Pergamon

Transpn Res.-C, Vol. 5, No. 5, pp. 313–331, 1997

0968-090X/97 $17.00+0.00

PII: S0968-090X(97)00016–8

# DEVELOPMENT AND EVALUATION OF NEURAL NETWORK FREEWAY INCIDENT DETECTION MODELS USING FIELD DATA

HUSSEIN DIA* and GEOFF ROSE
Institute of Transport Studies, Department of Civil Engineering, Monash University, Clayton, Victoria 3168, Australia

*(Received 7 October 1996; in revised form 31 July 1997)*

**Abstract**—This paper discusses a multi-layer feedforward (MLF) neural network incident detection model that was developed and evaluated using field data. In contrast to published neural network incident detection models which relied on simulated or limited field data for model development and testing, the model described in this paper was trained and tested on a real-world data set of 100 incidents. The model uses speed, flow and occupancy data measured at dual stations, averaged across all lanes and only from time interval $t$. The off-line performance of the model is reported under both incident and non-incident conditions. The incident detection performance of the model is reported based on a validation-test data set of 40 incidents that were independent of the 60 incidents used for training. The false alarm rates of the model are evaluated based on non-incident data that were collected from a freeway section which was video-taped for a period of 33 days. A comparative evaluation between the neural network model and the incident detection model in operation on Melbourne's freeways is also presented. The results of the comparative performance evaluation clearly demonstrate the substantial improvement in incident detection performance obtained by the neural network model. The paper also presents additional results that demonstrate how improvements in model performance can be achieved using variable decision thresholds. Finally, the model's fault-tolerance under conditions of corrupt or missing data is investigated and the impact of loop detector failure/malfunction on the performance of the trained model is evaluated and discussed. The results presented in this paper provide a comprehensive evaluation of the developed model and confirm that neural network models can provide fast and reliable incident detection on freeways. © 1997 Elsevier Science Ltd. All rights reserved

*Keywords*: freeway incident detection, artificial neural networks, freeway traffic management, intelligent transport systems, artificial intelligence.

## 1. INTRODUCTION

The high cost of congestion caused by incidents such as accidents, disabled vehicles, construction work and other events that result in a capacity reduction of the facility, has prompted a growing worldwide 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. For example, incidents are believed to constitute about 50–60% of the total delays on U.S. freeways (Lindley, 1987). 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 road safety and real-time traffic control. Motorists can be informed by providing them with real time traveller information regarding expected delays, potential bottlenecks and other information that can help them in choosing alternate routes during congested conditions. 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 procedures for detecting incidents which are both reliable and quick to respond. 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

*Author for correspondence: Fax: +61 3 9905 4944; e-mail: hussein.dia@eng.monash.edu.au

that maximise detection rate while only generating an operationally acceptable level of false alarms.

This paper first discusses the development of a relatively new-generation of algorithms for freeway incident detection using artificial neural networks (ANNs). In contrast to published neural network 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 emphasis of the paper, however, will be on the off-line performance evaluation of the trained ANN model and the investigation of the model's fault-tolerance under corrupt or missing data conditions. 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 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 results of a comparative evaluation between the ANN model and the model in operation on Melbourne's freeways is then presented. Finally, the impact of loop detector failure/malfunction on the performance of the ANN model is evaluated and discussed.

## 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.

## 3. 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) and the algorithm that was jointly developed by the Australian Road Research Board (ARRB) and the Victoria State Road Authority (VicRoads) (Luk and Sin, 1992); the McMaster algorithm (Persaud and Hall, 1989) and the time series algorithms, e.g. the auto-regressive integrated moving average (ARIMA) algorithm (Ahmed and Cook, 1982).

## 4. 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.

## 5. 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 incident intervals. 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 DR and 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.

## 6. DATA FOR THE DEVELOPMENT OF ANN INCIDENT DETECTION 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.

## 7. DATA COLLECTION

The data required for model development in this study were assembled from two data sources held at the VicRoads Traffic Control and Communications Centre (TCCC) in Melbourne, Australia. A schematic showing the sections of the Tullamarine Freeway where the data were collected is shown in Fig. 1. The first data source comprised information logged by operators monitoring traffic conditions regarding the incidents that occurred on the freeways. This information is also received from a variety of other sources including motorists or the currently operational ARRB/VicRoads incident detection model. However, when operators are busy managing incidents, it is not uncommon for important details to be left out from the records. This presented some difficulties when examining the incidents since in many cases the location of the incident, its direction or time of occurrence remained unknown from the record. A total of 385 incidents that were logged during the period from January 1992 to March 1994 were extracted from this database. The second data source comprised lane-by-lane loop-detector data consisting of speed, flow and occupancy measurements in 20-s cycles. For the freeway under consideration (Fig. 1), the detector station spacing ranged between 450 and 1070 m, with an average spacing of about 580 m for the 14 detector stations.

Out of the 385 incidents recorded by the operators in the log, only 120 incidents could be confirmed from the loop data. The rest either occurred outside the 8.5 km segment of the freeway or during light flow conditions and therefore had no effect on traffic conditions. Others could not be

