

1998

AN APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN FREEWAY INCIDENT DETECTION

Sujeeva A. Weerasuriya
University of South Florida

Follow this and additional works at: <https://scholarcommons.usf.edu/etd>



Part of the [American Studies Commons](#)

Scholar Commons Citation

Weerasuriya, Sujeeva A., "AN APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN FREEWAY INCIDENT DETECTION" (1998). *Graduate Theses and Dissertations*.
<https://scholarcommons.usf.edu/etd/2949>

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Graduate School
University of South Florida
Tampa, Florida

CERTIFICATE OF APPROVAL

Ph.D. Dissertation

This is to certify that the Ph.D. Dissertation of

SUJEEVA A. WEERASURIYA

with a major in Civil Engineering has been approved by
the Examining Committee on November 2, 1998
as satisfactory for the dissertation requirement
for the Doctor of Philosophy in Civil Engineering degree

Examining Committee:

Co-Major Professor: Jian John Lu, Ph.D., P.E.

Co-Major Professor: Ram M. Pendyala, Ph.D.

Member: William C. Carpenter, Ph.D., P.E.

Member: Rafael A. Perez, Ph.D.

Member: A. N. V. Rao, Ph.D.

Member: Steven E. Polzin, Ph.D., P.E.

AN APPLICATION OF
ARTIFICIAL NEURAL NETWORKS IN
FREEWAY INCIDENT DETECTION

by

SUJEEVA A. WEERASURIYA

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Civil Engineering
Department of Civil and Environmental Engineering
College of Engineering
University of South Florida

December 1998

Co-Major Professor: Jian John Lu, Ph.D., P.E.
Co-Major Professor: Ram M. Pendyala, Ph.D.

DEDICATION

To my loving parents, Punchisingo and Seelawathie, and my siblings, Thusantha, Deepa, and Amila, who believed in me. To my wife, Nilmini, for encouragement and unfailing love during times of sorrow and times of joy.

ACKNOWLEDGMENT

My deepest gratitude goes to my co-major advisors, Drs. Jian John Lu and Ram M. Pendyala, for their help and guidance throughout this research. I would also like to thank Drs. William C. Carpenter, Rafael A. Perez, A. N. V. Rao, and Steven E. Polzin for their valuable opinions, constructive criticism on this research and service on my graduate advisory committee. My special thanks goes to Dr. Edward A. Mierzejewski for his continuous encouragement throughout my research and serving as the chairperson of the final defense committee.

I would also like to extend my greatest appreciation to my supervisor at Center for Urban Transportation Research (CUTR), Mr. Michael Pietrzyk, who inspired me during my tenure at CUTR and encouraged me in my research work. My sincere appreciation also goes to CUTR and the team who helped me learn the real world working environment within a university system.

I thank my wife, Nilmini, for her continuous support, encouragement, and the help in preparing sketches for the dissertation. I would like to acknowledge Mr. Ahmed Aburahma, Division Manager, Manatee County, Florida, for his joint effort to get more data for my research and Dr. Karl Petty at the University of California, Berkeley for providing the data for the study. My appreciation also extends to Mr. John Cheney of Florida Department of Transportation, District 5, for his constructive comments during my research.

TABLE OF CONTENTS

	Page
LIST OF TABLES	ii
LIST OF FIGURES	iv
ABSTRACT	vi
CHAPTER 1	
INTRODUCTION	1
1.1 Problem Definition	1
1.2 Research Objectives	11
1.3 Organization of Document	12
CHAPTER 2	
BACKGROUND	14
2.1 Definition of an Incident	14
2.2 Traffic Flow under Incident Conditions	15
2.3 Incident Management Systems	18
2.4 Automatic Incident Detection Systems	23
2.5 Previous Models	25
2.5.1 Algorithm-Based Models	26
2.5.1.1 California Algorithm	26
2.5.1.2 Minnesota Algorithm	28
2.5.2 Artificial Neural Network-Based Models	30
CHAPTER 3	
DESCRIPTION OF ARTIFICIAL NEURAL NETWORKS	37
3.1 Behavior of a Single Neuron	38
3.2 Behavior of a Network of Neurons	42
3.2.1 Feed-Forward	43
3.2.2 Back-propagation	44
CHAPTER 4	
FREEWAY DATA	48
4.1 Data Source	48
4.2 Data Description	50

4.3	Data Verification	51
4.4	Data Preparation	52
CHAPTER 5		
	MODEL DEVELOPMENT	56
5.1	Model Inputs	56
	5.1.1 Traffic Flow Variables	57
	5.1.2 Geometric Variables	57
5.2	Model Output	58
5.3	ANN Software	59
5.4	Network Architectures	60
	5.4.1 Feed Forward Networks	62
	5.4.2 Recurrent Networks	73
CHAPTER 6		
	PERFORMANCE MEASURES AND TRAINING RESULTS	83
6.1	Performance Evaluators of AID Models	83
	6.1.1 Conventional Detection Rate	84
	6.1.2 Modified Detection Rate	85
	6.1.3 Other Performance Measures	87
	6.1.4 Persistence Checks	88
6.2	ANN Model Training Results	89
6.3	Calibration of Conventional Algorithms	94
	6.3.1 California Algorithm	99
	6.3.2 Minnesota Algorithm	99
	6.3.3 Discussion	100
6.4	Summary	101
CHAPTER 7		
	RESULTS OF MODEL EVALUATION	104
7.1	Evaluation Data	104
7.2	Neural Network Models	105
7.3	Comparative Evaluation of ANN with Conventional Algorithms	108
7.4	Comparison of Conventional and Modified Detection Rates	111
CHAPTER 8		
	SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	116
8.1	Summary	116
8.2	Conclusions	119
8.3	Recommendations	120
REFERENCES		122
VITA		End Page

LIST OF TABLES

	Page
1. Data Transformation for Geometric Variables	55
2. Model Input Variables and Nomenclature for Feed Forward Networks	64
3. Model Input Variables and Nomenclature for Recurrent Networks	74
5. Training Results for 3-F model with 12 Hidden Neurons	90
6. Selected Single Hidden Layer Models During Training	92
7. Selected Double Hidden Layer Models During Training	93
8. Calibration Results for California Algorithm No. 8	100
9. Calibration Results for Minnesota Algorithm	101
10. Evaluation Performance of Single Hidden Layer Feed Forward ANNs	106
11. Evaluation Performance of Single Hidden Layer Recurrent ANNs	107
12. Evaluation Performance of Double Hidden Layer Feed Forward ANNs	109
13. Evaluation Performance of Double Hidden Layer Recurrent ANNs	110
14. Evaluation Results for ANN Model and the Conventional Algorithms	112
15. Comparative Evaluation of DRIP and Conventional DR	115

LIST OF FIGURES

	Page
1. Queuing Diagram for Incident Occurrence in Freeway System	17
2. California Algorithm No. 8 with Five-minute Roll-wave Suppression Logic	27
3. Structure of Minnesota Algorithm	31
4. Model of a Nonlinear Neuron	40
5. Behavior of Piecewise-Linear Function	41
6. Behavior of Sigmoid Function	42
7. Single-Hidden-Layer Feed-Forward Neural Network	43
8. Schematic Diagram of the I-880 Study Area	49
9. Distribution of Afternoon NB Volume on I-880 on February 16, 1993	52
10. Distribution of Afternoon NB Speed of I-880 on February 16, 1993	53
11. Distribution of Afternoon NB Occupancy on I-880 on February 16, 1993	54
12. Model 1-F :- Feed Forward ANN with Single-Hidden Layer and 18 Input Variables	65
13. Model 2-F :- Feed Forward ANN with Single-Hidden Layer and 24 Input Variables	66
14. Model 3-F :- Feed Forward ANN with Single-Hidden Layer and 22 Input Variables	67
15. Model 4-F :- Feed Forward ANN with Single-Hidden Layer and 28 Input Variables	68
16. Model 5-F :- Feed Forward ANN with Two-Hidden Layers and 18 Input Variables	69
17. Model 6-F :- Feed Forward ANN with Two-Hidden Layers and 24 Input Variables	70
18. Model 7-F :- Feed Forward ANN with Two-Hidden Layers and 22 Input Variables	71
19. Model 8-F :- Feed Forward ANN with Two-Hidden Layers and 28 Input Variables	72
20. Model 1-R :- Recurrent ANN with Single-Hidden Layer and 18 Input Variables	75
21. Model 2-R :- Recurrent ANN with Single-Hidden Layer and 24 Input Variables	76
22. Model 3-R :- Recurrent ANN with Single-Hidden Layer and 22 Input Variables	77

23. Model 4-R :- Recurrent ANN with Single-Hidden Layer and 28 Input Variables . .	78
24. Model 5-R :- Recurrent ANN with Two-Hidden Layers and 18 Input Variables . . .	79
25. Model 6-R :- Recurrent ANN with Double-Hidden Layer and 24 Input Variables .	80
26. Model 7-R :- Recurrent ANN with Double-Hidden Layer and 22 Input Variables .	81
27. Model 8-R :- Recurrent ANN with Double-Hidden Layer and 28 Input Variables .	82
28. Performance Envelope for Single Hidden Layer Feed Forward Networks	95
29. Performance Envelope for Single Hidden Layer Recurrent Networks	96
30. Performance Envelope for Double Hidden Layer Feed Forward Networks	97
31. Performance Envelope for Double Hidden Layer Recurrent Networks	98
32. Performance Envelope for the California and Minnesota Algorithms.	102
33. Performance Envelope for ANN Models and Conventional Algorithms	113

AN APPLICATION OF
ARTIFICIAL NEURAL NETWORKS IN
FREEWAY INCIDENT DETECTION

by

SUJEEVA A. WEERASURIYA

An Abstract

Of a dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Civil Engineering
Department of Civil and Environmental Engineering
College of Engineering
University of South Florida

December 1998

Co-Major Professor: Jian John Lu, Ph.D., P.E.
Co-Major Professor: Ram M. Pendyala, Ph.D.

Non-recurring congestion caused by incidents is a major source of traffic delay in freeway systems. With the objective of reducing these traffic delays, traffic operation managers are focusing on detecting incident conditions and dispatching emergency management teams to the scene quickly. During the past few decades, a few number of conventional algorithms and artificial neural network models were proposed to automate the process of detecting incident conditions on freeways. These algorithms and models, known as automatic incident detection methods (AIDM), have experienced a varying degree of detection capability.

Of these AIDMs, artificial neural network-based approaches have illustrated better detection performance than the conventional approaches such as filtering techniques, decision tree method, and catastrophe theory. So far, a few neural network model structures have been tested to detect freeway incidents. Since the freeway incidents directly affect the freeway traffic flow, majority of these models have used only traffic flow variables as model inputs. However, changes in traffic flow may also be stimulated by the other features (e.g., freeway geometry) to a greater extent. Many AIDMs have also used a conventional detection rate as a performance measure to assess the detection capability. Yet the principle function of incident detection model, which is to identify whether an incident condition exists for a given traffic pattern, is not measured in its entirety by this conventional measure.

In this study, new input feature sets, including freeway geometry information, were proposed for freeway incident detection. Sixteen different artificial neural network (ANN) models based on feed forward and recurrent architectures with a variety of input feature sets were developed. ANN models with single and double hidden layers were investigated for

incident detection performance. A modified form of a conventional detection rate was introduced to capture full capability of AIDMs in detecting incident patterns in the freeway traffic flow. Results of this study suggest that double hidden layer networks are better than single hidden layer networks. The study has demonstrated the potential of ANNs to improve the reliability using double layer networks when freeway geometric information is included in the model.

Key Words: *Freeway, Incident, Detection, Artificial Neural Networks*

Abstract Approved: _____
Co-Major Professor: Jian John Lu, Ph.D., P.E.
Assistant Professor, Department of Civil
and Environmental Engineering

Date Approved: _____

Abstract Approved: _____
Co-Major Professor: Ram M. Pendyala, Ph.D.
Assistant Professor, Department of Civil
and Environmental Engineering

Date Approved: _____

CHAPTER 1

INTRODUCTION

1.1 Problem Definition

Traffic delay is a nightmare that every driver traveling on freeway systems may have to experience. These traffic delays are an increasing burden on safety, economy, productivity, and quality of life in countries across the world. Freeway traffic delay is a direct consequence of congestion in the freeway system. In the U.S. alone, congestion has grown throughout the 1980s and early 1990s, leading to a reduction in overall freeway performance. In 1981, about 25% of urban interstate freeway-miles were classified as highly congested; by 1993, that proportion had nearly doubled to 45% (ATA 1997). Though freeway congestion is a nationwide problem in scope, it is heavily concentrated in the largest urban areas such as Los Angeles, New York, San Francisco, Washington, D.C., Miami, Boston, Chicago etc. This congestion is spreading to many smaller urban communities and even to some rural interstate corridors.

Freeway congestion can be caused by either recurrent events or incidents (non-recurrent events). The recurrent congestion, which can be predicted, is caused by high traffic volumes using the freeway. Usually, high traffic volumes occur during morning and afternoon commute hours, special event times, emergency evacuations, etc. Incident congestion,

unpredictable in nature, is caused by events such as accidents, spilled loads, fallen debris, stalled vehicles, etc. The congestion induced by an incident depends on the duration of the incident, the number of closed lanes, and the traffic volume at the time.

The existing freeway systems are designed to operate acceptably within certain ranges of traffic demand. When an unpredictable event occurs, traffic backs up and creates a sudden and temporary decrease in the capacity of a particular section of the facility. As the demand volume exceeds the temporarily reduced capacity, the excess demand volume creates queues (traffic congestion), increases traffic delay, and develops a potential environment for more accidents (Hall 1993). Whether the congestion is caused by recurrent or non-recurrent events, traffic congestion leaves a noticeable footprint in the traffic flow. These footprints are noticeable in traffic flow variables such as volume, speed, and occupancy (expressed as the percent of time the vehicles occupied a particular location on a lane). Recurrent and incident (non-recurrent) congestions create subtle yet distinctive signatures in the traffic flow variables. At the onset of incident congestion, the capacity decrease is sudden and so is the change in traffic flow. On the other hand, when recurrent congestion occurs, such as under bottleneck conditions, the capacity decrease is gradual and so is the change in traffic flow.

Due to heavy traffic volume and limited access, incident induced congestion in a freeway system is more severe and creates higher overall traffic delays than that of arterial and other minor roadways. In 1984, almost 61% (766.8 million vehicle-hours) of urban freeway delays were incident related, and this delay is expected to increase up to 70% (4,857.5 million vehicle-hours) by year 2005 (Lindley 1987). Furthermore, the user cost of incident delay is expected to increase from 5.6 to 35.8 billion dollars during this 22-year period. These

projections are even without accounting for the effect of emotional distress of drivers and passengers that experience the traffic delays.

Freeway incident management systems provide a solution to the problem of incident congestion through coordinated activities designed to reduce the impact of freeway incidents on traffic flow. These incident management systems can provide several short-term and long-term benefits; minimize loss of time and productivity, reduction in fuel consumption, lower vehicle operating costs, minimize degradation of air quality, reduction in the cost of delivering goods, increase in highway safety, and increase in a region's economic competitiveness and quality of life. Accordingly, freeway incident management has become a top priority in Traffic Management Centers (TMC) across the country.

It is a well-known fact that incidents can be better managed if they are identified quickly. Therefore, identification of freeway incidents has become a vital characteristic of a successful freeway traffic management system. Incident detection using manual methods, such as probe vehicles, closed circuit television (CCTV) cameras, or special police patrol, can be quite costly, time consuming, and labor intensive. To contend with these problems, recent studies have focused on automating the incident detection process and by that reducing the time and the labor required. In general, these automatic incident detection (AID) systems utilize traffic data gathered from inductive loops buried under pavement. Signals from pairs of these loops are used to estimate traffic flow variables such as volume, speed, and occupancy. Historically, performance of AID systems has been measured by detection rate (percentage of incidents detected) and false alarm rates (percentage of time that the AID system detects nonexistent or false incident conditions).

Conventional AID systems developed over the past years have used different techniques including decision trees for pattern recognition (Payne et al. 1978), Kalman filters (Willsky et al. 1980), time series analysis (Ahmed and Cook 1982), catastrophe theory (Persaud and Hall 1989), and low pass filters (Stephanedes and Chassiakos 1993). The decision tree method (also known as California algorithm) introduced by Payne et al. (1978) is based on discontinuity in occupancy values between two adjacent loop detector stations. The algorithm has been developed in several different versions over the years. The California algorithm No. 8 (Payne et al. 1978) has a five minute roll-wave suppression logic that aims to reduce false incident detections due to shock waves approaching from downstream. It consists of five threshold values to determine whether traffic conditions have changed. These threshold values are predetermined with traffic data for which the traffic status (i.e., incident free, incident present, compression waves downstream, etc.) at each time interval is known. The algorithm can be first applied from a known traffic status at each freeway section. If a significant discontinuity in occupancy is detected, the change in the traffic status initiates an incident alarm. Several published literature have revealed that incident detection accuracy of the California algorithm is very low (Payne et al. 1978; Arceneaux et al. 1989; Cheu and Ritchie 1993). Moreover, the algorithm calibration is very lengthy and involves testing as many as several thousands of threshold value combinations.

Willsky et al. (1980) used dynamic models to estimate traffic flow parameters that may relate to incidents. The density and speed of freeway links were modeled in non-linear differential equations. Segment (link) capacity was used as an incident detection parameter. Using a Kalman filter, segment density and speed were estimated based on volume and

occupancy variables measured from loop detectors at both upstream and downstream ends of each segment. An incident is detected by a reduction in link capacity which is calculated from the speed equation.

Ahmed (1983) used a time series model to detect freeway incident conditions. Parameters of the time series model were estimated from an autocorrelation of single detector occupancies. The parameters are updated online. With a 95% confidence interval, occupancy at the single detector station is predicted for the next detection interval. If the actual occupancy falls outside the 95% confidence interval, an incident alarm is issued.

Persaud and Hall (1989) introduced catastrophe theory into freeway incident detection. This incident detection method, also known as the McMaster algorithm, uses volume and occupancy data from the fast lane of a loop detector station. The McMaster algorithm consists of location specific volume-occupancy plots that classify traffic data points into one of several areas. Incident congestion is detected if upstream and downstream traffic data points fall into certain areas in the location specific plots. An incident alarm is declared if incident congestion persisted for three consecutive intervals.

Low pass filtering algorithm (commonly known as the Minnesota algorithm) was used in incident detection by Stephanedes and Chassiakos (1993). Traffic occupancy at two loop stations were used with a short-term time averaging (or low pass filtering) to reduce adverse effects of short-term traffic fluctuations and impulsive noise in the detection process. The Minnesota algorithm traces the filtered spatial occupancy difference between adjacent detector stations through time, and issues an incident alarm when this difference changes

significantly in a short time. It uses one threshold value to detect traffic congestion and another threshold value to detect the onset of incident congestion.

These conventional algorithms possess several shortcomings. For example, in order to use the California algorithm No. 8 at a location, accurate traffic status should be known at the beginning. Despite the fact that all three traffic variables (volume, speed, and occupancy) are affected by incidents, the California and Minnesota algorithms use only occupancy related variables as inputs and the McMaster algorithm uses only volume and occupancy variables as inputs. Another potential disadvantage of the McMaster algorithm is that it only uses traffic variables from the fast lane. Modeling equations, such as those used by Willsky et al. (1980), may not always be satisfactory in replicating traffic flow in actual situations (Cheu and Ritchie 1994). Most conventional algorithms (except the Minnesota algorithm) can only detect 40-70% incidents occurring at low false alarm rates (0.01-0.90%). Simply put, despite development of these several algorithms, the conventional AID models developed so far possess low reliability levels (Stephanedes and Liu 1995). Consequently, traffic engineers are reluctant to depend on such low reliable models for freeway operations. Therefore, reliability and quick identification of incidents have become basic ingredients of a successful incident detection model.

Artificial neural networks (ANNs) are good at trend prediction, pattern recognition, modeling, control, signal filtering, noise reduction, image analysis, classification, evaluation etc. (Lawrence 1993). In fact, new uses for artificial neural networks are being found in a variety of research fields every day. However, every application of ANN shares the ability to make associations between known inputs and outputs by observing many examples or

patterns. Researchers in transportation engineering area have used artificial neural networks since late 80's. The application attempts includes simulating driver behavior (Yang et al. 1992; Dougherty and Joint 1992), estimating travel time (Nelson and Palacharla 1993; Hua and Faghri 1994), classifying pavement distress from video images (Hua and Faghri 1993), detecting vehicles from images (Mead et al. 1994; Belgaroui and Blosseville 1993), and detecting incidents.

During this decade a smaller number of ANN models have been developed to automate the freeway incident detection process. Wiederholt et al. (1993) developed two single-station feed forward ANN models to detect incidents. Using a dynamic traffic simulation model, traffic flow data for Highway 401 in Toronto, Canada were simulated. The simulation was based on observed traffic patterns from the Toronto section of Highway 401 between 2pm and 3pm on June 8, 1992. The simulated traffic flow data (volume, speed, and occupancy) averaged over all lanes at 20 second intervals during 61-hours were used as model inputs. Both models had a single-hidden layer and a single neuron in the output layer. The first model had link (section) number and traffic volume, speed, and occupancy variables measured at current time intervals from single detector station as inputs. The second model had the link number and the three traffic variables measured at current time intervals of up to two previous time intervals from a single detector station as model inputs. Performance of both models revealed an incident detection rate of 97% and a false alarm rate of 3%.

Cheu and Ritchie (1994) developed a two-station feed forward ANN model using simulated traffic flow variables. The simulated traffic volume and occupancy values averaged over all lanes during 30 second intervals were used in the model development. Only single

hidden layer networks were tested by the authors. The best model had a detection rate of 80% with a false alarm rate of 1.46% during an evaluation phase which also used simulated traffic data that was different from training data.

Hsiao (1994) developed a feed forward ANN model combined with fuzzy logic to detect freeway incidents. Traffic data collected at 14 loop detector stations on Highway 401 in Toronto, Canada in Spring 1993 was used in model training and evaluation. The traffic volume, speed, and occupancy variables were collected at single loop detector stations and averaged during 20 second intervals. These numeric input values were then translated into three linguistic categories ('low', 'medium', and 'high') using fuzzy logic and used as ANN model inputs. The author used the traffic data at current time intervals in the model. The best model during training exhibited a detection rate of 76.19% with a false alarm rate of 8.05%.

Stephanedes and Liu (1995) developed a feed forward ANN model based on two-station traffic data. The traffic data were collected at 14 loop detector stations during a 72-day period in 1989 in westbound I-35 in Minneapolis, Minnesota. Model inputs consisted of traffic volume and occupancy data at 10 consecutive 30 second intervals at adjacent loop stations. Randomly selected 425 traffic patterns (including both 'incident free' and 'incident present' condition) were used in the training. The authors tested the model using traffic data, part of which included the whole training data, collected during the 72-day period. The test results indicated a detection rate of 70% to 80% with a false alarm rate of 0.12% to 0.26%.

Abdulhai and Ritchie (1995) used a modified probabilistic neural network (PNN) approach to detect freeway incidents. Using the same simulated data set as in Cheu and Ritchie's (1993) study, the authors gained a detection rate of 100% with a false alarm rate of

4.77%. In order to get good performance when the model was applied to a new location, the model had to learn new traffic patterns during the implementation process.

The above mentioned ANN studies show an overall improvement in detection performance over conventional algorithms such as the California, Mc Master, and Minnesota algorithms. However, these studies still have some shortcomings. For example, the one hour traffic pattern used in Wiederholt et al. (1993) study may not be representative enough to simulate the variety of traffic conditions that exist in a freeway system. Many of ANN-based models described above (except the model developed by Stephanedes and Liu 1995) were trained based on simulated traffic data. Cheu and Ritchie (1994) stated that experience has shown that the modeling equations may not always satisfactorily replicate traffic flow in actual situations leaving some concern over the estimated traffic flow variables using dynamic modeling. Therefore, the performance of models developed using simulated traffic data may contain some sampling bias. ANN models developed by Cheu and Ritchie (1994), Stephanedes and Liu (1995), and Abdulhai and Ritchie (1995) used only two of the three traffic variables as model inputs despite all three traffic variables are affected by incidents. Following a historical trend, many ANN models still use a conventional detection rate that may not truly reflect the capability of detecting individual incident patterns as described later in this section. The majority of the ANN models illustrate that there is still some room for improvement in performance. The incident detection models developed based on PNN have its own disadvantages. Since a separate hidden layer neuron is required for separate input patterns of each example in the training data set, the PNN models can become quite large. Once trained, these PNN models take more time to run than back propagation networks

(Lawrence 1993). Furthermore, in problems such as incident detection with a large amount of complex data, back propagation is more accurate (Specht and Shapiro 1991).

When an incident occurs on a freeway section, the traffic speed downstream of the section increases while the traffic speed upstream of the section decreases. In contrast to the change in speed, the downstream occupancy decreases while the upstream occupancy increases due to the congestion created by the incident. Generally speaking, the ANN models are trained to identify such changes from one geometric location to the other, learn hidden relationships among traffic flow variables, and detect the onset and continuation of incidents. For freeways, vehicles enter through entrance ramps and leave through exit ramps that change the traffic flow. Changes in traffic flow from one end to the other end of a freeway segment are also apparent when the freeway segment expands (adding one or more lanes) or contracts (merging one or more lanes). Even under normal traffic conditions, then, change in laneage and presence of entrance and exit ramps (or geometric variables) affect traffic flow characteristics. In general, the ANN models are not station specific and a trained ANN model is applied to freeway segments with variety of geometric features. It can be easily conceived, therefore, the ANN models should be able to differentiate the changes in traffic flow between upstream and downstream sections from those caused by changes in geometric features and those caused by incidents alone. Should the ANN model has prior knowledge of the geometric changes, it can recognize these recurrent congestion patterns from incident congestion patterns and may reduce the false alarms. Therefore, these geometric variables should be considered when developing incident detection models.

Previous researchers have used a conventional ‘detection rate’ that assesses capability of AID models in detecting the number of incidents from an incident database. Traffic data at a predetermined interval (20 sec, 30 sec, or 60 sec, etc.) were used as ANN model inputs to detect the presence of incident conditions. Generally speaking, ANN models in the field are only used to classify traffic flow patterns at the predetermined interval into two states (or conditions): ‘incident free’, or ‘incident present’ states. Depending on the severity, an incident may contain tens, hundreds, or even thousands of ‘incident present’ traffic flow patterns. A properly trained/calibrated AID model should be able to identify the majority of the traffic flow patterns as ‘incident present’ patterns, not just at the onset of the incident but during the entire incident period as well. Therefore, when assessing the incident detection capability of AID models, a detection performance measure that assesses model output for detecting every possible ‘incident present’ pattern should be utilized to better reflect the field implementation conditions.

1.2 Research Objectives

The main objective of this research effort was to conduct an extensive study to search ways to develop an AID using ANN that exhibits increased reliability and to develop improved performance measures. Four different approaches were proposed to achieve the objective. First, freeway incidents affect all three measurable traffic flow variables (volume, speed, and occupancy). Therefore, all three traffic variables will be used in the developing AID models to detect incidents. Second, not just incident conditions but also other variables such as presence of entrance and exit ramps, lane expansion and merges can induce changes

in traffic flow variables between two locations on a freeway. To study the potential improvements of these new inputs, freeway geometric variables were also included in the model development process. Third, beyond models based on feed forward ANN architectures, new models based on recurrent ANN architecture were developed to investigate whether the new recurrent architecture adds any significance to the detection performance. Fourth, the performance of the best ANN model developed was validated with an independent traffic data with conventional AID algorithms to illustrate the reliability of ANN models developed in this study. These ANN models were developed with rich data from I-880 in California.

1.3 Organization of Document

Contents of the dissertation are distributed into eight chapters. Following is a brief description of the chapter contents.

- Chapter 1 The problem of freeway incident detection and research objectives are summarized.
- Chapter 2 Background on freeway incident detection, freeway incident management, an introduction to ANNs, existing conventional algorithms and ANN models are discussed.
- Chapter 3 This chapter gives a brief introduction to ANNs and describes the way neurons behave in a network setting. Procedure of information processing for a feed forward ANN with backpropagation capabilities is discussed.
- Chapter 4 Describes the source of the freeway traffic flow data, incident database, analysis of the traffic data, and process of data preparation for model inputs.

- Chapter 5 Discusses the development of various ANN models for incident detection, introduces traffic flow variables and the freeway geometric variables as model inputs.
- Chapter 6 Discuss ANN model training and testing in detail. A modified performance measure is introduced.
- Chapter 7 ANN models and conventional AID methods are compared using evaluation data. The results are presented in both tabular and graphical form. Both conventional and modified performance measures are compared.
- Chapter 8 Summarizes the research findings and conclusions.

CHAPTER 2

BACKGROUND

2.1 Definition of an Incident

As various research fields have grown and branched away specializing in their respective areas of interest, traffic engineering has been branching away into ‘traffic operations’ and ‘traffic safety’ arenas. Traffic operation personnel may view incident-provoked bottlenecks differently than traffic safety personnel. Several different descriptions of incidents exist in the literature (Stephanedes and Liu 1995; Abdulhai and Ritchie 1995 and 1997). Depending on the perspective of who interprets, differences in the definitions arise. From traffic safety viewpoint, an incident is any non-recurring event that could pose a hazard to motorists (Abdulhai and Ritchie 1995 and 1997). The events such as disabled vehicles, spilled loads, accidents, and temporary maintenance and construction activities can be included in this category. From a traffic operations viewpoint, the interpretation would be similar but limited to the events causing ‘unexpected’ congestion shockwaves, queues, and delays. Usually the maintenance or the construction activities on freeways are planned ahead of time and are even informed to the travelers through message signs well ahead (days and even weeks ahead) of such an activity. Therefore, these activities are not ‘unexpected’ and needs to be excluded in the definition (Abdulhai and Ritchie 1995 and 1997). As the number

of lanes in a freeway section increases, a minor incident on a lane may have little or no effect on the traffic. These minor incidents that do not cause congestion and traffic delays are also excluded from the definition as well. AID is primarily a traffic management tool. Therefore, defining an incident as “any *unexpected non-recurring* event that disrupts the normal traffic behavior, producing congestion shockwaves, queues, and traffic delays” seems logical in the context of freeway traffic management system.

2.2 Traffic Flow under Incident Conditions

The most severe congestion problems occur at freeway bottlenecks which can be generally defined as a portion of the freeway with lower capacity than the incoming section of the freeway. This reduction in capacity can originate from a variety of sources such as a decrease in the number of through traffic lanes, and reduced shoulder widths, and presence of temporal traffic obstruction. In general, two classes of traffic bottlenecks can be identified; recurring bottlenecks and incident provoked bottlenecks. Recurring bottlenecks occur where the freeway itself limits capacity by, for example, a physical reduction in the number of lanes. Such bottlenecks result from typical recurring traffic flows that exceed the restrictive vehicular capacity of the bottleneck area. In contrast, incident-provoked bottlenecks occur as a result of vehicle breakdown, spilled loads, or accidents that effectively reduce freeway capacity by restricting the through movement of traffic. Since incident-provoked bottlenecks are unexpected and temporary in nature, they have features that distinguish them from recurring bottlenecks. For example, an accident may affect traffic flow to come to a sudden

stop or slow down immediately while recurrent bottleneck may slow down traffic flow at much reduced rate.

The events that occur during an incident can be pictured in traffic operational viewpoint through a queuing diagram as shown in Figure 1. Suppose an incident occurs at a freeway section at time t . If a constant arrival rate of vehicles (λ in vehicles per hour) is assumed for the study period, Figure 1(a) presents the necessary input requirements to solve the problem. Under normal conditions, service rate (μ - in vehicles per hour) of the freeway section exceeds the arrival rate (λ). Since the normal service rate (μ) of the freeway section exceeds the arrival rate, queues would not normally form under 'incident free' conditions. However, when an incident occurs at time t , it may effectively block one or more lanes of the freeway section reducing the maximum number of vehicles that can pass through the section (or reducing the service rate). When this reduced service rate (μ_R) falls below the arrival rate, traffic queues are formed and the effect of the incident begins to spread. Let the reduced service rate lasts for t_R hours.

