WORKING PAPER
ITS-WP-96-17

Impact of Data Quality on the Performance of Neural Network Incident Detection Models

by
H Dia
and
G Rose

INSTITUTE OF TRANSPORT STUDIES
The Australian Key Centre in Transport Management
Monash University
Clayton Vic 3168

ISSN-015602116

No: Working Paper ITS-WP-96-17

Title: Impact of Data Quality on the Performance of Neural Network Incident Detection Models

Author: Mr Hussein Dia
PhD Candidate
Department of Civil Engineering
Monash University

Dr Geoff Rose
Senior Lecturer
Department of Civil Engineering
Monash University

Source: This paper was submitted for the 3rd World Congress on Intelligent Transport Systems, Florida, USA, October, 1996.

Contact: Institute of Transport Studies (Monash)
The Australian Key Centre in Transport Management
Department of Civil Engineering
Monash University
Clayton Vic 3168
Tel. No. (03) 9905 9627
Fax. No. (03) 9905 4944
Email :itsinfo@eng.monash.edu.au

Date: October, 1996

Abstract:

One of the challenges in using field data for the development of neural network incident detection models is to be able to train models that can handle the noisy nature of the loop detector data. The noise in the field data, which may be the result of either a systematic or random error, can have an adverse effect on the performance of an incident detection model, especially in terms of false alarm rate. This paper describes a number of procedures for evaluating the impact of data quality on the performance of a neural network incident detection model that was trained and tested on field data (comprising speed, flow and occupancy measurements) collected from a number of freeways in Melbourne, Australia. Since this model was developed for implementation in an actual system, the paper also reports on a number of techniques and procedures for evaluating the model's performance in the case of missing or incorrect data either during training or after implementation. In addition to the original research findings reported in this paper, the described procedures are also of interest to practitioners since they address many issues relevant to the implementation of incident detection systems and the quality of the detector data. These issues include evaluating the impact of detector failures, communications malfunction and missing or incorrect data on the model's performance. In addition, the paper also describes how the same procedures can be used to evaluate the impact of speed data on the model's performance (ie. the impact of using dual-loop instead of single- loop detectors).

Keywords:

incident detection models, systematic error, random error, freeway incidents, AID systems, input clamping, communications malfunction, model performance, noise levels, random noise, ANN model.

INTRODUCTION

1. The high contribution of freeway incidents to urban congestion, pollution and deteriorated safety conditions has prompted road authorities around the world to implement automatic incident detection (AID) systems on urban freeways. A number of AID models with varying structures and data requirements have been developed over the last two decades. Common measures for the evaluation of incident detection algorithms, which are independent of the theoretical foundations of the algorithms, have also been formulated. These are detection rate (DR), false alarm rate (FAR) and mean time to detect (MTTD). Few of the developed algorithms, however, have been implemented in practice due to various limitations and varying operational levels in terms of the incident detection performance criteria, ie. DR, FAR and MTTD.

2. One promising approach for the implementation of AID systems involves the application of Artificial Neural Networks (ANNs). These are also referred to as parallel distributed processing systems or connectionist systems and have been implemented within recent years as a paradigm of computation and knowledge representation. The results reported in this paper are based on research work that was undertaken to develop artificial neural network freeway incident detection models using real world data. An incident data set of 100 incidents were collected from the Tullamarine Freeway in Melbourne, Australia. A training (calibration) data set of 60 incidents were used to develop an AID system based on ANNs (Dia and Rose, 1995) while the remaining 40 incidents (validation-test data set) were used to validate the performance of the ANN model (Dia, 1996; Dia *et al.* 1996). In the development of the ANN models, speed, flow and occupancy data were used to train the models.

3. Once the ANN incident detection model was developed, it was important to evaluate the relative importance of the various input parameters of the model and provide an insight into how the model works. This is an important issue from a practical perspective because it is necessary to evaluate the incident detection model's performance when certain input features become unavailable, either during training or after implementation in the field. The issue of data unavailability could arise when the detector technology used in the field does not capture all the data needed to train the full model (eg. the use of single inductive loops that do not measure speed or other detector technologies that may only provide a limited set of data) or due to detector failure that could prevent the detectors from providing the full data needed for the model's operation. It is also important to be able to evaluate the model's performance in the case it is provided with incorrect data. This paper will therefore discuss and apply some methods reported in the literature for addressing these practical issues.