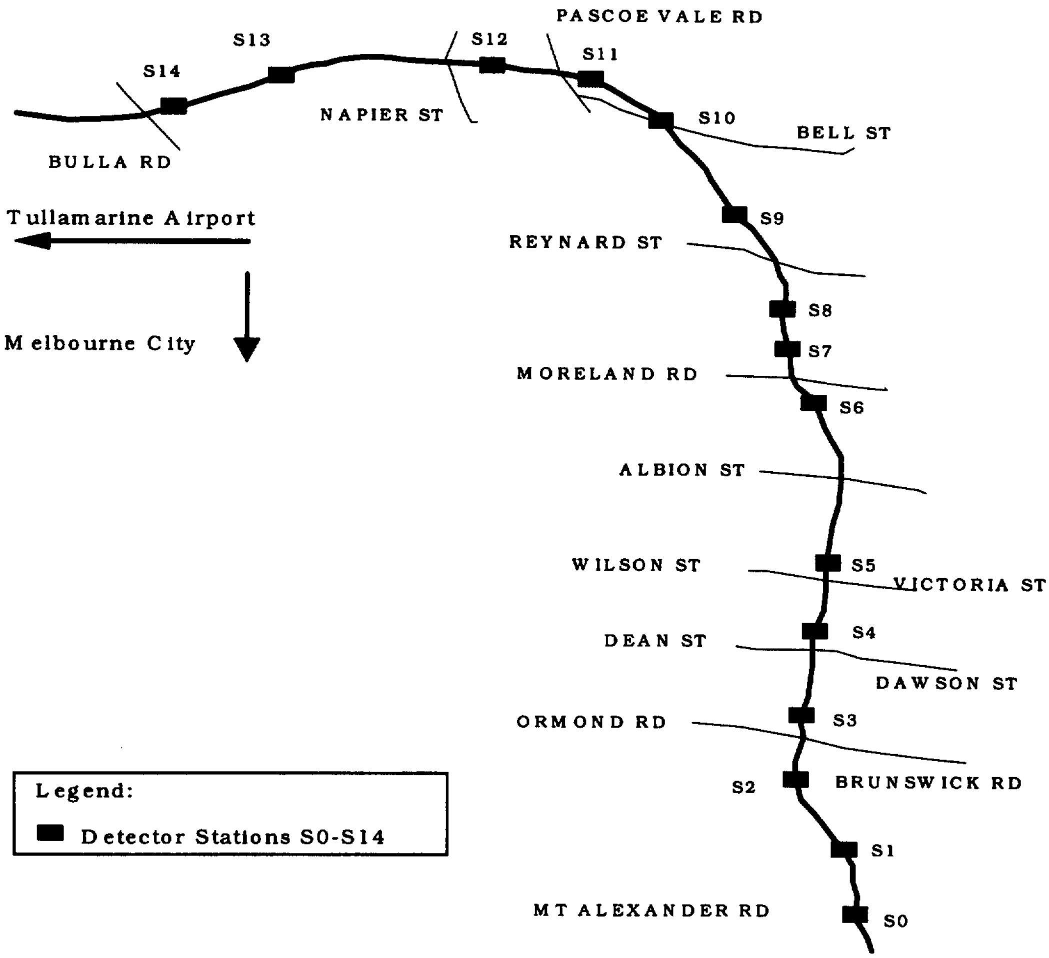


Fig. 1. Schematic of Tullamarine Freeway.

confirmed due to missing information about the location or time of incident. However, for sixty of these confirmed incidents, the detector data at the upstream and/or downstream stations were faulty or corrupted. These incidents, although confirmed, cannot be used for model development in this study since data of good quality from all lanes at both stations upstream and downstream of the incident need to be used. Therefore, out of the 385 incidents logged by the operators, only 60 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.

## 8. 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.

In the case of simulated data (Ritchie and Cheu, 1993), the incident start and end times are known precisely. However, when 'real world' data are being used, the incident start and end times are rarely, if ever, known precisely. Estimates of the time of occurrence and clearance of these incidents were compiled from the data sources described previously. The times provided in the operator log, however, were reported as the times when the operator detected or confirmed the incidents and not the times when the incidents actually occurred. Using the log time as a guide, the incident was deemed to have occurred in the first of three 20-s consecutive intervals (immediately before the start time reported in the log) when there was at least a 20% variation in flow, speed or occupancy values (Dia, 1996). It is important to emphasise here that the use of the 20% threshold variation to establish the incident start time did not pre-condition the data in any manner that could influence the training of the ANNs. This is due to the fact that there were many occasions under incident-free conditions when fluctuations greater than 20% were observed. Therefore, applying the 20% variation in flow, speed or occupancy criteria to the intervals immediately before the start time reported in the log resulted in the determination of the specific 20-s interval representing the start of the incident. A similar procedure was adopted for determining the incident end time. Once these times were determined, all observations between the start and end of incidents were assigned output state 2 {1}. In addition, at least 15 min of incident-free data prior to incident occurrence and after incident clearance were also included in the training. All incident-free data were assigned an output state 1 {0}.

## 9. 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), location on the freeway and incident duration. Two data sets, each comprising 30 incidents, were then selected randomly from the 60 incidents to form the master training and training-test data sets (Dia and Rose, 1995). The detailed characteristics of the 60 incidents are provided in Dia (1996). For incident severity, one incident resulted in blocking one lane; 42 incidents resulted in blocking two lanes and 17 incidents resulted in blocking three lanes immediately upstream of the incident. Only one incident occurred during light flow conditions (i.e. flows less than 700 $vh^{-1}\ l^{-1}$), 39 occurred during heavy flow conditions (i.e. flows exceeding 1550 $vh^{-1}\ l^{-1}$) and 20 incidents occurred during moderate flow conditions (i.e. flows between 700 and 1550 $vh^{-1}\ l^{-1}$). In addition, a total of 20 incidents occurred during peak-hour conditions. A total of 12 incidents lasted for less than 30 min, 24 lasted between 30 and 60 min, 16 lasted between 60 and 90 min and 8 incidents lasted for more than 90 min.

## 10. 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 models. This data set should be independent of the data sets used for model training. The training data sets described previously (covering the period January 1992 to April 1994) were collected in April 1994 for the purpose of training the ANN models. In March 1995, another set of incident data, the validation-test data set, was collected from two freeways in Melbourne for validating the performance of the trained ANN models. One of the freeways (Tullamarine) is the same freeway that was used for collecting the training data sets which did not comprise any data from the other freeway (South Eastern Freeway). A total of 90 incidents that occurred on the two freeways between January 1992 and March 1995 were extracted from the operator's log for examination. A total of 50 incidents were discarded because the loop detectors upstream and/or downstream of the location of incidents were faulty or the data from these detectors were corrupted. This resulted in the compilation of a validation-test data set comprising 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.