Figure 1(b) shows the number of cumulative vehicles arrived and departed the section as the time passed by. The arrivals are shown as a straight line passing through the origin with a positive slope equivalent to the arrival rate (λ). During the first period, the service line follows the arrival line until the incident occurs at time t . Once the incident occurs, the service rate becomes equivalent to μ_R and maintains a flatter slope until the incident is removed. If it is a severe incident that completely blocks ('zero' service rate) all the traffic moving in that direction, it makes the flatter slope to a horizontal line. When the incident is cleared, the service rate increases up to μ , and the service line gets a steeper slope.

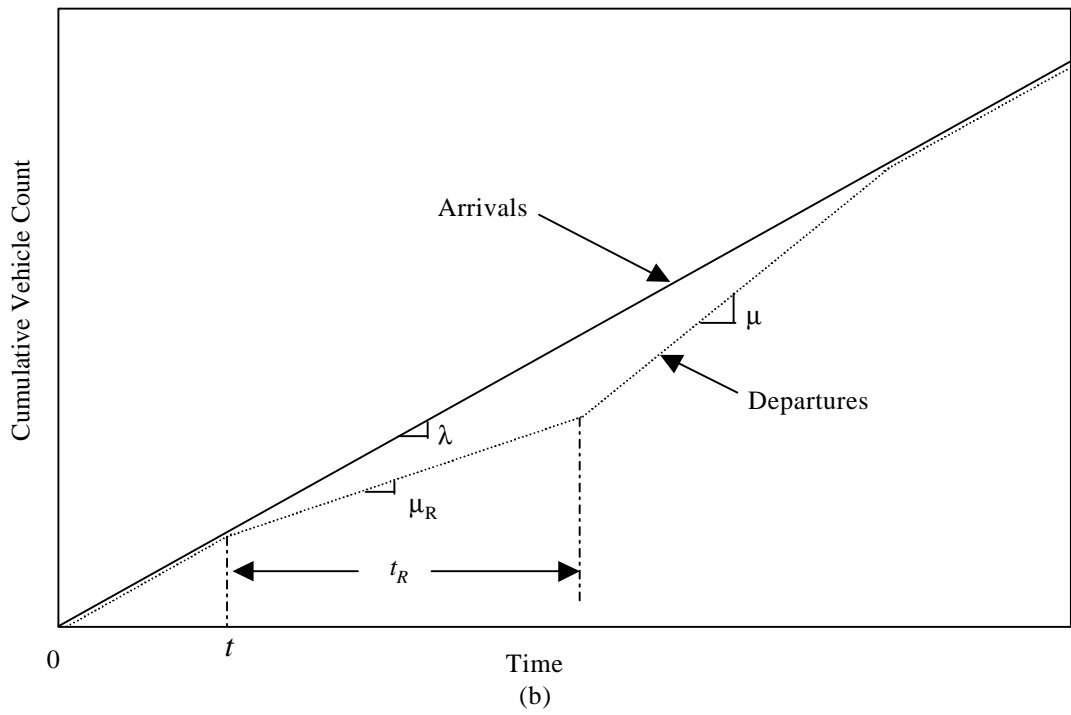
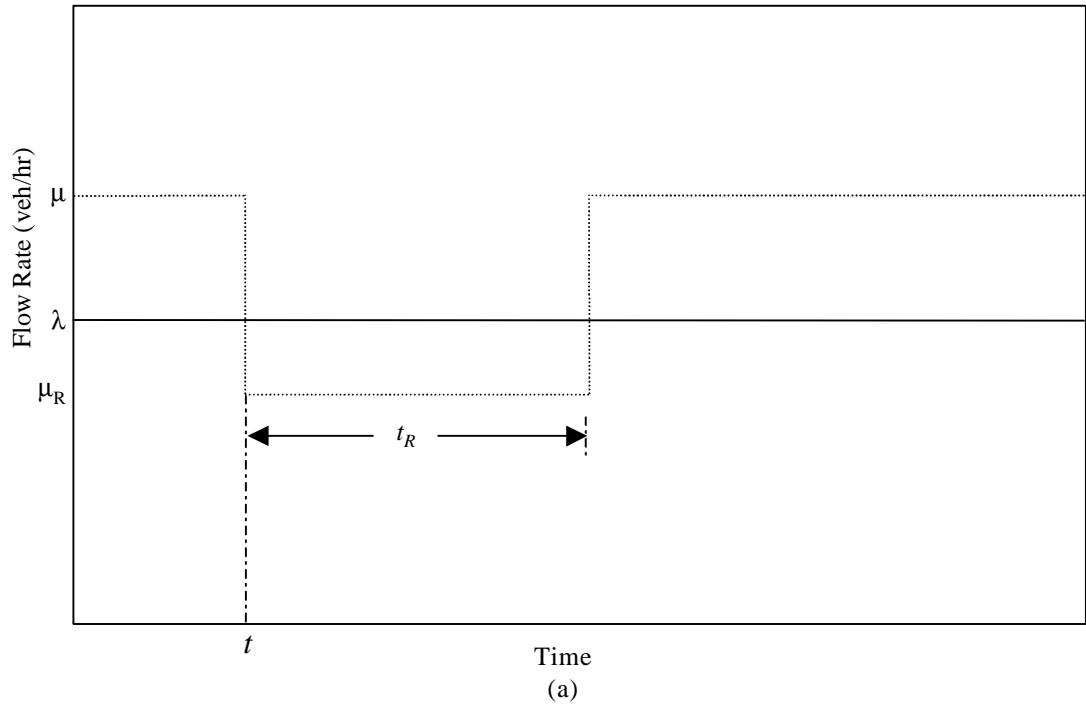


Figure 1. Queuing Diagram for Incident Occurrence in Freeway System

This continues until the arrival line and the service line intercept, at which time the service line once again overlays the arrival line.

Although traffic flow under incident conditions can be simulated using modeling software such as INTRAS (Integrated Traffic Simulation) and INTEGRATION, formulating analytical models to detect freeway incidents have not been much successful. Willsky et al. (1980) have developed dynamic models to estimate traffic flow parameters that may relate to incidents. Using non-linear differential equations, the density and speed of freeway links (or sections) have been modeled. Section capacity was expressed as a function of equilibrium density (the highest traffic density when the equilibrium traffic speed starts to decrease), equilibrium speed, and jam density. Traffic volume and occupancy measured at every 5 second intervals from loop detectors at upstream and downstream have been used to estimate the density and speed by means of a Kalman filter. An incident causes to decrease the equilibrium density of the section which decreases the capacity. This reduction in the capacity was used to detect incidents. Experience has shown that modeling equations may not always replicate actual traffic flow satisfactorily (Cheu and Ritchie 1994).

2.3 Incident Management Systems

Providing a solution to non-recurrent congestion problem through coordinated activities designed to reduce the impact of incidents on traffic is generally described as incident management. The coordinated activities include use of personnel and equipment resources from one or more emergency management agencies in mitigating the non-recurrent congestion. Primary objectives of incident management are to reduce traffic delay that

incident condition creates and to increase the traveler safety. Therefore, rapid detection, response, and clearance of incidents are necessary characteristics for an incident management system to be successful. In general, an incident management system can include several components.

(i) *Detection*: The determination that an incident of some nature has occurred at a location, for which an appropriate response can be formulated, is the detection component. Freeway incidents can be detected either by manual methods or automated methods. Manual methods are based on visual processes that include traffic surveillance CCTV cameras, cellular phone calls, roadside call boxes, routine police patrol, etc. Automated methods are mainly based on traffic flow data gathered from inductive loops buried under pavement or road side detectors and include conventional algorithms and artificial neural network-based models.

(ii) *Verification*: The determination of the approximate location with a certain accuracy (i.e., between two adjacent exit and entrance ramps or between two traffic observation stations [e.g., inductive loop stations, CCTV camera locations, etc.]) and the nature of the incident is the verification component of the incident management system.

(iii) *Emergency response*: Once an incident is detected and verified, the process of initiation and transportation of the appropriate emergency management personnel and equipment to and from the incident location falls under emergency response component. Traditionally, the communication of incident information is handled by police dispatches, but an increasing

number of local governments and states are building special multipurpose traffic management centers to coordinate variety of traffic, incident, and other emergency communications.

(iv) *Site management*: Once at the scene, the use of appropriate traffic control measures at the incident location and control of emergency resources is known as the site management. This is in fact one of the crucial tasks that must be managed with great care to ensure safety of the emergency personnel and traveling public and to reduce overall traffic delay. Depending on the severity of the casualty, one or more lanes may be temporarily closed to facilitate landing and taking off of emergency medical service helicopters.

(v) *Clearance*: The clearance is the removal of debris, spilled materials, and wreckage to restore the roadway to its normal condition until the section capacity is regained. Depending on the severity of the incident, one or more lanes may be temporarily closed to clear the roadway from debris and perhaps to repair the damages to the highway infrastructure.

(vi) *Travel advisory*: The circulation of accurate and timely information to the traveling public concerning traffic conditions at the incident location and suggested alternate routes is known as the travel advisory. An increasing number of local governments and states utilize highway advisory radio, variable message signs, Internet, and other communication methods to convey congestion information and to encourage the drivers to take alternate routes. An effective travel advisory can essentially reduce the traffic queue and travel delay to a greater extent. Many local and state governmental agencies are creating partnership with private agencies

such as TV Cable service providers, commercial radio stations, and companies with traffic information services to transmit real time traffic conditions to the traveling public.

Incident management systems are deployed in several cities across the country including Atlanta, Boston, Charlotte, Chicago, Dallas / Fort Worth, Denver, Detroit, Los Angeles, Minneapolis / St. Paul, New York City, Phoenix, Sacramento, Seattle, and Washington DC (ATA 1997). Nine of these metropolitan areas currently have automatic detectors and the remaining metropolitan areas are planing to install automatic detectors. All these metropolitan areas have variables message signs to issue advance warning to travelers about freeway traffic conditions ahead.

Incident Detection Methods

The incidents are detected and reported by several different means. Currently, the incidents are reported via routine police patrols, call boxes or motorist aid phones, and cellular phones, citizen's broadcast radios (CBR), etc. The incidents are detected via inductive loop detectors, closed circuit television cameras (CCTV's), and infrared video imaging, etc. The incident detection process can be divided into manual and automatic detection methods. Examples for the manual detection methods include routine police patrol, manual inspection of CCTV monitors, distress calls made through call boxes, cellular phones, and CBRs. The automatic detection methods, at a minimum, can utilize real time traffic data gathered from inductive loop detectors, infrared video images, and other road side detectors to detect incidents.

Traditionally, routine police patrol and public using call boxes, citizen's broadcasting radios, or cellular phone calls reported incidents. The location description from some of these eye witness counts (especially the information gathered from distress calls from affected parties, cellular callers, and citizen's broadcasting radios, etc.) may be quite imprecise (Ivan 1994; Yim and Ygnance 1995). Although these traditional sources can provide information of incidents anywhere in a freeway system, there is no guarantee that each incident will be reported promptly, or will be reported at all. Traffic Management Centers (TMC) cannot just rely on the fact that every incident will be promptly reported to them by the citizens. Therefore, developing less labor and resource intensive incident detection methods are essential for managing traffic on our freeways today.

With the invent of advanced technologies and implementation of intelligent transportation systems, more agencies are seeking solutions through of state of the art technologies and equipments to identify traffic incident conditions and the incident locations on freeways. Several researches have developed methods to automate the incident detection process through algorithms (Payne et al. 1978; Arceneaux et al. 1989; Presaud and Hall 1989; Hall et al. 1991; Stephanedes and Chassiakos 1993) and artificial neural network models (Wiederholt et al. 1993; Hsiao 1994; Cheu and Ritchie 1994; Stephanedes and Liu 1995; Abdulhai and Ritchie 1995 and 1997) that are based on the traffic data gathered from inductive loop detectors.

The ideal freeway traffic management system would include automated incident detection methods with higher reliability before resorting to manual detection via closed circuit television systems. The camera system can be used to confirm and classify incidents,

and to aid operator dispatch process, and is not intended to be a primary incident detection tool (Wiederholt et al. 1993). Operating a TMC with AID methods can be more efficient and cost effective because fewer operators will be needed to monitor larger areas of the freeway. Using a combination of AID method and CCTV observation, incidents can be classified more reliably and in greater detail. Therefore, more detailed information can be provided to the public via the radio, television, variable message signs, and other advanced traffic information system devices.

2.4 Automatic Incident Detection Systems

Freeway traffic incidents may be caused by excessive speed differences among individual vehicles, abrupt lane changing vehicles, slow-moving vehicles, spilled loads on roadway, weather conditions, road surface conditions, other geometric features of the freeway, etc. Whatever the cause may be, these incidents leave subtle footprints on traffic volume, speed, and occupancy patterns. To declare an incident condition, every AID method relies on identification or classification of these footprints from the traffic flow data. However, distinguishing these footprints left behind by traffic incidents from those left behind by common bottlenecks have been a very troublesome challenge as echoed by low performance levels of conventional AID methods. To increase overall performance, many AID methods utilized a 'threshold' concept. According to the threshold concept, if a certain parameter of the detection system, estimated from traffic data, is greater than a predefined threshold value, an incident alarm is declared. These threshold values are predetermined for best performance by trial and error using traffic data with known 'incident present' and

‘incident free’ traffic flow conditions. Detection systems may even use multiple threshold values (e.g., California algorithm no. 8) depending on the design of particular AID method.

The AID methods can obtain traffic data from sources such as inductive loop detectors, video images, or roadside to vehicle communication (e.g., Advanced Vehicle Identification [AVI]) systems. In freeways where traffic data gathering is based on inductive loops, almost every lane is equipped with pairs inductive loops that are buried under the pavement. Usually, each pair in each lane is aligned together and buried perpendicular to the traffic lanes on sections along the freeway. The spacing between adjacent loop pairs, in general, depends up on the freeway geometric features. Based on number of locations being used to gather traffic data for each traffic condition detection attempt, the AID systems can be grouped into two major categories (Busch and Fellendorf 1990) as follows:

- (i) Single station-based systems: That is the traffic data utilized in each traffic condition detection attempt is exclusively extracted at a single loop station. A few single station incident detection systems (Hall et al. 1993; Hsiao 1994; Antoniadis and Stephanedes 1996) exist in the literature.
- (ii) Multi-station based systems: That is the traffic data utilized in each traffic condition detection attempt is extracted from at least two adjacent loop stations. Majority of the detection methods found in literature utilize traffic data gathered at adjacent two loop stations.

Both groups can be subdivided further, by the type of input data the system requires, into microscopic and macroscopic systems. Microscopic systems need detailed data from individual vehicles, whereas the macroscopic systems rely on aggregated data of a group of

vehicles during a predetermined interval (for example, average speed over a 30-second period). Video image and AVI-based systems are some examples of microscopic systems, whereas many algorithm and neural network-based systems are examples of macroscopic systems. Artificial neural network model developed by Hsiao (1994) is an example of single station model, and the model developed by Cheu and Ritchie (1994), and Stephanedes and Liu (1995) are examples of multi-station models.

2.5 Previous Models

Automatic incident detection systems can be divided into two categories: algorithm-based systems and neural network-based systems. The algorithm-based systems are mainly computer programs in the form of “IF.. THEN..” logics that can distinguish different scenarios using predetermined thresholds. These ‘algorithm-based’ incident detection systems utilize linear/polynomial equations and/or combination of graphs based on empirical rules/observations to detect incidents. Neural network systems on the other hand utilize artificially created intelligence (knowledge) to detect incidents. The most critical difference between these two systems is that algorithm techniques require the empirical relationships/conditions between the input data and the conclusions (outputs). On the contrary, the neural network systems do not need to have the empirical relationships between inputs and outputs already established since it can capture those relationships directly from the data (Pietrzyk and Perez 1996).

2.5.1 Algorithm-Based Models

Several attempts have been made over the past three decades to automate freeway incident detection process. Of these attempts, algorithm-based models were the first models developed (California Algorithm in 1976) to detect freeway incidents automatically. Other algorithm-based models such as McMaster algorithm (Presaud and Hall 1989; Hall et al. 1991) and Minnesota algorithm (Stephanedes and Chassiakos 1993) were developed with the intention of improving reliability of automated incident detection. The California and Minnesota algorithms are two-station algorithms whereas McMaster algorithm is a single-station algorithm. Since the ANN models developed in this study is a two-station model, only California algorithm and Minnesota algorithms are used for comparative evaluation of ANN model results and are discussed in detail in the following sections.

2.5.1.1 California Algorithm

Payne et al. (1976) introduced the California algorithm based on discontinuity in occupancy values between two adjacent loop detector stations. It utilizes 60 second average occupancy, from adjacent loop detectors, transformed in to four different estimates to be used as algorithm inputs. The inputs are in the form of absolute, relative and temporal differences in occupancy values between two stations. California algorithm no. 8 was selected in this study since it consists of a five minute roll-wave suppression logic that aims to reduce false alarms due to shock waves approaching from the downstream. In the algorithm no. 8, traffic status is divided into 8-different states. Figure 2 shows the input features, description of traffic status, and structure of the algorithm. Starting from a known traffic status, it uses “IF

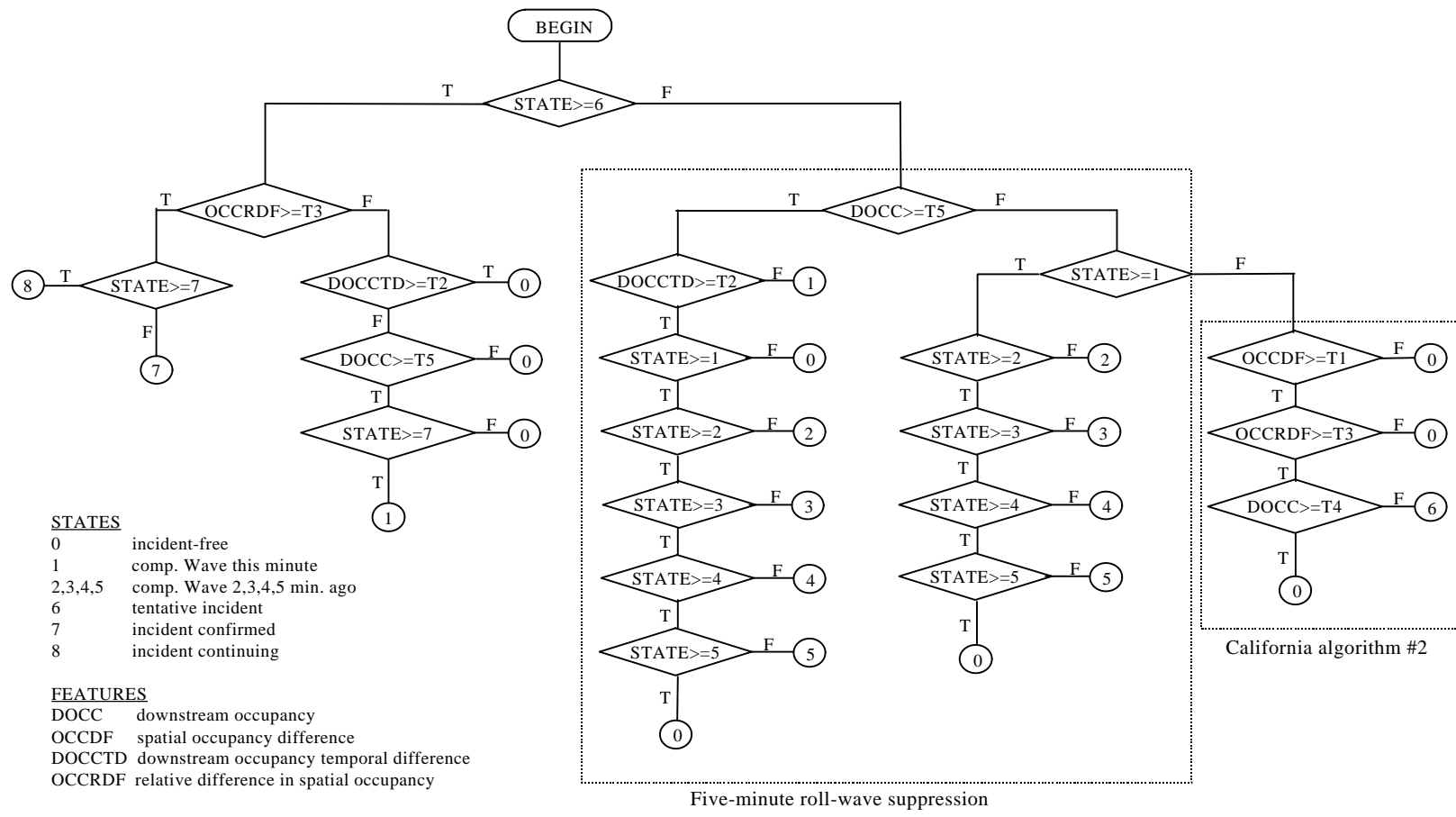


Figure 2. California Algorithm No. 8 with Five-minute Roll-wave Suppression Logic

.... THEN” logic (or decision tree) to determine the change in the traffic status according to changes in occupancy values between two stations. The algorithm has five threshold values in the logic to determine if traffic status was changed. If a significant discontinuity in occupancy is detected, change in the traffic status initiates an incident alarm.

Arceneaux et al. (1989) reported calibration accuracy of algorithm no. 8 as 50-20% detection rate and 0.125 to 0.003% false alarm rates using traffic data gathered from Los Angeles freeway system. Al-Deek et al. (1994) showed that the calibration of California algorithms 7, 8, and 10 was lengthy process that involved testing as many combinations of threshold values as possible.

2.5.1.2 Minnesota Algorithm

Stephanedes and Chassiakos (1993) introduced this algorithm which is also based on identifying discontinuities in traffic occupancy values. The Minnesota algorithm is built with a moving average filter that uses 30 second occupancy data from two adjacent loop stations. At time t , upstream station’s occupancy value, o_t^u , and downstream station’s occupancy value, o_t^d , are used to calculate the spatial occupancy difference ($o_t^u - o_t^d$). The hypothesis tested assumes that an incident occurred at time $t-5$ and compares the average spatial occupancy value changes between upstream and downstream stations before the incident and after incident.

Average spatial occupancy differences from time $t-5$ to t , and from $t-15$ to $t-6$ intervals are calculated as follows:

$$y_t^a = \frac{1}{6} \sum_{k=0}^5 (o_{t-k}^u - o_{t-k}^d) \quad (1)$$

where y_t^a is the average occupancy difference between upstream and downstream stations during the previous six consecutive time intervals.

$$y_t^b = \frac{1}{10} \sum_{k=6}^{15} (o_{t-k}^u - o_{t-k}^d) \quad (2)$$

where y_t^b is the average occupancy difference between upstream and downstream stations during the previous sixth and fifteenth consecutive time intervals.

To increase the transferability potential, these spatial variations are transformed into two different ratios as follows:

$$\text{RAT1} = \frac{y_t^a}{m_t} \quad (3)$$

and

$$\text{RAT2} = \frac{y_t^a - y_t^b}{m_t} \quad (4)$$

where RAT1 ratio is to test the discontinuity in spatial occupancy during the last six time steps, RAT2 ratio is to evaluate the change in spatial occupancy between after and before the incident occurs, and m_t is a normalization factor to account for traffic conditions prior to the incident and is defined as follows:

$$m_t = \frac{1}{10} \max \left\{ \sum_{k=6}^{15} o_{t-k}^u ; \sum_{k=6}^{15} o_{t-k}^d \right\} \quad (5)$$

The structure of this algorithm is shown in Figure 3. The two threshold values ‘Thr1’ and ‘Thr2’ used in the algorithm are predetermined from trial and error process using traffic data for known ‘incident present’ and ‘incident free’ traffic conditions. The algorithm includes two tests: congestion detection test and an incident detection test. Traffic congestion is detected if RAT1 is found to be greater than the threshold ‘Thr1.’ Once a congested traffic condition is detected, an incident is declared if RAT2 is greater than the threshold ‘Thr2.’ Continuation of the incident is detected by presence of traffic congestion after incident is detected.

2.5.2 Artificial Neural Network-Based Models

The performance of these conventional algorithms has not provided acceptable performance levels (Stephanedes and Liu 1995), as they have been hindered by excessive false alarm rates at higher detection rates. During the past two decades, as the technology became more advanced, many researchers took advantage of the powerful and efficient computer systems and exposed artificial neural networks applications in numerous speciality fields.

Artificial neural networks (ANNs) are good at trend prediction, pattern recognition, modeling, control, signal filtering, noise reduction, image analysis, classification, evaluation etc. (Lawrence 1993). In fact, new uses for artificial neural networks are being found in a

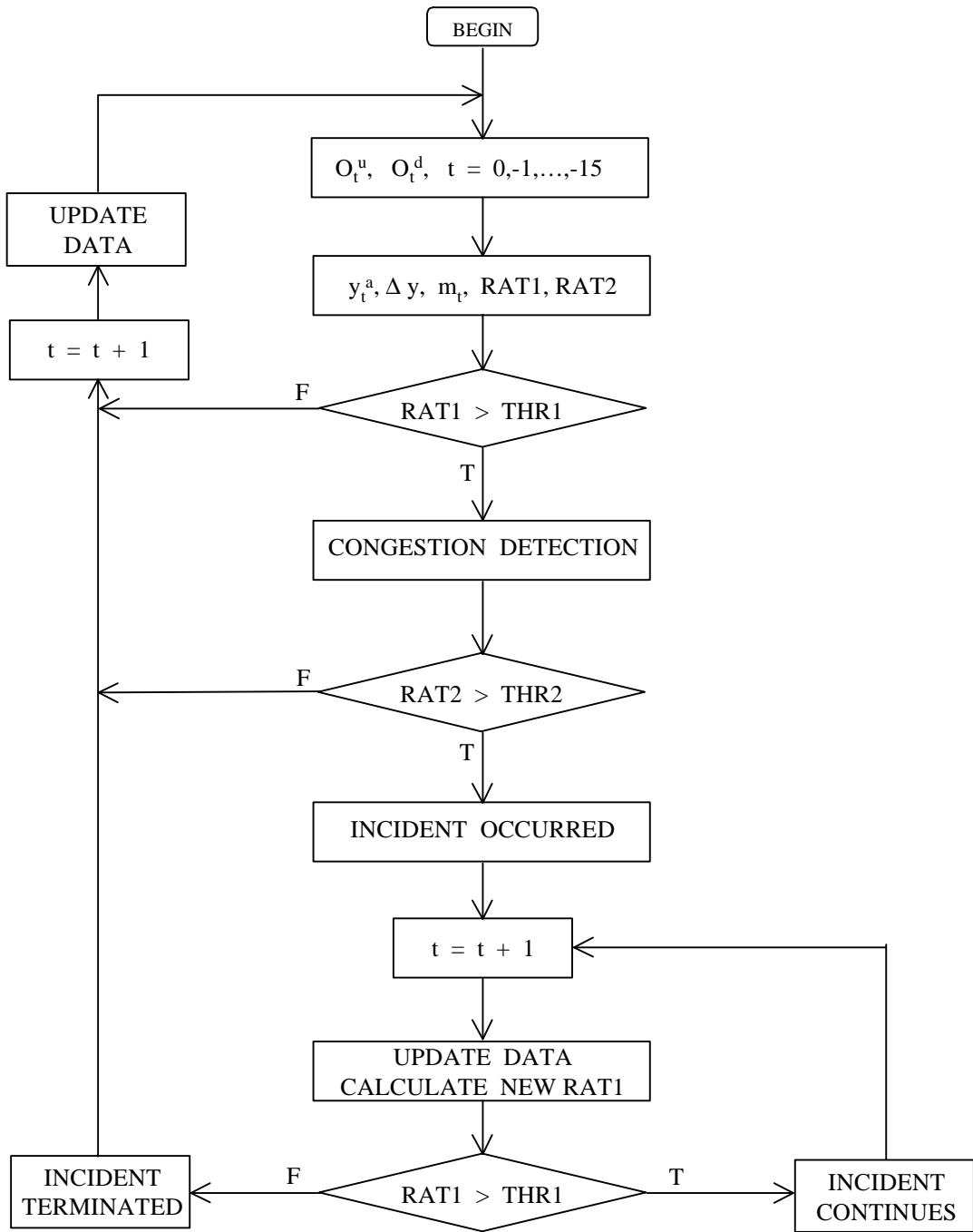


Figure 3. Structure of Minnesota Algorithm

variety of research fields every day. However, every application of ANN shares the ability to make associations between known inputs and outputs by observing many examples. Researchers in transportation engineering have used artificial neural networks since late 80's. The application attempts included simulating driver behavior (Yang et al. 1992; Dougherty and Joint 1992), estimating travel time (Nelson and Palacharla 1993; Hua and Faghri 1994), classifying pavement distress from video images (Hua and Faghri 1993), detecting vehicles from images (Mead et al. 1994; Belgaroui and Blosseville 1993), and detecting incidents. Ann-based freeway incident detection studies, found through an extensive literature survey, are briefly discussed in chronological order.

Wiederholt et al. (1993) developed two single-station ANN models to detect incidents. Traffic flow variables (speed, volume, and occupancy) were simulated for Highway 401 in Toronto, Canada, using a calibrated INTEGRATION model. INTEGRATION is a dynamic traffic network and controller simulating software. Observed traffic patterns from Toronto section of the highway 401 between 2pm and 3pm on June 8, 1992 were used to generate the simulated traffic flow. The simulated output data were averaged over all lanes at 20 second intervals for sixty-one different one-hour traffic scenarios. Traffic data from ten links (or sections) were divided into training and testing data sets in the ANN model development. The two ANN models were based on single loop station traffic data. Inputs for the first model included link number, and traffic data (speed, volume, and occupancy) at current time interval. Inputs for the second model included link number, and historical traffic data from current time interval up to two previous time intervals. Both models consisted of a single hidden layer and a single output neuron. Model outputs were transformed into binary

format with '1' indicating 'incident present' condition and '0' indicating 'incident free' condition. The two models were tested with a fixed number of hidden layer neurons (10 and 12 hidden layer neurons in the first and second models, respectively). Learning rate of 1.5 and a momentum rate of 0.9 were used in training both models. Performance of both models after training and testing resulted a detection rate of 97% and false alarm rate of 3%.

Cheu and Ritchie (1994) developed a two-station ANN model to automate the detection of freeway incidents. Model training and testing were performed using simulated freeway traffic flow data from INTRAS. INTRAS is a microscopic freeway traffic simulation software. Traffic data from eight detector stations in a 5-mile section on westbound SR-91 Riverside freeway in Orange County, California were gathered in the study. Traffic volume from 4:45am to 7:30pm on January 3, 1991 were used to simulate the traffic flow. Simulated volume and occupancy values averaged over all lanes during 30 second intervals were used in model development. The model inputs included normalized volume and occupancy values from upstream station up to four previous time intervals and downstream station up to two previous time intervals. The architecture that provided highest performance had an input layer with 16 nodes, a hidden layer with 9 neurons, and an output layer with a single neuron. Output of the network was translated into binary format with '1' to suggest 'incident present' condition and '0' to suggest 'incident free' condition. Model performance estimates indicated a detection rate of 80% with a false alarm rate of 1.46% during an evaluation phase using simulated traffic data.

Combining fuzzy logic and artificial neural networks, Hsiao (1994) developed an automatic incident detection system. Traffic data collected at 14 loop detector stations on

Highway 401 in Toronto, Canada, from February 5 through April 7, 1993, were used in model training and evaluation. The model inputs included traffic volume, occupancy, and speed averaged during 20-second intervals. Fuzzy logics were used to classify each input variable (in numeric format) into 'low', 'medium', and 'high' linguistic categories representing '0', '1', and '2' respectively. Model output included two possibilities; incident 'possible' and incident 'impossible.' The model is a single station model with 3 input layer nodes, double hidden layers connected in a feed-forward format, and an output layer with a single neuron. Traffic data at current time interval were the only input variables used in the model. Performance rates of the best model during the training estimated a detection rate of 76.19% and a false alarm rate of 8.05%.

