curve (PECA) and the slope of the curve (particularly for low values of FAR). The PECA for a perfect discriminator (Masters, 1993) will be 10,000 (ie. 100×100). This procedure is particularly useful since it helps with evaluating the model's performance based on the total picture of DRs, FARs and DTs by using a single index, ie. the area under the PEC.

6. RESULTS OF TRAINING

A total of 500 models consisting of different input parameters and hidden units were designed and trained in this study (Dia, 1996). In order to compare the performance of the different groups of models and determine which model types had better incident detection performance, statistical analysis techniques were used to investigate the trade off in performance between these groups of models. The results from the statistical analysis and refinement of a selected number of models (Dia, 1996) revealed that the architecture of the best performance model was similar to that shown previously in *Figure 1* (model MLF). This model uses dual stations, traffic data (speed, flow and occupancy) averaged across all lanes and only from the current time interval *t*. The optimal number of hidden PEs for this model was found to be 14.

7. PERFORMANCE EVALUATION UNDER INCIDENT CONDITIONS

The evaluation of the ANN model under incident conditions will be based on the application of model MLF to the validation-test data set, which comprised 40 incidents. This evaluation will give an indication of the prediction ability of the model in detecting incidents that the model had not previously seen. For the purpose of this study, an incident will only be declared as detected if the algorithm takes no longer than 5 minutes for its detection. Therefore, the failure to detect an incident could be either due to the inability of the algorithm to detect the incident at all or the inability of the algorithm to detect the incident within 5 minutes.

Investigation of several techniques with the potential to reduce the FAR (Dia, 1996) revealed that model MLF is best applied using a DT of (0.640) and a two-interval persistence test. The incident detection performance of this model, which will be referred to as model MLF1, is shown in *Table 1* and *Figure 3* below. It should be pointed out here that a FAR of 0.1% corresponds to about 4.3 false alarms per day per section (0.1% × 3 decisions every minute × 60 minutes/hour × 24 hours/day).

Table 1: Incident detection performance of model MLF based on the validation-test data set

Data Set	Model ID	Decision Threshold (DT)	Incident Detection Performance: Detection Number	Detection Rate (%)	False Alarm Rate (%)	Time to Detect (Second)
Validation -Test Set (40 incidents)		0.300	36/40	90.0	0.755	156
		0.400	36/40	90.0	0.442	170
		0.500	35/40	87.5	0.273	181
		0.620	33/40	82.5	0.091	199
	MLF1	0.640	33/40	82.5	0.065	203
		0.650	30/40	75.0	0.026	205
	MLF2	0.695	20/40	50.0	0.013	216
		0.700	18/40	45.0	0.013	216
		0.710	13/40	32.5	0.000	200

