Development of a Neural Network Pedestrian Gap Acceptance Model for Enhanced Operation of Midblock Signalled Pedestrian Crossings

GLENN LYONS*, JOHN HUNT+ AND FRASER MCLEOD*

*Transportation Research Group, University of Southampton,
Southampton SO17 1BJ, United Kingdom
G.Lyons@soton.ac.uk [corresponding author]

+Cardiff School of Engineering, Cardiff University, The Parade, PO Box 686,
Cardiff CF2 3TB, United Kingdom
Hunt@Cardiff.ac.uk

Keywords: Traffic; Pedestrians; Neural Networks

Abstract

UK transport policy has shifted dramatically in recent years. The new policy direction seeks to promote walking as an alternative to the car for short trips. Midblock signalled pedestrian crossings are a common method of resolving the conflict between pedestrians and vehicles.

This paper considers alternative operating strategies for midblock signalled pedestrian crossings that are more responsive to the needs of pedestrians without adding to the costs of business by increasing delay to motorists and freight traffic. Specifically the paper considers the development of a pedestrian gap acceptance model that would be required in the proposed strategies. A succession of artificial neural network models are developed and factors influencing performance of the models both in terms of accuracy and processing requirements are considered in detail. The paper concludes that a feedforward artificial neural network using backpropagation can deliver a gap acceptance model with a high degree of accuracy within the context and constraints under which it has been developed.

1 Introduction

1.1 UK transport policy

Until recently transport policy emphasised the role of the private car as a symbol of personal freedom. During the 1990s transport policy entered into a state of flux with a realisation that, with rising levels of traffic and congestion and a deterioration in the quality of the urban environment for pedestrians, it was no long acceptable to continue to concentrate on the needs of the motorist. In 1998 the publication of the first Transport White Paper for some 20 years set out a new policy agenda build around the need to reduce car use and promote alternative means of travel. For local journeys, walking is seen as a suitable alternative mode of transport and the Government wishes to make walking a more viable, attractive and safe option. The White Paper 'A New Deal for Transport – Better for Everyone' (DETR, 1998) sets an expectation for local authorities to give more priority to walking. The aims include reducing waiting times for pedestrians at traffic signals and giving them priority in the allocation of time at junctions where this supports more walking.

This improvement for pedestrians needs to be achieved without increasing delays (and costs) to essential road users. Congestion and unreliability of journeys add to the costs of business, undermining competitiveness particularly in our towns and cities where traffic is worst. The CBI has put the cost to the British economy at about £15 billion every year, some estimates are lower but agree that the cost to the nation runs into billions of pounds every year and is rising (DETR, 1998). Hitherto, provision of new capacity was the accepted means to address these issues. However, current and future policies are likely to focus attention on making better use of the existing infrastructure. Existing infrastructure is not only for use by goods distribution vehicles and private cars. New policies raise the profile and importance of provision for pedestrians and public transport vehicles (primarily buses). Goods logistics in particular is based increasingly on just-in-time deliveries. Overall journey time and more specifically journey time reliability are critical factors. Continued increase in congestion and further competition for roadspace from pedestrians and public transport could pose serious problems for the freight industry. The research described in this paper endeavours to reconcile the conflicting requirements of motorised traffic and pedestrian movements.

1.2 Existing pedestrian crossing facilities

In the United Kingdom most signalled midblock pedestrian crossings are of the Pelican (<u>Pe</u>destrian <u>Li</u>ght <u>Con</u>trolled) type. The United Kingdom has more than 11000 Pelican crossings (County Surveyors' Society, 1994) installed between junctions to provide pedestrians with signal protected periods in which to cross streams of traffic between junctions. The operational strategies for Pelican crossings (DoT, 1995) are based on default priority for vehicles with pedestrian right of way available, on demand, at times and with frequencies that are consistent with minimising delay to vehicle occupants. At most crossings vehicle precedence is retained during periods when vehicle flow is sufficiently low for pedestrians to identify gaps in vehicle arrivals in which they can cross the road. Current operating strategies impose substantially higher delays on pedestrians who obey the signal indications. There is thus a strong incentive for pedestrians to seek to reduce their delay by crossing in gaps in traffic flow when the red man is showing. Pedestrians become particularly impatient when a red man continues to be shown during periods of low vehicle flow. The reduction of unnecessary delay for pedestrians who obey the signal indications would encourage more pedestrians to use the facility correctly and reduce pedestrian annoyance and risk taking. Although no clear relationship has been established between pedestrian delay and casualties, a more balanced and responsive approach to the allocation of time at Pelican crossings has the potential to make a substantial contribution to a decrease in pedestrian casualties as well as improving pedestrian amenity.