In 1995, Stephanedes and Liu developed a two-station based feed-forward neural network to detect freeway incidents. Traffic data collected at 14 detector stations along a 5.5 mile corridor in westbound I-35 in Minneapolis, Minnesota, during a 72-day period in 1989 were used in the model development process. It consisted of one minute traffic volume and occupancy data updated every 30 second intervals and averaged over all lanes during 4:00pm to 6:00pm. The network had 40 input layer nodes (excluding the node for the threshold bias unit), 30 hidden layer neurons, and a single output layer neuron. Model inputs consisted of traffic volume and occupancy data at 10 consecutive 30 second time intervals from pairs of adjacent loop stations. The training data contained 425 input patterns including 'incident free' and 'incident present' conditions. During the training, model output had either '0' or '1' to indicate 'incident free' or 'incident present' conditions. The testing data included all the traffic data collected during 140 hours over 72 day period. During model testing, the output

consisted of a continuous value between '0' and '1'. When the output was greater than a predefined threshold value (0.5 in this case) the traffic condition was interpreted as 'incident present' and 'incident free' otherwise. The model had a detection rate of 70 to 80% with a false alarm rate of 0.12 to 0.26% during the model testing process.

Abdulhai and Ritchie (1995 and 1997) used a modified probabilistic neural network (PNN) to automate the freeway incident detection process. The modified PNN includes an input layer, three hidden layers, and an output layer. One of the three hidden layers, known as a pattern layer stores input patterns. The number of neurons in the pattern layer is equivalent to the number of training patterns in the network. Abdulhai and Ritchie (1997) used the same data that Cheu and Ritchie (1994) used in their study. A modified form of the Bayesian-based method was employed in the model development process. The universality or transferability concept of AIDMs was introduced and described in detail. The study considered an AIDM is transferable if a trained model can be directly applied, without recalibration of any parameters, to a new geometric location or the same geometric location after significant laps in time. Model was tested with simulated data for westbound I-35 and the performance results yielded an impressive detection rate of 100% with a false alarm rate of 4.77% (Abdulhai and Ritchie 1995).

All the AID systems based on ANN studies discussed above exhibit an overall improvement in incident detection performance over the conventional algorithms (e.g., California and Minnesota algorithms). However, these ANN-based studies still have some short comings. For example, one hour traffic pattern used in Wiederholt et al. (1993) study may not be a representative enough to simulate variety of traffic conditions that exists in a

freeway system. Many ANN-based models described above (except model developed by Stephanedes and Liu 1995) used simulated traffic data to train ANN models. Cheu and Ritchie (1994) stated that the experience have shown the modeling equations may not always be satisfactory in replicating traffic flow in actual situations, leaving some concern over estimated traffic flow variables using dynamic modeling. Therefore, the model performance may contain some sampling bias. ANN models developed by Cheu and Ritchie (1994), Stephanedes and Liu (1995), and Abdulhai and Ritchie (1995) used only two of the three traffic variables as model inputs despite all three traffic variables are affected by incidents. Still every model used a conventional detection rate that may not truly reflect the capability of detecting individual incident patterns as described later in this section. Since a separate hidden layer neuron is required for separate input pattern for each example in the training data set, PNN models can become quite large. Once trained, these PNN models take more time to run than back propagation networks (Lawrence 1993). Furthermore, in problems such as incident detection with large amount of complex data, back propagation is more accurate (Specht and Shapiro 1991).

CHAPTER 3

DESCRIPTION OF ARTIFICIAL NEURAL NETWORKS

The most complex biological network ever known to this date is the human brain. It consists of hundreds of billions of special cells, known as neurons, which are connected together in complex form. These neurons send information back and forth to each other through their connections. A network of this kind can perform intelligent functions such as learning, analysis, prediction, and recognition. The functions of these neurons in human brain are mimicked in artificial neural networks using computers. However, the neurons in human brain are much more complicated than the neurons used in the artificial neural networks. Hecht-Nielsen (1990) defines an artificial neural network (ANN) as,

“... a parallel, distributed information processing structure consisting of processing elements (which can possess a local memory and carry out localized information processing operations) interconnected via unidirectional signal channels called connections. Each processing element has a single output connection that branches into as many collateral connections as desired; each carries the same signal - the processing element output signal. The processing element output signal can be of any mathematical type desired. The information processing that goes on within each processing element can be defined arbitrarily with the restriction that it must be

completely local; that is, it must depend only on the current values of the input signals arriving at the processing element via impinging connections and on values stored in the processing element's local memory."

The neurons in most common ANNs are usually organized in three type of layers: input, hidden, and output layers. Neurons in an input layer just act as information feeding points and do not involve in any sort information processing. Therefore, they are referred as input nodes in this study. An ANN may contain one or more hidden layers depending on the problem and the best number of neurons in each layer is determined by trial and error. In the literature, preferred method of referring to an ANN architecture is by the number of hidden layers. Therefore, these terminologies were followed throughout this dissertation. The number of input layer neurons is equivalent to the number of input variables in a problem. Usually, an ANN has a single output layer and the number of output layer neurons is equal to the number of outputs required in the solution. The number of input nodes and output neurons are dictated by the problem itself. Behavior of a larger ANN can be easily explained once the behavior of a single neuron is clearly understood. The purpose of this chapter is to give a general understanding of the process of information flow in an ANN with feed-forward structure with backpropagation capability process information during training and actual implementation.

3.1 Behavior of a Single Neuron

A neural network is composed of several neurons that are connected to one or more neurons by synapses or links. These connections are characterized by a strength or weight of

their own. A weight is positive if the associated synapse is excitatory and is negative if the synapse is inhibitory.

Consider a single neuron with I input nodes as shown in Figure 4. When the neuron processes information, input signal (x_i) of i^{th} input node ($i=1,2,3, \dots, I$), which is connected by a synapse to the neuron, is multiplied by the synaptic weight w_i . Every input signal weighted by the respective synapses is summed at the neuron. Activation (or squashing) function limits the amplitude of the output (the summation) of the neuron and yields the non-linearity feature of the information processing. Typically, the range of normalized amplitude of the neuron's output is set to be a closed unit interval $[0, 1]$ or $[-1, 1]$. For the single nonlinear neuron (as in Figure 4), the following equations show this data processing operation mathematically:

$$\mathbf{u} = \sum_{i=1}^I w_i x_i \quad (6)$$

where u is the net input to the neuron and x_i is the input signal from an input node i , and w_i is the connection weight from an input node i to the neuron.

$$y = \varphi(\mathbf{u} - \theta) \quad (7)$$

where y is the neuron's output signal (also known as its activation), $\varphi(\cdot)$ is the activation function for the neuron, and θ is the bias threshold used to offset the net input to the neuron.

The activation function defines the output of a neuron in terms of the activity level of its input. Depending on the application, a variety of activation functions are being used by researchers. In this study, a piecewise-linear function and a sigmoid function were used

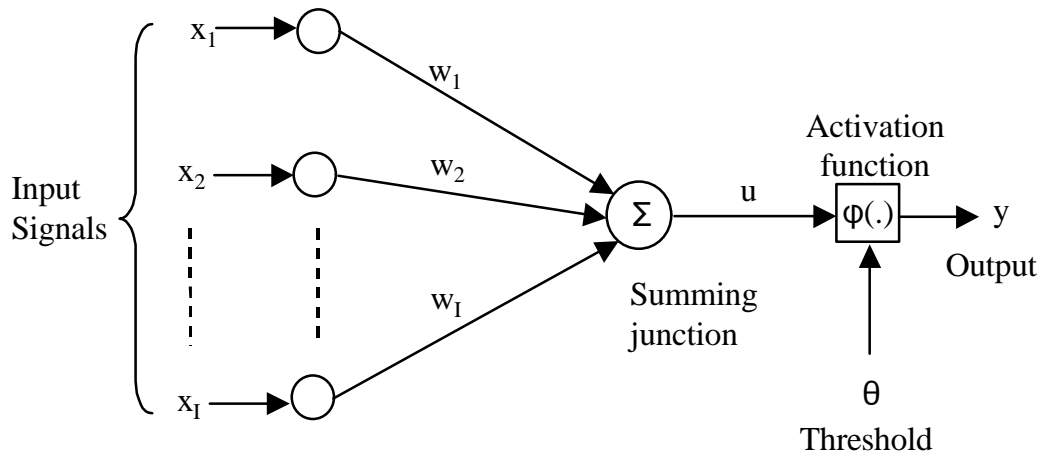


Figure 4. Model of a Nonlinear Neuron

in the ANN models. Basic features of these two activation functions are briefly discussed next.

A piecewise-linear function provides continuously varying values in $[-0.5, 0.5]$ input range and constant values ('0' or '1') on either side of the input range. The activation function is defined by:

$$\varphi(u) = \begin{cases} 1 & \text{if } u \geq 1/2 \\ u + 1/2 & \text{if } 1/2 > u > -1/2 \\ 0 & \text{if } u \leq -1/2 \end{cases} \quad (8)$$

where u is the total summation at a neuron and $\varphi(u)$ is the activation function. The behavior of the piecewise linear function is shown in Figure 5.

The sigmoid or logistic function is by far the most common form of activation function used in the development of artificial neural networks. It is a strictly increasing

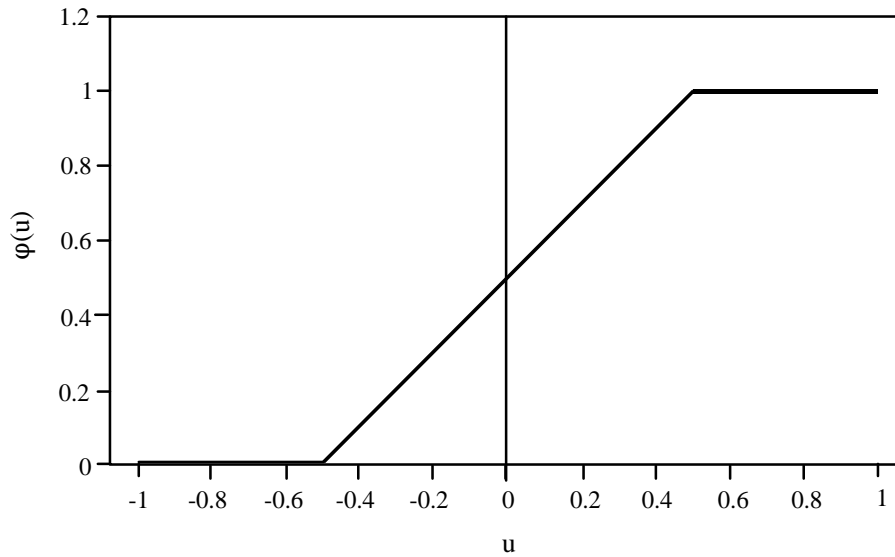


Figure 5. Behavior of Piecewise-Linear Function

function that exhibits smoothness and asymptotic properties. With a positive slope parameter, a , the sigmoidal function is defined by:

$$\phi(u) = \frac{1}{1 + e^{-au}} \quad (9)$$

By varying the parameter a , different slopes of sigmoid function can be obtained as shown in Figure 6. Rumelhart and McClelland (1986) discussed the features of sigmoid function in great detail and stressed the importance of unit activation functions in the development of artificial neural networks.

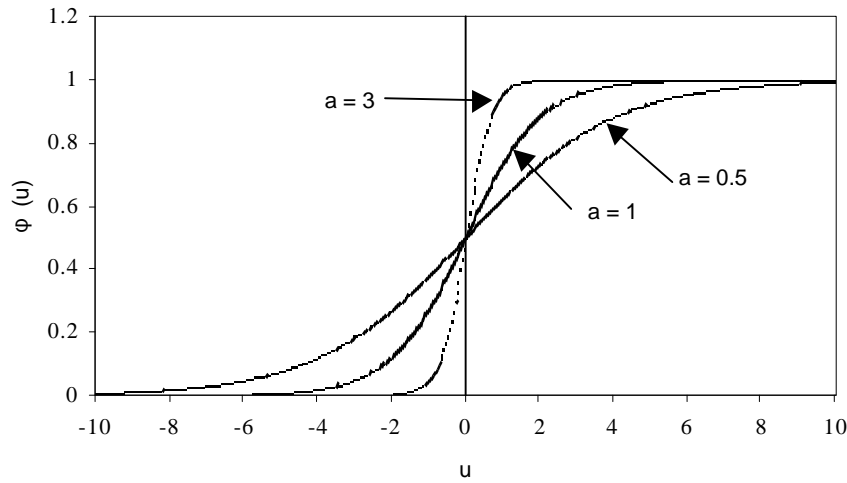


Figure 6. Behavior of Sigmoid Function

3.2 Behavior of a Network of Neurons

An ANN is built up of several neurons in hidden and output layers. These neurons in a network function in the fashion explained in the previous section. To understand how an ANN process information, consider a simple single-hidden-layer feed forward network with backpropagation capabilities as shown in Figure 7. Nodes in input layer are represented by subscript i and the neurons in hidden and output layers are represented by subscripts j and k respectively. The inputs to the network are denoted by x_i (where $i = 0, 1, 2, \dots, I$), and the output from hidden layer and output layer neurons are denoted by y_j (where $j = 0, 1, 2, \dots, J$) and z_k (where $k = 1$ in this case) respectively. Synaptic weight w_{j0} (corresponding to a fixed input $x_0 = -1$), equivalent to the bias threshold θ_j , is added to the j^{th} neuron in the hidden layer (where $j = 1, 2, \dots, J$). Similarly, synaptic weight w_{k0} (corresponding to a fixed input $y_0 = -1$), equivalent to the bias threshold θ_k , is added to the output layer neuron (where $k = 1$).

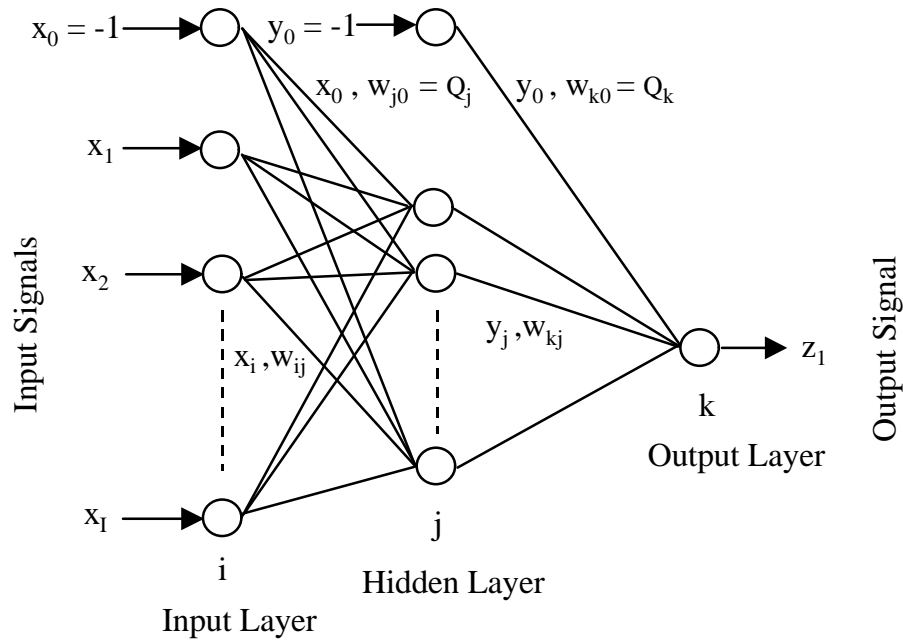


Figure 7. Single-Hidden-Layer Feed-Forward Neural Network

The purpose of the input layer node $i = 0$ and the hidden layer node $j = 0$ is to represent the bias threshold values for hidden layer neurons $j = 1, 2, 3, \dots, J$ and output layer neuron $k = 1$. This representation makes the formulation of learning process of the network much easier. The training processing of the network consists of two distinct steps; a feed-forward (forward pass) step and a backpropagation (backward pass) step. (Backpropagation is the learning technique used in this study.)

3.2.1 Feed-Forward

In the feed-forward step, the array of input data is presented to the network, and the current connection weight and bias threshold values (θ_j 's and θ_k 's - values used to offset the

net input to each neuron) are used to calculate the resulting output value. The input signals are processed as follows.

- (i) For each neuron in the hidden layer, every input signal is multiplied by the weight of the connection from that input node to the hidden layer neuron.
- (ii) The total of these weighted input signals and the bias threshold is the net input to the hidden layer neuron. This net input becomes the argument for the activation function which determines the hidden layer neuron's output signal y_j according to equation 7.

For the neural network with a single hidden layer as shown in Figure 7, the above process can be mathematically derived. The output from j^{th} hidden layer neuron is:

$$y_j = \varphi_j \left(\sum_{i=0}^I w_{ji} x_i \right) \quad (10)$$

These hidden layer outputs y_j (where $j = 1, 2, \dots, J$) and the bias threshold y_0 multiplied by their corresponding weight values w_{kj} (where $k = 1$ and $j = 0, 1, 2, \dots, J$) are fed into the output layer neuron as inputs. This set of hidden layer outputs, in a feed forward network, produces the following output from the output layer neuron:

$$z_k = \varphi_k \left(\sum_{j=0}^J w_{kj} y_j \right) = \varphi_k \left(\sum_{j=0}^J w_{kj} \left[\varphi_j \left(\sum_{i=0}^I w_{ji} x_i \right) \right] \right) \quad (11)$$

3.2.2 Back-propagation

The output (z_k) calculated for each input pattern is compared with the corresponding desired output. Gradient steepest descent, using the square of the difference between the desired and observed output as an error function, is used to adjust the connection weights and

bias thresholds so that the network can return an output closer to the desired output when it is presented to the network next time. According to general leaning rule (delta rule) in correcting any connection weight at the end of t cycles, the value of the weight for the cycle $t+1$ is estimated as the summation of weight at cycle t and weight correction applied at the end of the cycle t . This step is shown in the following equation.

$$w(t + 1) = w(t) + \Delta w(t) \quad (12)$$

The weight correction, $\Delta w(t)$, is estimated as proportional to the decrease in error function with respect to the connection weight. Mathematically, this can be written as,

$$\Delta w(t) = -\eta \frac{\partial E}{\partial w} \quad (13)$$

The value ‘ η ’ is the leaning rate of the neural network. Square error function, E , for the network output is one half of the sum of squared differences between the desired and predicted output of the output layer neurons for a single input pattern. This can be represented in the following equation:

$$E = \frac{1}{2} \sum_k (d_k - z_k)^2 \quad (14)$$

where, ‘ d_k ’ is the desired output of the k^{th} output neuron (in this case $k=1$).

In gradient steepest descent method, the partial derivative of the error function is calculated with respect to each connection weight for an output layer to estimate the weight correction as follows:

$$\frac{\partial E}{\partial w_{kj}} = -(d_k - z_k) \frac{\partial[\varphi_k(\sum_{j=0}^J w_{kj} \varphi_j(\sum_{i=0}^I w_{ji} x_i))]}{\partial w_{kj}} = -\delta_k \varphi_j(\sum_{i=0}^I w_{ji} x_i) \quad (15)$$

where δ_k is given by the following equation.

$$\delta_k = (d_k - z_k) \varphi_k'(\sum_{j=0}^J w_{kj} \varphi_j(\sum_{i=0}^I w_{ji} x_i)) \quad (16)$$

Similarly, the gradient descent for the hidden layer can be formulated by taking the partial derivative of the error function with respect to each connection weight for the hidden layer neurons as follows:

$$\frac{\partial E}{\partial w_{ji}} = -\sum_k (d_k - z_k) \frac{\partial[\varphi_k(\sum_{j=0}^J w_{kj} \varphi_j(\sum_{i=0}^I w_{ji} x_i))]}{\partial w_{ji}} = -\delta_j x_i \quad (17)$$

where δ_j is given by the following equation:

$$\delta_j = \varphi_j'(\sum_{i=0}^I w_{ji} x_i) \sum_{k=0}^J \delta_k w_{kj} \quad (18)$$

Each partial derivative in equations (15) and (17) is the basis for calculating the appropriate adjustment for the corresponding connection weight. The smaller the learning rate, η , the smaller the change to the synaptic weights in the network will be from one iteration to the next. As the change in synaptic weights gets smaller, the trajectory in weight

space will be smoother. This improvement, however, is attained at the cost of a slower rate of learning. If, on the other hand, a larger learning rate (too large to speed up the learning rate) is used, the resulting large changes in the synaptic weights assume such a form that the whole network may become unstable (i.e., the error term oscillates and ceases to converge to a single value). Rumelhart and McClelland (1986) used a simple method to increase the rate of learning and yet avoid the danger of instability through a modified delta rule by introducing a momentum term α . This modified correction with the momentum term for the output layer is given by the following equation:

$$\Delta w_{kj}(t) = \eta \delta_k \phi_j \left(\sum_{i=0}^I w_{ji} x_i \right) + \alpha \Delta w_{kj}(t-1) \quad (19)$$

The modified correction for the hidden layer can be represented by the following:

$$\Delta w_{ji} = \eta \delta_j x_i + \alpha \Delta w_{ji}(t-1) \quad (20)$$

where Δw_{kj} and Δw_{ji} are the changes computed for the connection weights from hidden layer neuron j to output layer neuron k and input layer unit i to hidden layer neuron j respectively, and δ_j is the error propagated backward through neuron j .

The network training continues until the error stops decreasing. During the training, the network works to fit the training data with which it is presented, and the direction or extent of the adjustments to the connection weights is not controlled. If it is allowed to train too long, it will over-fit the training data, or tailor its weight vector too closely to input patterns which may not represent conditions the network will face in operation (Caudill 1990). This problem can be resolved by specifying a maximum training cycles (epoches).

CHAPTER 4

FREEWAY DATA

The majority of AID models developed in previous studies were trained (or calibrated) using simulated freeway incident conditions. Of the ANN models developed, Cheu and Ritchie (1994), Stephanedes and Liu (1995), and Abdulhai and Ritchie (1997) used actual freeway data to evaluate the models. When simulating traffic flow conditions, the traffic is assumed to follow traffic flow equations and other conditions inherent to the simulation package. Especially under traffic congestion, these theoretical assumptions and equations may not best explain the behavior of actual traffic flow. Therefore, real-world traffic data were used in this research to subdue these negative effects in AID model training.

4.1 Data Source

Real-world traffic data from a freeway (I-880) in Hayward, California were used in this study. Both traffic and incident data were collected and made available by the PATH program at the University of California at Berkeley, California. The study site is 9.2 miles long and has 3 to 5 lanes in each direction at various locations of the freeway. A layout of freeway geometry is shown in Figure 8.

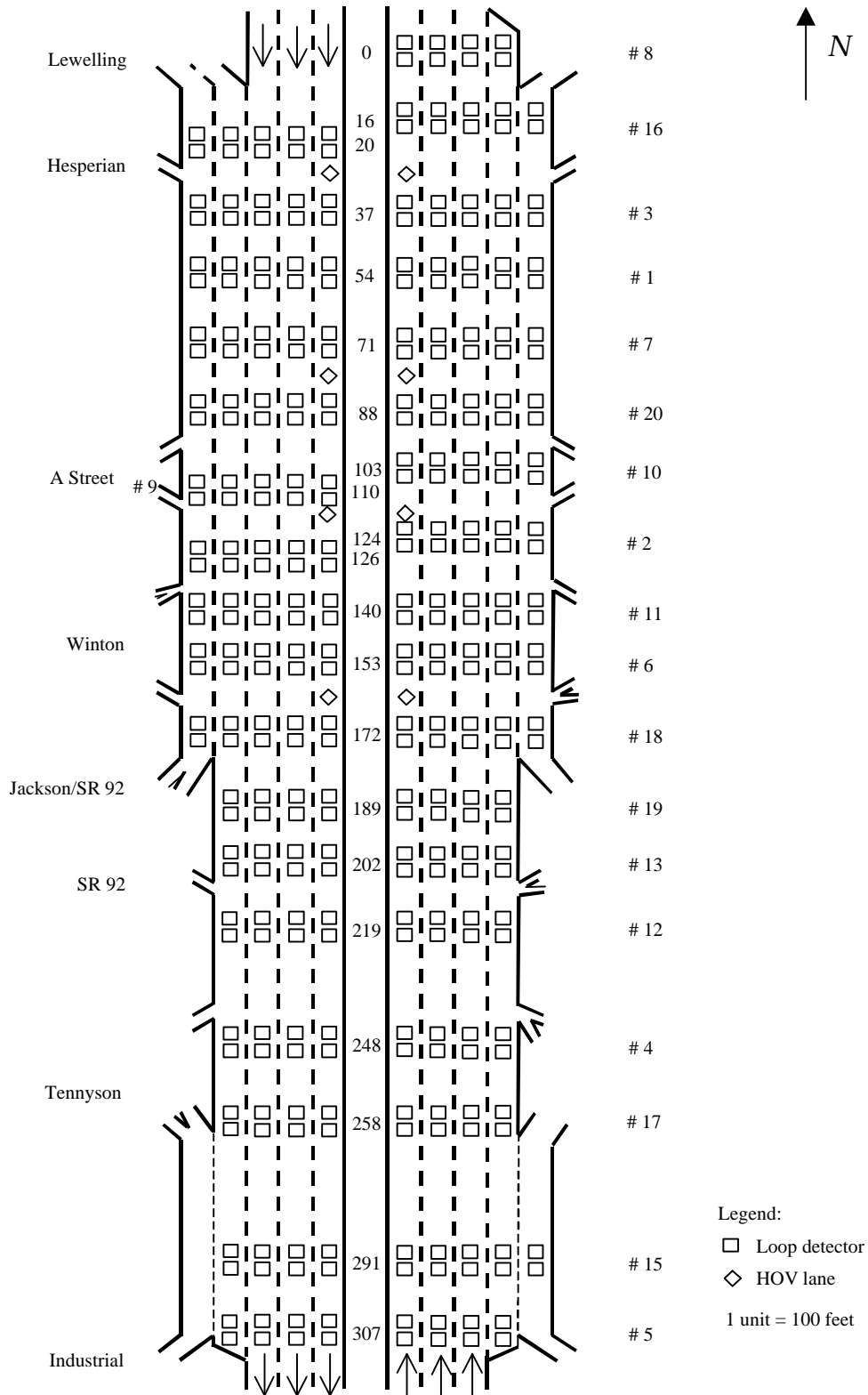


Figure 8. Schematic Diagram of the I-880 Study Area

4.2 Data Description

Between Industrial and Lewelling exits, northbound section of the freeway is divided into 17-cross-sections (or stations) while southbound section is divided into 16-cross-sections as shown in Figure 8. At each cross section of the freeway, a pair of inductive loops is buried under the pavement on each lane as shown in small squares. Spacing between adjacent loop stations ranges from 1000 feet to 3300 feet. Entrance and exit ramps are instrumented with single loop detectors. Loop data were collected from 5am - 10am and 2pm - 8pm. Probe vehicles were used to collect real time incident data. These probe vehicles were in operation from 6:30am - 9:30am and 3:30pm - 6:30pm during weekdays collecting incident data. The signal from loop detectors were in non-ASCII format and had been processed to obtain traffic volume, speed, and occupancy of each lane at each station using a software developed by the University of California at Berkeley. These traffic flow variables averaged over all lanes during 30 second interval were used in this study.

Incident data were collected through probe vehicles which were operating with a 7 minute headway in the study corridor. When a probe vehicle encountered an incident, the driver transmitted information such as the incident location, vehicle type, direction, and number of vehicles involved to a command center located near the study corridor. This information was then processed to create a comprehensive incident database. The incident database collected during February 16 - March 19, 1993 has 1210 incident records and that collected during September 27 - October 29, 1993 has 971 incident records.

Besides the traffic flow variables, freeway geometric data such as presence of entrance and exit ramps and lane expansion and merger information were also considered as model

inputs. The northbound has 5-entrance and 6-exit ramps while the southbound has 6-entrance ramps and 5-exit ramps in the study area. The northbound has 2-lane expansions and 2-lane mergers while the southbound has 1-lane expansion and 2-lane mergers.

4.3 Data Verification

In this study, daily traffic flow data from every mainline loop station in both directions were inspected for consistency and continuity. The inspection was conducted by preparing 3-dimensional graphs of volume, speed, and occupancy each using the Matlab software. Abnormal and missing data were then identified and removed from the model development phase. Three samples of such 3-dimensional graphs for northbound afternoon traffic on February 16, 1993 are shown in Figures 9, 10, and 11. The minimum and maximum ranges of each traffic flow variable were also obtained to establish a data normalization criterion.

The location information of some incidents was found inaccurate. To fix this problem, the incident database was compared with the probe vehicle database (or car database). The driver of the probe vehicle presses a key of an onboard computer when an incident is encountered. The car database consists of odometer reading and the time when the driver passed the key. This information was compared with the freeway layout and the incident database to determine the correct location of the incident.

Once the traffic flow data and incident database were scrutinized, traffic flow data were subjected to a final visual verification. This was necessary to confirm continuity of traffic flow data between two stations and exact location of incidents to be used in the model

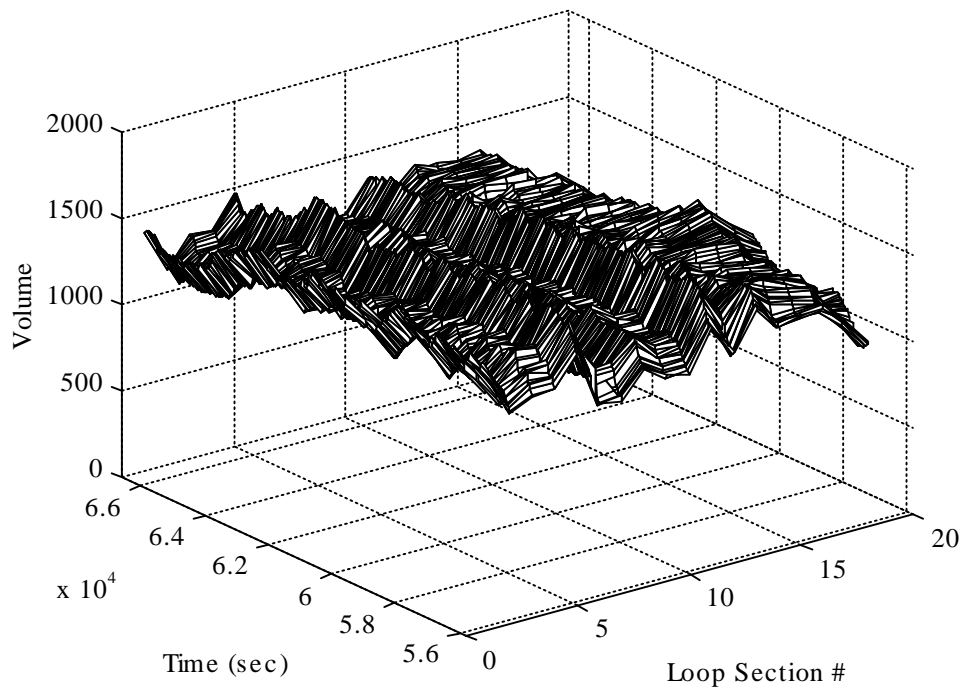


Figure 9. Distribution of Afternoon NB Volume on I-880 on February 16, 1993

development process. The traffic volume, speed, and occupancy data were plotted against time using the MS Office Excel. Traffic data from two adjacent loop stations were plotted one pair at a time for the visual analysis.