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. It should be mentioned here that investigation of the impact of sample size on model performance (Dia, 1996) revealed that an 'adequate' performance could be obtained by training an ANN model with about 12 to 25 incidents that comprise about 5000 to 11000 observations, provided that incident and non-incident conditions are proportionately represented in these observations.

## 11. A FRAMEWORK FOR AUTOMATED INCIDENT DETECTION USING ANNS

The general framework for automated incident detection using ANNs is shown in Fig. 2 (Rose and Dia, 1995). Consider the section of freeway shown in Fig. 2(a) which is defined by upstream and downstream detector stations. A corresponding ANN incident detection model structure is shown in Fig. 2(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 (averaged across all lanes in 20-s 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 earlier success, especially in real-time pattern recognition problems, and 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 Fig. 2(b), has been chosen for this study (Dia and Rose, 1995).

## 12. 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).

## 13. ANN FEATURES/PARAMETERS

The next step after arriving at the structure of the models to be investigated was the selection of an appropriate set of ANN features and parameters to use in the appraisal of these models. Table 1 below lists the ANN features which were arrived at after considerable investigation and used throughout the designed experiments (Dia, 1996).

The logicon projection network (Wilensky and Manukian, 1992) is a variant of the back-propagation network. The basic motivation behind the development of this model was the desire to build a faster and more streamlined network by combining the positive features of closed and open boundary networks (NeuralWare, 1993). Closed boundary networks, e.g. Adaptive resonance

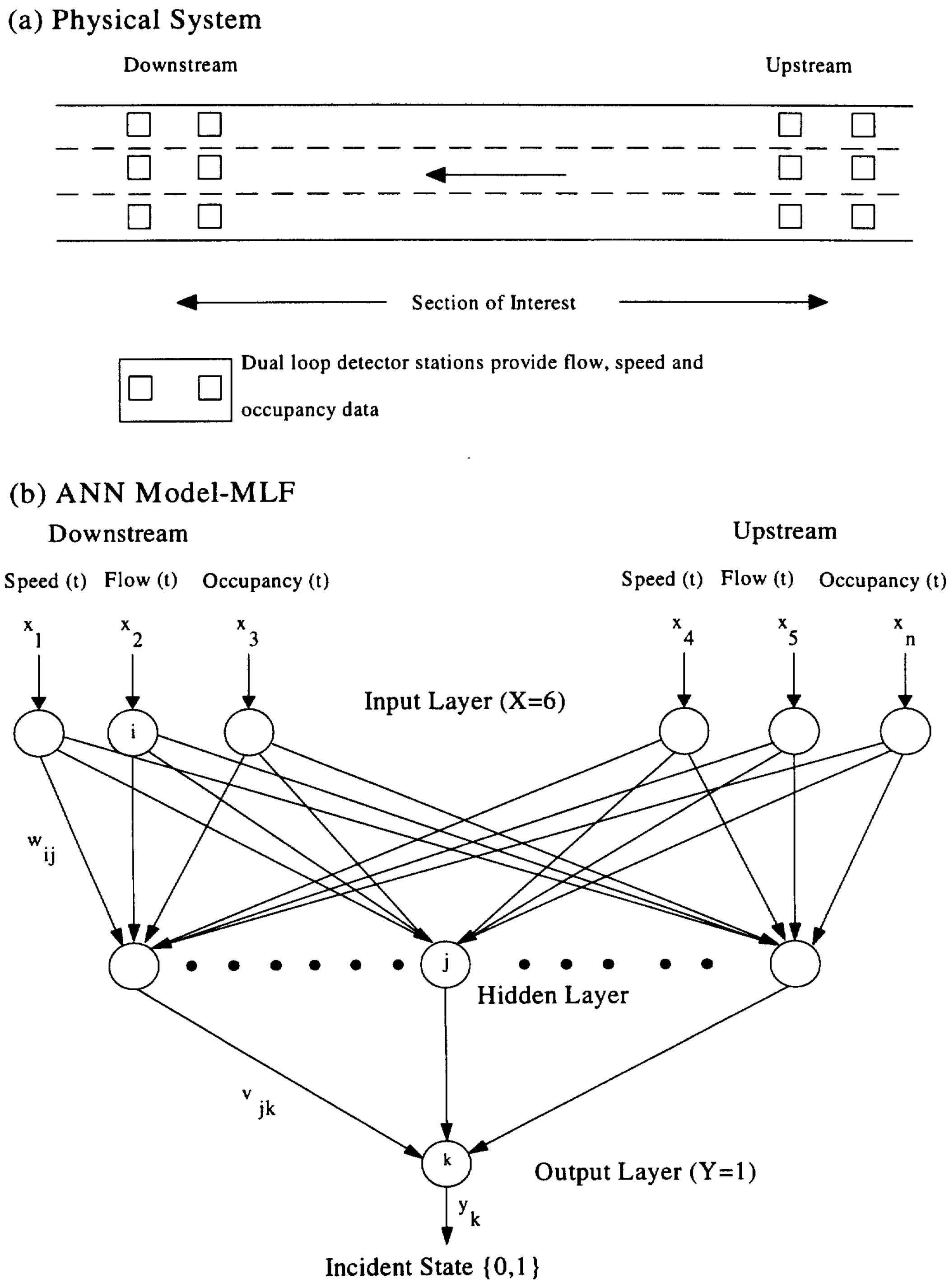


Fig. 2. ANN modelling framework.