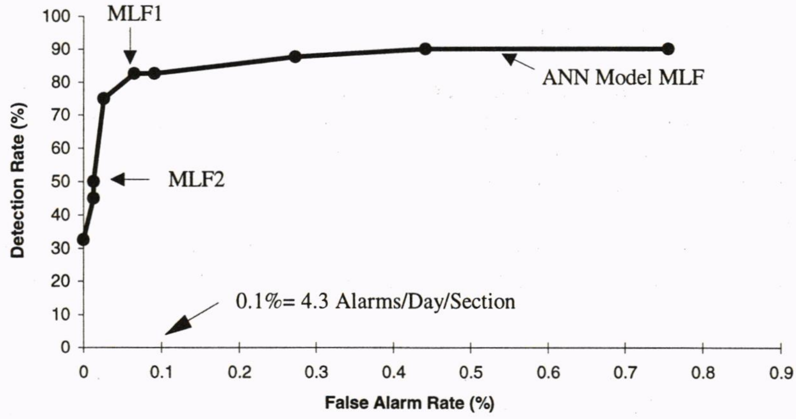

Figure 3- PEC for model MLF based on the validation-test set of 40 incidents

Table 1 and Figure 3 also show the performance measures for model MLF based on the application of a two-interval persistence test and a wider range of DTs. The tradeoff in performance using different DTs is best illustrated by comparing the DR and FAR of model MLF1 with model MLF2 which uses a higher DT of (0.695). The application of a higher DT clearly results in a reduction of both DR and FAR. Model MLF1 was selected for the evaluation of the incident detection performance based on the results shown in *Table 1* and Figure 3. The results for model MLF1 will also be used as an indicator of the expected incident detection performance of the model when it is implemented in the field. The results for model MFL1 based on the Tullamarine (Tulla) and South Eastern Freeway (SEF) data sets are shown in *Table 2.*

Table 2: Incident detection performance of model MLF1 based on the 40 incidents

Data Set	Number of Incidents	Incident Detection Performance			
		Detection Rate		False Alarm Rate	Mean Time-to-Detect
		No.	%	(%)	(Second)
Validation Test Set-Tulla	25	19/25	76.0	0.09	188
Validation Test Set-SEF	15	14/15	93.3	0.00	224
Total Set	40	33/40	82.5	0.065	203

These results indicate that the DR and FAR performance of model MLF1 on the validation-test data sets is generally consistent with its expected performance in *Table 1* (82.5% DR, 0.065% FAR and 203 second MTTD). The results also show that the reported FARs vary according to the incident data set under consideration. One possible explanation for this is related to the fact that not all incidents in the data sets had similar characteristics in relation to the 'noise' or general 'cleanliness' of the incident. The presence of a larger number of 'noisy' incidents in a certain data set would therefore affect the performance of the model (especially in terms FAR) on that data set.

The performance of the model was also examined by segmenting the results on the basis of incident severity (one, two or three lanes blocked), prevailing flow conditions (light, moderate or heavy) and period of day (peak or off-peak). The detailed results from this analyses are provided in Dia (1996). However, these results have shown that:

(a) The poorest performance was achieved for incidents which only involved a single lane blockage. The DR of the model in this case was 66.7% with a corresponding FAR of 0.0%.

(b) Even under light flow conditions the model was able to detect 75.0% of the incidents with a corresponding FAR of 0.0%. The false alarms generated by the algorithm were primarily associated with heavy flow conditions.

(c) The model's performance is consistent across the periods of the day with respect to DR (82.4% in the off-peak versus 83.3% in the peak) and MTTD (201 & 216 seconds in the off-peak and peak respectively). The FAR results, however, showed more variability and were ten times as high during the peak period than the off-peak period (0.25% versus 0.029%).

The performance of the ANN model according to the time taken by the algorithm to detect incidents is also of interest. The model detected 72.7% of incidents in less than 4 minutes, 39.4% in less than 3 minutes and 18.2% in less than 2 minutes. Inspection of the operators' log files revealed that 2 incidents were detected by the operators before their impact on traffic was confirmed from the detector data. It should be mentioned, however, that when operators are busy managing incidents, it is not uncommon for important details (including the time of detection of incidents) to be left out. Only 18.4% (7 incidents) of the remaining 38 incidents were detected by the operators within 3.0 minutes of their occurrence. The average time taken by the operators to detect the 38 incidents was 6.9 minutes after their estimated occurrence times. The results reported in this study for the MTTD of model MLF1 (*Table 1*) was 3.4 minutes (203 seconds). This suggests that the ANN model has the potential to provide a 50% improvement in efficiency compared to the average time taken by the operators to detect incidents.

8. PERFORMANCE EVALUATION UNDER NON-INCIDENT CONDITIONS

The FAR results reported previously for model MLF1 were based on the non-incident conditions immediately before/after the occurrence/clearance of incidents. The unstable traffic conditions during these periods may cause the algorithm to generate a higher number of false alarms than would be expected during normal traffic conditions. In addition, it is also desired that the model's FAR performance under daily traffic conditions (especially during peak-hour periods) is evaluated and reported. To investigate the off-line FAR performance of the model over an extended period of time, Section S4-S5 of the Tullamarine Freeway (*Figure 4*) was continuously video-taped (on 24-hour basis) for a period of 33 days (a total of 141,655 observations). The location of the video camera and the field of view captured through the camera is also shown in *Figure 4*.