1.3 Alternative operating strategies for pedestrian crossing facilities

The proposed alternative strategies are based on the concept that it is possible to identify, in advance, a sequence of vehicle arrivals at a midblock signalled pedestrian crossing which the average pedestrian would perceive as a crossing opportunity. This information can then be applied to provide an operating strategy that is more responsive to pedestrian demand by providing an earlier change to pedestrian precedence than is possible currently. The enhanced strategies will minimise disbenefit, which may otherwise be associated with a system which is more responsive to pedestrian demand, to vehicle occupants by making more effective use of periods of low vehicle flow. The enhanced strategies depend on developing a

technique to identify vehicle arrival patterns which represent a gap crossing opportunity for the average pedestrian.

The study of pedestrian response to gaps in vehicle flow is well established. Traditional models (Tanner, 1951) of pedestrian road crossing delay are based on strategies in which pedestrians respond to symmetrical nearside and farside gaps. In practice pedestrian response in interpreting and responding to gaps is more complex. Artificial neural networks (ANNs) provide a tool for developing an appropriate on line system that can interpret a greater number of possible explanatory variables for crossing behaviour in an efficient way with a potential enhancement in performance. The potential of ANN techniques continues to be assessed in a wide and diverse range of transport applications (Dougherty, 1995). This paper considers the use of a feedforward ANN trained using backpropagation to develop an ANN gap acceptance model for pedestrian crossing opportunities. The backpropagation learning algorithm, used with a feedforward ANN (Multi-Layer Perceptron), is the most widely applied arrangement in both research and development (Taylor, 1994). The Learning Vector Quantization paradigm (an alternative ANN model form used for classification tasks) was also assessed during the research but failed to improve on the results presented in this paper. Model development has made use of the ANN software simulation package Predict (NeuralWare, 1995). In addition to the primary aim of assessing the viability of the proposed approach, the focus of the research has been the data used in model development with an aim to determine and minimise appropriate input parameters for the gap acceptance model. A pragmatic approach has been adopted whereby a succession of increasingly improved models are developed based on reasoned decisions concerning options, settings, parameters and forms of data representation for preceding models and their associated performance with the available data.

2 Development of Gap Acceptance Models Using Site Data

2.1 Description of the Data Set

In a previous study of pedestrian gap acceptance (Hunt and Griffiths, 1991), more than 13,000 pedestrian crossing manoeuvres were analysed and described by a digital representation of key events. Each pedestrian manoeuvre consisted of a number of rejected crossing opportunities followed by an accepted opportunity. Each opportunity is represented as an example in the database including the crossing decision of the pedestrian for each opportunity. At a single point in time a pedestrian can be presented with more than one crossing opportunity in terms of how he/she elects (or not) to cross the road (see Figure 1). Each *crossing situation* therefore comprises a set of opportunities of this type. Table 1 shows an example of a set of crossing opportunities presented at a single point in time. For the example shown in the Table the pedestrian elects to cross between the 4th and 5th farside vehicles.

Figure 1

Table 1

2.2 Look-up Table Model

Considering the variables Near, Far1 and Far2 as the only determinants of gap acceptance for an average pedestrian, the data set can be used to produce matrices showing the probability of acceptance by pedestrians according to time gaps tabulated using intervals for Near, Far1 and Far2. Figure 2 shows the resulting tables of probabilities of a crossing opportunity being rejected for a 10.0m wide road based on data from multiple sites. It must be noted that the matrices present probabilities of rejecting a particular crossing opportunity and not, as individual values, the probability of electing whether or not to cross the road at a particular instant in time when presented with a *set* of crossing opportunities (as illustrated in Figure 1).

Figure 2

2.3 ANN Models

The first stage of this research was to determine how well an ANN model could perform when presented with crossing opportunity data from the observed data sets and required to determine the gap acceptance decision for each *opportunity*. The data set consisted of a series of sub-sets representing data collected from several sites in South Glamorgan, Wales (the matrices in Figure 2 are based on these and other sites from different regions). The Predict software was used to develop backpropagation-based ANN models.

The data sub-set for a single site was first used (with the assumption that modelling performance might be better when using site specific data to develop the model). The data were divided in two with half the examples used to train and test the ANN and the remaining data used for validation.

For the single site the overall accuracy of classification of gap acceptance decisions was 95 % (for the validation data). An ANN developed subsequently using the multiple site data set produced a classification accuracy of 92 % suggesting that site specific characteristics were not markedly influencing gap acceptance decision making.

In view of these promising results, a more detailed ANN approach was considered that involved presenting the ANN with raw data i.e. vehicle arrival times as they would be recorded in practice.

3 Development of Gap Acceptance Models Using Simulated Data

The raw data were not available for the observed data sets used previously. An approach was therefore adopted in which the raw vehicle arrival data were generated using an algorithm and the corresponding crossing decisions in practice *made by an average pedestrian* were determined based on the look-up tables. It should be noted that model performance now reflects that of modelling the average pedestrian and not the ability to model decisions of individual pedestrians on street.