Table 1. ANN features used in model development

| ANN feature | Description |
|---|---|
| ANN model | Logicon projection network |
| Objectiive function | Classification rate (the average of the correctly classified states) |
| Learning rule | Quickprop |
| Transfer function | Sigmoid |
| Output ranges | 0.2–0.8 (instead of 0–1) |

theory networks (ART), are fast learning because they properly initialise the network weights and thresholds to prototypes of the training data set. Open boundary networks (e.g. the back-propagation), on the other hand, have the advantage of minimising the output error through gradient descent. The Logicon projection network combines these two basic types of networks by initialising the weights and thresholds as prototypes such that the network's output error is already close to a desirable minimum. The network is subsequently treated as a standard back-propagation network and trained using a gradient descent learning algorithm which further reduces the output error. This strategy can result in faster training times by avoiding the flat 'plateaus' and local minima on the error curve that are often experienced by gradient descent algorithms. This, in addition to using the Quickprop learning rule (which uses a quadratic estimation heuristic to determine the direction and step size for updating the connection weights), were found to result in faster learning times (Dia, 1996).

## 14. TRAINING STRATEGY

After a model is designed, it is trained on the training data set for a maximum of 100 cycles. Each cycle of training involves the random presentation of all the observations in the training file to the network. Therefore, the training of the model for 100 cycles meant that each vector in the training file was presented to the ANN model exactly 100 times. All other training parameters were held constant during training. At the end of each cycle, the trained model is tested on the training-test data set. If the classification rate (the objective function) on the training-test data set improves, the model is saved and training continues. If the classification rate does not improve for any consecutive 100 tests, the training is stopped and the last model saved constitutes the best model for the given input features and ANN parameters.

As was mentioned earlier, the number of training cycles required for the Logicon Projection Network are generally lower than those for other back-propagation networks. The choice of 100 cycles was the result of considerable investigation (Dia, 1996) in which it was found that this number of cycles, along with the training strategy described above, were sufficient conditions for the network to learn the general patterns needed to produce the correct classification for the training-test data set.

## 15. MODEL EVALUATION USING PERFORMANCE ENVELOPES

The basic principle behind the ANN incident detection model being developed in this study is the correct classification of the traffic flow input parameters, provided in 20-s cycles, into either an incident-free $\{0\}$ or incident $\{1\}$ state. Assuming no persistence tests are applied, an incident is then detected when the traffic state changes from $\{0\}$ to $\{1\}$.

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 processing element (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 Fig. 3) is called the Performance Envelope Curve (PEC) of the network.

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 of 0.8. The upper-right corner (100%,100%) will be the other endpoint of the PEC. This point corresponds to a DT of 0.2. 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 Fig. 3). At the ideal performance DT, the DR would be 100% and the FAR would be 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

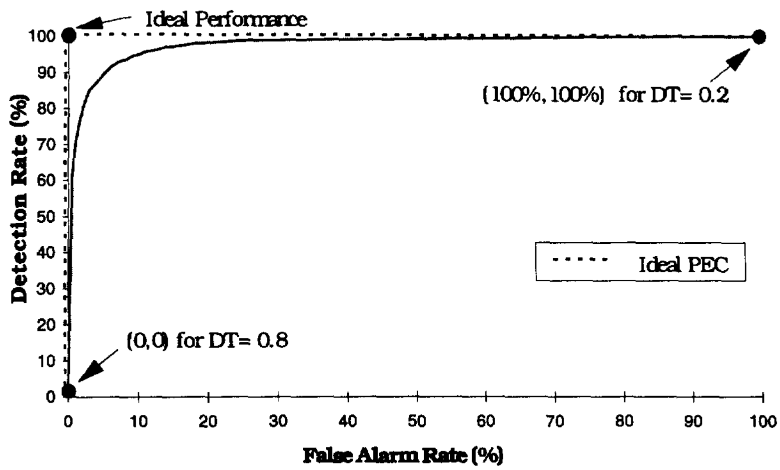


Fig. 3. A performance envelope curve.

(i.e. 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, i.e. the area under the PEC.

## 16. 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 revealed that the architecture of the best performance model was similar to that shown previously in Fig. 2 (model MLF). This model uses speed, flow and occupancy data collected from dual stations, 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.

## 17. PERFORMANCE EVALUATION OF THE ANN INCIDENT DETECTION MODEL

Investigation of several techniques with the potential for reducing 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 2 and Fig. 4 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 min × 60 min $h^{-1}$ × 24 h $day^{-1}$). Table 2 and Fig. 4 also show the performance measures for model MLF based on the application

Table 2. Incident detection performance of model MLF based on the validation-test data

| Data set | Model ID | Decision threshold (DT) | Incident detection performance: Detection Number | Detection Rate(%) | False alarm rate (%) | Pesistence test = 2: Time to detect (s) |
|---|---|---|---|---|---|---|
| Validation-test data 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 |

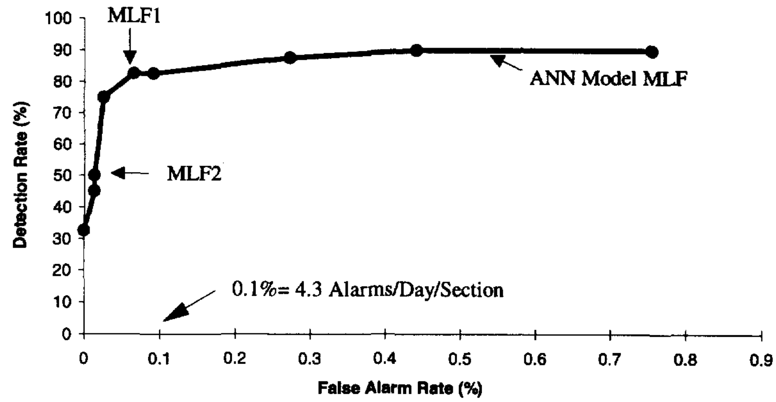


Fig. 4. PEC for model MLF based on the validation-test set of 40 incidents.

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 2 and Fig. 4. 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.

When assessing the performance of an incident detection model, it is important to conduct the assessment in the context of both incident and non-incident conditions. In addition, it is also useful to compare the performance of the model under consideration with other incident detection algorithms such that the trade off in performance using different theoretical techniques is also evaluated. These separate performance assessments are reported in the sections which follow.