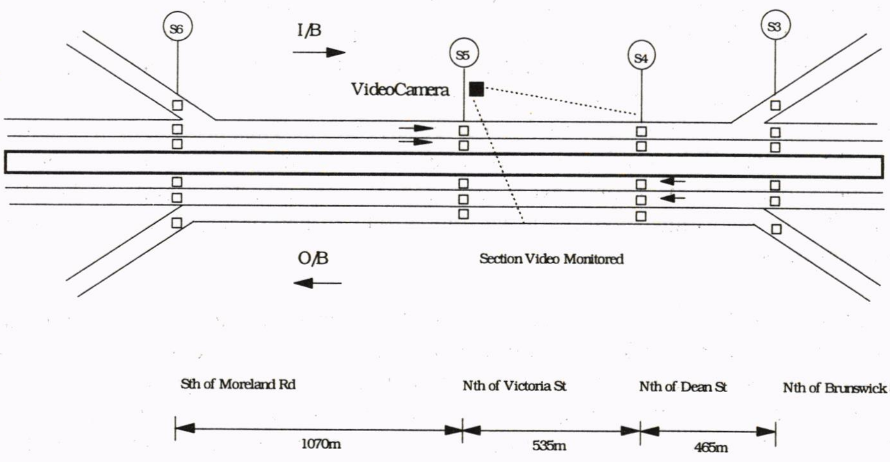

Figure 4- Schematic of section S4-S5 on the Tullamarine Freeway used for video monitoring and evaluation of the off-line FAR

A time lapse video recorder was used to monitor traffic conditions within the section. In addition to the video data, traffic data from the detector stations at S4 and S5 were also collected for the duration of the period. In this way, the model could be run on the detector data and when an alarm was raised, the video could be advanced to that same time period to see what traffic conditions had caused the alarm.

8.1 Analysis of Detector and Video Data

The results from an initial appraisal of the off-line FAR indicated that a large number of alarms were generated during peak-hour periods. Therefore, it was decided to divide each day into five periods, as shown in *Table 3*, such that the performance of model MLF is evaluated separately for each period.

Table 3: Classification of periods of the day for the off-line FAR evaluation

Period of Day	Time Frame	Total Number of 20-Second Observations for the 33-Day Test Period
A	00:00-06:00	35,446
B	06:00-10:00	23,642
C	10:00-16:00	35,394
D	16:00-20:00	23,714
E	20:00-24:00	23,459
Total	00:00-24:00	141,655

For the purpose of evaluating the off-line FAR, model MLF was applied to the 33 days of detector data using the two DTs described previously: 0.640 (ie. model MLF1) and 0.695 (ie. model MLF2). The two models were run in 'prediction' mode using the 33-day detector data. In this mode, the models were presented with only the input values (ie. the speed, flow and occupancy data at both the upstream and downstream stations). The output of the model in response to these inputs represented either an incident or incident-free condition for the specific time period under consideration. This procedure was adopted because the model will be run in this mode when it is eventually implemented in the field. The model's output or decisions (every 20-seconds) were monitored and all the alarms generated by the algorithm, along with the time at which they occurred, were written to a file for later inspection.

The next step in the evaluation procedure involved the examination of the individual alarms generated by the models and investigating the causes of these alarms by checking the video tapes using the time-lapse video recorder. This was made possible because the video tapes included a time stamp that was synchronised with the time stamp in the detector data files. A total of five incidents occurred within section S4-S5 during the 33 days of off-line evaluation. These incidents were confirmed using the video tapes which also allowed for the determination of the true start and end times of the incidents. The incident detection performance of model MLF1 on these incidents consisted of a DR of 100.0% (5/5) and a MTTD of 108 seconds. The performance of model MLF2, however, comprised a DR of 60.0% (3/5) and a MTTD of 146 seconds.