3.1 Vehicle Arrival Data

When standing at the kerbside waiting to cross a road a pedestrian's immediate consideration is the first vehicle to approach on the nearside and the first vehicle to approach on the farside of the carriageway. In addition to the first farside vehicle, subsequent farside vehicles are relevant as the pedestrian may choose to leave the nearside kerb and then allow one or more farside vehicles to cross his/her path before crossing to the farside kerb. This is addressed using the look-up tables by considering a set of crossing opportunities associated with different pairs of farside vehicles.

Inputs to the ANN model had to effectively include all farside vehicle arrivals that would be included in a set of crossing opportunities. Further to this, an ANN must have a fixed number of inputs. An appropriate number of inputs was selected

that would generally represent all relevant crossing opportunities. However, some examples will have fewer associated crossing opportunities, i.e. when fewer farside vehicles are in the pedestrian's field of view. For such examples an input value must still be assigned to farside vehicles that are not 'present'.

Based on the assumption of vehicle detectors in practice positioned 150m from the crossing it was determined that 5 farside vehicle arrivals would be considered for ANN input and that a default value equal to a gap of 10s would be assumed for a vehicle that was not 'present'. The value of 10s was chosen because it would be large enough to always be interpreted as an acceptable gap and would minimise any degradation of ANN performance. The ANN normalises all inputs based on the minimum and maximum values found in the training database associated with each input. A very large default value could effectively distort the normalised values by reducing the differences between the real values for that input (although Predict does have built in capability to separate out outlying clusters of input data).

The input variables (all with units of seconds) considered were:

N   (the arrival time of the first nearside vehicle)

F1   (the arrival time of the first farside vehicle)

F2   (the gap between the first and second farside vehicles)

F3   (the gap between the second and third farside vehicles)

F4   (the gap between the third and fourth farside vehicles)

F5   (the gap between the fourth and fifth farside vehicles)

Vehicle arrival data were generated for two directions of travel using the displaced negative exponential distribution and a random number generator (Salter and Hounsell, 1996). Random headways were produced based on an assumed traffic flow of 600 veh/h/direction. To produce pedestrian crossing decision examples the following steps were then taken using these data.

For each second in time and for each direction of travel:

1. Calculate values for N, F1, F2, F3, F4 and F5 based on vehicle arrivals at the nominal crossing point. The random generation of gaps is assumed to take place at the detector location for each direction of travel. It is then assumed in turn that calculation of vehicle arrivals at the crossing point are based on a cruise journey time from each detector of 15s.

2. Translate the N, F1, F2, F3, F4 and F5 data into a set of (Near, Far1, Far2) crossing opportunities.

3. Use Figure 2 to obtain an associated probability of each opportunity in the set being rejected. If at least one opportunity in the set has a probability of less than 0.5 then it is assumed that when presented with this set of crossing opportunities at a given point in time the average pedestrian elects to cross the road.

4. Output N, F1, F2, F3, F4 and F5 plus the crossing decision determined from steps 2 and 3.

Data were generated for a 1 hour period - this produced 7148 examples (= 3600 seconds x 2 directions - 52 examples with no vehicles between detectors and crossing location). The mix of data generated is dependent on the selected flow of 600 veh/h. To check that a representative mix of examples had been generated, frequencies corresponding to the cells of the matrices in Figure 2 were determined. All cells were found to contain an acceptable number of examples.

3.2 ANN Model Results

Data were split between train/test data (first 3598 examples) and validation data (remaining 3550 examples). Backpropagation-based ANN models were developed using Predict.

An ANN trained using all 6 input parameters achieved a prediction accuracy of 90 % on the validation data. It was subsequently found, when considering the relative importance of the available input parameters that by only using N and F1 for the same model development data sets, a comparable prediction accuracy of 89 % was achieved. This suggested initially that most of the causal information associated with the gap acceptance decision is provided by these two variables. To determine the validity of this finding the validation data examples were considered according to the input variables N and F1 as shown in Figure 3.

Figure 3

Figure 3 highlights some very clear classification boundaries in the data. It must be noted that this clarity has arisen in part because the data represent decision outcomes for an average pedestrian. In practice it is likely that, for example, 'accept' decisions will be made by some pedestrians faced with a situation defined by N<5s. However, all such situations for the average pedestrian are rejected as shown in

Figure 3. The typical walking speed of a pedestrian can be taken as 1.2 m/s. At this speed it would take just over four seconds to cross the nearside lane of a 10.0m wide road. The decision making resulting from the model is therefore intuitively sensible. When N>6s and F1<6s there is not a single gap acceptance decision for the average pedestrian, with both reject and accept decisions being made depending upon the values of F2-F5.

The Figure broadly indicates that only situations that are represented by N>6s and F1<6s have, collectively, an inconclusive outcome. The majority of examples can be explained by N and F1 alone. Visualisation of the data and model performance as a 2-dimensional space portrayed in Figure 3 suggests that the predominance of examples whose associated outcome is determined by the N =5-6s boundary has saturated the model development process. This has effectively prevented the more subtle distinctions between examples in the bottom right corner from being recognised during model development.