## 18. PERFORMANCE EVALUATION UNDER INCIDENT CONDITIONS

The evaluation of the ANN model under incident conditions will be based on the application of model MLF1 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 min 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 min. The results for model MFL1 based on the Tullamarine and South Eastern Freeway (SEF) data sets are shown in Table 3.

These results indicate that the DR and FAR performance of model MLF1 on the validation-test data set is generally consistent with its expected performance in Table 2 (82.5% DR, 0.065% FAR and 203 s 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.

Table 3. Incident detection performance of model MLF1 based on the total set of 40 incidents

| Data set | Number of incidents | Incident detection performance | | | |
|---|---|---|---|---|---|
| | | Detection rate | | False alarm rate (%) | Mean time-to-detect (s) |
| | | No. | % | | |
| Validation test set–Tulla marine | 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 |

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 and 216 s 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%).

*18.1. Distribution of incidents detected according to the time to detect incidents*

The ability of an incident detection algorithm to detect incidents quickly has become the most critical requirement of an efficient freeway management system. In this context, the performance of the ANN model according to the time taken by the algorithm to detect incidents is of interest. The model detected 72.7% of incidents in less than 4 min, 39.4% in less than 3 min and 18.2% in less than 2 min. It should be mentioned here that model MLF1 implements a two-interval persistence test which results in a minimum delay of 40 s before an incident condition is declared.

As was mentioned previously, the operators' log files obtained from VicRoads included the times when the operators detected or confirmed the incidents. Inspection of these log times revealed that two incidents were detected by the operators before their impact on traffic was confirmed from the detector data. Only 18.4% (seven incidents) of the remaining 38 incidents were detected by the operators within 3 min of their occurrence. The average time taken by the operators to detect the 38 incidents, however, was 6.9 min after their estimated occurrence times. The results reported in this study for the MTTD of model MLFI (Table 2) was 3.4 min (203 s). 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.

## 19. PERFORMANCE EVALUATION DURING 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 (Fig. 5 overleaf) was continuously video-taped (on 24-h 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 Fig. 5. 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.

*19.1. 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, each day was divided into five periods, as shown in Table 4, such that the performance of the model is evaluated separately for each period.

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 (i.e. model MLF1) and 0.695 (i.e. model MLF2). The two models were run in 'prediction' mode using the 33-day detector data.

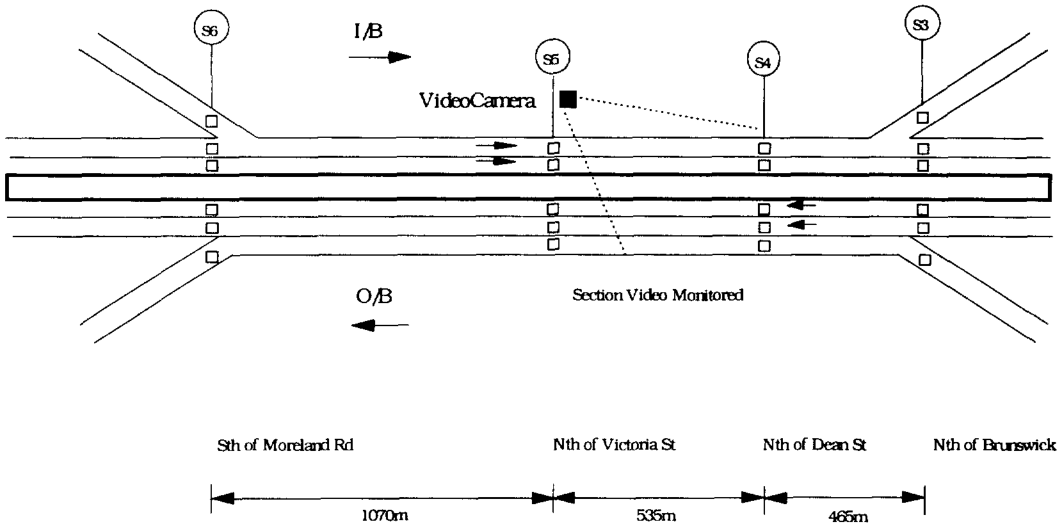


Fig. 5. Schematic of section S4–S5 on the Tullamarine Freeway used for video monitoring and evaluation of the off-line FAR.

Table 4. Classification of periods of the day for the off-line FAR evaluation for section S4–S5

| Period of day | Time frame | Total number of 20-s observations | Average volume ($vh^{-1}l^{-1}$) | |
|---|---|---|---|---|
| | | | Inbound direction | Outbound direction |
| A | 00:00-06:00 | 35,446 | 340 | 314 |
| B | 06:00-10:00 | 23,642 | 1548 | 847 |
| C | 10:00-16:00 | 35,394 | 1643 | 1125 |
| D | 16:00-20:00 | 23,714 | 1705 | 1396 |
| E | 20:00-24:00 | 23,459 | 882 | 692 |
| | 00:00-24:00 | 141,655 | 1185 | 849 |

In this mode, the models were presented with only the input values (i.e. 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 s) 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 s. The performance of model MLF2, however, comprised a DR of 60.0% (3/5) and a MTTD of 146 s.