4.4 Data Preparation

Traffic volume and speed data were normalized to increase the efficiency in model development process. This normalization was performed by dividing each flow value by a predetermined maximum value for the respective flow variable. A maximum freeway lane volume of 2,200 vphpl was used to normalize the traffic volume data. A few individual

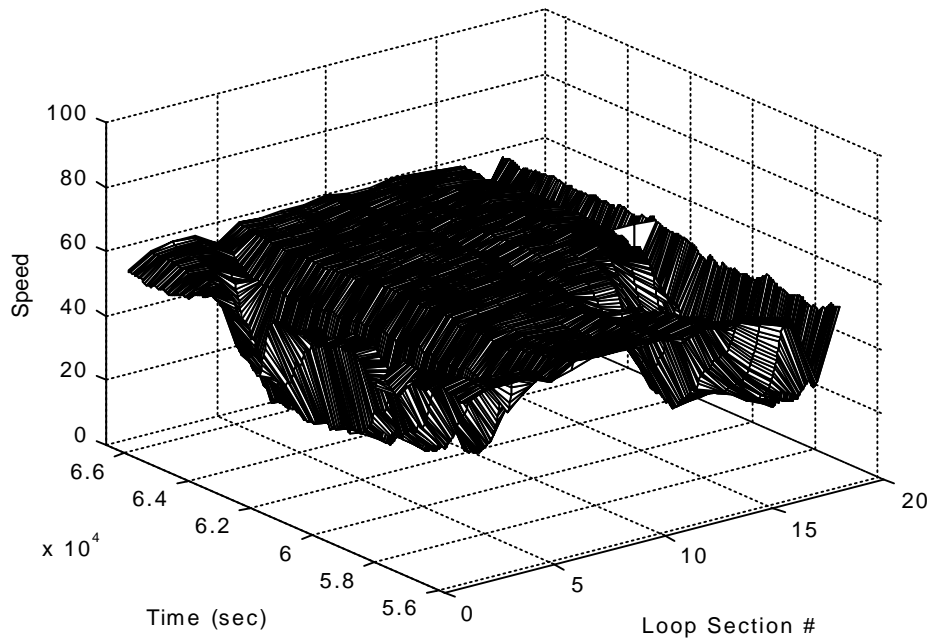


Figure 10. Distribution of Afternoon NB Speed of I-880 on February 16, 1993

vehicles could cruise above 100 mph on a freeway at any given moment. However, a fleet of vehicles may not travel above 100 mph passing a single freeway section for an extended period. Since the speed of vehicles during 30 second interval was taken for the average speed calculation, the average speed could be expected to be less than 100 mph for the majority of time. Therefore, a maximum speed of 100 mph was used in normalizing 30 second average speed data. Since occupancy is expressed as a percentage of time that loops were occupied by vehicles, the average occupancy during 30 second interval was used without any modifications in the model development process.

The four freeway geometric variables, entrance and exit ramps and lane expansion and merger information, considered were included as four different inputs. Possible scenarios

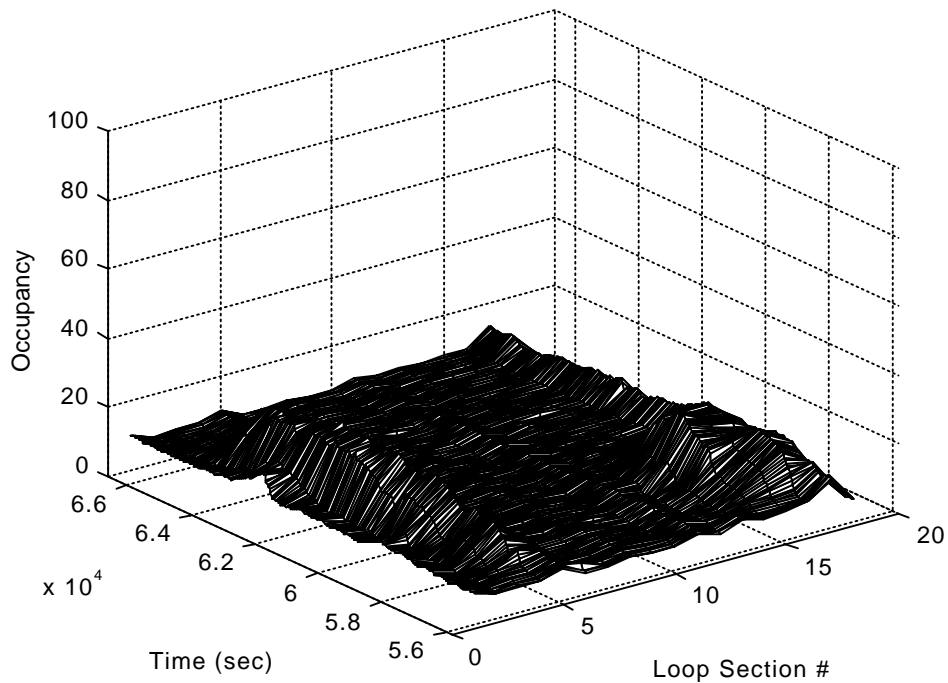


Figure 11. Distribution of Afternoon NB Occupancy on I-880 on February 16, 1993

for the freeway geometric variables were divided into two categories: ‘present’ (Yes), and ‘not present’ (No). For example, if an entrance ramp is ‘present’ between two adjacent loop stations, the linguistic value of the entrance ramp input variable was taken as ‘Yes.’ On the other hand, if an entrance ramp is ‘not present’ between the two loop stations, the linguistic value of the entrance ramp input variable was taken as ‘No.’ This linguistic code for the freeway geometric information was then transferred into binary code as shown in Table 1.

Once the accuracy of the freeway incident data was established, all the incidents occurred and traffic data gathered during September 27 - October 29, 1993 were selected for model development. Both morning and afternoon traffic data were used from station pairs

Table 1. Data Transformation for Geometric Variables

Geometric Variable Between Two Stations	Type of Code	
	Linguistic	Binary
Entrance ramp	Yes	1
	No	0
Exit ramp	Yes	1
	No	0
Lane expansions	Yes	1
	No	0
Lane merges	Yes	1
	No	0

where actual incidents occurred on a given day. From this pool of traffic data, about 90 percent was selected for training on a purely random basis. Total of 13,718 training traffic patterns were selected of which 3,218 belonged to ‘incident present’ patterns. The other 10 percent (1,840 patterns) of the data were used to test the model during the training process. The testing data included 532 incident present patterns.

Traffic flow and incident data collected during the February 16 through March 19, 1993 were used in the evaluation and comparison of the ANN model results. The data for the evaluation and comparison phase were extracted and verified in a similar manner as previously described. Total of 9,818 traffic flow patterns were selected of which 2,231 belonged to ‘incident present’ traffic conditions.

CHAPTER 5

MODEL DEVELOPMENT

5.1 Model Inputs

The California and Minnesota algorithms identified occupancy as the single most important variable in automatic incident detection system. Nevertheless, recent studies (Presaud and Hall 1989; Cheu and Ritchie 1994; Stephanedes and Liu 1995) have shown that including traffic volume in the detection system increased the overall accuracy and reliability of the AID systems. Cheu and Ritchie (1994) reported that adding traffic speed variable in his models did not increase the accuracy considerably. However, Hsiao (1994) used traffic volume, speed, and occupancy as inputs in a model that was based on fuzzy and neural network theories. Apparently, several researchers have used different combination of traffic variables in AID systems based on their professional judgment in the input variable selection process. It is conceivable that traffic speed changes between upstream and downstream ends of an incident. This means that the traffic speed can be selected as yet another input for the AID systems. Therefore, traffic volume, speed, and occupancy variables were utilized as model inputs in this study. Moreover, freeway geometric information was also used as inputs on selected models.

5.1.1 Traffic Flow Variables

When an incident occurs on a freeway section, vehicle carrying capacity of the section temporarily reduces. This reduced capacity creates a hindrance to the normal traffic flow by changing volume, speed, and occupancy both upstream and downstream of the incident location. Therefore, the traffic flow variables from two stations are the best candidates (compared to data from a single station) for incident detection model inputs.

The traffic flow data from each lane were averaged across all lanes at each loop station during 30 second intervals. The data from adjacent loop stations were normalized as mentioned in section 4.4 before utilizing in the model development process. This was necessary to expedite the ANN model training process. Under normal traffic conditions, a vehicle traveling at 40 mph on the freeway takes about 85 seconds to pass two farthest (maximum spacing of 3300 feet) loop stations. That means, upstream traffic flow at time $t-2$ and downstream traffic flow at time t best includes a continuous mass of traffic flow. Therefore, traffic flow data at least from time intervals t , $t-1$, and $t-2$ at upstream station were used in this study. In some model architectures, however, traffic data up to time interval $t-4$ at upstream station and up to time interval $t-2$ at downstream station were used. This line of reasoning is consistent with the research conducted by Cheu and Ritchie (1994) and Abdulhai and Ritchie (1997).

5.1.2 Geometric Variables

When a freeway section expands, the added capacity after the expansion may yield reduced average lane volume and occupancy across all lanes. Under normal traffic flow

conditions, average downstream volume and occupancy will be less than that at upstream. In contrast, when a freeway section contracts (by merging lanes), the reduced capacity after the merger may yield higher average volume and occupancy across all lanes. This reduced capacity close to downstream station may create recurrent congestion (or bottleneck) between the two stations under heavy traffic conditions. Under normal traffic flow conditions, the average downstream lane volume and occupancy will be greater than that at upstream. This illustrates the fact that traffic flow can be affected by means other than incident conditions. Therefore, the effect of these geometric variables on traffic flow should be characterized in the AID models. Effect of entrance and exit ramps, in between two adjacent stations, on traffic flow was represented in the models developed in this study. Earlier models (Hsiao 1994; Stephanedes and Liu 1995), however, did not have these geometric information as model inputs.

This line of reasoning may be expanded to include other freeway and environmental data such as gradient, horizontal and vertical curves, weather condition (dry vs rainy), light condition, mix of traffic (i.e., percentage of passenger cars, trucks, and other vehicle types), etc. However, an extensive database that includes these variables could not be found at the present time to investigate the suitability of these as model inputs.

5.2 Model Output

In this research, ANN were designed to identify traffic conditions from the model inputs, and classify the conditions either as ‘incident present’ or ‘incident free.’ During supervised model training, this linguistic desired output was transformed into a binary output.

That is 'incident present' and 'incident free' conditions were translated as '1' and '0,' respectively. Since the desired model output is in binary format, only a single output layer neuron was needed in each of the ANN architectures developed. While preparing independent data sets for the model development (training, testing, and evaluation), 'incident present' condition (or '1') was used during the entire incident period as reported in the incident database while 'incident free' condition (or '0') was used at other times.

A linear activation function was used in the output layer neuron in all the models. Since the desired output is either '0' or '1', the model was forced to yield values within [0,1] range. Actual output value less than or equal to 0.5 was interpreted as '0' (or 'incident free') and outputs greater than 0.5 was interpreted as '1' (or 'incident present') during model development.

5.3 ANN Software

ANN software used in the model development was "Professional Version Basis of AI Backprop." The software can model networks with feed forward and recurrent features. It is programmed in C++ language under 'Linux' operating system. The software can read input pattern files saved in ASCII text format with values separated by a single space. Input patterns need to be saved in different lines in the text file. The known limitation on the number of input patterns is that each line containing input pattern values has to be less than 255 characters. The professional version of the software uses 32-bit binary arithmetic which is about four times faster than 16-bit student version. Several activation functions are

supported in the software: smooth Sigmoid, tanh, x and y (where x runs from -1 to 1 and y runs from 0 to 1), linear, and Gaussian functions.

5.4 Network Architectures

A best ANN model should be able to store hidden relationships common to a group of inputs, which may not be seen by a human eye. These relationships in ANN are stored in a form of weights and bias information. This phenomena is also known as network ‘generalization.’ Generalizations are very important in applications when a network is also required to make predictions for input patterns that are not in the training data. Several researches have discussed advantages of overdetermined ANN models that generalizes (Carpenter and Hoffman 1997). In contrary to the generalization, a network that is too complex may fit the noise, not just the signal patterns, leading to a situation known as ‘overfitting.’ Overfitting is especially not warranted in an application of this sort since it can easily lead to ANN predictions that are far beyond the range of the training data, and can even yield wild predictions (SAS 1998). In the incident detection scenario, the training data can include so little information and different patterns that inputs may not contain all possible traffic and freeway geometric conditions. Therefore, generalization of the ANN models is very important when using the models under a variety of situations.

The problem in hand more-or-less specifies the number of inputs and outputs required. During model development process, many ANN researchers face with two basic questions: how many hidden layers are needed? and how many hidden layer neurons are needed? Despite few suggestions and guidelines on selecting number of hidden layers and hidden layer neurons

proposed by different researchers, a universal rule or set of equations that dictates the ideal number of hidden layers or hidden layer neurons in an ANN model does not exist to date. Haykin (1993) recommends using double hidden layer ANNs claiming that the first hidden layer extracts local features while the second hidden layer extracts global features of the input patterns. Few other researches (Sontag 1992; Surkan and Singleton 1990) also argue that double hidden layer ANNs provide better results. On the other hand, Lawrence (1993) argue to use a single hidden layer stating that more than a single hidden layer may significantly increase the training time. Experimental work by Villiers and Barnard (1992) showed that double hidden layer networks are only more prone to fall into bad local minima. Due to these controversial opinions on the best number of hidden layers, both single hidden layer and double hidden layer networks were developed on a trial and error basis.

ANN with too few hidden layer neurons will not be able to learn enough from training data. On the other hand, too many hidden neurons will allow the network to memorize the training data without generalizing from the training set for unforeseen patterns. Some text books and articles offer “rule of thumb” for choosing number of hidden neurons. Lawrence (1993) recommends starting with a number of hidden layer neurons that is less than the number of inputs plus the number of outputs divided by two. Then after training the network should be tested with a testing data which is independent from training data. Lawrence (1993) recommends adding a few hidden neurons at a time and repeat the process until the training error is decreased to a point where no further decrease in error can be achieved. General consensus among many ANN professionals recognize that these rules may fail in many situations than they succeed (SAS 1998; Tvetter 1997). At present the only sure way to

know the best number of hidden layers and number of hidden layer neurons is the trial and error approach as utilized by Cheu and Ritchie (1994) and Abdulhai and Ritchie (1995).

In this study, two different ANN groups were tested for incident detection accuracies. The first network group tested included eight different network architectures that transmit input information from input layer to its output layer (feed forward). This information flow occurs only in the forward direction. The second network group tested included eight different network architectures with a combination of transmitting information from input layer to output layer through hidden layers (feed forward structure) and from output layer directly back into the input layer (recurrent loop structure). The recurrent loop was setup from the output layer neuron to an input layer node. Both network architectures were trained with a backpropagation error correction method using the “Professional Basis of AI Backprop” software discussed in section 5.3. The sigmoid activation function was used in every hidden layer neuron and the linear activation function was used in the output layer neuron.

5.4.1 Feed Forward Networks

The behavior of feed forward networks were discussed in section 3.2.1. A feed forward network can have as many hidden layers as desired. Cheu and Ritchie (1994) and Stephanedes and Liu (1995) only considered single hidden layer networks. However, no credible reasoning was given as to why networks with more than one hidden layer were not tested for incident detection. In this research, ANN models with single and double hidden layers were tested to investigate whether this additional hidden layer would enhance detection

performance and eventually lead to a better incident detection model. Each model consisted of single input layer, single or double hidden layers, and single output layer. Total of eight ANN model architectures with feed forward structures were developed. Each model architecture represented a different input feature set or different hidden layer combinations as shown in Table 2. They were named by the model number and the way network process information.

The training process started with a single hidden layer networks with 18 input neurons, a minimum of 8 hidden layer neurons, and a single output neuron. Model weights and bias were initialized using a random number generator that has a mean of '0' and a standard deviation of '1.' The ANN was trained with training data up to 5000 epoches (or iterations or cycles) using backpropagation technique available in the AI software. While the model training was in progress, model error was calculated using the test data, which were set a side during the data preparation, at every 10 cycles. Training was continued until both training and testing errors kept on decreasing. The training was stopped when the testing error started to increase while the training error was still on the decreasing trend. The model weights and bias were initialized again using a random number generator and the model was trained again as explained. This training process with different initial weights and biases was repeated until the standard deviation of the trained model errors remained unchanged. The rest of the single hidden layer models were trained following the procedure described above. Then a second hidden layer was added to the earlier model architectures and repeated the training process. To select the best ANN model for each model type listed in Table 2, a total of 568 feed forward ANN models were developed during the training process.

Table 2. Model Input Variables and Nomenclature for Feed Forward Networks

	Model Inputs	Time Interval	Single-Hidden Layer Networks				Two-Hidden Layer Networks			
			1-F	2-F	3-F	4-F	5-F	6-F	7-F	8-F
Upstream	Volume, Speed, and Occupancy	t	x	x	x	x	x	x	x	x
		t-1	x	x	x	x	x	x	x	x
		t-2	x	x	x	x	x	x	x	x
		t-3	-	x	-	x	-	x	-	x
		t-4	-	x	-	x	-	x	-	x
Downstream	Volume, Speed, and Occupancy	t	x	x	x	x	x	x	x	x
		t-1	x	x	x	x	x	x	x	x
		t-2	x	x	x	x	x	x	x	x
Geometric Information	Entrance Ramp		-	-	x	x	-	-	x	x
	Exit Ramp		-	-	x	x	-	-	x	x
	Lane Added		-	-	x	x	-	-	x	x
	Lane Merger		-	-	x	x	-	-	x	x

Legend: x - Input was included in the model
 (-) - Input was not included in the model
 F - Feed forward network

The architectures of these non-linear feed forward ANN models developed are shown in Figures 12 through 19. For a given initial random weights and bias, a single hidden layer model with 18 input units and 8 hidden neurons took about 8-hours to be trained in a Pentium 166 Mhz with a 32 MB of RAM memory. The training time increased exponentially as the number of network connections (weights and bias) increased.

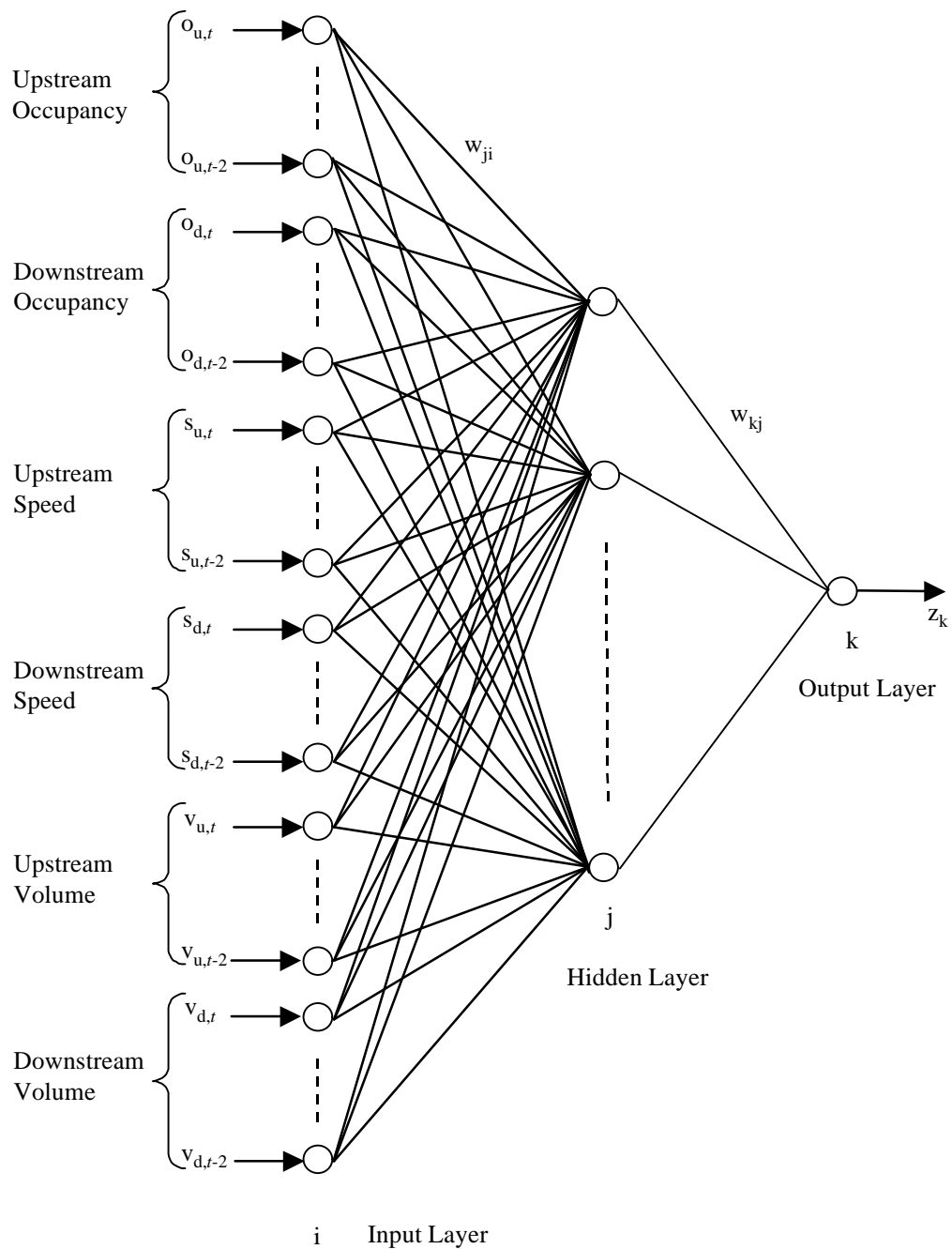


Figure 12. Model 1-F :- Feed Forward ANN with Single-Hidden Layer and 18 Input Variables.

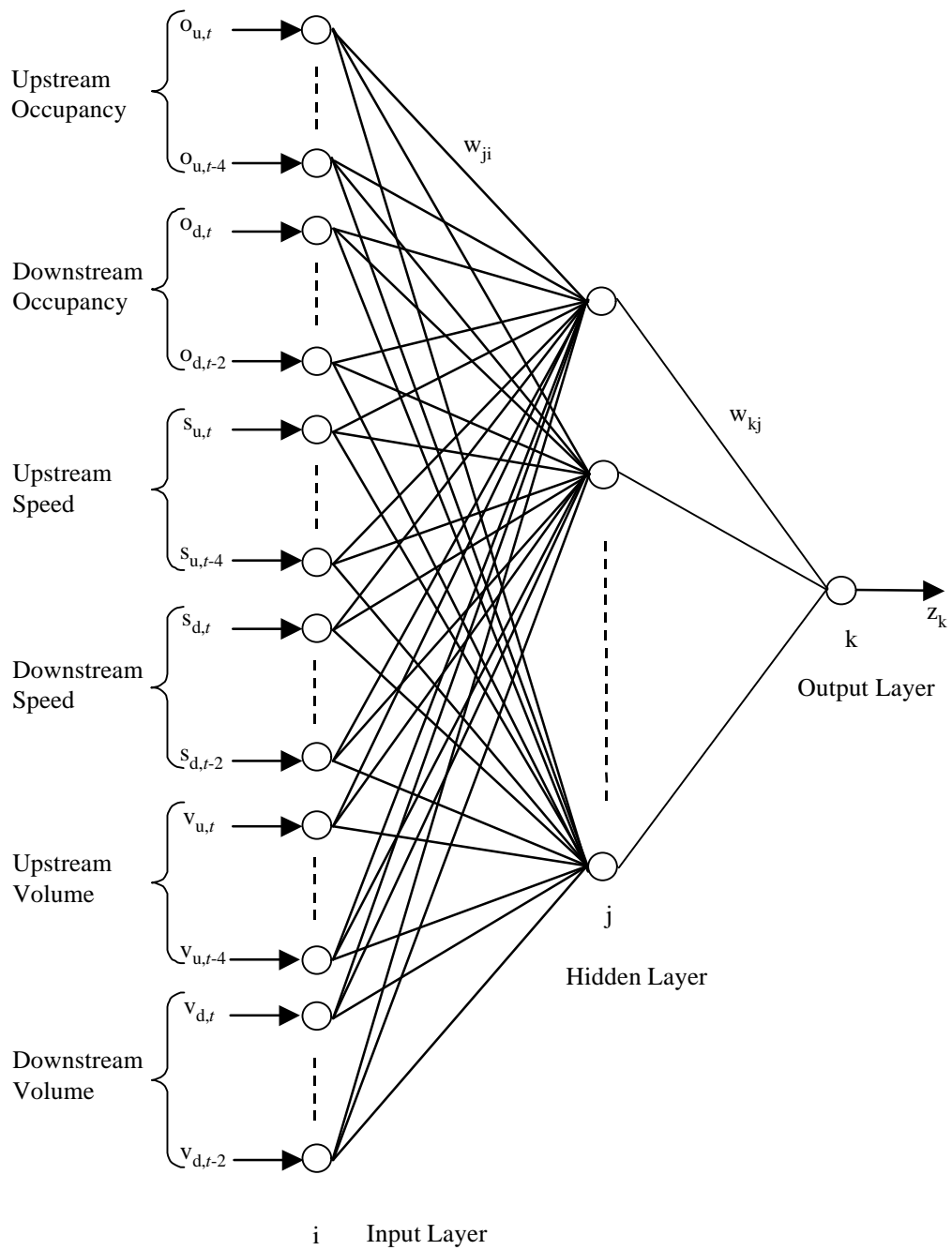


Figure 13. Model 2-F :- Feed Forward ANN with Single-Hidden Layer and 24 Input Variables.

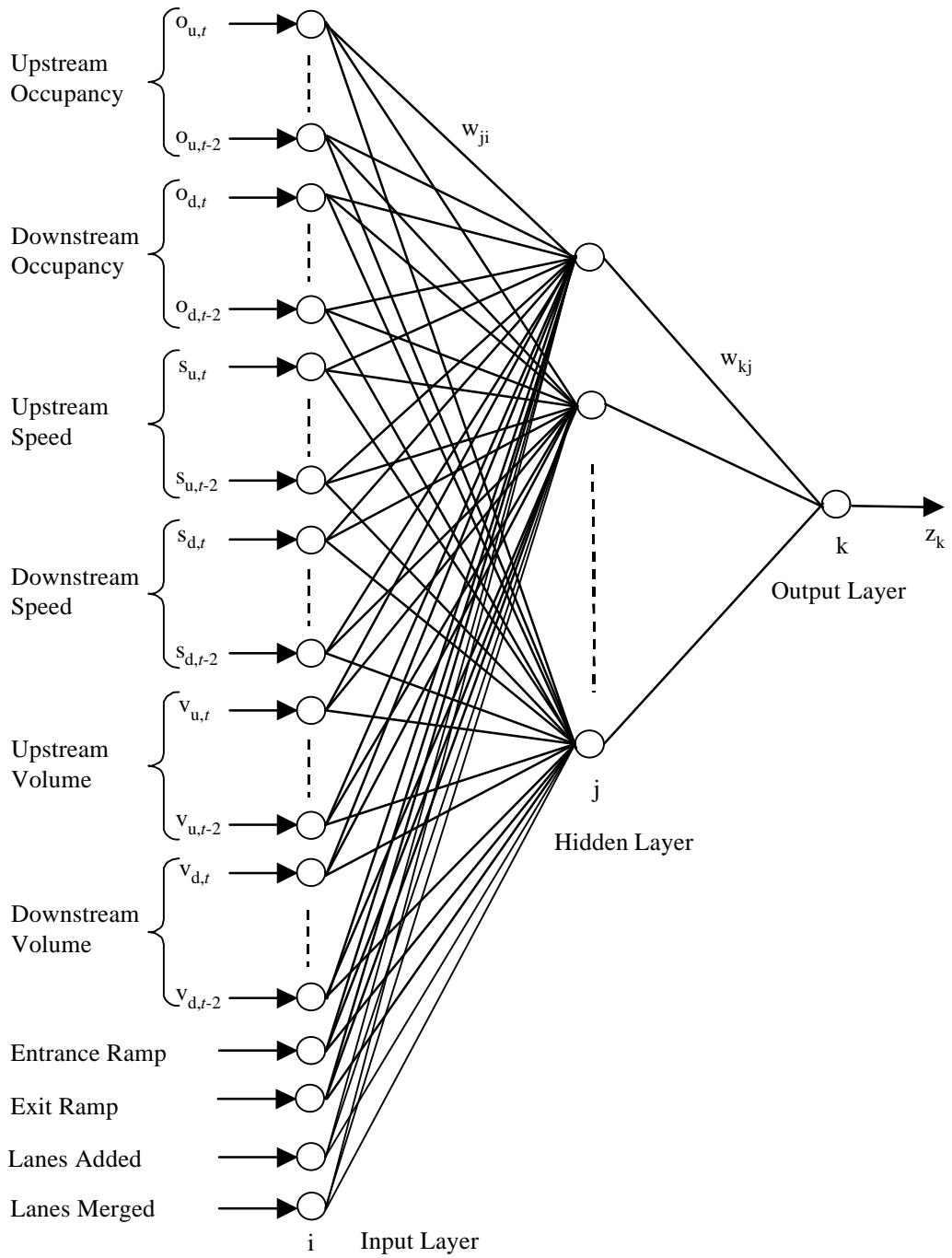


Figure 14. Model 3-F :- Feed Forward ANN with Single-Hidden Layer and 22 Input Variables.

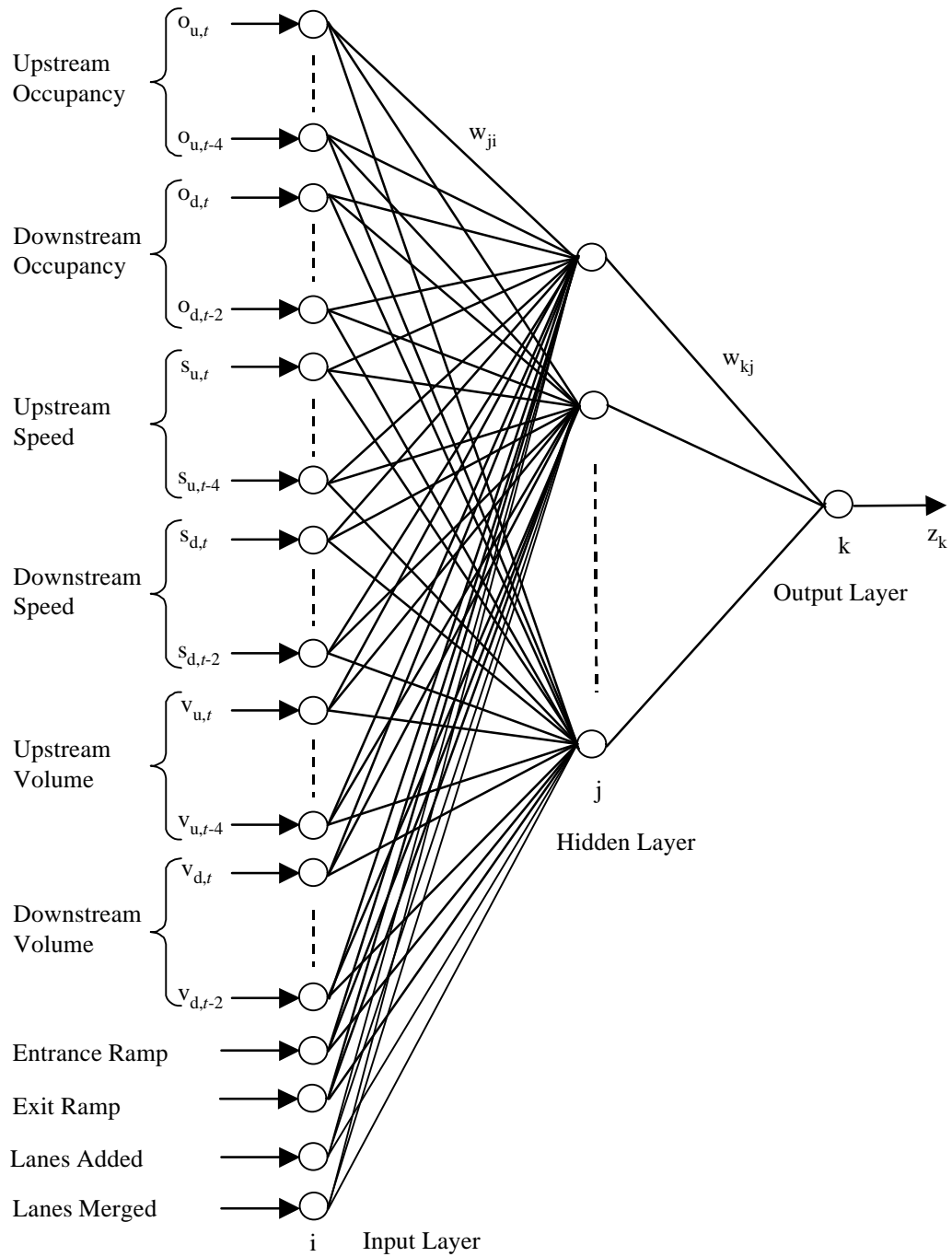


Figure 15. Model 4-F :- Feed Forward ANN with Single-Hidden Layer and 28 Input Variables.