3.2.1 Data sub-set analysis

To determine the importance of the F1-F5 variables a series of further models were built. For the building of these models different (sub)sets of input variables were available and different data sets were used.

Table 2

Table 2 highlights the predominant use, when the entire data set is used, of the N input in model development with only marginal improvements in response accuracy when any or all of F1-F5 are also included as inputs. Indeed the model development process completely rejects F3 and F5 as usefully contributing inputs. However, when the subset of data is considered (N>6s, F1<6s) the variables F1-F5 become much more important. Model 4 rejects the use of F1. This is intuitively sensible for F1<6 and for a 10.0m wide road since F1 will not be used by the average pedestrian as a gap through which to cross. The reduction in misclassifications in Model 4 (for N>6s, F1<6s examples) appears to be more-or-less uniform across the input environment space. To check this, a further subdivision of the data was made according to whether F2 was more or less than six seconds. Two separate models were trained using the subdivided data but only a marginal difference in performance was found.

Table 3 compares the prediction errors of two models. The first is a pure ANN model (Model 1) that is trained and tested using all the data. The second (Model 4) is effectively a hybrid rule-based/ANN model. If N<6s then the situation is rejected by the average pedestrian. If N>6s and F1>6s then the situation is accepted by the average pedestrian. If, however, N>6s and F1<6s then an ANN trained and tested using only examples corresponding to this situation (Model 4) is used to determine the average pedestrian's decision. The Table reveals a marked improvement in classification performance between the first and second model for examples where N>6s and F1<6s. Overall model performance is also improved with nearly 94 % of situations being correctly classified.

Table 3

3.2.2 No data transformation

Predict, by default, considers input data transformations in an attempt to improve correlations between the input and output parameters. This process can lead to improved model performance but not necessarily. If transformations are identified the process augments the number of input parameters and hence the pre-processing and processing requirements. Model 4 considered above included one additional input resulting from inclusion of a transformation. A new model was developed using the same data and available input parameters as Model 4 but with no data transformation permitted. The resulting model performance was marginally better than that of Model 4 with 79% correct classification for the validation data. This was repeated but with F1 removed (this parameter provides no useful information as explained above). The model (Model 8) correctly classified 81 % of validation examples. Using this ANN as the ANN in the hybrid model described earlier provides a model that correctly classifies 94.5 % of validation examples. Avoidance of transformed inputs means that the model can effectively take input data directly from detector output. Since performance has not been adversely affected, it was determined not to allow input transformations during development of subsequent models.

3.3   *Minimising Network Size*

Network size dictates the processing requirements of the deployed ANN model. In this application it is desirable to minimise processing requirements and hence network size. There are two means of reducing network size:

1. reduce the size of the hidden layer in the ANN; and
2. reduce the number of inputs to the model.

3.3.1   Reducing the size of the hidden layer

The backpropagation ANN Model 8 had 17 hidden layer PEs. Predict uses a technique called Cascade Correlation during the network building stage of model development. This process effectively increases the size of the hidden layer incrementally until no further performance improvements can be achieved. Two new ANNs were developed in which the maximum size of the hidden layer was restricted to 10 and 5 PEs. The models were the same as Model 8 in all other respects. The performance results are shown in Table 4. With the same number of input parameters for all three models, reducing the size of the hidden layer from 17 does compromise performance.

Table 4

3.3.2   Reducing the number of inputs to the model

It has already been determined that the input parameter F1 can be excluded without any detriment to model performance. The simulated data for model development assumed the inclusion of five farside vehicles for model input data with a default gap size of 10s for missing values when there were less than five vehicles between the detector and the crossing location. Although Predict has elected to include all farside input parameters except F1 an attempt was made to see if any further parameters could be removed without excessive detriment to performance. Figure 4 shows, for data where N>6s and F1<6s, the number of examples in the combined train/test/validation data sets where a default value of 10s was used for each farside input parameter.

Figure 4

Nearly 89 % of examples have only four farside vehicles between the detector and the crossing, hence F5 in all these cases has a fixed value of 10s. Consequently an ANN identical to Model 8 but with input parameter F5 removed was developed (Model 11). The model results are compared in Table 4. The new model has a marginally better performance than Model 8 and, probably as a consequence of having fewer inputs, also has far fewer hidden layer PEs than Model 8 despite having the same limitation on hidden layer size during training.

4 Final Model Specification

Figure 5 shows the hybrid rule-based/ANN pedestrian gap acceptance model that has been produced. When presented with unseen validation data the model correctly classifies 94.7 % of examples overall. The model represents decisions made by the average pedestrian when crossing a 10.0m wide road. The model has been developed using simulated input data. Output classifications are based on observed data processed into probability matrices.