The major causes of alarms generated by the models were found to fall into three main categories as shown in Fig. 6 overleaf: 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:

1. 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

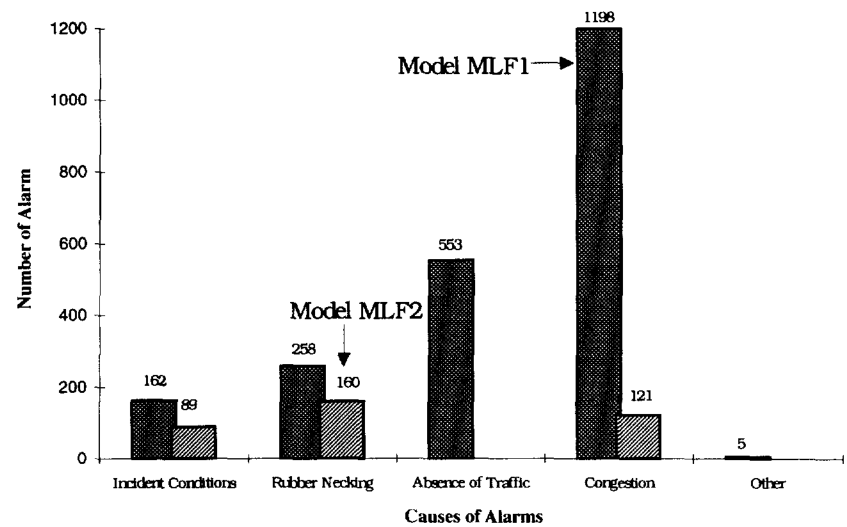


Fig. 6. Causes of alarms generated during video testing.

incident conditions shown in Fig. 6 is due to the fact that the ANN models were trained to distinguish between incident and non-incident conditions and would therefore continue to generate alarms as long as incident conditions are detected. Therefore, the repeated alarm declarations during an incident are not considered as false alarms. In practice, it may be possible to automatically turn the algorithm off (for the specific section) after detecting the incident and switching it back on after detecting the end of the incident. This, however, will require the development, testing and evaluation of algorithms for detecting the end of the incident, i.e. detecting when the traffic state changes from State 2 to State 1. 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.

2. 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. From a practical perspective, it would be possible to implement procedures that would validate the integrity of the data before they are presented to the model. If the data for a specific 20-s interval are found to be 'invalid', it is discarded and the model does not make a decision on that data. Alternatively, data from the immediate upstream station could be substituted until the faulty detectors are fixed.

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 and in some cases for the periods between 10:00 and 12:00. The detector data showed a significant drop in speed 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.

3. Non-Incident Related/False Alarm Category
   The few alarms that were generated within this category 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 Fig. 6 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.

*19.2. 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 to 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 Fig. 7 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 (i.e. 553 + 1198 + 5) for model MLF1 to 206 (i.e. 5 + 198 + 3) for the combined models (by about 88%). Table 5 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

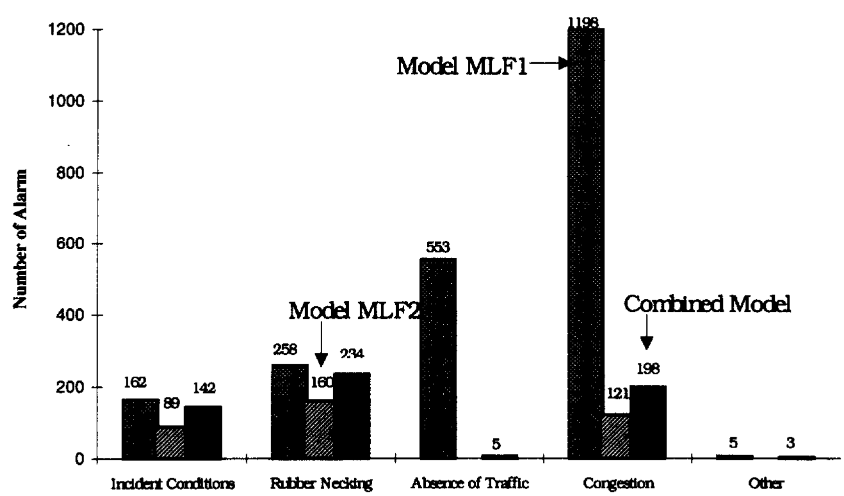


Fig. 7. Causes of alarms for the combined model strategy.

Table 5. 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:<br>—incident alarms<br>—rubber-necking alarms | 1756/282890<br>0.621% | 121/283061<br>0.043% | 206/282934<br>0.073% |
| (2) All alarms except:<br>—incident alarms<br>—rubber-necking alarms<br>—absence of traffic alarms | 1203/282337<br>0.426% | 121/283061<br>0.043% | 201/283286<br>0.071% |
| (3) All alarms except:<br>—incident alarms<br>—rubber-necking alarms<br>—absence of traffic alarms<br>—congestion alarms | 5/281139<br>0.0018% | 0/282994<br>0.000% | 3/282864<br>0.001% |

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 5 clearly demonstrate the effectiveness of implementing a combined model strategy. The expected DR using this technique 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, e.g. implementing a variable DT based on the traffic volumes during the previous 15 min. 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 Fig. 5 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. Based on this scenario, it is expected that about 87 false alarms per day would be generated on the Tullamarine Freeway facility (the 14 sections under study).

## 20. COMPARATIVE EVALUATION OF THE ANN AND ARRB/VICROADS MODELS

The next step in the evaluation process involved comparing the performance of the ANN model with an existing incident detection model (ARRB/VicRoads model) based on the independent validation-test data set of 40 incidents which was not used in the training or calibration of either model. The basic logic behind the ARRB/VicRoads model (Luk and Sin, 1992) is to compare the traffic data between adjacent stations and adjacent lanes and declare an incident if the differences exceed pre-determined threshold values. In addition to the conservation of flow principle, where the loss of traffic flow is considered a good indicator of an incident, three sets of algorithms are also used for identifying an incident: adjacent station comparison, adjacent lane comparison and time series differencing. Each algorithm represents a certain condition that must be met before an alarm is raised (Snell *et al.*, 1992).

### *20.1. Recalibrated ARRB/VicRoads model*

A comparative performance evaluation of any two incident detection algorithms is only meaningful if both algorithms are calibrated and tested on the same data sets. The ARRB/VicRoads model was initially calibrated using a small set of data (available within the implementation time frame) that was collected from the South Eastern Freeway (Luk and Sin, 1992). The data collected for this study have allowed the model to be recalibrated using the 60 incidents in the training data set. The incident detection performance results for the recalibrated model based on the validation-test data set are shown in Table 6 below (Dia *et al.*, 1996).