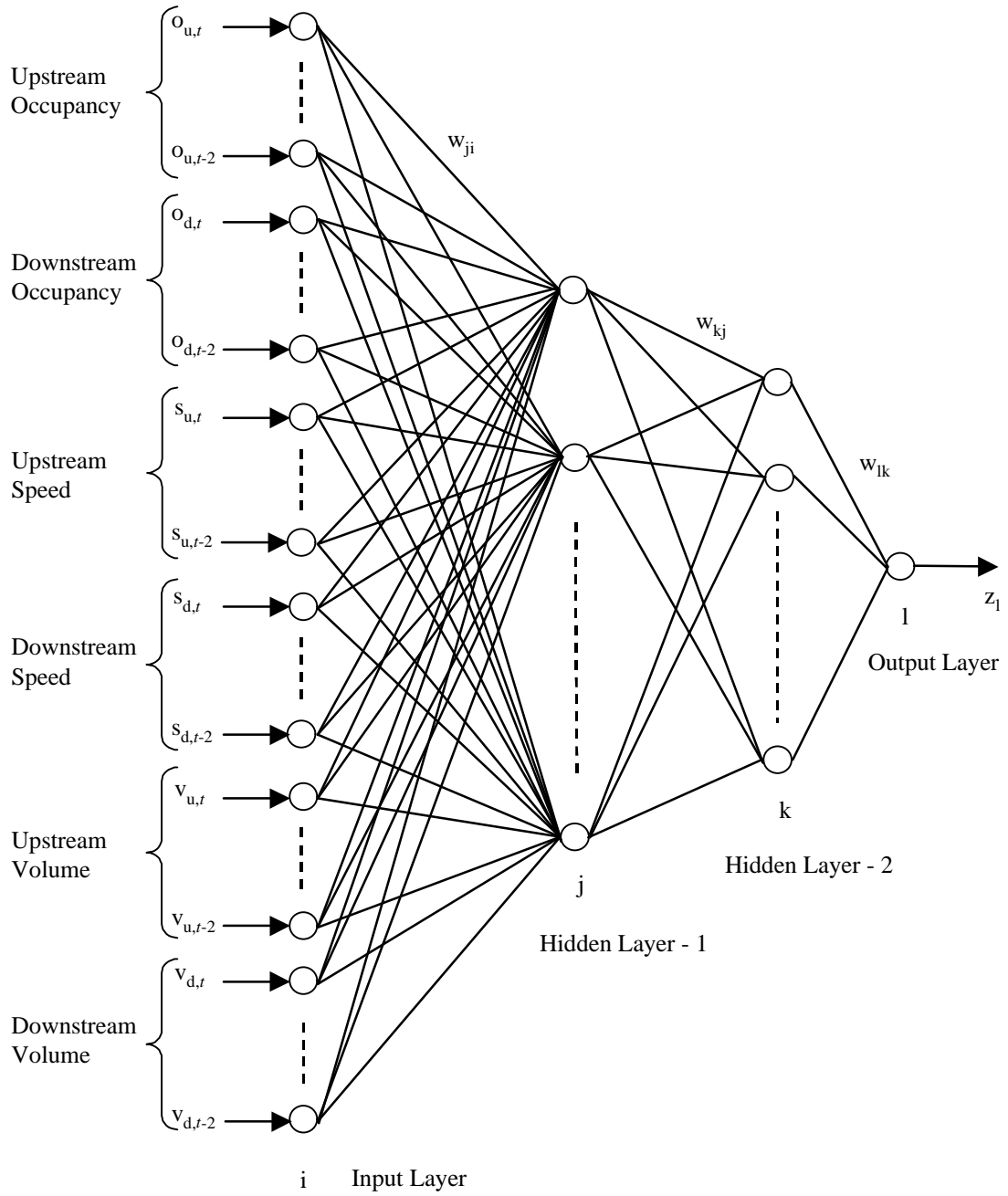


Figure 16. Model 5-F :- Feed Forward ANN with Two-Hidden Layers and 18 Input Variables.

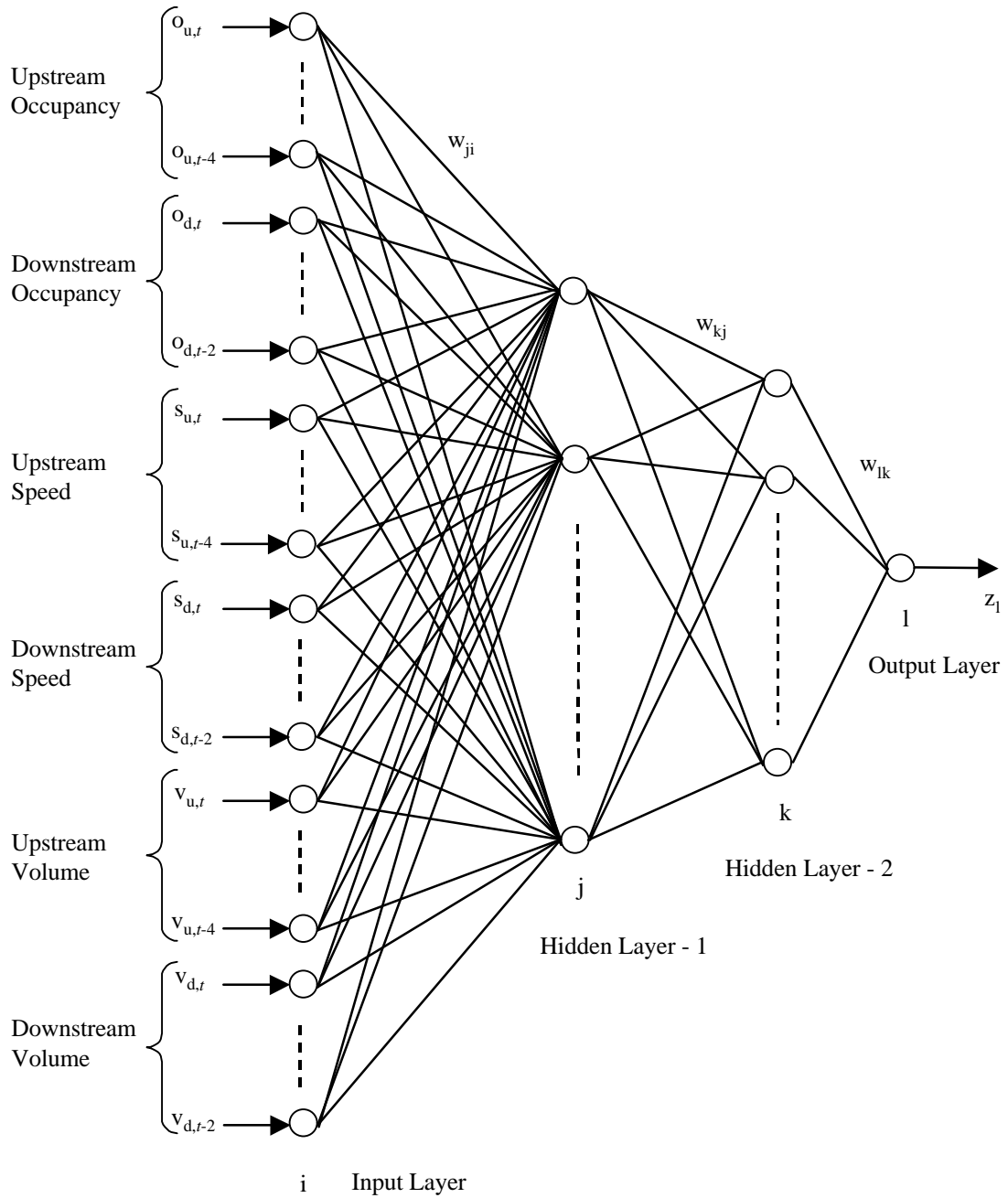


Figure 17. Model 6-F :- Feed Forward ANN with Two-Hidden Layers and 24 Input Variables.

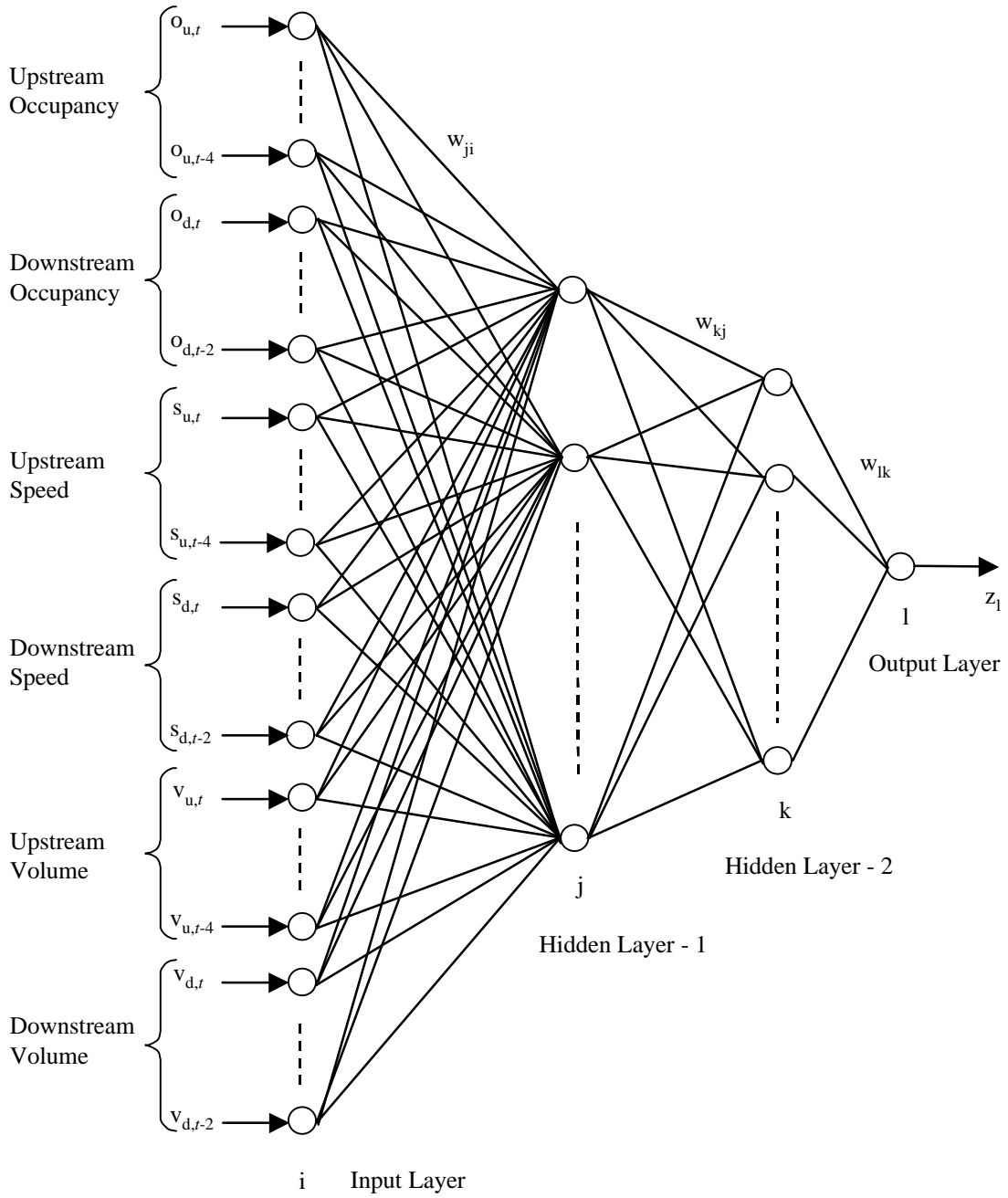


Figure 18. Model 7-F :- Feed Forward ANN with Two-Hidden Layers and 22 Input Variables.

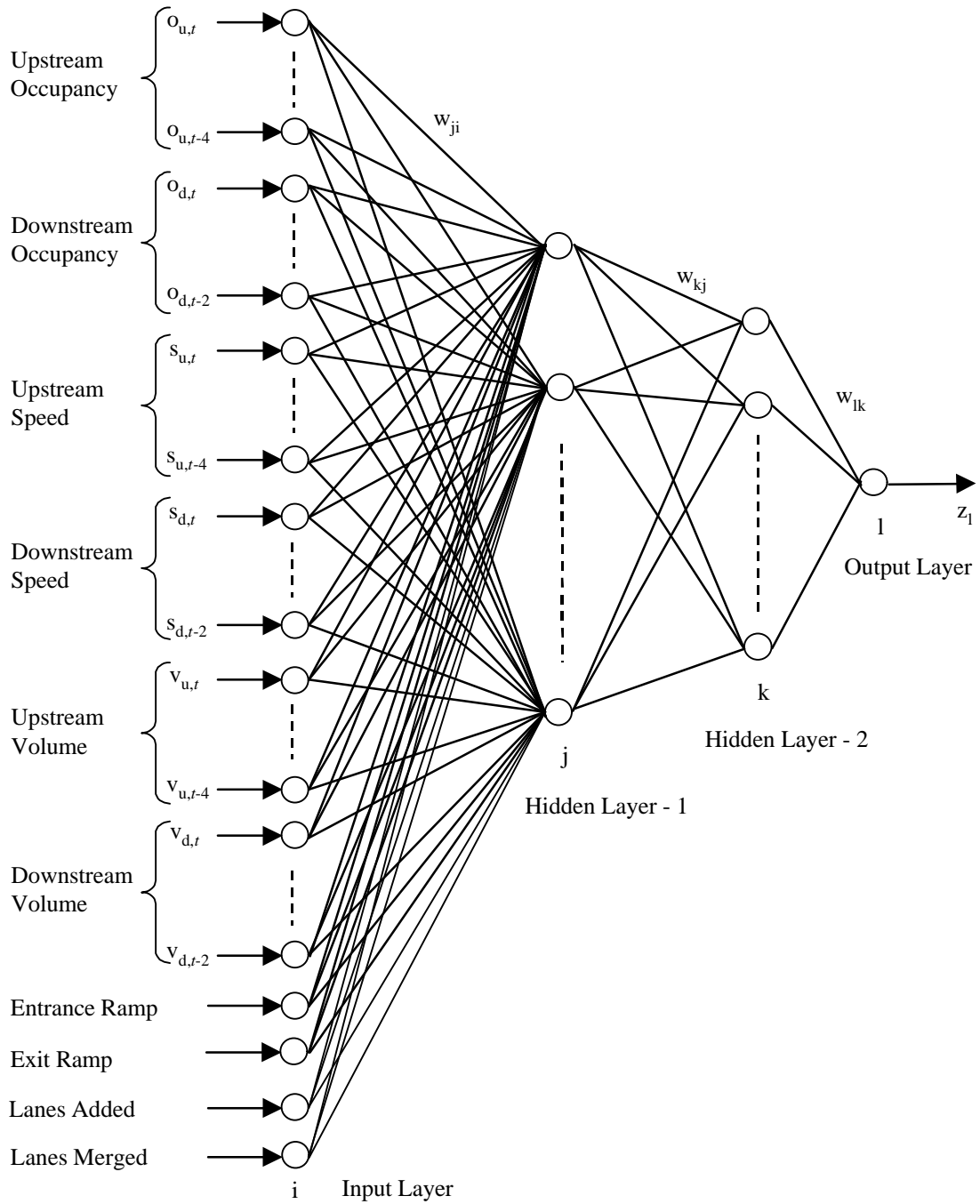


Figure 19. Model 8-F :- Feed Forward ANN with Two-Hidden Layers and 28 Input Variables.

5.4.2 Recurrent Networks

Recurrent networks have been in use in other areas of transportation engineering such as automated car controls (Neusser et al. 1991), and modeling schedule deviation of buses (Kalaputapu and Demetsky 1995). Effect of traffic condition ('incident free' or 'incident present') during previous time interval on that at current time interval was studied in this network architecture. The network was a combination of feed forward structure and a recurrent loop that ran from the output layer neuron to a new input layer neuron. ANN's prediction on traffic condition based on the input patterns at previous time step was rerouted from the output neuron to the input layer through the recurrent loop.

The recurrent networks developed in this study had an extra input node compared to its counterpart feed forward networks that were discussed in the previous section. ANN models with recurrent loop architecture were trained and tested utilizing the same procedure discussed in the previous section. The input combinations used in the model development and their nomenclature are listed in Table 3. The recurrent models developed in the study are shown in Figures 20 through 24.

Table 3. Model Input Variables and Nomenclature for Recurrent Networks

	Model Inputs	Time Interval	Single-Hidden Layer Networks				Two-Hidden Layer Networks			
			1-R	2-R	3-R	4-R	5-R	6-R	7-R	8-R
Upstream	Volume, Speed, and Occupancy	t	x	x	x	x	x	x	x	x
		t-1	x	x	x	x	x	x	x	x
		t-2	x	x	x	x	x	x	x	x
		t-3	-	x	-	x	-	x	-	x
		t-4	-	x	-	x	-	x	-	x
Downstream	Volume, Speed, and Occupancy	t	x	x	x	x	x	x	x	x
		t-1	x	x	x	x	x	x	x	x
		t-2	x	x	x	x	x	x	x	x
Geometric Information	Entrance Ramp		-	-	x	x	-	-	x	x
	Exit Ramp		-	-	x	x	-	-	x	x
	Lane Added		-	-	x	x	-	-	x	x
	Lane Merger		-	-	x	x	-	-	x	x

Legend: x - Input was included in the model
 (-) - Input was not included in the model
 R - Feed forward network

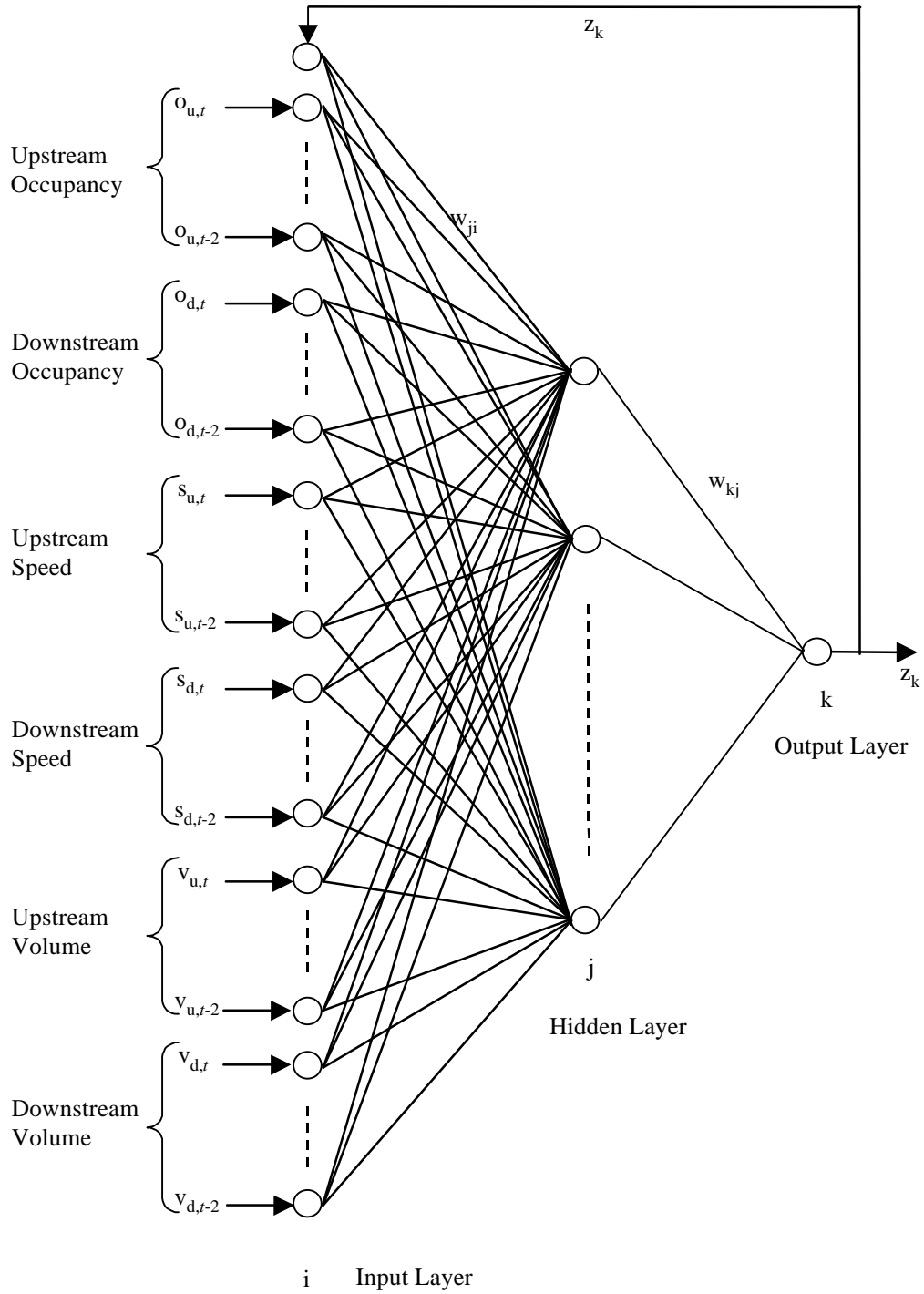


Figure 20. Model 1-R :- Recurrent ANN with Single-Hidden Layer and 18 Input Variables.

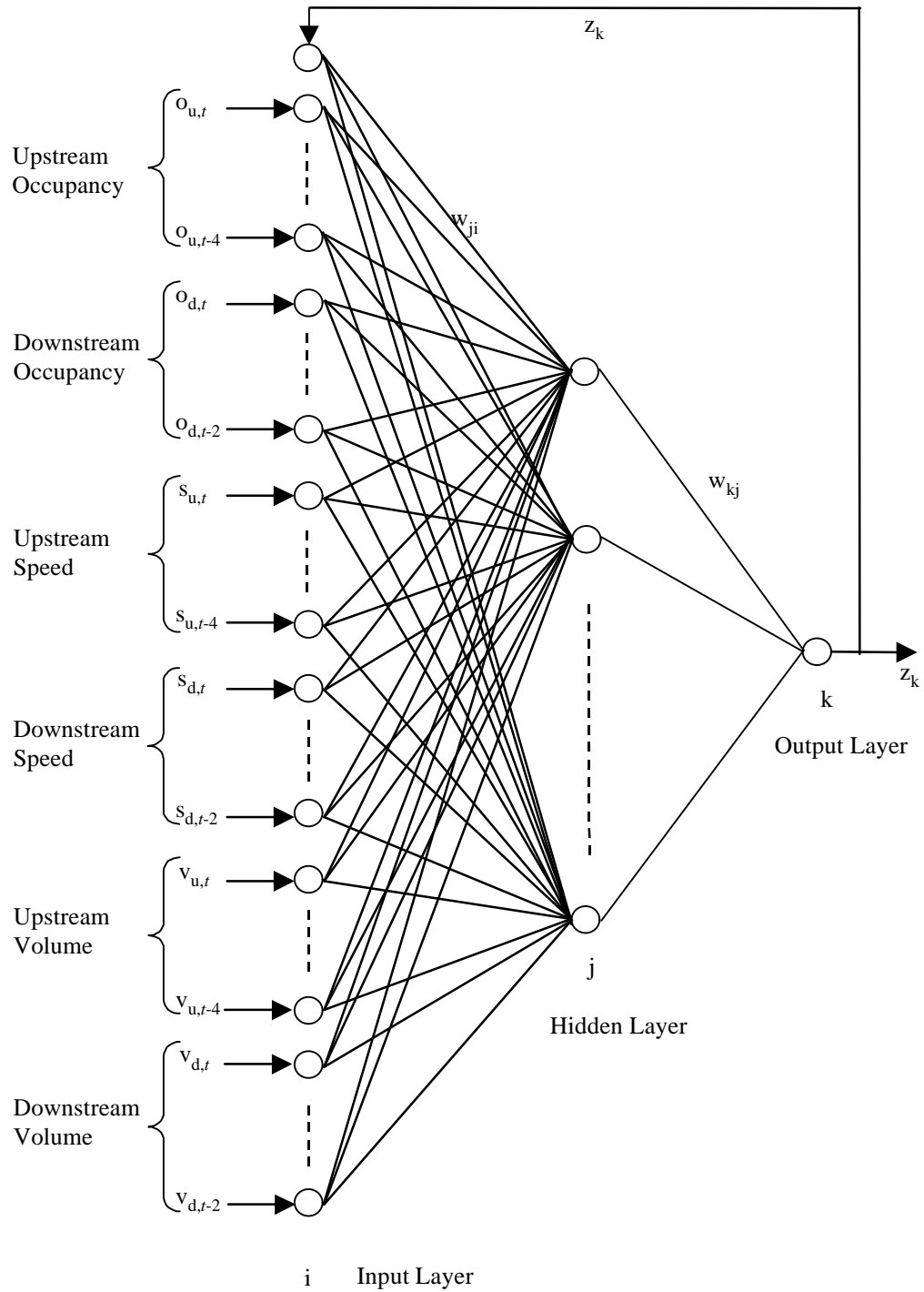


Figure 21. Model 2-R :- Recurrent ANN with Single-Hidden Layer and 24 Input Variables.

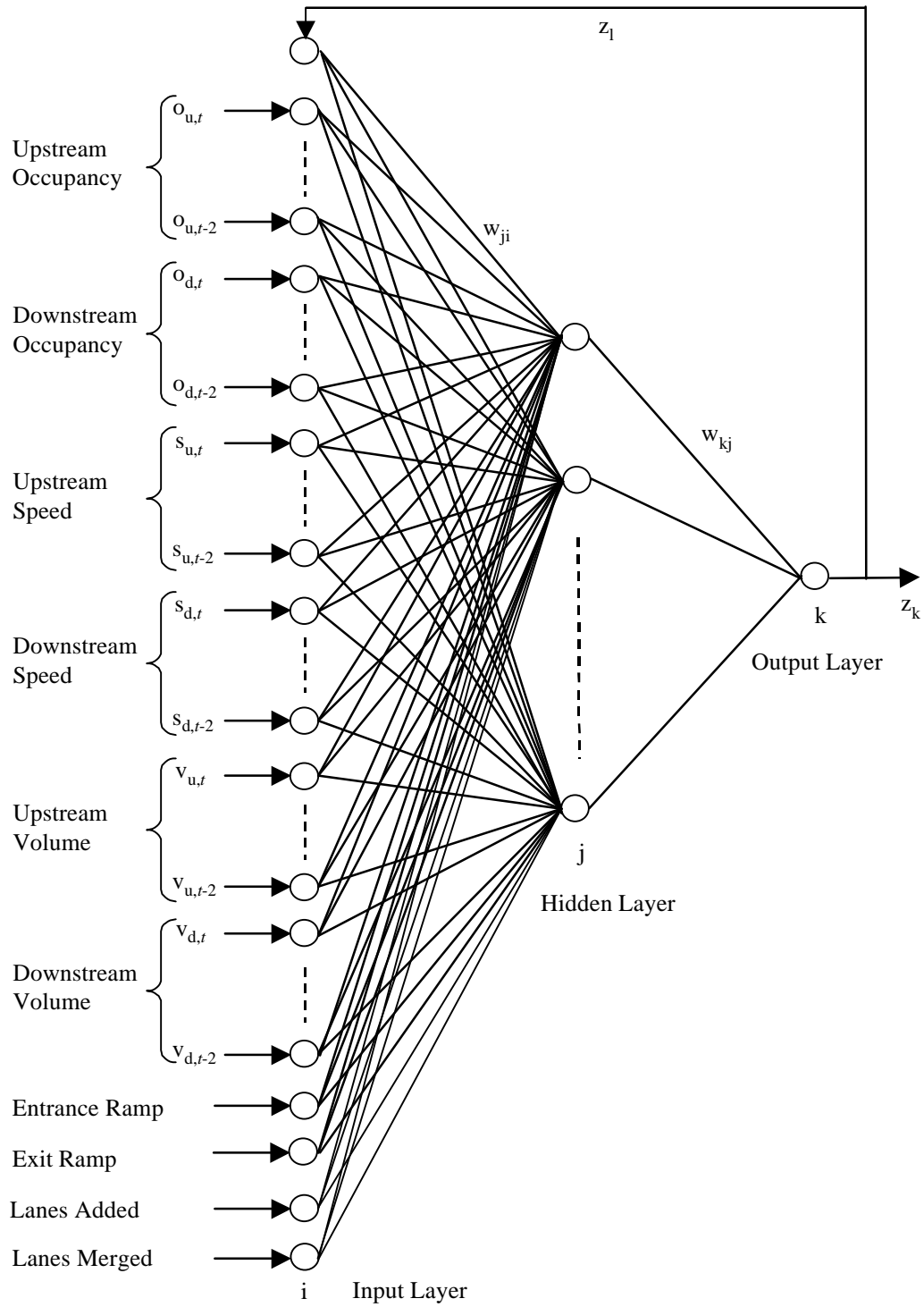


Figure 22. Model 3-R :- Recurrent ANN with Single-Hidden Layer and 22 Input Variables.

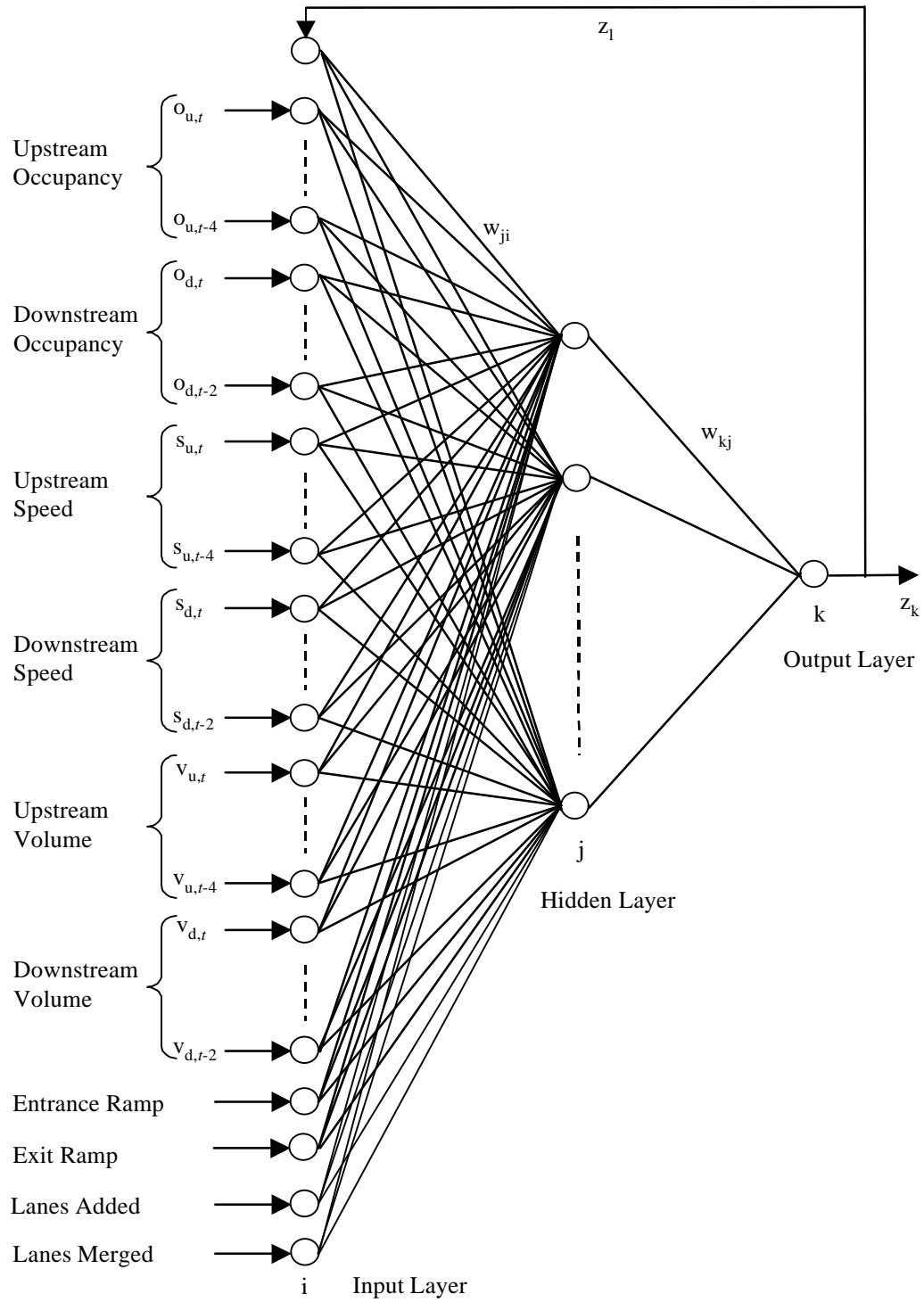


Figure 23. Model 4-R :- Recurrent ANN with Single-Hidden Layer and 28 Input Variables.