Figure 5

5 Conclusion and Discussion

This research has produced a pedestrian gap acceptance model that interprets input data that could in practice be provided, with minimal processing, from vehicle detectors. The model's performance in the context of the model development data used is commendable.

The classification performance figures quoted in the report must be set into context. The figures for the complete data sets are dependent on the mix of examples in the data sets. More importantly the results represent the model's ability to model decision making of an average pedestrian. In practice individual pedestrians will make different decisions to those of the average pedestrian. Approximately 80 % of individual crossing decisions in practice would concur with those of the average pedestrian (non N>6s, F1<6s examples). Hence it is likely that Model 11 which correctly classified 94.7 % of average pedestrian examples would, in practice, correctly classify approximately 75 % of individual pedestrian decisions. Without input variables such as age and gender to be able to distinguish between individual pedestrians the only course of action is to model the average pedestrian.

The ANN-based model has produced a classification performance nearly as good as that produced by the use of probability matrices. The use of the probability matrices in practice is inhibited by the data storage and processing requirements that would be imposed compared to the ANN alternative. In this feasibility study the two approaches are linked by the data used to develop the ANN model. It can be argued that this has in fact prevented an ANN approach from performing to its full potential. The observed data upon which the matrices are based only considers the first nearside vehicle. In practice further nearside vehicle arrivals may provide more causal

information to the model. The outcome classifications used for model development are derived directly from the probability matrices. The matrices reflect crossing opportunities as opposed to crossing situations (comprising a set of opportunities). In practice the acceptance or rejection of an individual opportunity is likely to depend upon the other opportunities that present themselves in a given crossing situation. In this context the probability values in the matrices are governed by the particular crossing situations that were observed. There is some potential for confusion in this mixing of opportunity based data and crossing situation modelling. Ideally new observed data would need to be collected to meet the requirements of the modelling approach adopted to determine the model's true performance ability in practice.

One issue that arose during model development was the use of default input values when no value was available in a particular example. The effect of doing this has not been explored explicitly in this study. The default values were selected with the intention of their effect on the model's output being negligible. The overall performance of Model 11 appears very good. However, it is possible that the default values may adversely affect model performance in some situations. This is not an issue unique to this study but remains something that merits separate investigation.

It must be noted that the gap acceptance decisions used in this study are based on observed crossings at random crossing locations. Nonetheless these data are of a similar form to those for midblock signalled crossings and the successful development of an ANN based model for the random crossing data clearly demonstrates the viability of the technique. Taking into account the preceding discussion, this feasibility study strongly indicates that it is possible to model pedestrian gap acceptance decisions with a high degree of accuracy. The resulting model should therefore satisfy the first requirement of the proposed enhanced operating strategies for pedestrian crossings.

References

County Surveyors' Society, 1994. TCUG Traffic signals survey 1994. County Surveyors' Society Environmental Committee, Traffic Management Working Group.

DoT, 1995. The Design of Pedestrian Crossings. Local Transport Note 2/95.

DETR, 1998. A New Deal for Transport: Better for Everyone, Transport White Paper, TSO, London.

Dougherty, M., 1995. A review of neural networks applied to transport. Transportation Research 3C (4) 247-260.

Hunt, J.G. and Griffiths, J.D., 1991. Pedestrian crossing criteria research - Random crossing model. TRL Contractor Report 248, TRL, Crowthorne, UK.

NeuralWare, 1995. NeuralWorks Predict - Complete Solution for Neural Data Modeling. NeuralWare, Inc., Pittsburgh, PA.

Salter, R.J. and Hounsell, N.B., 1996. Highway Traffic Analysis and Design. Macmillan Press Ltd.

Tanner J.C., 1951. The delay to pedestrians crossing a road, Biometrika 38 383-392.

Taylor, J.G., 1994. An International Perspective On Neural Networks. Proc. Neural Networks, Neuro-fuzzy and other Learning Systems for Engineering Applications and Research, EPSRC / DRAL, April, 73-87.

3rd crossing opportunity:

Far2   Far1

2nd crossing opportunity:

Far2   Far1

1st crossing opportunity:

Far2   (Far1=0)

Near
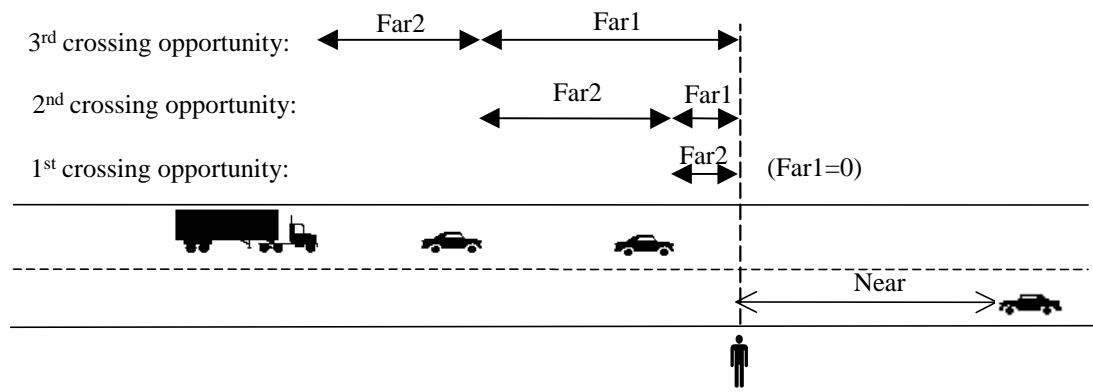
Figure 1. Representation of gaps defining a set of crossing opportunities for a given crossing situation.