The incident detection performance of model MLF based on the validation-test data set was presented previously in Table 2. The results of the comparative evaluation of the ANN and ARRB/VicRoads models, based on Tables 2 and 6, are shown in Fig. 8. The PECs shown in Fig. 8 clearly demonstrate the trade off in performance using these models. These results also demonstrate the substantial improvement in incident detection performance obtained by the ANN model over the ARRB/VicRoads model for the validation-test data set used in this study. Further details regarding the evaluation of the ARRB/VicRoads model and the comparative evaluation of the two models can be found in Dia *et al.* (1996).

## 21. THE IMPACT OF DATA QUALITY ON MODEL PERFORMANCE

The performance evaluation of the ANN model during incident and non-incident conditions have so far been based on valid and correct data. In practice, however, the loop detectors are susceptible to damage, failure or malfunction. It is therefore desirable that the impact of these factors on the performance of the trained ANN model be evaluated. These performance assessments are reported in the sections which follow.

## 22. IMPACT OF DETECTOR FAILURE/MALFUNCTION ON MODEL PERFORMANCE

A simple and relatively effective method for assessing the impact of an input parameter on model performance is sensitivity analysis by input clamping (Masters, 1993). With this technique, the network's performance is evaluated with one of the inputs clamped to a fixed value for the entire validation-test data set. The importance of that input can be gauged by its effect on the area under the performance envelope curve (PECA) of the ANN model (Masters, 1993). As was mentioned earlier, an incident detection model with an ideal performance of (DR = 100% and FAR = 0%) has a PECA of 10,000. In the input clamping method, if the PECA for a certain model

Table 6. Summary of ARRB/VicRoads model performance based on the validation-test data set

| Model | Persistence test | Incident detection performance | | | | |
|---|---|---|---|---|---|---|
| | | Detection | | False alarms | | Time to detect |
| | | Number | Rate (%) | Number | Rate (%) | (s) |
| ARRB/VicRoads | 0 | 22/40 | 55.0 | 38 | 0.47 | 176 |
| model-recalibrated | 1 | 17/40 | 42.5 | 13 | 0.16 | 202 |
| | 2 | 12/40 | 30.0 | 7 | 0.09 | 193 |
| | 3 | 10/40 | 25.0 | 5 | 0.06 | 204 |

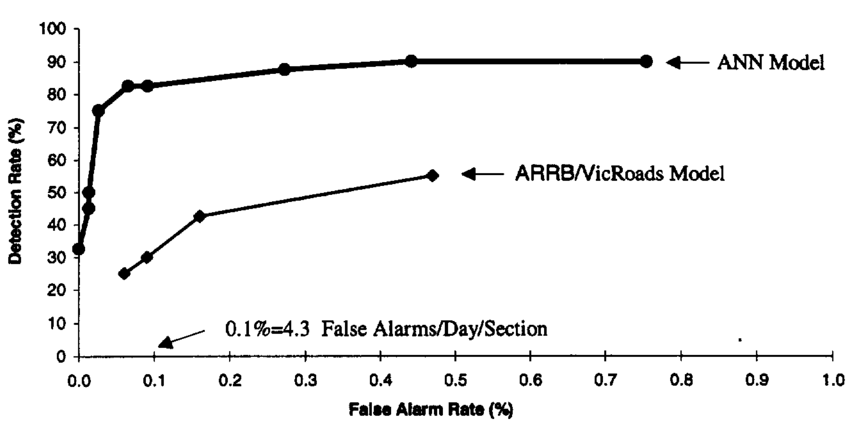


Fig. 8. Comparative evaluation of the ANN and ARRB/VicRoads model based on the validation-test data set of 40 incidents.

decreases substantially with that input clamped, then that input is important for incident detection. The value at which an input variable is clamped is typically selected to be within the range of actual values expected for that input variable when the network is eventually implemented in the field. Extreme values are avoided, where possible, since these can cause other inputs to be ignored (Masters, 1993).

One of the practical applications for sensitivity analysis by input clamping is the investigation of the impact of detector failure or communications malfunction on incident detection performance. The quality of the received data is affected by these failures which can have a profound effect on the performance of the algorithm, especially in terms of the FAR. The causes of error in the data are attributed to a variety of factors including detector lock-up due to high levels of vibration from heavy vehicles or external interference, loops damaged during road works, excavation or pavement resurfacing or electrical failure of detector cards (Snell *et al.*, 1992). These errors were reflected in the data files used in this study in two ways. The first type of error, detector failure, was observed when a particular station (consisting of all detectors in all lanes for a given direction) failed to transmit any data back to the TCCC. When this occurred, the three inputs (speed, flow and occupancy) for that station were reported as zeros in the entire data file. The other type of error (communications malfunction) was observed when a particular station did not transmit the traffic data (speed, flow and occupancy) for a certain number of 20-s intervals (data for these intervals were also transmitted as zeros), but the communications system reset itself afterwards and the data transmission was resumed. In both cases, this resulted in a false alarm being raised in the ANN incident detection model.

The impact of detector failure and communications malfunction on incident detection performance can therefore be investigated by clamping the input variable under consideration using a value of zero (Dia and Rose, 1996). If all inputs are clamped to zero, then this results in a DR of zero and a FAR of 100%. Table 7 below shows the sensitivity analysis results obtained by clamping one input at a time for the validation-test data set in a descending order of PEC area. These results, however, are only reported here based on the application of a zero persistence test. Although some of the results in Table 7 are of less practical relevance, they nevertheless reveal that the variation in model performance was substantial for the cases where the upstream speed, downstream speed or both were clamped. Compared to the original model with all inputs free, clamping the upstream speed input to zero resulted in just over a 30% decrease in the overall performance. Clamping the downstream speed input to zero, however, resulted in less than a 10% decrease in the overall performance of the model. The resulting incident detection performance in both cases, however, was not acceptable.