The major causes of alarms generated by the models were found to fall into three main categories as shown in *Figure 5* below: incident related (incident conditions and rubber necking), traffic flow related (absence of traffic and congestion) and 'other' (non-incident/false alarm related). The nature of the alarms generated within each of these categories and their causes are described below.

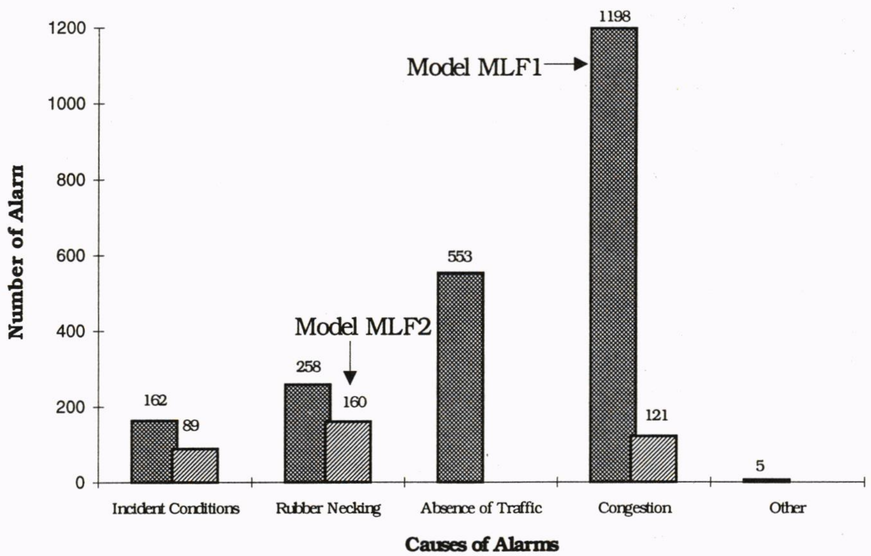

Figure 5: Causes of alarms generated during video testing

I. Incident Related Category

The alarms that were generated within this category fall into two sub-categories: incident conditions and rubber necking conditions. As was mentioned earlier, a total of five incidents occurred within section S4-S5 during the 33 days of off-line evaluation. The high number of incident conditions shown in *Figure 5* is due to the repeated alarm declarations during an incident and are therefore not considered as false alarms. During the analysis of the detector data, it was noticed that the ANN models generated many alarms within the test section when an incident occurred on the other side of the carriageway. When the video tapes were examined, 'rubber necking' conditions became evident when it was observed that drivers were slowing down due to the presence of an incident on the other side of the freeway.

II. Traffic Related Category

The alarms that were generated within this category also fall into two sub-categories: absence of traffic and congested conditions. The alarms generated for the absence of traffic category were easy to discern since the detector stations were not providing any data during these periods due to the absence of traffic. Data for these periods consisted of zero values for all input parameters for both the upstream and downstream stations of the section. This caused the models to generate a false alarm. For these periods, the video tapes did not show any abnormal traffic conditions. In fact, about 50% (278/553) of model MLF1 alarms were generated during the period from 00:00 to 06:00 in the morning when the traffic volumes inbound were light. The 'absence of traffic' alarm conditions cannot be considered as legitimate false alarms because they were basically due to the lack of 'valid' detector data.

As for the congested conditions, these were identified by checking the prevailing traffic flow conditions on the video tape for the designated alarm periods. A major part of these alarms was detected on the inbound direction of the section for the periods between 06:00 and 10:00. The detector data showed a significant drop in the speeds and flow; and an increase in the occupancy values during these periods for which the ANN models generated a series of false alarms. Investigation of the video tapes clearly showed congested conditions and slow moving vehicles for these periods which comprised the major part of the morning peak-hour for the inbound direction.