Far2	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10+	Far1
10+	100	100	100	100	100	99	99	97	97	99	65	
9-10	100	100	100	100	100	100	100	99	97	94	65	
8-9	100	100	100	100	100	100	100	99	97	95	62	
7-8	100	100	100	100	100	100	100	98	97	95	60	
6-7	100	100	100	99	100	100	100	98	97	95	76	
5-6	100	100	100	100	100	99	99	98	96	95	80	6+
4-5	100	100	100	100	100	100	99	98	96	95	83	
3-4	100	100	100	100	100	99	98	97	96	95	94	
2-3	100	100	100	100	100	100	100	100	99	99	98	
1-2	100	100	100	100	100	100	100	100	98	100	100	
0-1	100	100	100	100	100	100	100	100	100	100	100	
10+	100	100	100	100	95	88	86	73	33	18	7	
9-10	100	100	100	100	97	78	60	41	23	4	8	
8-9	100	100	100	100	97	81	65	49	33	17	6	
7-8	100	100	100	100	97	84	70	57	43	30	17	
6-7	100	100	100	100	98	87	76	65	54	42	25	
5-6	100	100	100	97	98	90	81	72	64	55	30	4-6
4-5	100	100	100	98	99	93	86	80	74	68	54	
3-4	100	100	100	100	99	96	92	88	84	81	78	
2-3	100	100	100	100	100	99	97	96	95	93	96	
1-2	100	100	100	100	98	100	91	100	100	100	99	
0-1	100	100	100	100	100	100	100	100	100	100	100	
10+	100	100	100	99	92	74	64	35	22	29	4	
9-10	100	100	100	100	88	62	37	11	0	0	0	
8-9	100	100	97	100	89	67	45	23	2	0	7	
7-8	100	100	100	100	90	72	54	36	18	0	12	
6-7	100	100	100	100	91	76	62	48	33	19	26	
5-6	100	100	100	100	92	81	70	60	49	38	46	2-4
4-5	100	100	100	100	93	86	79	72	65	58	63	
3-4	100	100	100	99	94	90	87	84	80	77	83	
2-3	100	100	100	99	95	95	95	96	96	97	99	
1-2	100	100	100	100	100	100	100	100	100	100	100	
0-1	100	100	100	100	100	100	100	100	100	100	100	
10+	100	100	100	99	98	85	57	29	25	14	4	
9-10	100	100	100	100	82	60	38	16	0	0	19	
8-9	100	100	100	100	84	66	48	29	11	0	19	
7-8	100	100	100	100	87	72	57	42	27	12	15	
6-7	100	94	100	100	90	78	67	55	44	32	39	
5-6	100	100	98	100	92	84	76	68	60	52	58	0-2
4-5	100	100	100	100	95	91	86	81	77	72	82	
3-4	100	100	100	99	98	97	96	94	93	92	95	
2-3	100	100	100	100	100	100	100	100	100	100	100	
1-2	100	100	100	100	100	100	100	100	100	100	100	
0-1	100	100	100	100	100	100	100	100	100	100	100	
10+	100	100	99	95	79	48	23	16	8	4	2	
9-10	100	100	100	82	63	43	24	5	0	0	10	
8-9	100	100	98	85	69	53	36	20	4	0	8	
7-8	100	99	100	88	75	62	49	35	22	9	18	
6-7	100	100	99	92	81	71	61	50	40	30	38	
5-6	100	100	99	95	88	80	73	65	58	51	72	=0
4-5	100	99	100	98	94	89	85	80	76	71	94	
3-4	100	100	100	100	100	98	97	95	94	92	99	
2-3	100	100	100	100	100	100	98	100	100	100	100	
1-2	100	100	100	100	100	100	100	100	100	100	100	
0-1	100	100	100	100	100	100	100	100	100	100	100	
Far2	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10+	Far1