4. In evaluating the impact of data quality on the model's performance, two methods are used. In the first method, sensitivity analysis techniques are used to test the prediction power of the model (after training is complete) when certain variables that were included in the model's development become unavailable (Masters, 1993). This method has practical implications in incident detection and can be used to evaluate the impact of detector failure and the impact of missing speed data on model performance. The second method is used to evaluate the model's formulation or mis-specification error by assessing the importance of including or excluding a certain variable from the model during training. The value of omitting a relevant variable or including an irrelevant variable (from the training) can therefore be determined. This has practical implications in incident detection and can be used to evaluate the model's performance when the detector technology does not capture all the data needed for training the full model. In addition, this paper also addresses the impact of systematic and random errors on model performance which also have practical implications and can be used to evaluate the impact of incorrect data or different detector technologies on model's performance.

SENSITIVITY ANALYSIS BY INPUT CLAMPING

6. As was mentioned previously, the information obtained from sensitivity analysis can be used for a number of practical applications and is particularly important in real-time systems, such as incident detection models, where the computational requirements of the system need to be minimised. One method of doing this is through reducing the size of the network by determining which inputs can be eliminated without allowing the model's performance to deteriorate. This information can also be used to determine the impact of faulty detectors (where related input values are received as zeros from the communications system) and the impact of missing speed data on the performance of the ANN model that was trained using speed data.

7. A simple and relatively effective method for assessing the importance of inputs is sensitivity analysis by input clamping (Masters, 1993). The network's performance is evaluated with one of the inputs clamped to a fixed value for the entire validation-test data set. The importance of that input can be gauged by its effect on the area under the Performance Envelope Curve (PECA) of the ANN model (Masters, 1996). In the case of incident detection, a model with an ideal performance of (DR=100% and FAR=0%) has a PECA of 10,000 (Dia, 1996). In the input clamping method, if the PEC area for a certain model decreases substantially with that input clamped, then that input is important for incident detection. The results from this analysis are only meaningful if one input is clamped at a time (Masters, 1993) due to the fact that the network was trained based on all inputs and hence the absence of many inputs at the same time can drive the network into providing distorted responses to the remaining inputs (Masters, 1993). The value at which an input variable is clamped is typically selected to be within the range of actual values expected for that input variable when the network is eventually implemented in the field. Extreme values are avoided, where possible, since these can cause other inputs to be ignored (Masters, 1993).

Impact of Detector Failure/Communications Malfunction on Model Performance

8. One of the practical applications for sensitivity analysis is the investigation of the impact of detector failure or communications malfunction on incident detection performance. The quality of the received data is affected by these failures which can have a profound effect on the performance of the algorithm, especially in terms of the false alarm rate. The causes of error in the data are attributed to a variety of factors including detector lock-up due to high levels of vibration from heavy vehicles or external interference, loops damaged during road works, excavation or pavement resurfacing or electrical failure of detector cards (Snell *et al.*, 1992).

9. These errors were reflected in the data files used in this study in two ways. The first type of error, detector failure, was observed when a particular station (consisting of all detectors in all lanes for a given direction) failed to transmit any data back to the Traffic Control and Communications Centre. When this occurred, the three inputs (speed, flow and occupancy) for that station were reported as zeros in the entire data file (for the whole 24-hour data file). The other type of error (communications malfunction) was observed when a particular station did not transmit the traffic data (speed, flow and occupancy) for a certain number of 20-second intervals (data for these intervals were also transmitted as zeros), but the communications system reset itself afterwards and the data transmission was resumed. In both cases, this resulted in a false alarm being raised in the ANN incident detection model. The impact of detector failure and communications malfunction on incident detection performance can therefore be investigated by using a value of zero for clamping the value of the input variable under consideration. If all inputs are clamped to zero, then this results in a DR of zero and a FAR of 100%. Table 1(a) and Figure 1(a) show the sensitivity analysis results obtained by clamping one input at a time for the validation-test data set (40 incidents) in a descending order of performance (in terms of the PEC area). Although some of the results in Table 1(a) are of less practical relevance, they nevertheless reveal that the variation in model performance was substantial for the cases where the upstream speed, downstream speed or both were clamped. Compared to the original model with all inputs free, clamping the upstream speed input to zero resulted in just over a 30% decrease in the overall performance. Clamping the downstream speed input to zero, however, resulted in less than a

10% decrease in the overall performance of the model. Although the percentage decrease in performance as a result of clamping the downstream speed was not as substantial as for the upstream speed, the resulting incident detection performance was still not acceptable.