III. Non-Incident Related/False Alarm Category

The few alarms that were generated within this category ('other' in *Figure 5*) were those that could not be accounted for or justified from the detector or video data and were therefore considered false alarms. A total of five false alarms were generated for model MLF1 and three false alarms for model MLF2.

The results shown in *Figure 5* clearly demonstrate that, based on the 33 days of video testing, the application of a higher DT (0.695 instead of 0.640) resulted in the elimination of all the alarms that were due to the absence of traffic (553 alarms). It also resulted in the reduction of the alarms that were due to recurrent congestion from 1198 to 121 (by about 90%). This, however, was at the expense of detecting only three of the five incidents that occurred during video testing.

9. FAR ANALYSIS USING THE COMBINED MODELS MLF1 AND MLF2

As was discussed previously, model MLF2 was shown to have the potential to reduce the large number of alarms that were generated during peak-hour conditions (by about 90%). The expected DR of model MLF2, however, was only 50% (*Table 2*). Model MLF1, on the other hand, was shown to have a superior incident detection performance with a DR of 82.5%.

Analysis of the results reported for model MLF1 during the video testing period revealed that about 87% of the alarms were raised during the periods A, B and C (from 00:00-16:00) and that about 75.3% of the alarms occurred during moderate to heavy flow conditions (Dia, 1996). This suggested that an overall improvement in incident detection performance may be obtained by applying model MLF2 to periods A, B and C while applying model MLF1 to periods D and E. It should be pointed out here that models MLF1 and MLF2 have exactly the same parameters. The only difference between the two models is the DT value used to declare an incident alarm. The parameters of model MLF (and hence models MLF1 and MLF2) were determined using the 60 incidents in the training set. The combined model strategy can therefore be thought of as a method to 'fine-tune' the performance of the trained ANN incident detection model. The alarms generated using this scenario of model combinations are shown in *Figure 6* below. The implementation of the combined models MLF1 and MLF2 resulted in the detection of all the five incidents and the reduction of the total number of 'non-incident related' alarms from 1756 for model MLF1 to 206 for the combined models (by about 88%).

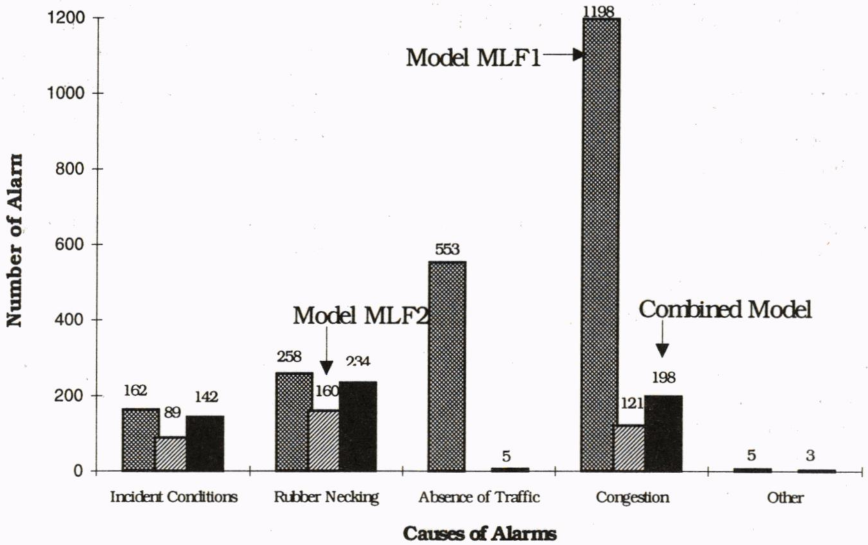

Figure 6: Causes of alarms for the combined model strategy

Table 4 presents FAR calculations for models MLF1, MLF2 and the combined models that are based on three different scenarios. In the first scenario, the FAR is calculated by excluding the legitimate incident-related alarms which included incident and rubber-necking alarms. In the second scenario, the FAR is calculated by excluding the incident-related alarms along with the alarms that were caused by the absence of traffic. As was mentioned previously, the alarms that were generated due to the absence of traffic cannot be considered as legitimate false alarms because they were basically due to the lack of 'valid' detector data. In the third and final scenario, the FAR calculations are based only on the alarms that could not be accounted for from the video and detector data and were therefore deemed as false alarms.