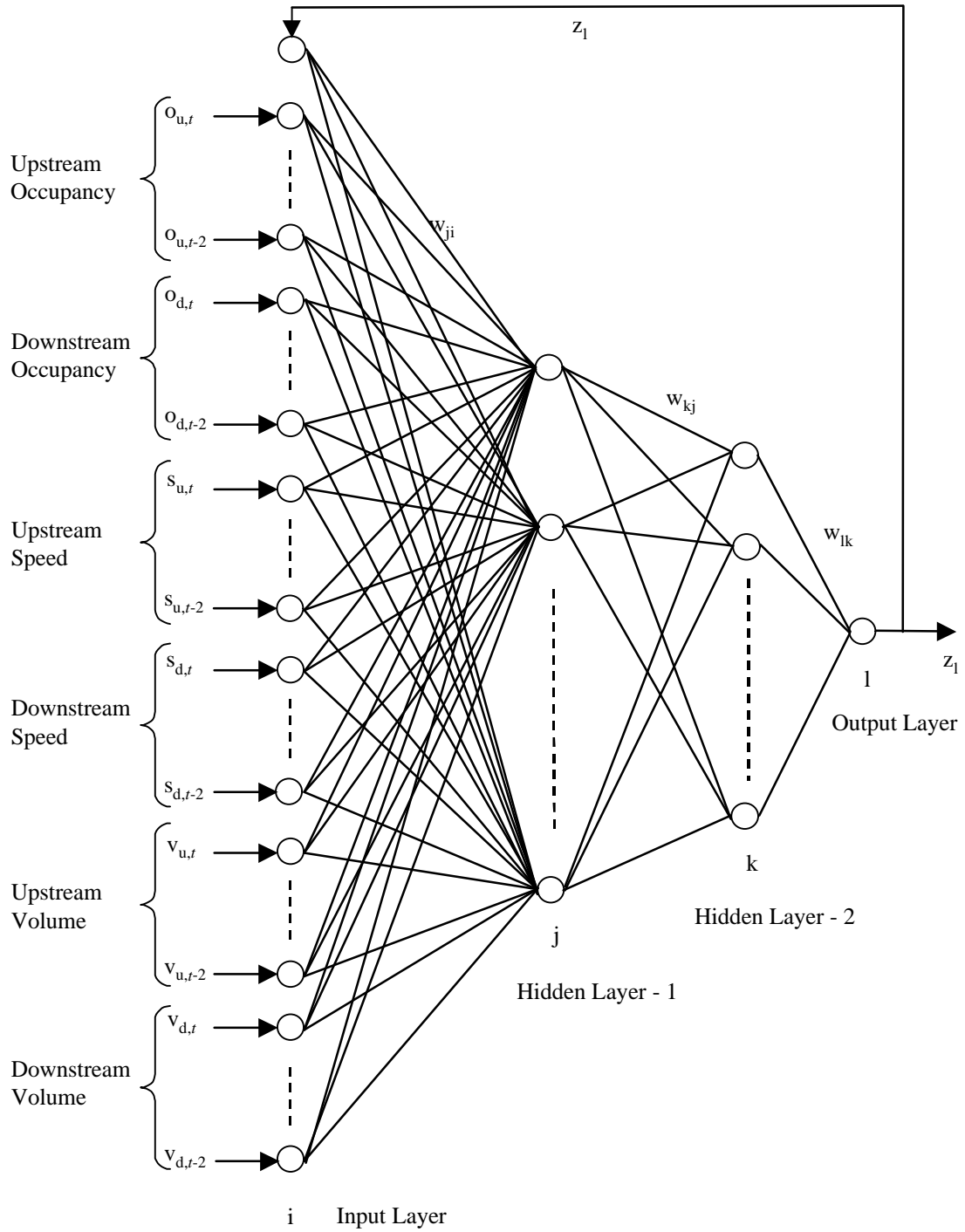


Figure 24. Model 5-R :- Recurrent ANN with Two-Hidden Layers and 18 Input Variables.

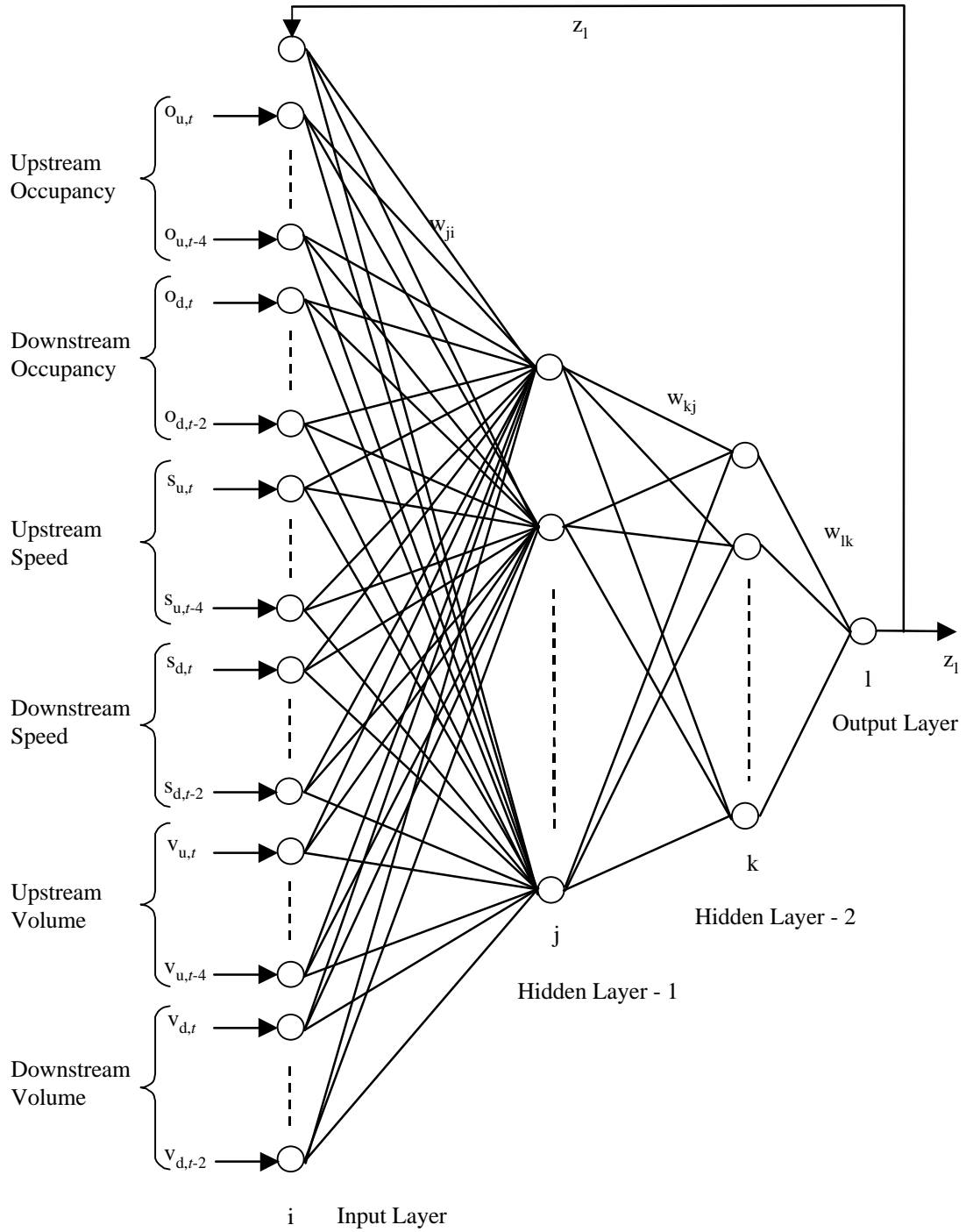


Figure 25. Model 6-R :- Recurrent ANN with Double-Hidden Layer and 24 Input Variables.

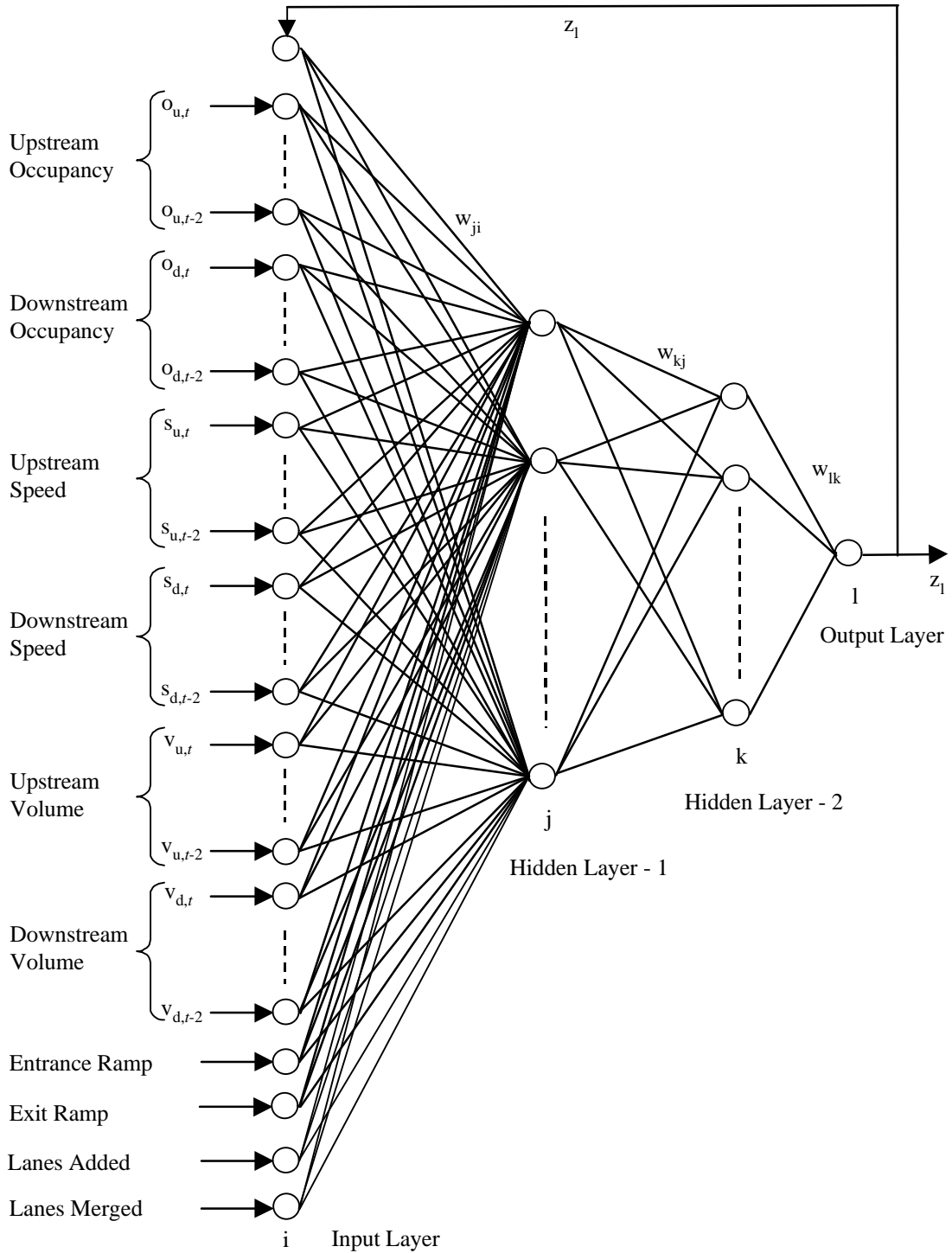


Figure 26. Model 7-R :- Recurrent ANN with Double-Hidden Layer and 22 Input Variables.

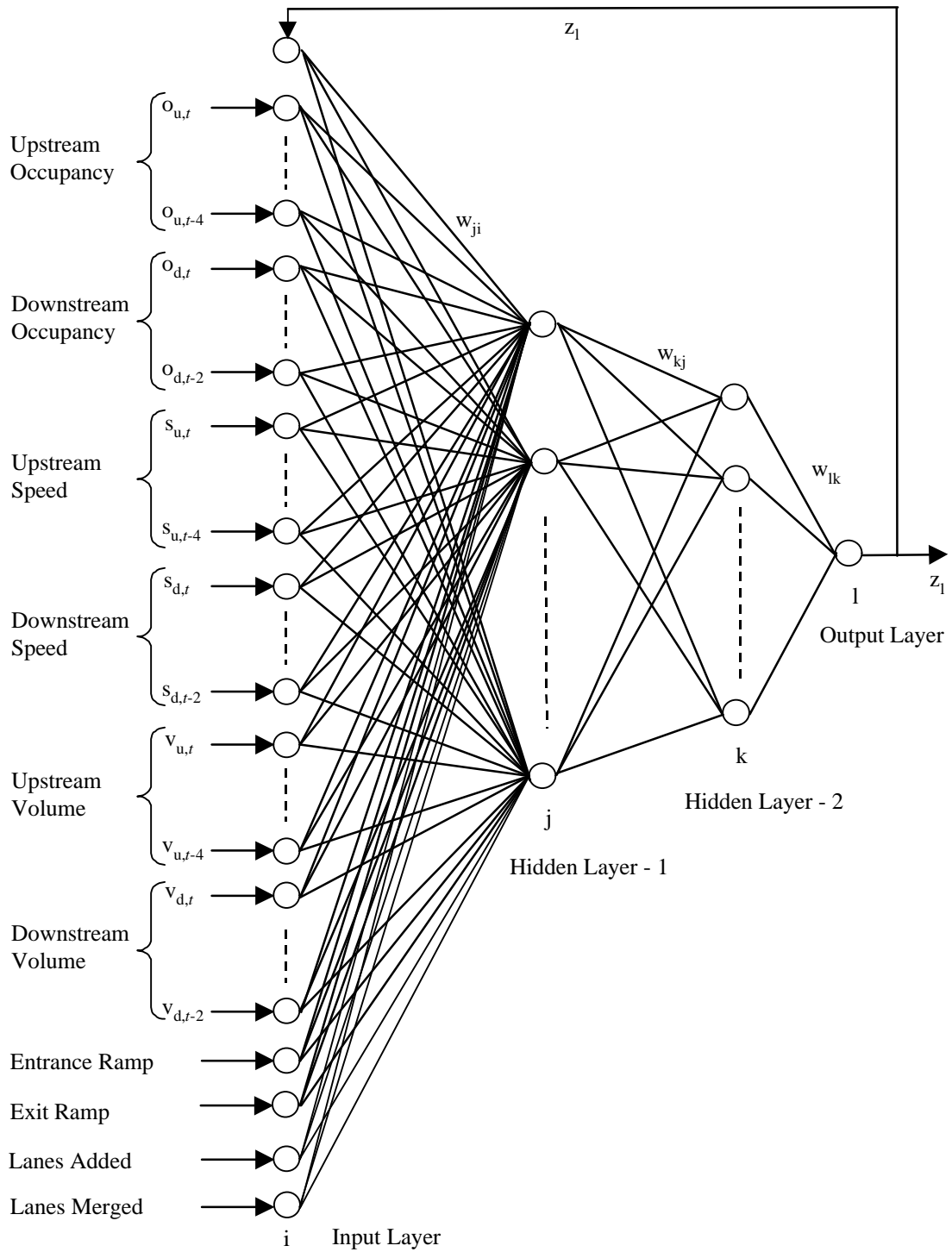


Figure 27. Model 8-R :- Recurrent ANN with Double-Hidden Layer and 28 Input Variables.

CHAPTER 6

PERFORMANCE MEASURES AND TRAINING RESULTS

6.1 Performance Evaluators of AID Models

AID models developed so far primarily depend on traffic data collected from inductive loop detectors buried under pavements. The data collected during predetermined time intervals (e.g., 20 sec, 30 sec, 60sec, etc.) is then averaged at a Traffic Management Center (TMC). Depending on the detection model requirements, the traffic data may be averaged for each individual lane or over all lanes in each direction. Each set of traffic data pertaining to a single time interval is referred to as a traffic pattern. When the processed traffic pattern and other types of data (for example; information about presence of exit and entrance ramps, change in number of lanes, or weather condition, etc.) are used as model inputs, the input data combination is referred as an input pattern in the discussion.

A trained automatic incident detection method (AIDM) [whether the method is either a network model such as ANN or an algorithm such as California algorithm] has stored signatures on ‘incident present’ and ‘incident free’ conditions. Once an input pattern is produced to such an AIDM, it determines whether the given input signature pattern belongs or closely resembles to either ‘incident present’ or ‘incident free’ signature conditions. If the AIDM’s determination is that the input signature belongs or closely resembles to a signature

with ‘incident present’ condition, then the output of AIDM is interpreted as a ‘1,’ (as in the this study) indicating incident traffic condition. Otherwise, the output of AIDM is interpreted as a ‘0,’ indicating an incident free normal traffic condition. (Different methods may use other ways of representing the traffic condition: ‘2’ instead of ‘1’ to represent ‘incident present’ condition and ‘1’ instead of ‘0’ to represent ‘incident free’ condition.) One may choose to issue an incident alarm just as the AIDM detects the incident condition. If an incident condition exists in the actual traffic flow and the AIDM recognizes the incident condition, the AIDM is said to have detected the incident pattern (or signature). On the other hand, if the actual traffic flow is incident free and AIDM recognizes it as an incident condition, the AIDM is said to have detected a false incident. The goal of any AIDM is to detect as many incident patterns as possible while lowering the number of false incidents detected. Typically, the performance of AIDM is measured by detection rate, false alarm rate, and mean time to detect incidents. The detection rate measures the ability of AIDM to detect actual incidents. The false alarm rate measures the rate at which AIDM makes false incident detection. Speed and efficiency of AIDM in detecting actual incidents are measured by the mean time to detect incidents.

6.1.1 Conventional Detection Rate

A careful review of the literature revealed that previous researchers have used a conventional detection rate with the following definition:

$$DR = \frac{TID}{TISD} \quad (21)$$

where DR is the detection rate, TID is the total number of incidents detected by the AIDM, and TISD is the total number of incidents in the sample database.

Even though this definition implies a capability of an AIDM to detect incidents, it may not characterize the true ability and performance of AIDM in detecting incident patterns. For example, if the AIDM detects incident conditions at least once in 90 of 100 incidents, the estimated detection rate according to the above definition is 90%. An ideal AIDM should be able to distinguish every incident pattern from incident-free patterns. If the goal of training an AIDM is just to identify incident conditions at least once sometime during an incident, the model may not be able to distinguish majority of the incident patterns during training. Moreover, the AIDM may not be able to detect majority of incident patterns during implementation as well. To overcome this dilemma, DR can be modified to evaluate AIDM's performance in detecting incident patterns rather than just a mere number of incidents.

6.1.2 Modified Detection Rate

It should be noted that for a given input pattern, AIDMs can only detect whether an incident condition exist or not. They (at least the AIDMs that have been developed so far) cannot detect whether single or multiple incidents have occurred in the freeway section. Therefore, when measuring the AIDM's performance of detecting incidents, a more realistic gauge should be utilized. Depending on the severity of an incident, traffic may be affected for a few minutes to several hours. Traffic data collected during this period would yield tens, hundreds, or even thousands of traffic patterns that reflect incident conditions. When AIDM is being used in a TMC, traffic operators will receive an incident alarm as an incident is

detected. Since the AIDM is designed to detect incident patterns and issue incident alarms, the AIDM should at least continue to issue incident alarms until the incident is cleared (in this case until traffic becomes normal). To the traffic operator, continuous alarms would indicate the duration and severity effects of the incident.

Therefore, a modified detection rate based on detected incident patterns that may better reflect the actual application of an AIDM in a TMC environment was proposed in the study. The proposed performance measure would capture the traffic operator's experience with an AIDM in a more realistic manner. The modified detection rate is defined by:

$$DRIP = \frac{TIPD}{TIPSD} \quad (22)$$

where DRIP is the detection rate of incident patterns, TIPD is the total number of incident patterns detected by the AIDM, and TIPSD is the total number of incident patterns in the sample database.

The importance of this definition can be illustrated with the earlier example. Suppose the 100 incidents have 5,000 incident patterns in total and that the AIDM detects a single incident pattern each (just enough to issue incident alarms) in 90 incidents, the detection performance based on the modified ratio, DRIP, is estimated to be 1.8%. However, if the model detects 40 incident patterns each in 90 incidents, the detection performance is increased to 72%. Even though the conventional DR in both situations would be 90%, the DRIP can capture the detection performance measure in a more realistic manner. That shows that the DRIP truly measures the detection performance of the model.

6.1.3 Other Performance Measures

The other important measures in use to assess AIDM performance are the false alarm rate (FAR) and mean time to detect incidents (MTD). A false alarm is said to have occurred if the AIDM issues an incident alarm when an incident free condition is reported in the database. The FAR is defined as follows:

$$\text{FAR} = \frac{\text{TFA}}{\text{NTA}} \quad (23)$$

where FAR is the false alarm rate of the AIDM, TFA is the total number of false alarms issued by the AIDM, and NTA is the total number of times the AIDM was applied.

Cheu and Ritchie (1994) proposed a slightly different version of FAR. Instead of estimating the false alarm rates for the entire database with several incidents, the author proposed an estimate of false alarms per detected incident (FAPDI), and used the FAPDI to find best ANNs for a given number of hidden layer neurons in the author's study. To be consistent with the performance measures in model training, testing, and evaluation phases, the FAR defined in the equation 23 was used to estimate false alarm rates throughout this study.

Upon receiving an incident alarm from the AIDM, if the TMC operators dispatch emergency response team (police, fire rescue, ambulance, and tow trucks, etc.) to the field and finds that the incident alarm is fault, the time and cost associated with the effort is totally wasted. Moreover, the TMC operators and those involved in the emergency rescue team would have less confidence on the AIDM should it results more and more false alarms.

Therefore, false alarms are considered more serious type of error than the error of not detecting an actual incident condition.

The time between the actual occurrence of incident and the time it is detected by an AIDM is a measure of how fast the AIDM is in detecting incidents. It is important that this time gap is kept to a minimum for an efficient incident response. The mean time to detect (MTD) incidents is estimated by summing up the time an AIDM takes from actual occurrence of an incident to issue an incident alarm and then dividing it by the total number of detected incidents in the database as follows.

$$MTD = \frac{STT}{TISD} \quad (24)$$

where MTD is the mean time to detect an incident in the field, STT is the summation of total time AIDM takes to detect each incident, and the TISD is the total number of incidents in the sample database.

6.1.4 Persistence Checks

Often times, when an AIDM detects an incident condition in the traffic flow, researchers prefer AIDM to wait for a certain time interval to issue an incident alarm. If the detected incident condition persists for a predetermined time period (e.g., for 60, 90, or 120 seconds), then the AIDM is allowed to issue an incident alarm. During the predetermined period, AIDM may issue either incident free or tentative incident condition. This process of delaying incident alarms is also known as persistence checks. Persistence check is devised

essentially to reduce false alarms. As a by-product, however, persistence check reduces the detection rate as well.

6.2 ANN Model Training Results

Sixteen different ANN architectures were developed with traffic flow and geometric variables as inputs. The learning rate of the models was varied between 0.04 and 0.06 and the momentum term was varied between 0.9 and 0.7 in different training sessions to optimize training of each model. The training coefficients are in the same order of magnitude compared to previous ANN studies on incident detection. When selecting a best model for a given network architecture (i.e., for a given number of input nodes, hidden layers, and neurons in each hidden layer) the model was trained with different initial weights that were randomly assigned. Depending upon the error curve of the model and the location of the initial weights on the curve, training may land the model in a local minima (despite using a momentum term) or in the global minima (the desired location for the model to land at the end of training). To ensure that the model results did not come from a trained model stuck in a local minima, each model was trained with different initial weight combinations until the error of the trained model became stable. The model training with different initial weights continued until the standard deviation of the error of trained models became stable. For model 3-F with 12 neurons in a single hidden layer, this result is listed in Table 5.

The first column indicates the number of model training trials with different initial weights. The number of each instance when the model did not detect an actual incident pattern or detected an incident condition in absence of an actual incident condition was

Table 5. Training Results for 3-F model with 12 Hidden Neurons

Model Number	Conventional Detection Rate (DR)	Detection Rate of Incident Patterns (DRIP)	False Alarm Rate (FAR)	Error Rate	Standard Deviation of Error Rate	Coefficient of Variation of Error Rate
1	100%	90.37%	1.12%	0.0338	-	-
2	100%	89.68%	1.17%	0.0359	-	-
3	100%	89.81%	1.06%	0.0346	-	-
4	95.65%	89.43%	0.95%	0.0343	-	-
5	100%	90.15%	1.14%	0.0345	-	-
6	100%	89.19%	0.95%	0.0348	0.0007	0.0207
7	100%	90.02%	1.06%	0.0340	0.0007	0.0201
8	100%	89.96%	1.06%	0.0341	0.0007	0.0192
9	100%	90.09%	1.08%	0.0340	0.0006	0.0185
10	100%	90.06%	1.15%	0.0348	0.0006	0.0178
11	100%	90.21%	1.14%	0.0343	0.0006	0.0169
12	100%	89.93%	1.09%	0.0345	0.0006	0.0162
13	100%	90.06%	1.12%	0.0345	0.0005	0.0155
14	100%	89.96%	1.09%	0.0344	0.0005	0.0149

estimated. This number was divided by the number of times that the model was applied to calculate the error rate for each model trained with different initial weights. Since model training was to continue until the standard deviation of the error rate became stable and the mean of error rate could change from n models to $n+1$ models, establishing a common ground was necessary to compare the change in the standard deviation as training continued with more initial weight combinations. Therefore, instead of comparing standard deviation of the error rate itself, a coefficient of variation was estimated for comparison purposes. The coefficient of variation of the error rate in the n^{th} row under the ‘model number’ column (in Table 5) means that the coefficient was estimated from 1st through n^{th} models. The coefficient of variation from 1st through 6th models and 1st through 14th models remains very small. Furthermore, the standard deviation of the error rate remains very small from 1st through 6th models and 1st through 14th models. The model (in this case model 1) with the minimum error rate was selected as the best model for the single hidden layer with 12 neurons.

As discussed in section 5.3, each model type (e.g., model type 3-F) was tested with different number of hidden layer neurons until model performance did not indicate a significant gain with the added neurons. For a given model type, the best network for each hidden layer neuron combination was selected by comparing the error rate as described above. Tables 6 and 7 list the selected models from single and double hidden layer architectures, respectively. Each row in the Tables 6 and 7 represents the best hidden layer neuron combination for each model type that exhibited the best performance. For example, in Table 6 the best 4-F network selected had 18 hidden layer neurons with a DRIP of 91.64% and FAR of 2.30%.

Table 6. Selected Single Hidden Layer Models During Training

Model Architecture	No of Hidden Layer Neurons	DRIP	FAR	Error Rate
Feed Forward Networks				
1-F	10	59.79%	0.01%	0.0945
2-F	20	81.82%	0.46%	0.0472
3-F	12	90.37%	1.12%	0.0338
4-F	18	91.64%	2.30%	0.0426
Recurrent Networks				
1-R	10	91.30%	2.22%	0.0426
2-R	12	90.65%	1.41%	0.0361
3-R	12	90.02%	1.16%	0.0350
4-R	15	90.43%	1.61%	0.0386

In general, a model may experience two types of errors: detecting incident patterns when there is no incident, and not detecting an incident pattern when there is an incident pattern. These model errors cause reliability problems in practical applications. Therefore, when selecting a best ANN model, both errors should be taken into consideration to increase the AIDM's overall reliability. Estimated model error rates in Tables 6 and 7 reflect both of these error types. Selection of the best model was based on the error rate. A model with minimum error rate which maintains a reasonable DRIP was selected as a best ANN model to detect freeway incidents.

The output of an ANN model is usually processed with a threshold or tolerance. If the output was greater than a certain threshold (this value was taken as 0.5 in this study), an 'incident present' condition was said to have detected. By varying this threshold value,

Table 7. Selected Double Hidden Layer Models During Training

Model Architecture	No of Hidden Layer Neurons		DRIP	FAR	Error Rate
	1 st Layer	2 nd Layer			
Feed Forward Networks					
5-F	8	2	90.49%	0.67%	0.0290
6-F	16	4	92.14%	1.14%	0.0299
7-F	22	2	94.22%	1.35%	0.0270
8-F	12	2	94.38%	1.14%	0.0246
Recurrent Networks					
5-R	10	2	90.21%	2.49%	0.0479
6-R	22	2	91.45%	1.06%	0.0306
7-R	12	3	88.97%	0.93%	0.0351
8-R	8	3	97.17%	2.40%	0.0306

different performance levels can be estimated for AIDMs. The affect of the threshold value on model results can be analyzed in a form of curves known as performance envelops. For the models developed in this study, the performance curves at the end of training process are shown in Figures 25 through 27. These curves were developed by gradually varying the threshold value and analyzing the traffic condition output from each model. This calculation was performed using a program written in ‘C-language’ once the model outputs were available for the training data.

From the training results for single hidden layer ANNs shown in Table 6 and Figure 28, the 3-F model exhibits the overall lowest error rate and best performance. The feed forward model 1-F with upstream and downstream traffic data up to $t-2$ time intervals did not perform well during the training. Adding the upstream traffic data up to $t-4$ time intervals

have shown a little improvement in model 2-F over model 1-F. Adding geometric variables in the model 3-F seems to have improved the model performance. However, including upstream traffic data up to $t-4$ time intervals and geometric variables in model 4-F have not shown much improvement in overall performance. Similar observations can be made from the Table 6 for single hidden layer models with recurrent loops. Yet the best training performance is exhibited by model 3-F.

Table 7 indicates that adding geometric inputs in the model had a positive impact on the detection. The lowest error rate in double hidden layer architecture was exhibited by the feed forward model 8-F that utilize all traffic and geometric variables. Further, the added recurrent loop did not have any significant improvement in the recurrent models (except models 1-R and 2-R). Comparison between single and double hidden layer networks portrays that the double hidden layer networks perform better than the single hidden layer networks in freeway incident detection.

6.3 Calibration of Conventional Algorithms

California algorithm no. 8 and Minnesota algorithm utilize occupancy data from upstream and downstream stations and set of threshold values to detect incidents. Therefore, these two algorithms were used to compare the performance results of ANN models. In order to avoid any bias for the threshold values from previous publications and the data used, the same training data used to train ANN models were used to calibrate the two algorithms. To keep the algorithms on the same level of application, these algorithms were calibrated as non station specific algorithms.