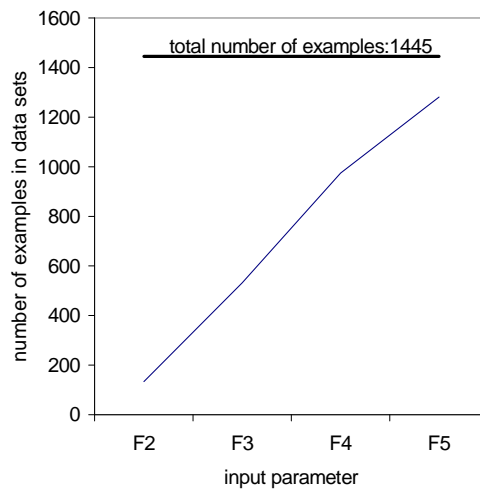Near

Figure 2. Probability matrices for rejection of crossing opportunities for 10.0m wide roads (Hunt and Griffiths, 1991).
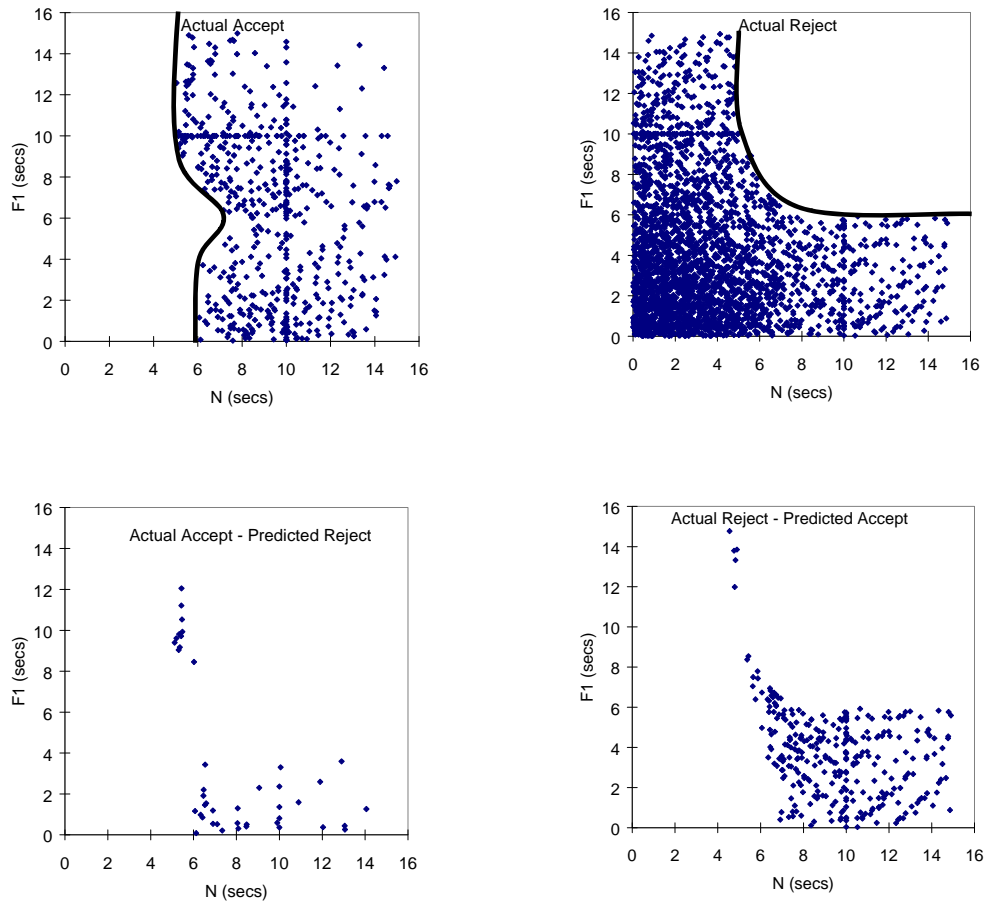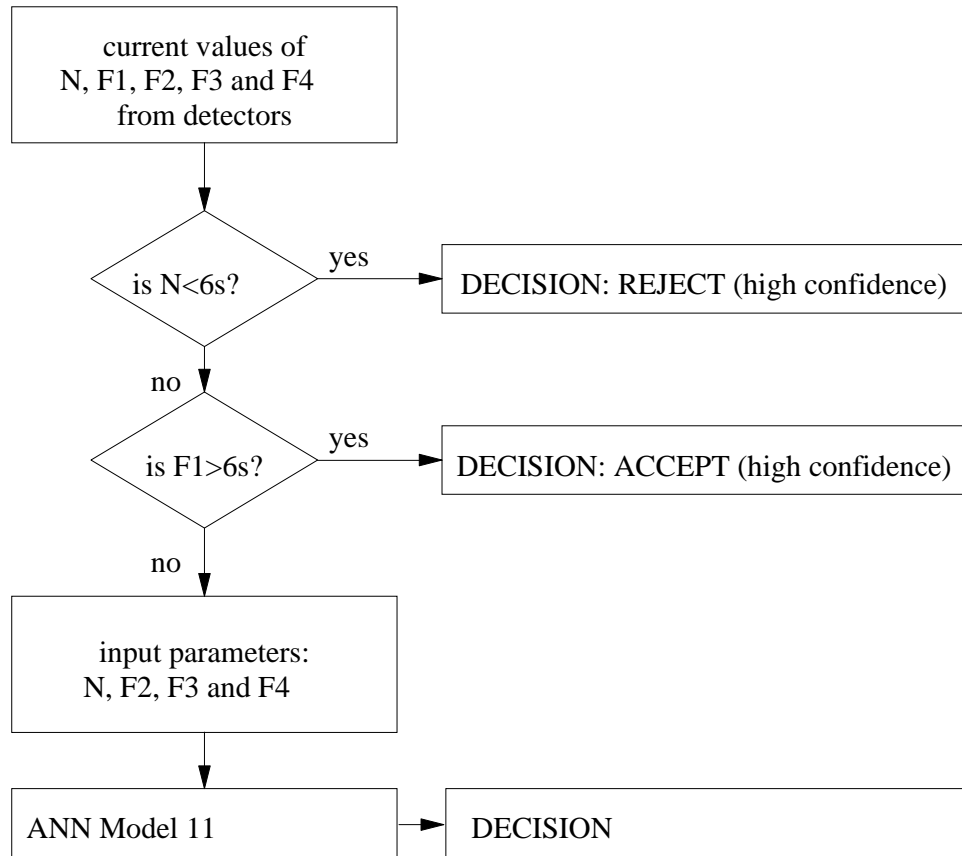
Figure 3. Validation data sets with ANN misclassifications and classification boundaries.