The FAR results shown in *Table 4* clearly demonstrate the effectiveness of implementing a combined model strategy. The expected DR using this technique (based on the 40 incidents in the validation-test data set) would be between 50% and 82.5%, although based on the five incidents that occurred during the off-line evaluation, the DR was found to be 100%. The improved FAR results obtained by using two different DTs suggest that additional improvements in performance may be obtained by implementing a dynamic DT, eg. implementing a variable DT based on the traffic volumes during the previous 15 minutes. There are, however, many issues that need to be investigated in this regard such as the number of dynamic DTs to use, the trade-off in performance between the DR and FAR using these thresholds and a more detailed analysis of the volume categories to which variable DTs need to be applied. These important issues are currently being investigated by the authors. The FAR calculations shown in *Table 4* reveal that the implementation of a combined model strategy would result in a worst case scenario of 0.073% FAR which is equivalent to about 6.2 false alarms per day (206/33 days) on both directions of section S4-S5.

Table 4: Summary of FAR calculations for the 33-days of off-line video testing on section S4-S5

Basis for FAR Calculations	FAR for Model MLF1 (DR=82.5%)	FAR for Model MLF2 (DR=50.0%)	FAR for Combined Models (DR=50%-82.5%)
(1) All alarms except: - incident alarms - rubber-necking alarms	1756/282890 0.621%	121/283061 0.043%	206/282934 0.073%
(2) All alarms except: - incident alarms - rubber-necking alarms - absence of traffic alarms	1203/282337 0.426%	121/283061 0.043%	201/283286 0.071%
(3) All alarms except: - incident alarms - rubber-necking alarms - absence of traffic alarms - congestion alarms	5/281139 0.0018%	0/282994 0.000%	3/282864 0.001%

10. CONCLUSIONS

The results presented in this paper have demonstrated the feasibility of using 'real-world' data for developing ANN incident detection models. These results provide a comprehensive evaluation of the ANN models and confirm that these models can provide fast and reliable incident detection on freeways. The performance of the selected model, especially in terms of FAR, was substantially improved by implementing a combined model strategy where higher decision thresholds were applied to the peak-periods of the day. Efforts are currently underway to further improve the model's performance by implementing a dynamic decision threshold based on prevailing traffic flows rather than periods of the day. The upcoming implementation of the ANN model on Melbourne's freeways will provide a unique opportunity for evaluating the on-line performance of the model over an extended period of time. There is also scope in future research efforts to compare the performance of the ANN model with other overseas models (eg. California, McMaster or Minnesota algorithms) based on the same data set of 100 incidents used in this study.

ACKNOWLEDGMENTS

The support provided by VicRoads for this project is gratefully acknowledged. In particular, the authors acknowledge the generous assistance of Mr. Francis Sin and Mr. Anthony Snell.

REFERENCES

Ahmed S.R. and Cook A.R. (1982). Application of time-series analysis techniques to freeway incident detection. *Transportation Research Record*, **841**, 19-21.

Cheu, R.L. and Ritchie, S.G. (1995). Automated detection of lane-blocking freeway incidents using artificial neural networks. *Transportation Research,* 3**C**(6), 371-388.

Dia, H. (1996). Artificial Neural Network Models for Automated Freeway Incident Detection. Ph.D. Dissertation. (Monash University: Clayton).