## 23. IMPACT OF MISSING SPEED DATA ON MODEL PERFORMANCE

From a practical perspective, it is important to evaluate the impact of missing speed data at both the upstream and downstream stations on the performance of the ANN model that was

Table 7. Impact of input clamp (input set to zero) on model performance

| ANN model's input | | | | | | Incident detection performance Decision threshold = 0.5, Persistence test = 0 | | | Validation-test data set | |
|---|---|---|---|---|---|---|---|---|---|---|
| Upstream | | | Downstream | | | | | | | |
| Speed | Flow | Occ | Speed | Flow | Occ | DR (%) | FAR (%) | MTTD (s) | PEC Area | % decrease in PEC area |
| √ | √ | √ | √ | √ | √ | 95.0 | 2.1 | 135 | 9963[a] | — |
| √ | √ | √ | √ | √ | × | 92.5 | 2.2 | 138 | 9963 | 0.00 |
| √ | × | √ | √ | √ | √ | 100.0 | 7.3 | 94 | 9922 | 0.41 |
| √ | √ | × | √ | √ | √ | 92.5 | 1.5 | 158 | 9814 | 1.49 |
| √ | √ | √ | √ | × | √ | 92.5 | 2.1 | 140 | 9806 | 1.58 |
| √ | √ | √ | × | √ | √ | 62.5 | 1.1 | 142 | 9012 | 9.54 |
| × | √ | √ | × | √ | √ | 42.5 | 8.45 | 117 | 7527 | 24.50 |
| × | √ | √ | √ | √ | √ | 100.0 | 91.6 | 24 | 6722 | 32.50 |

√;Free input. ×;clamped input.
[a]Represents the initial model with all inputs free.

trained using these speeds. This analysis can provide an insight into the performance of the trained ANN model should it be implemented on a facility that only uses single loop detectors. In order to investigate this, the upstream and downstream speeds were clamped to zero values and the procedure repeated. The results of this evaluation are also shown in Table 7. When both the upstream and downstream speeds are clamped (representing a situation where the model is implemented on a facility that only uses single loop detectors), the overall performance of the ANN model based on the Performance Envelope Curve Area deteriorates by about 25%.

It should be pointed out here that these results are for a model that was originally trained with speed data and was then tested on a facility where the speed data was not available. Table 7 clearly indicates that failure to provide speeds at the upstream station of a section can result in a significant deterioration of incident detection performance. In the event of detector failures at the upstream station, measures could be implemented such that speed data from the immediate upstream station are provided until the detector problem is fixed. The satisfactory incident detection results obtained in this study from modelling on longer sections (about 1070 m) of the Tullamarine Freeway (Dia, 1996) suggest that it is feasible for such a strategy to be implemented provided that detectors are typically spaced at 500 m. It is also recommended that automated procedures be implemented for identifying detector failures at the traffic control centre and consequently implementing these strategies until the detectors are fixed.

## 24. 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 reported results of the comparative performance evaluation between the ANN and ARRB/VicRoads models clearly demonstrated the substantial improvement in incident detection performance obtained by the ANN model over the ARRB/VicRoads model.

The results from the input clamping technique clearly showed that failure to provide speed data at a station can result in a significant deterioration of model performance within that section. In this event, measures could be implemented such that speed data from the immediate upstream station are provided until the detector problem is fixed. It is also recommended that automated procedures be implemented for identifying detector failures at the traffic control centre.

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. Although this study provided a preliminary assessment of the transferability potential of the ANN model based on the 15 incidents that were collected from the South Eastern Freeway, it is recommended that a larger incident data set be collected from other freeways in Melbourne and/or other cities in Australia or overseas for investigating the transferability potential of the model. There is also scope in future research efforts to compare the performance of the ANN model with other overseas models (e.g. California, McMaster or Minnesota algorithms) based on the same data set of 100 incidents used in this study.

*Acknowledgements*—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. The authors also wish to thank the anonymous referees for their valuable comments and suggestions.

## 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* **3C**(6), 371–388.

Dia, H. (1996) Artificial neural network models for automated freeway incident detection. Ph.D. dissertation, Monash University, Clayton, Victoria, Austria.

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. (1996) Impact of data quality on the performance of neural network incident detection models. *Accepted for the Third World Congress on Intelligent Transport Systems '96*, Orlando, FL, October 14–18 1996.

Dia, H., Rose, G. and Snell, A. (1996) Comparative performance of freeway automated incident detection algorithms. *Proceedings Roads 96: Joint 18th ARRB Transport Research Conference & Transit New Zealand Land Transport Symposium*, Christchurch, New Zealand, September 1996.

Hoose, N. (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.

Lindley, J. A. (1987) Urban freeway congestion: quantification of the problem and effectiveness of potential solutions. *Institute of Transportation Engineers Journal* **57**(1), 27–32.

Luk, J. Y. K. and Sin, F. Y. C. (1992) *The Calibration of Freeway Incident Detection Algorithms*. Australian Road Research Board Working Document No. WD TE 92/001. ARRB, Vermont South, Victoria, Australia.

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

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.

NeuralWare (1993) *NeuralWorks Professional II/Plus and NeuralWorks Explorer*. NeuralWare Inc., Pittsburgh, PA.

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* **1**( Melbourne, Australia), 123–140.

Snell, A., Sin, F. Y. C. and Luk, J. Y. K. (1992) Freeway incident management in Melbourne: an initial appraisal. *Proceedings of the 16th Australian Road Research Board Conference* **16**(5), 301–313.

Stephanedes, Y. J. and Liu, X. (1995) Artificial neural networks for freeway incident detection. *Transportation Research Record* **1494**, 91–97.

Wilensky, G. and Manukian, N. (1992) The projection neural network. International Joint Conference on Neural Networks. *1992* **Vol. II**, 358–367.