Figure 4. Number of examples where N>6s and F1<6s in which input parameter values are missing and are defaulted to a value of 10s.

```
                ┌─────────────────────┐
                │   current values of │
                │   N, F1, F2, F3 and F4 │
                │    from detectors   │
                └─────────────────────┘
                           │
                           ▼
                         ╱ ╲           yes      ┌────────────────────────────────────┐
                        ╱   ╲ ─────────────────▶│ DECISION: REJECT (high confidence) │
                        ╲ is N<6s? ╱            └────────────────────────────────────┘
                         ╲   ╱
                          ╲ ╱
                    no     │
                           ▼
                         ╱ ╲           yes      ┌────────────────────────────────────┐
                        ╱   ╲ ─────────────────▶│ DECISION: ACCEPT (high confidence) │
                        ╲ is F1>6s? ╱           └────────────────────────────────────┘
                         ╲   ╱
                          ╲ ╱
                    no     │
                           ▼
                ┌─────────────────────┐
                │   input parameters: │
                │   N, F2, F3 and F4  │
                └─────────────────────┘
                           │
                           ▼
            ┌─────────────────────┐  ┌────────────────────────────────┐
            │   ANN Model 11      │─▶│ DECISION                       │
            └─────────────────────┘  └────────────────────────────────┘
```

Figure 5. Hybrid rule-based/backpropagation ANN model specification for gap

acceptance decisions for an average pedestrian crossing a 10.0m wide road.

Decision	Near (s)	Far1 (s)	Far2 (s)
Reject	20.0	0.0	4.1
Reject	20.0	4.1	3.2
Reject	20.0	6.3	1.7
Reject	20.0	8.0	3.3
Accept	20.0	11.3	8.0

Table 1. A 'set' of crossing opportunities presented to a pedestrian at a single point in time according to the surrounding pattern of vehicle arrivals.

ANN model number	available input parameters	rejected input parameters	total no. of inputs including trans-formations	data set examples used	correct classification rate in validation (%)
1	N,F1,F2,F3,F4,F5	F3,F5	4	all	90
2	N,F1	none	6	all	89
3	N	none	3	all	85
4	N,F1,F2,F3,F4,F5	F1	7	N>6s, F1<6s	78
5	N,F1	none	3	N>6s, F1<6s	61
6	N	none	2	N>6s, F1<6s	52

Table 2. Performance based on available input parameters and data.

ANN model number	misclassifications/total number of examples					overall classification error in validation (%)
	$N < 6s$ $F1 < 6s$	$N < 6s$ $F1 > 6s$	$N > 6s$ $F1 > 6s$	$N > 6s$ $F1 < 6s$	overall	
1	0/1848	22/702	16/298	309/702	347/3550	9.77
4	0[*]/1848	41[*]/702	20[*]/298	157/702	218/3550	6.14

[*]NN Model 4 was derived using only a subset of the data (N>6s, F1<6s) - the classification is taken to be the majority verdict decision

Table 3. Comparison of pure ANN model and hybrid rule-based/ANN model performance for validation data.

ANN model number	input parameters	imposed hidden layer size limit (PEs)	resulting hidden layer size (PEs)	correct classification rate in validation (%)
8	N,F2,F3,F4,F5	30	17	81.1
9	N,F2,F3,F4,F5	10	7	75.9
10	N,F2,F3,F4,F5	5	5	76.9
11	N,F2,F3,F4	30	10	81.9

Table 4. Models' performance with varying network size.