Dia, H. and Rose, G. (1995). Development of artificial neural network models for automated detection of freeway incidents. *Paper presented at The 7th World Conference on Transport Research,* Sydney, Australia, July 1995 (forthcoming in the selected conference proceedings).

Dia, H. and Rose, G. (1997). Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. Forthcoming in the special issue of Transportation Research Part C 'Applications of Neural Networks in Transport', Dougherty, M. (guest editor).

Hoose, Neil (1992). IMPACTS, an image analysis tool for motorway surveillance. *Traffic Engineering and Control.* **33**(3), 140-147.

Levin M. and Krause G.M. (1979). Incident detection algorithms. Part 1: off-line evaluation; Part 2: on-line evaluation. *Transportation Research Record*, **722**, 49-64.

Masters, T. (1993). *Practical Neural Network Recipes in C++.* (Academic Press Inc.: San Diego).

Michalopoulos, P.G., Jacobson, R.D., Anderson, C.A. and DeBruycker, B. (1993). Automatic incident detection through video image processing. *Traffic Engineering and Control*, **34**(2), 66-75.

Persaud, B.N. and Hall, F.L. (1989). Catastrophe theory and patterns in 30-second freeway traffic data - implications for incident detection. *Transportation Research*, **23A**(2), 103-113.

Ritchie, S.G. and Cheu, R.L. (1993). Simulation of freeway incident detection using artificial neural networks. *Transportation Research,* **1C**(3), 203-217.

Rose, G. and Dia, H. (1995). Freeway automatic incident detection using artificial neural networks. *Proceedings of the International Conference on Application of New Technology to Transport Systems*, Vol. **1**, Melbourne, Australia, 123-140.

Stephanedes, Y.J. and Liu, X. (1995). Artificial Neural Networks for Freeway Incident Detection. *Transportation Research Record*, **1494**, 91-97.

The 2nd Conferencs of the Eastern Asia Society for Transportation Studies
organized by
Korean Society of Transportation
and supported by

EASTS (TOKYO)

Air Safety Foundation
Air Traffic Control Association, Japan
Air Traffic Service Research Institute
Airport Environment Improvement Foundation
Airport Security Business Center
All Japan Air Transport and Service Association
All Japan Airport Refueling Service Association
All Japan Airport Terminals Association Inc.
Almec Corporation
Association of Air Transport Engineering and Research
Chubu International Airport Research
Coastal Development Institute of Technology
Commuter Aerodrome Support and Aeronautical Service Foundation
East Japan Railway Culture Foundation
Expressway Technology Center
Hanshin Expressway public Corporation
Honshu-Shikoku Bridge Authority
Institute of Behavioral Sciences
Institute of Transportation Economics
JAL Foundation
Japan Aeromedical Research Center
Japan Aeronautic Association
Japan Aeronautical Engineers' Association
Japan Aircraft Pilot Association
Japan Civil Aviation Promotion Foundation
Japan Non-government Railways Association
Japan Pilot Training Promotion Association
Japan Radio Air Navigation Systems Association
Japan Railway Technical Service
Japan Road Association
Japan Society of Traffic Engineers
Japan Traffic Culture Association
Japan Transport Cooperation Association
Japan Transport Economics Research Center
Japan Transportation Planning Association
Kikaku Kaihatsu Inc.
Metropolitan Expressway Company
Overseas Coastal Area Development Institute of Japan
Railway Technical Research Institute
Reliability Engineering Foundation for Air Navigation Facilities
Scheduled Airlines Association of Japan
Technology Center of Metropolitan Expressway
The Japan Port and Harbour Association
Waterfront Vitalization and Environment Research Center

KST (KOREA)

Korean Ministy of Construction and Transportation
The Korea Transport Institute
Korea Research Institute of Human Settlements
Seoul Development Institute
Road Traffic Safety Association
Korea Institute of Construction Technology
Asiana Airline
Association of Korea Professional Transport Engineering
Korea Airport Construction Authority
LG Industrial Systems Co. Ltd.