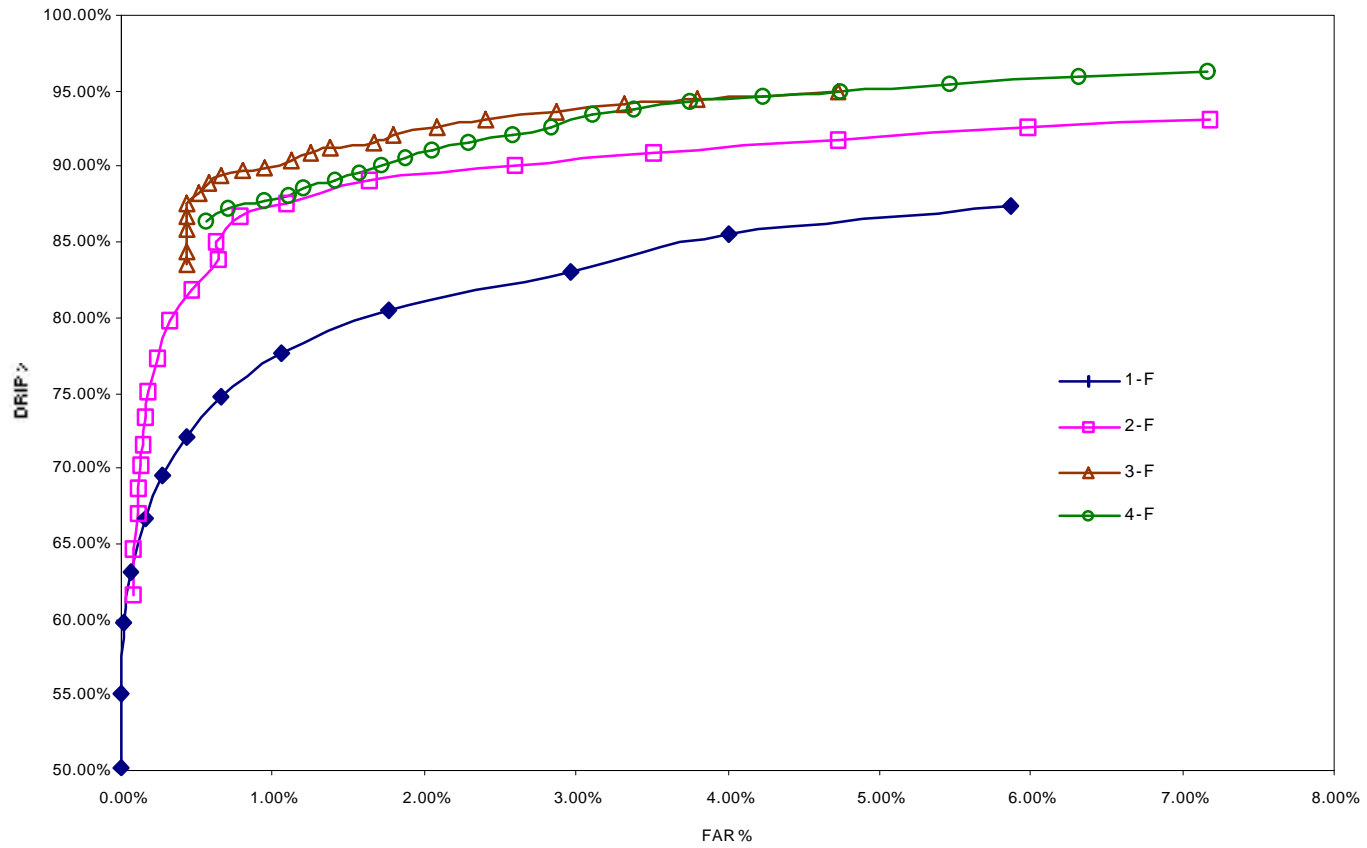


Figure 28. Performance Envelope for Single Hidden Layer Feed Forward Networks

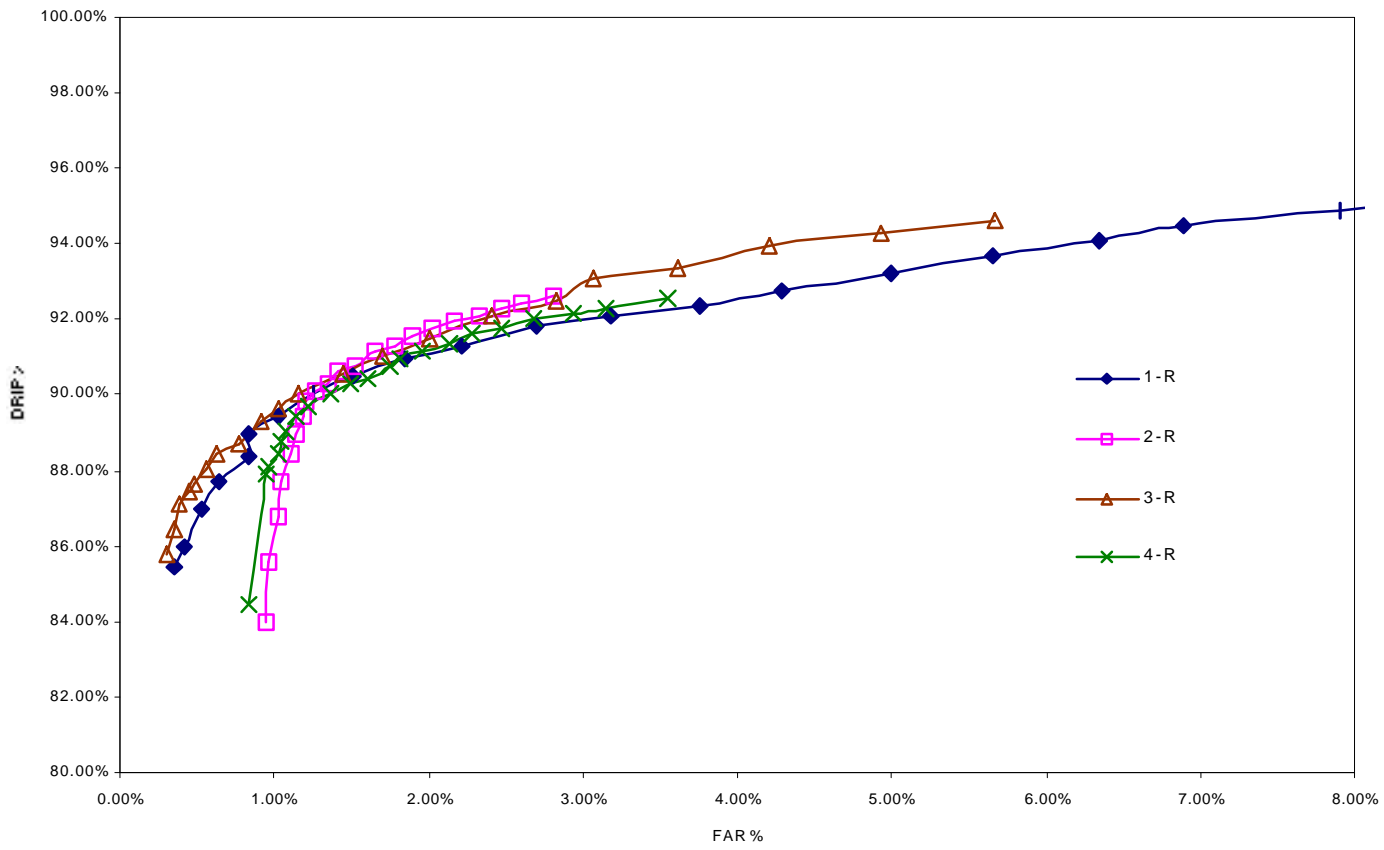


Figure 29. Performance Envelope for Single Hidden Layer Recurrent Networks

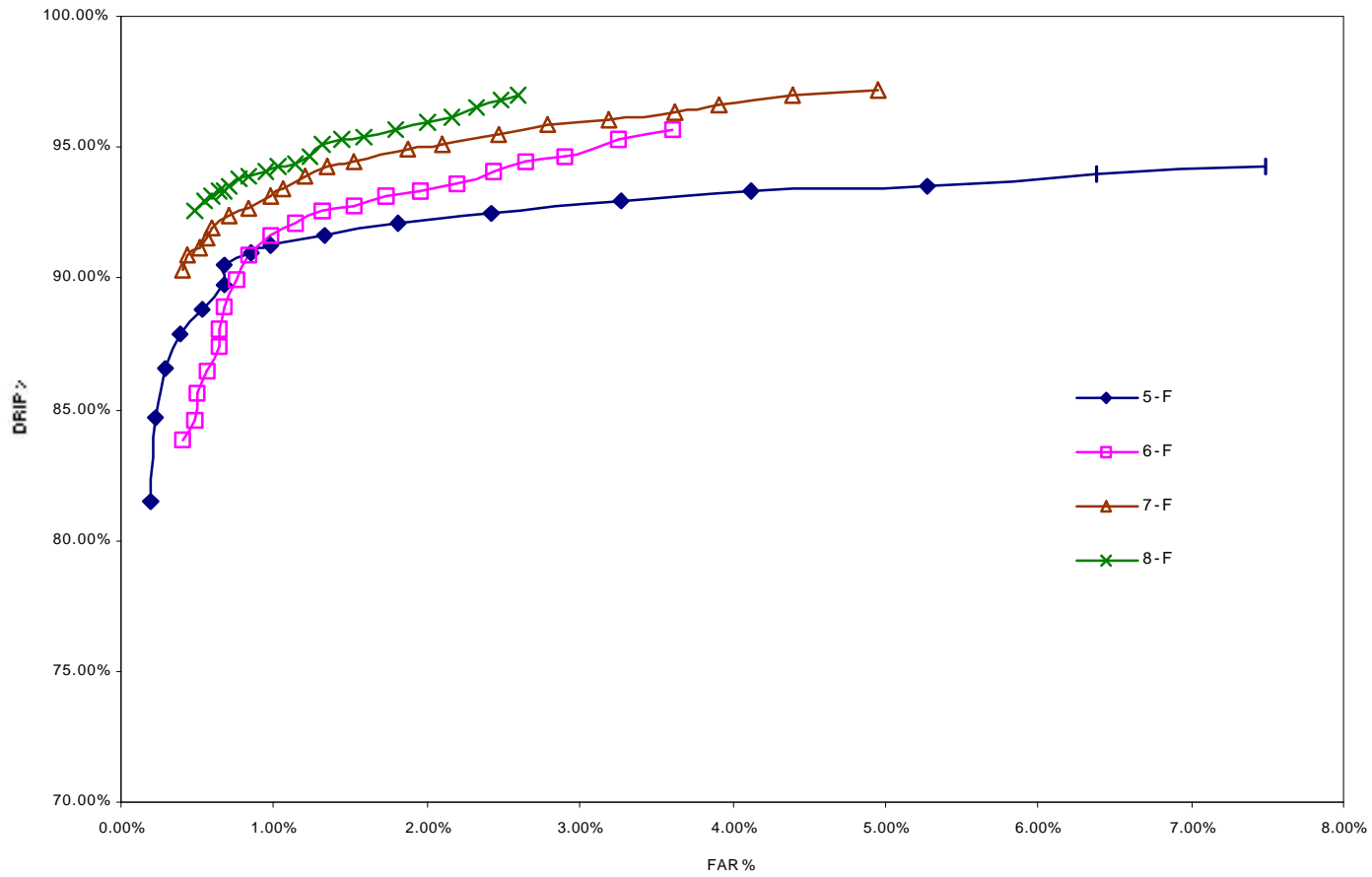


Figure 30. Performance Envelope for Double Hidden Layer Feed Forward Networks

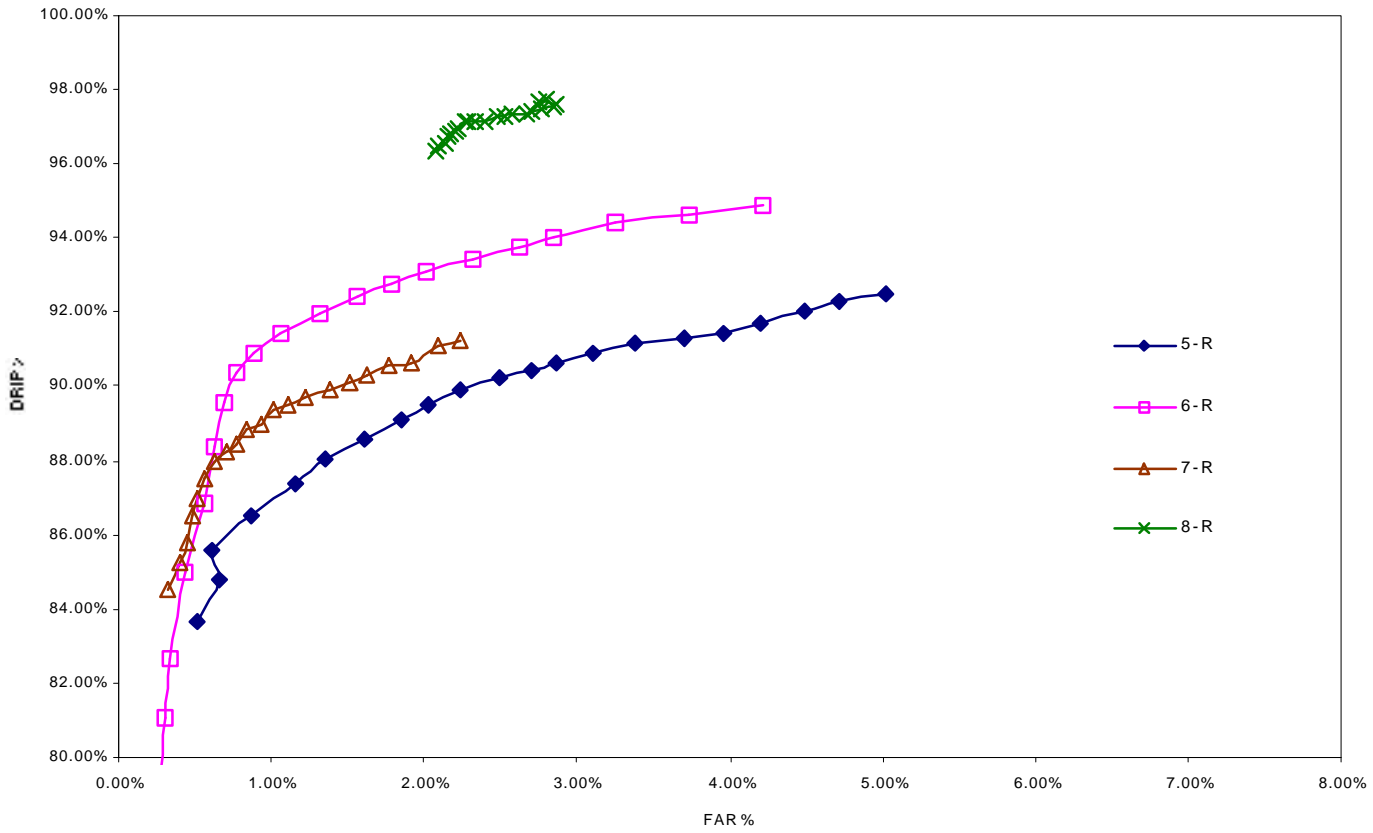


Figure 31. Performance Envelope for Double Hidden Layer Recurrent Networks

6.3.1 California Algorithm

2:07 PM To calibrate the California algorithm, 60 second occupancy averages were used at 30 second intervals. Since the algorithm has five thresholds, a large number of threshold values were required to cover possible combinations including and beyond ranges and magnitudes reported by Payne and Tignor (1978) and Arceneaux, et al. (1990). A program was written in 'C- language' to calibrate the algorithm using the training data used in ANN model training process. The threshold combinations that produced the lower error rates are listed in the Table 8. Threshold combination that yielded the lowest error rate was selected for model evaluation.

6.3.2 Minnesota Algorithm

Since the Minnesota algorithm only utilize two threshold values, the search for threshold value combinations was less arduous than for the California algorithm. Traffic data from ANN training was used to calculate the RAT1 and RAT2 values. Threshold values reported by Stephanedes and Chassiakos (1993) was used as a guide to establish a range of possible threshold combinations. A program was written in 'C-language' to search for various possible threshold combinations and the results are listed in the Table 9. The threshold combination that yielded the lowest error rate was selected for evaluating performance results with ANN models.

Table 8. Calibration Results for California Algorithm No. 8

No.	Threshold Sets					DRIP	FAR	Error Rate
	T1	T2	T3	T4	T5			
1	10.0	-0.500	0.010	20.0	20	75.57%	2.58%	0.0831
2	9.8	0.635	0.078	12.2	30	55.81%	2.41%	0.1278
3	13.4	-0.286	0.312	15.8	30	52.81%	2.58%	0.1367
4	5.0	-0.500	0.010	10.0	20	50.12%	2.57%	0.1427
5	10.0	-0.500	0.010	10.0	20	44.93%	2.17%	0.1509
6	15.0	-0.500	0.010	20.0	20	41.55%	2.26%	0.1597
7	15.8	0.645	0.248	15.2	30	37.94%	1.28%	0.1583
8	15.0	-0.500	0.010	10.0	20	34.34%	1.87%	0.1727

6.3.3 Discussion

During the calibration process of both algorithms, several thousands of threshold sets were tested for incident detection performance. Some threshold sets yielded similar performance measures in terms of DRIP. Under these circumstances, the threshold sets that resulted in lower FAR were selected for the evaluation process discussed in Chapter 7. The performance envelope for California algorithm and Minnesota algorithm during the calibration process for the selected threshold sets is shown in Figure 32. Table 9 and Figure 32 suggest that the Minnesota algorithm performs better than the California algorithm during calibration. These two algorithms with the listed threshold sets were used in the ANN model evaluation to establish the performance gain of ANN models.

Table 9. Calibration Results for Minnesota Algorithm

Threshold Sets			DRIP	FAR	Error Rate
No.	T1	T2			
1	0.20	0.20	76.35%	1.15%	0.0670
2	0.25	0.15	73.93%	0.51%	0.0663
3	0.25	0.20	71.47%	0.24%	0.0693
4	0.25	0.25	68.71%	0.24%	0.0758
5	0.30	0.15	64.67%	0.09%	0.0838
6	0.30	0.20	62.37%	0.09%	0.0892
7	0.30	0.25	59.91%	0.09%	0.0949
8	0.30	0.30	59.23%	0.08%	0.0964

6.4 Summary

Total of sixteen artificial neural networks were developed for freeway incident detection. Of the total, eight were based on feed forward architecture with single and double hidden layer models. The remaining eight models were based on recurrent architecture with single and double hidden layers. Eight different input combinations were used in the sixteen different models. The input combinations included traffic volume, speed, and occupancy data from t to $t-4$ time intervals at upstream stations, traffic volume, speed, and occupancy data from t to $t-2$ time intervals at downstream stations, and freeway geometric information (presence of entrance and exit ramps and lane expansions and mergers).

The models were trained and tested using a “A1 Backprop” software. Model training was started with random initial weights and the training was continued testing error started

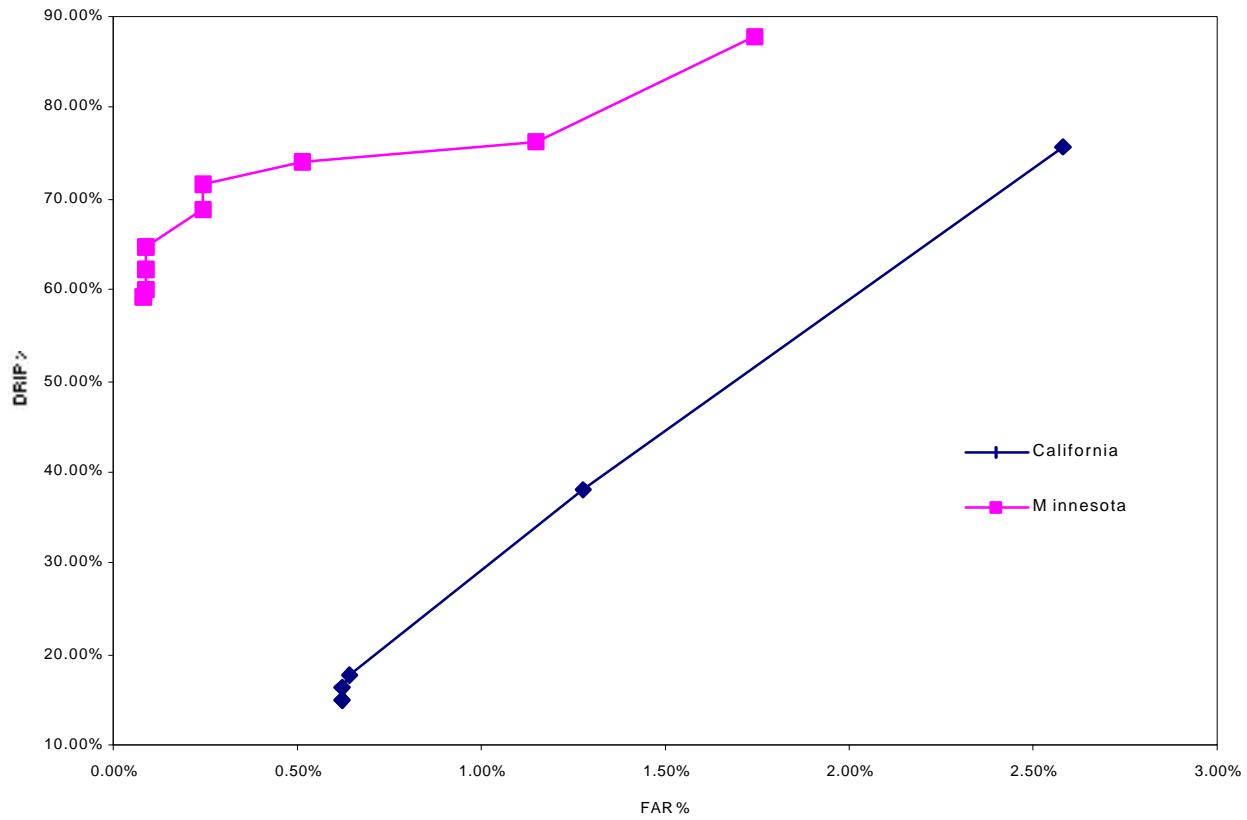


Figure 32. Performance Envelope for the California and Minnesota Algorithms.

to increase while training error was still in the decreasing trend. Each architecture type was trained with varying the number of hidden layer neurons. The best models was selected based on the lowest error rate calculated during the training process. Model training results suggested that model 8-F had the overall best performance with 94.38% of incident patterns detected at a 1.14% false alarm rate.

The California and Minnesota algorithms were also calibrated with the data used to train ANN models. Large number of possible threshold sets were tested during this calibration process. The threshold sets that resulted in lower overall error rate were selected for comparative evaluation of ANN model results.

CHAPTER 7

RESULTS OF MODEL EVALUATION

During model development process, a training data set was used to train ANN models and a testing data set was used to test each model as the training progressed. Once different ANN models were developed, the model that exhibited a better overall performance was selected as the best candidate model for freeway incident detection. The elaborate process of model development through best model selection is discussed in great detail in sections 5.4 and 6.2. Third phase of this model development procedure is the model evaluation (or validation). In the context of this study, model evaluation is referred to as the process by which a trained AIDM is assessed for its incident detection performance using a data set that is independent of the training and testing data set. In this chapter, evaluation process of the models developed in Chapter 5 is discussed.

7.1 Evaluation Data

A data set different from that used in training and testing process was used in the model evaluation process. Traffic data collected during February 16 - March 19, 1993 from I-880 freeway as described in Chapter 4 were scrutinized for any abnormalities in traffic flow data. The incident database corresponding to the February 16 through March 19, 1993 period

had 1,210 incident records collected during 6:30am-9:30am and 3:30pm-6:30pm. A total of 9,818 traffic patterns were used in the evaluation process.

7.2 Neural Network Models

The best recurrent ANN models in each architecture type developed in Chapter 6 were evaluated for their performance with an independent field data set. Performance measures such as DRIP, FAR, error rate, and MTD were estimated for the models listed in Tables 6 and 7. Moreover, model performance was estimated with 1 to 3 persistent checks.

Table 10 lists the performance results for single hidden layer feed forward models with persistent checks. Comparison of these models revealed that 1-F, 2-F, and 3-F had lower error rates than during training. Model 3-F exhibited an increase in both DRIP and FAR levels. Even though the FAR of model 3-F increased during evaluation, it could be further reduced from 0.98% to 0.61% with application of three persistence checks. As can be seen from the Table 10, when persistence check was introduced to the model output, both FAR and DRIP decreased across the board while MTD increased. On the other hand, model 4-F exhibited more errors than during training. The model 4-F exhibited the worst performance with 9.79% false alarms at while detecting 98.16% of incident patterns. Despite the use of persistence checks, performance of the model 4-F did not improve considerably.

Except 1-R recurrent model listed in Table 11, all the other single hidden layer networks did not perform well with the evaluation data. Particularly, the 2-R model with no persistent check had an extremely high false alarm rate compared to any ANN model in the evaluation process. Once the persistence check was introduced to the model output, the false

alarm rate significantly reduced from 17.81% to 5.47%. Following this drop, the MTD increased dramatically from 5 seconds to 153 seconds. When comparing the DRIP, only 4-R model had lower DRIP values during the evaluation.

Table 10. Evaluation Performance of Single Hidden Layer Feed Forward ANNs

Model	Persistent Check	DRIP	FAR	Error Rate	MTD (sec)
1-F	0	68.85%	0.07%	0.0715	392.1
	1	68.39%	0.04%	0.0712	430.7
	2	67.99%	0.01%	0.0709	458.6
	3	67.57%	0.00%	0.0708	488.6
2-F	0	79.02%	4.40%	0.0917	280
	1	78.94%	3.56%	0.0832	293.4
	2	78.85%	2.84%	0.0761	306.7
	3	78.76%	2.22%	0.0699	321.7
3-F	0	95.25%	0.98%	0.0206	116.7
	1	95.22%	0.83%	0.0190	140
	2	95.18%	0.71%	0.0179	170
	3	95.13%	0.61%	0.0169	200
4-F	0	98.16%	9.79%	0.1021	20
	1	98.15%	9.07%	0.0948	33.3
	2	98.14%	8.50%	0.0891	51.7
	3	98.13%	8.07%	0.0848	71.7

Table 11. Evaluation Performance of Single Hidden Layer Recurrent ANNs

Model	Persistent Check	DRIP	FAR	Error Rate	MTD (sec)
1-R	0	94.04%	2.67%	0.0402	111.7
	1	93.98%	2.45%	0.0381	145
	2	93.93%	2.29%	0.0365	175
	3	93.87%	2.14%	0.0349	205
2-R	0	94.98%	17.81%	0.1895	5
	1	94.75%	6.86%	0.0800	95
	2	94.61%	6.13%	0.0727	123.3
	3	94.68%	5.47%	0.0661	153.3
3-R	0	94.40%	5.08%	0.0636	105
	1	94.34%	2.27%	0.0354	145
	2	94.28%	2.16%	0.0343	175
	3	94.23%	2.06%	0.0333	205
4-R	0	85.70%	2.75%	0.0600	46.7
	1	83.95%	1.45%	0.0470	160
	2	83.67%	1.22%	0.0447	186.7
	3	83.37%	1.02%	0.0427	213.3

Tables 12 and 13 list evaluation performance measures for double hidden layer networks. From the results shown in Table 12, models 5-F and 8-F performed better with the evaluation data. Model 8-F detected more incident patterns during the evaluation than model 5-F did and even had lower false alarms. It detected incidents at an average of 83 seconds without any persistence check. Once the persistence check was applied to model 8-F, the false alarm rate considerably reduced from 0.99% to 0.01% while the detection of incident patterns

did not deteriorate that much. Performance of model 5-F was not quite the same. The false alarm rate did not reduce considerably even if persistence check was applied. Persistence checks can effectively filter out majority false detections attributable to random fluctuations in the traffic flow. The persistence check also delays the speedy detection of actual incident patterns since the AIDM would not issue an incident alarm until incident conditions are detected throughout a predetermined number of time intervals has passed.

Double hidden layer recurrent models were also examined with the evaluation data. Results of the evaluation are listed in Table 13. Model 5-R had a slightly lower error rate and a higher false alarm rate than during training. Among the recurrent networks, model 6-R exhibits the best performance values during evaluation. Persistent checks had reduced the FAR of model 6-R from 0.16% to 0.06%.

7.3 Comparative Evaluation of ANN with Conventional Algorithms

Performance of the best single and double hidden layer ANN models were compared with those of the California algorithm no. 8 and Minnesota algorithm discussed in Chapter 2. The same data set that was used in the evaluation of ANN models was used to evaluate the performance of these two conventional algorithms. For this comparative evaluation, the threshold sets listed in Table 8 (in section 6.3) were used for California algorithm and threshold sets listed in Table 9 (in section 6.3) were used for Minnesota algorithm. Table 14 lists the DRIP, FAR, error rate, and MTD of the ANN models 3-F and 8-F with the California algorithm no. 8 and Minnesota algorithm during the evaluation. From the Table 14, it can be seen that both ANN models detected more incident patterns than the two

Table 12. Evaluation Performance of Double Hidden Layer Feed Forward ANNs

Model	Persistent Check	DRIP	FAR	Error Rate	MTD (sec)
5-F	0	92.74%	0.95%	0.0260	116.7
	1	92.68%	0.80%	0.0245	143.3
	2	92.61%	0.69%	0.0234	178.3
	3	92.54%	0.62%	0.0227	208.3
6-F	0	96.68%	2.63%	0.0338	53.3
	1	96.66%	2.30%	0.0306	73.3
	2	96.64%	2.00%	0.0275	93.3
	3	96.62%	1.75%	0.0251	116.7
7-F	0	98.07%	4.59%	0.0503	61.7
	1	98.06%	4.24%	0.0468	80
	2	98.05%	3.97%	0.0441	101.7
	3	98.04%	3.77%	0.0421	125
8-F	0	95.43%	0.99%	0.0203	83.3
	1	95.40%	0.61%	0.0165	103.3
	2	95.37%	0.30%	0.0133	133.3
	3	95.33%	0.01%	0.0105	156.7

conventional algorithms. Therefore, the ANN models 3-F and 8-F could detect incident free patterns more accurately compared to the two conventional algorithms. Both ANN models had lower error rates than the two conventional algorithms.

Performance envelopes for the ANN models 3-F and 8-F and the California and Minnesota algorithms are shown in Figure 33. This figure clearly shows that the models 3-F and 8-F had highest DRIP, and essentially the best overall performance of any AIDM in

Table 13. Evaluation Performance of Double Hidden Layer Recurrent ANNs

Model	Persistent Check	DRIP	FAR	Error Rate	MTD (sec)
5-R	0	92.92%	3.14%	0.0475	138.3
	1	92.86%	2.96%	0.0457	168.3
	2	92.80%	2.82%	0.0443	198.3
	3	92.73%	2.70%	0.0431	228.3
6-R	0	89.87%	0.16%	0.0246	178.3
	1	89.71%	0.11%	0.0241	210
	2	89.62%	0.08%	0.0238	240
	3	89.53%	0.06%	0.0236	270
7-R	0	91.62%	1.96%	0.0386	120
	1	91.55%	1.83%	0.0374	150
	2	91.48%	1.79%	0.0370	180
	3	91.41%	1.76%	0.0367	210
8-R	0	98.21%	6.01%	0.0642	65
	1	98.20%	5.23%	0.0563	80
	2	98.19%	4.74%	0.0514	96.7
	3	98.18%	4.35%	0.0476	115

terms of DRIP, FAR, and error rate values. The fact that both ANN models performed better than the conventional algorithms suggest that performance of ANN models are in accordance with the findings during the training. The best threshold value that provides best DRIP and FAR (determined by different traffic operational personnel) can be found for each locality.

Few observations can be made about the Minnesota algorithm during the evaluation. The Minnesota algorithm had the lowest MTD compared to the other models listed in Table

14. Despite a higher false alarm rate and lower DRIP, the Minnesota algorithm was sensitive to the changes in occupancy values. During onset of an incident, such changes in occupancy values between upstream and downstream stations are apparent. This may explain the reason for the Minnesota algorithm's relatively faster incident detection than other models even at high false alarm rates.

7.4 Comparison of Conventional and Modified Detection Rates

Previous AIDM studies have been accustomed to using conventional detection rate as described in Chapter 6 to evaluate incident detection performance. Since it does not directly measure the capability of AIDM in detecting individual incident patterns, a modified version of the detection rate was introduced in Chapter 6.

Table 15 lists the modified detection rate (DRIP) and the conventional detection rate (DR) calculated for the ANN models 3-F and 8-F and California and Minnesota algorithms using the evaluation data. Model 3-F detected 95.25% of incident patterns while model 8-F detected 95.45% of incident patterns correctly with no persistence check employed. However, both ANN models detected every incident occurrence in the incident database used in the evaluation process. Therefore, according to the definition of DR in equation 16 (Chapter 6), DRs for both models were calculated to be 100%. Clearly, the incident detection performance of the ANN models can only be differentiated using estimated the modified detection rate. If one were to use the conventional detection rate, in this case, the DR cannot differentiate the incident detection performance of the two models. The California and Minnesota algorithms were not able to detect every incident occurrence in the data base. The

Table 14. Evaluation Results for ANN Model and the Conventional Algorithms

Persistence or Threshold Set No.	DRIP	FAR	Error Rate	MTD (sec)
3-F				
0	95.25%	0.98%	0.0206	116.7
1	95.22%	0.83%	-	140
2	95.18%	0.71%	-	170
3	95.13%	0.61%	-	200
8-F				
0	95.43%	0.99%	0.0203	83.3
1	95.40%	0.61%	-	103.3
2	95.37%	0.30%	-	133.3
3	95.33%	0.01%	-	156.7
California Algorithm No. 8				
1	80.82%	11.11%	0.1547	190.9
5	69.97%	10.76%	0.1758	197.1
7	46.44%	9.34%	0.2151	164.2
Minnesota Algorithm				
1	88.93%	9.83%	0.1234	42.4
2	80.19%	7.99%	0.1249	47.6
3	58.40%	2.44%	0.1190	287.6

California algorithm detected the least number of incidents out of all the models, and had a range of DRs for its different threshold sets as indicated in Table 15. On the other hand, the Minnesota algorithm detected majority of incident occurrences using the thresholds listed. Nevertheless, the Minnesota algorithm had trouble detecting the prolong continuation of

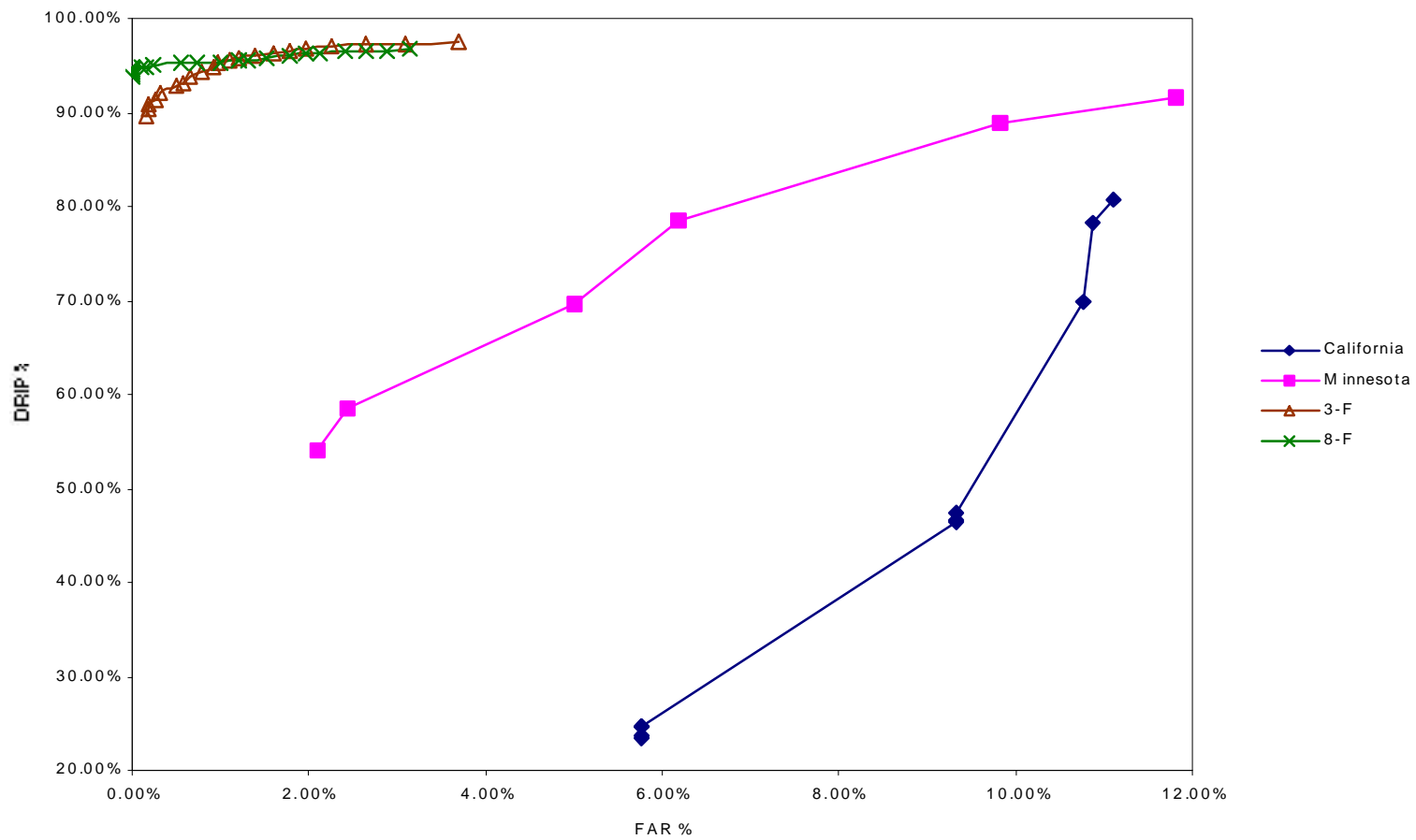


Figure 33. Performance Envelope for ANN Models and Conventional Algorithms

incident conditions in many occasions. Even though the algorithm could detect incidents relatively faster, it seemed to have problems in detecting majority of incident patterns during the latter part of each incident. From the Table 15, this scenario can be clearly observed with DRIP ranging from 88.95% to 58.40% for the Minnesota algorithm with the three different threshold sets. Even though the DR values for the California and Minnesota algorithm are different for different threshold sets, DR values did not measure the true capability of algorithm's in detecting incident patterns. The major function of AIDMs are to detect whether incident condition exists based on individual traffic flow patterns. Therefore, the efficiency of AIDMs in performing this task can be better measured by using the modified detection rate as the evidence indicated. In summary, DRIP is clearly a true measure of AIDM's ability to detect incident patterns during model development as well as in actual implementation than the conventional detection rate.

7.5 Summary

The best models selected in each of the sixteen different architectures were put through an evaluation test. Evaluation test was conducted to validate model performance with an independent data set. The evaluation results indicated that model 3-F had the best overall performance among the single hidden layer networks while model 8-F had the best overall performance results among the double hidden layer networks.

The models 3-F and 8-F were used in comparative evaluation with the California and Minnesota algorithms. Performance measures for the two algorithms were calculated using the same evaluation data. Comparative analysis of the performance results suggested that

Table 15. Comparative Evaluation of DRIP and Conventional DR

Persistence or Threshold Set No.	DRIP	Conventional DR	FAR
3-F			
0	95.25%	100%	0.98%
1	95.22%	100%	0.83%
2	95.18%	100%	0.71%
3	95.13%	100%	0.61%
8-F			
0	95.43%	100%	0.99%
1	95.40%	100%	0.61%
2	95.37%	100%	0.30%
3	95.33%	100%	0.01%
California Algorithm No. 8			
1	80.82%	78.95%	11.11%
5	69.97%	68.42%	10.76%
7	46.44%	47.37%	9.34%
Minnesota Algorithm			
1	88.93%	89.47%	9.83%
2	80.19%	89.47%	7.99%
3	58.40%	89.47%	2.44%

both ANN models outperformed the conventional algorithms. Of the two ANN models, the model 8-F exhibited a better overall performance.

CHAPTER 8

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

8.1 Summary

Traffic incidents are a major contributor to the congestion that cause traffic delays in freeway systems. To minimize traffic delays caused by incidents, traffic operational managers focus on detecting incident conditions and dispatching incident management teams as quickly as possible. During the past few decades, various conventional algorithms and ANN models have been proposed to automatically detect incident conditions on freeways. Many of the published AID systems possess several shortcomings. Even though the incidents affect the traffic flow, only a combination of traffic flow variables have been used in many of these AID systems as model inputs. These AID systems were trained to identify changes in traffic flow data that represent incident conditions. However, changes in freeway geometry may also affect the traffic flow. This effect may be represented in new models by including geometric variables as additional inputs to reduce false alarms. In addition, the majority of AID models based on ANN utilize simulated traffic data to train models. A conventional detection rate was utilized in the AID systems as a performance measure to assess the detection capability. However, the principle function of incident detection models is to identify whether an incident condition exists for a given traffic pattern. To measure the true detection performance, a modified ratio that utilizes the number of detected incident patterns should be used instead.

In this study, sixteen models based on two different ANN architectures (feed forward and recurrent) were developed. Models in each architecture group had either single or double hidden layers. Based on the ANN architecture (i.e., feed forward and recurrent architectures) and the number of hidden layers (i.e., single or double layers) these models can be subdivided into four different model groups. Each model group had a combination of traffic flow variables and freeway geometric variables. Real life traffic data were used to train, test, and evaluate each of the sixteen ANN models. Both upstream and downstream volume, speed, and occupancy data for up to four consecutive time intervals were included in the model input combination. The freeway geometric information (presence of entrance and exit ramps and laneage expansions and mergers between the upstream and downstream) were also included in the model input combination to test their affect on incident detection performance. Traffic data gathered from I-880 during September 27 - October 29,1993 was used for model training and testing, while traffic data gathered during February 16 - March 19, 1993 were used for model evaluation. As the models were trained, testing data were used to ensure that models were generalized and not overtrained. A modified detection rate based on detection of incident pattens was introduced to measure the true capability of AID systems in detecting incident patterns. Additionally, performance of the ANN models developed were evaluated and compared with that of conventional algorithms such as the California and Minnesota algorithms.

Performance measures such as DRIP, DR, FAR, and error rate of each model developed were estimated at the end of training process. Single hidden layer feed forward models 1-F through 4-F had a wider range for the error rate during the training process.

Recurrent models 2-R, 3-R, and 4-R had similar DRIP and FAR values. Of the single hidden layer models, the models 3-F and 3-R exhibited better performance during training within the respective model architectures. In fact, both models had almost the same input feature sets except for an added recurrent loop in the model 3-R.

The feed forward models with double hidden layers exhibited the lowest overall error rate among all the models developed in this study. During the training, the best overall performance was exhibited by model 8-F with double hidden layers. The model 8-F utilized traffic flow and freeway geometric variables as model inputs. To be more specific, upstream traffic variables up to $t-4$ time intervals, downstream traffic variables up to $t-2$ time intervals, presence of entrance and exit ramps between two loop detector stations, and presence of lane expansions and mergers between two loop detector stations were used as the model inputs. The results of this research have demonstrated that new changes introduced to ANN inputs can indeed improve freeway incident detection performance.

Two commonly used conventional algorithms, the California algorithm no. 8 and the Minnesota algorithm, were also calibrated, using the same data used to train ANN models, for performance comparison with the ANN models. A large number of threshold sets were used in calibrating both algorithms. The best performing threshold sets were selected for comparative evaluations.

The best ANN models in each of the sixteen model architectures selected during training were used in the evaluation process. Model 3-F had the best performance among single hidden layer networks and model 8-F had the best performance among double hidden layer networks. During the comparative evaluations of the ANN models, the performance

envelope for models 3-F, 8-F, California algorithm, and Minnesota algorithm reiterated that model 8-F performances as observed during the training. Model 8-F detected more incident patterns than any other AIDM with a lower false alarm rates. The comparison between the modified detection rate (DRIP) and conventional detection rate (DR) for the models revealed that DRIP provided the necessary performance index to compare incident detection capability of different AIDMs.

8.2 Conclusions

A freeway traffic management system that quickly detects and removes incidents from the freeway increases the safety of the travelers and decreases overall traffic delay. Accordingly, an automated incident detection system is designed to ease the process of identifying freeway incidents and expedite the disposal of emergency management team to the scene much faster. In this research, ANN-based models were developed to accomplish this task. The ANN models were trained and tested with actual field data. Developed ANN models were validated using an independent data set.

Several important observations can be drawn from the analysis of training and validation results. Freeway geometric features have proven to increase the model performance. The modified detection rate introduced proved its potential to capture the true ability of AIDMs to detect freeway incident patterns. ANN models have far superior performance over the conventional algorithms. The ANN model 8-F with the input combination of upstream traffic data up to $t-4$ time intervals, downstream traffic data up to $t-2$ time intervals, and freeway geometric data was proven to yield the best incident detection

performance during both training and validation processes. The false alarm rate of model 8-F could still be reduced with three persistence checks. Therefore, the ANN model 8-F with three persistence checks can be concluded to be a more practical model for traffic management implementation. This is a model that a traffic operations manager can use not only to detect the onset of an incident but the continuation of incident condition as well. The modified detection rate depicts this in prospective.

From the ANN model development stand point, this study has also proved that double hidden layer feed forward networks are better for freeway incident detection compared to either the single hidden layer feed forward networks or both the single and double hidden layer recurrent networks investigated.

8.3 Recommendations

Previous ANN models developed to detect freeway incidents were based on traffic data gathered at inductive loop detector stations. The models were developed with the assumption that the ANN models can detect the relationships among the traffic flow variables and can detect sudden changes in traffic flow that are attributable to incidents. However occasional it may be, sudden adverse weather conditions (such as a down pouring of rain or snow) may also induce similar sudden changes in traffic flow such that ANN models may incorrectly detect as incident conditions. To reduce these false alarms resulting from such weather conditions, the weather information may be included as future model inputs. Future highway infrastructure improvements may include weather sensors along freeways that can transmit weather data for incident detection purposes. Further field testing with rich source

of data from different locations may be required to study true transferability potential of the 8-F model.

REFERENCES

- Abdulhai, B. and Ritchie, S. G. (1995). "Performance of Artificial Neural Networks for Incident Detection in ITS." *Transportation Congress: Civil Engineers-Key to the World Infrastructure*, New York ASCE, Vol. 1, pp. 227-238.
- Abdulhai, B. and Ritchie, S. G. (1997). "Development of Universally Transferable Freeway Incident Detection Framework." *Transportation Research Board, 76th Annual Meeting*, Paper no. 971121, January 12-16, Washington, D.C.
- Ahmed, S. A. (1983). "Stochastic Processes in Freeway Traffic Part II. Incident Detection Algorithm." *Traffic Engineering and Control*. Vol. 24, No. 6-7, pp 309-314.
- Al-Deek, H., Garib A., and Radwan, A. E. (1994). "New Method for Estimating Freeway Incident Congestion." *Transportation Research Record No. 1494*, pp. 30-39.
- Antoniades, C. N. and Stephanedes, Y. J. (1996). "Single-Station Incident Detection Algorithm (SSID) for Sparsely Instrumented Freeway Sites." *Proceedings of the International Conference on Applications of Advanced Technologies in Transportation Engineering*, ASCE, New York, NY, pp. 218-221.
- Arceneaux, J., Smith, J., Dunnett, A., and Payne H. (1990). "Calibration of Incident Detection Algorithms for Operational Use." *Traffic Control Methods, Proceedings of The Fifth Engineering Foundation Conference on Traffic Control Methods*, Engineering Foundation, Santa Barbara, California, pp. 17-32.
- ATA Foundation, Inc. (1997). "Incident Management." *Executive Summary*, January, Alexandria, VA.
- Aultrnan-Hall, L., Hall, F. L., Shi, Y., and Lyall, B. (1991). "Catastrophe Theory Approach to Freeway Incident Detection." *Proceedings of the Second International Conference*

of Applications of Advanced Technologies in Transportation Engineering, Edited by Stephanedes, Y. J. and Sinha, K. C., American Society of Civil Engineers, pp. 373-377.

Behbahanizadeh, K. and Hidas, P. (1996). "Modeling Traffic Incidents in Urban Arterial Networks." ITS World Congress, Orlando, Florida, October.

Belgaroui, B. and Blosseville, J. M. (1993). "A Road Traffic Application of Neural Techniques." Recherche Transports, Securite, English Issue, No 9, pp. 53-65.

Blosseville, J. M., Morin, J. M., and Lochegnies, P. (1993). "Video Image Processing Application: Automatic Incident Detection on Freeways." Proceedings of the ASCE Third International Conference of Applications of Advanced Technologies in Transportation Engineering, Edited by Hendrickson, C. T. and Sinha, K. C., American Society of Civil Engineers, pp. 77-83.

Busch, F. and Fellendorf, M. (1990). "Automatic Incident Detection on Motorways." Traffic Engineering and Control, Vol. 31, No. 4, April, pp. 221-227.

Carpenter, W. C. and Hoffman, M. E. (1997). "Guidelines for the Selection of Network Architecture." Artificial Intelligence for Engineering Design, Analysis, and Manufacturing, 11, pp. 395-408.

Chang, E. C. (1992). "Freeway Incident Detection using Advanced Technology." Proceedings of 3rd Workshop on Neural Networks, February, pp. 317-321.

Chang, E. C. and Huarng K. (1993). "Incident Detection Using Advanced Technologies." Paper No. 930943, The 72nd Annual Meeting, Transportation Research Board, Washington, D.C., January.

Chang, E. C. and Huarng, K. (1993). "Freeway Incident Management Expert System Design." Paper No. 930945, The 72nd Annual Meeting, Transportation Research Board, Washington, D.C., January.

- Chang, E. C. and Wang, S-H. (1994). "Improved Freeway Incident Detection using Fuzzy Set Theory." *Transportation Research Record*, No. 1453, December, pp. 75-82.
- Chassiakos A. P. and Stephanedes Y. J. (1993). "Smoothing Algorithms for Incident Detection." *Transportation Research Record* No. 1394, pp. 8-16.
- Chen, C-H. (1993). "A Dynamic Real-Time Incident Detection System for Urban Arterials: Framework and Methodology." *Advanced Traffic Management Conference*, St. Petersburg, FL, Large Urban Systems, Washington D.C.: Federal Highway Administration, pp. 185-193.
- Cheu, R. L. and Ritchie, S. G. (1994). "Neural Network Model for Automated Detection of Lane-Blocking Freeway Incidents." *Proceedings of the International Conference on Advanced*, Singapore, May, pp. 245-252
- Cheu, R. L. and Ritchie, S. G. (1995). "Automated Detection of Lane-Blocking Freeway Incidents using Artificial Neural Networks," *Transportation Research Part C: Emerging Technologies* Vol. 3, No. 6, December, pp. 371-388.
- Cohen, S. and Ketselidou, Z. (1993). "A Calibration Process for Automatic Incident Detection Algorithms." *Microcomputers in Transportation Proc. 4 Int. Conf. Microcomput. Transp.*, ASCE, New York, New York, pp. 506-515.
- Corby, M. J. and Saccomanno, F. F. (1997). "Analysis of Freeway Accident Detection." *Transportation Research Board 76th Annual Meeting*, Washington, DC, January.
- Dia, H. and Rose G. (1996). "Impact of Data Quality on the Performance of Neural Network Incident Detection Models." *ITS World Congress*, Orlando, Florida, October.
- Dougherty, M. and Joint M. (1992). "A Behavioral Model of Driver Route Choice using Neural Networks." *Proceedings, International Conference on Artificial Intelligent Applications in Transportation Engineering*, San Buenaventura, CA.
- Dougherty, M. (1995). "A Review of Neural Networks Applied to Transport." *Transportation Res.- C. Vol. 3, No. 4*, pp. 247-260.

- Fambro, D. B. and Ritch, G. P. (1980). "Evaluation of an Algorithm for Detecting Urban Freeway Incidents During Low-Volume Conditions." Transportation Research Record No. 773, pp. 31-39.
- Fazio, J. (1990). *Modeling Safety and Traffic Operations at Freeway Weaving Sections*. Ph.D. Dissertation, University of Illinois at Chicago.
- Forbes, G. J. (1992). "Identifying Incident Congestion." ITE Journal, Vol. 62, No. 6, pp. 17-22.
- Goldblatt, R. B. (1980). "Investigation of the Effect of Location of Freeway Traffic Sensors on Incident Detection." Transportation Research Record No. 773, pp. 24-30.
- Hall, F. L., Yong, S., and George, A. (1993). "On-Line Testing of the McMaster Incident Detection Algorithm Under Recurrent Congestion." Preprint no 930330, 72nd Annual Meeting, Transportation Research Board, Washington, D.C.
- Han, Lee D. (1995). "Arterial Incident Simulation and Visualization." ITE 65th Annual Meeting, pp. 260-267.
- Han, L. D. and May, A. D. (1989). "Artificial intelligence Approaches for Urban Network Incident Detection and Control." Traffic Control Methods, Proceedings of the Engineering Foundation Conference, Santa Barbara, California, Feb. 26 - Mar. 3, pp. 159-176.
- Hattan, D. E., Finch J., Lipp L., Corcoran L. J., and Sumpter L. (1993). "Implementing an Areawide Incident Management and IVHS Program." ITE International Conference, pp. 83-89.
- Haykin, S. (1994). *Neural Networks, A Comprehensive Foundation*. Macmillan College Publishing Company, Inc., New York.
- Hecht-Nielsen, R. (1990). *Neurocomputing*. Addison Wesley, California.

- Hertz, J. A. and Palmer, R. G. (1991). *Introduction to the Theory of Neural Network Computation*. Addison Wesley, California.
- Hourdakis, J. and Chassiakos, A. P. (1996). "Preliminary Features of a Decision Support System for Incident Detection." Proceedings of the International Conference on Applications of Advanced Technologies in Transportation Engineering, ASCE, New York, NY, pp. 227-232.
- Hsiao, C-H. (1996). "The Artificial Intelligence Incident Detection Algorithms." ITE International Conference, pp. 31-35.
- Hsiao, C-H., Lin, C-T., and Cassidy, M. J. (1993). "The Application of Fuzzy Logic and Neural Networks to Automatically Detect Freeway Traffic Congestion." Presented at 72nd Annual Meeting, Transportation Research Board, Washington, D. C.
- Hsiao, C-H. (1994). *The Application of Fuzzy Logic and Neural Networks to Freeway Incident Detection*. Ph.D. Dissertation, Purdue University, Lafayette, Illinois.
- Hua, J. and Faghri, A. (1993). "Traffic Mark Classification using Artificial Neural Networks." Proceedings, Pacific Rim Conference, Seattle, WA.
- Hua, J. and Faghri, A. (1994). "Applications of Artificial Neural Networks to Intelligent Vehicle-Highway Systems." Transportation Research Record No. 1453, pp. 83-90.
- Ivan, J. N. (1996). "Neural Network Representations for Arterial Street Incident Detection Data Fusion." 75th Annual Meeting of the Transportation Research Board, Washington, DC, January.
- Kuhne, R. D. and Immes, S. (1993). "Freeway Control Systems for Using Section-Related Traffic Variable Detection." Proceedings of the ASCE International Conference on Applications of Advanced Technologies in Transportation Engineering, ASCE, New York, NY, pp. 56-62.
- Ivan, John N. (1994). *Real-Time Data Fusion for Arterial Street Incident Detection Using Neural Networks*. Ph.D. Dissertation, Northwestern University, IL, June.

- Judycki, D. C. and Robinson, J. R. (1992). "Managing Traffic during Nonrecurring Congestion." ITE Journal, Vol. 62, No. 3, pp. 21-26.
- Kalaputapu, R. and Demetsky, M. (1995). "Modeling Schedule Deviations of Buses Using Automatic Vehicle Location Data and Artificial Neural Networks." Preprints no. 950465, Transportation Research Board Conference, January, Washington, D.C.
- Kaneko, Y., Kawashima, H., and Yamamoto, M. (1996). "Incident Detection at Curves Using Neural Network." ITS World Congress, Orlando, Florida, October.
- Katakura, M. (1996). "A Fast Detection Method of the Changes of Traffic Condition Based on Pulse Data of Vehicle Detectors." ITS World Congress, Orlando, Florida, October.
- Khattak, A. J. (1991). *Driver Response to Unexpected Travel Conditions: Effect of Traffic Information and Other Factors*. Ph.D. Dissertation, Northwestern University.
- Lawrence, J. (1993). *Introduction to Neural Networks - Design, Theory, and Application*. California Scientific Software Press.
- Levin, M. and Krause, G. M. (1978). "Incident Detection: A Bayesian Approach." Transportation Research Record No. 682, TRB, National Research Council, Washington, D. C., pp. 52-58.
- Levin, M. and Krause, G. M. (1979a). "A Probabilistic Approach to Incident Detection on Urban Freeways." Traffic Engineering and Control, Vol. 20, No. 3, pp. 107-109.
- Levin, M. and Krause, G. M. (1979). "Incident-Detection Algorithms Part 1. Off-Line Evaluation and Part 2. On-Line Evaluation." Transportation Research Record No. 722, TRB, National Research Council, Washington, D.C., pp. 49-64
- Lindley, J. A. (1987). "Urban Freeway Congestion: Quantification of the Problem and Effectiveness of Potential Solutions." ITE Journal, Vol. 57, No. 1, pp. 27-32.

- Madanat, S. M., Teng, H-L, and Liu, P-C. (1996). "A Sequential Hypothesis-Testing Based Freeway Incident Response Decision-Making System." Proceedings of the International Conference on Applications of Advanced Technologies in Transportation Engineering, ASCE, New York, New York, pp. 286-291.
- Magee, M., Franke E., Kinkler, S., and Seida, S. (1996). "A New Way to Look at Traffic Incidents." ITS World Congress, Orlando, Florida, October.
- Mead, W. C., Fisher, H. N., Jones, R. D., Bisset, K. R., and Leopold, A. L. (1994). "Application of Adaptive and Neural Network Computational Techniques to Traffic Volume and Classification Monitoring." Transportation Research Board, Washington, D.C., TRR 1466, pp. 116-123.
- Moody, J. and Darken, C. J. (1989). "Fast Learning in Networks of Locally-Tuned Processing Units." Neural Computat. Vol 1, pp 281-294
- Nathanail, T. and Zografos, K. (1994). "Simulation of Freeway Incident Restoration Operations." Vehicle Navigation and Information Systems Conference, IEEE, Piscataway, NJ, 94CH35703, pp. 229-232.
- Neusser, S., Hoefflinger, B., Nijhuis, J., Siggelkow, A., and Spaanenburg, L. (1991). "A Case Study in Car Control by Neural Networks." Proceedings, International Symposium on Automotive Technology and Automation, Florence, Italy, May, pp. 607-613.
- Parkany, E. and Bernstein, D. (1995). "Design of Incident Detection Algorithms using Vehicle-to-Roadside Communication Sensors." Transportation Research Record, No. 1494, July, pp. 67-74.
- Payne, H. J. and Tignor, S. C. (1978). "Freeway Incident-Detection Algorithms Based on Decision Trees With States." Transportation Research Record No. 682, TRB, National Research Council, Washington, D. C., pp. 30-37.
- Persaud, B. N., Hall, F. L., and Hall, L. M. (1990). "Congestion Identification Aspects of the McMaster Incident Detection Algorithm." Transportation Research Record. No. 1287, TRB, National Research Council, Washington, D. C., pp. 167-175.

- Petty, K. F., Skabardonis, A., and Varaiya, P. P. (1996). "Methodology for Estimating the Impacts of Incident Management Measures." Third World Congress on Intelligent Transport Systems, Orlando, Florida, October.
- Pietrzyk, M. C. and Perez, R. A. (1996). "Solving Transportation Problems with Artificial Neural Networks." Presentation, ITS World Congress, Orlando, FL, October 14-18.
- Pike, R. W. (1986). *Optimization For Engineering Systems*. pp. 208-233.
- Roseman, D. and Skehan, S. (1995). "Automated Arterial Incident Detection Santa Monica Freeway Smart Corridor." ITE Compendium of Technical Papers, pp. 27-31.
- SAS, Inc (1998), "Frequently Asked Questions." Web Site, www.sas.com, January.
- Skabardonis, A., Petty, K., Noeimi, H., Rydzewski, D., and Varaiya, P. P. (1996). "The I-880 Field Experiment: Database Development and Incident Delay Estimation Procedures." 75th Annual Meeting, Transportation Research Board, January, Washington, D.C.
- Sontag, E. D. (1992). "Feedback Stabilization using Two-Hidden Layer Nets." IEEE Transactions on Neural Networks, 3, pp. 981-990.
- Stephanedes, Y. J. and Chassiakos, A. P. (1993). "Application of Filtering Techniques for Incident Detection." Journal of Transportation Engineering, ASCE, Vol. 119, No. 1, pp. 13-26.
- Stephanedes, Y. J., and Chassiakos A. P. (1991). "A Low Pass Filter for Incident Detection." Proc. 2nd Int. Conf.-Application on Advanced Technologies in Transportation Engineering, Minneapolis, Minn., pp. 378-382.
- Stephanedes, Y. J. and Liu, X. (1995). "Artificial Neural Networks for Freeway Incident Detection." TRR 1494, pp 91-97.

- Sugiyama, M., Kasai, K., Namai, T., and Hasegawa, T. (1996). "Detection of Traffic Congestion by Automatic Incident Detection System," ITS World Congress, Orlando, Florida, October.
- Surkan, A. J. and Singleton, J. C. (1990). "Neural Networks for Bond Rating Improved by Multiple Hidden Layers." IJCNN, San Diego, June 17-21, Vol. 2, IEEE Press, pp. 157-162.
- Taniguchi, E., Hayama, A., Morikawa, K., and Sugawara, K. (1996). "Deployment of Automatic Incident Detection System in Expressway." ITS World Congress, Orlando, Florida, October.
- Terano, T., Asai, K., and Sugeno, M. (1992). *Fuzzy Systems Theory and Its Applications*, Academic Press Inc.
- Tignor, S. C. and Payne, H. J. (1977). "Improved Freeway Incident Detection Algorithms." Public Roads Vol. 41, No. 1, Engineering Societies Library, New York, June, pp. 32-40.
- Tveter, D. (1998), "Professional Basis of AI Backprop." Web Site, www.dontveter.com, July.
- Villiers, J. de and Barnard, E. (1992). "Backpropagation Neural Networks with One and Two Hidden Layers." IEEE Transactions on Neural Networks, Vol 4, No. 1, January, pp. 136-141.
- Vogl, T. P., Mangis, J. K., Rigler, A. K., Zink, W. T, and Alkon, D. L. (1988). "Accelerating the Convergence of the Back-Propagation Method." Biological Cybernetics, Vol 59, pp. 257-263.
- Wang, M. H. (1991). *Modeling Freeway Incident Clearance Time*. M.S. thesis, Northwestern University.
- Wiederholt, L., Okunieff, P., and Wang, J. (1993). "Incident Detection and Artificial Neural Networks." Advanced Traffic Management Conference, St. Petersburg, FL, Large Urban Systems, Washington D.C.: Federal Highway Administration, pp. 195-206.

- Willsky, A. S., Chow, E. Y., Gershwin, S. B., Greene, C. S., Houpt, P. K., and Kurkjian, A. L. (1980). "Dynamic Model-Based Techniques for the Detection of Incidents on Freeways." *IEEE Transactions on Automatic Control*, Vol. 25, No. 3, pp. 347-360.
- Yim, Y. and Ygnance, J. (1995). "A Comparative Analysis of Cellular Phone User Profiles and Travel Behavior in the San Francisco Bay Area and the Paris Region." *TRB 74th Annual Meeting*, January.
- Zhang, H. and Ritchie, S. G. (1994). "Real-Time Decision-Support System for Freeway Management and Control." *Journal of Computing in Civil Engineering*, Vol. 8, No. 1, January, pp. 35-51.

VITA

Sujeeva A. Weerasuriya

- 1990 B.Sc. in Civil Engineering (Honors), University of Moratuwa, Sri Lanka.
- 1993 M.Sc. in Civil Engineering (Water Resources), Clarkson University, Potsdam, New York.
Thesis: "Oil Spills In Ice Covered Waters"
Professor Poojitha Yapa
- 1993 - 1994 Graduate Research Assistant, Department of Civil and Environmental Engineering, University of South Florida, Tampa, Florida.
- 1994 - 1998 Graduate Research Assistant, ITS Team, Center for Urban Transportation Research, University of South Florida, Tampa, Florida.
- 1998 Ph.D. in Civil Engineering (Transportation and Traffic Engineering), University of South Florida, Tampa, Florida.
Dissertation: "An Application of Artificial Neural Networks in Freeway Incident Detection"
Professors Jian John Lu and Ram Pendyala