Impact of Missing Speed Data on Model Performance

10. It was previously mentioned that the results from the input clamping technique are only meaningful if one input is clamped at a time. From a practical perspective, however, it is important to evaluate the impact of missing speed data at both the upstream and downstream stations on the performance of the ANN model that was trained using these speeds. This analysis can provide an insight into the performance of the trained ANN model should it be implemented on a facility that only uses single loop detectors (without retraining). In order to investigate this, the upstream and downstream speeds were clamped to zero values and the procedure repeated for the validation-test data set. The results of this evaluation are also shown in Table 1(a) and Figure 1(a). When both the upstream and downstream speeds are clamped (representing a situation where the model is implemented on a facility that only uses single loop detectors), the performance of the ANN model, based on the validation-test data set, deteriorates by about 25%.

11. It should be pointed out here that these results are for a model that was originally trained with speed data and was then tested on a facility where the speed data was not available. Figure 1(a) clearly indicates that failure to provide speeds at the upstream station of a section can result in a significant deterioration of incident detection performance. In the event of detector failures at the upstream station, measures could be implemented such that speed data from the immediate upstream station are provided until the detector problem is fixed. The satisfactory incident detection results obtained in this study from the modelling on the longer sections of the Tullamarine Freeway (about 1070 meters) suggest that it is feasible for such a strategy to be implemented in the case of detector failures provided that detectors are typically spaced at 500 meters. Since the failure to provide speed at the upstream station of a section is more serious for detector failures than communications malfunction, it is recommended that automated procedures be implemented for identifying detector failures at the traffic control centre and consequently implementing these strategies until the detectors are fixed.

VALUE OF INFORMATION

12. As was mentioned earlier, including or excluding a certain variable from the model during training can be used to determine the value of omitting a relevant variable or including an irrelevant variable from the training. This has practical implications in incident detection and can be used to evaluate the model's performance when the detector technology does not capture all the data needed for training the full model, especially in the case of using single loop detectors that do not provide speed measurements.

13. The method described here for evaluating the impact of using single-loop detectors on model performance, 'the backward elimination method', is similar to the methods used in regression analysis for the selection of variables and model building. In 'the backward elimination method', the training is started with all of the inputs being used, then one input is eliminated at a time and the network is retrained. The input whose elimination causes the least decrease in performance is removed. The procedure is repeated until there is no remaining input whose elimination can be tolerated (Masters, 1993).

14. The ANN model incident detection performance is measured using the area under the performance envelope curve (PEC). The results for the 'backward elimination' procedure are summarised in Table 1(b) in a descending order of performance (in terms of PEC area). Examination of these results reveal that the deterioration in performance was substantial for five of the models, namely models 7, 16, 20, 19 and 12. This is reflected in the low values of the PEC area for these models, the high value of false alarm rates for the shown decision threshold and the large percentage decrease in the PEC area compared to the original model.

15. It is important to mention here that when the value of information is being evaluated from a practical perspective, some of the combinations of eliminated inputs in Table 1(b) become less relevant. What is important from a practical point of view, however, is the impact of using single loop detectors (which do not provide speed measurements) on the model's performance. To investigate this, the upstream and downstream speed inputs were eliminated from the initial model and the model was retrained. The results of this procedure are shown for model 21 in Table 1(b). These results indicate that training ANN incident detection models using data from single loop detectors (instead from dual loops which provide speed measurements) would result in only a 0.18% deterioration in performance of the ANN model. It should be emphasised, however, that the best ANN performance is only achieved by training the ANN model on all the available data (speed, flow and occupancy from both the upstream and downstream stations).

16. Table 1(b) shows that the elimination of the upstream speed alone, compared to the initial model, resulted in a decrease of 0.83% only (model 1). However, the deterioration in performance increased significantly when the upstream speed was eliminated along with other inputs. The most significant deterioration in performance was observed for models 7, 16, 20, 19 and 12 with a decrease of about 13.97%, 27.67%, 33.35%, 35.55% and 35.70%, respectively. All of these models had their upstream speed and occupancy inputs eliminated.

IMPACT OF INPUT NOISE LEVELS ON MODEL PERFORMANCE

17. The ANN incident detection model developed in this study was trained on 'real world' data to detect incidents. The traffic data collected from the freeways were found to be inherently noisy (Rose and Dia, 1995). Therefore, a very useful indicator of the power of the network to detect an incident is to determine how much noise can be tolerated by the network at the input layer (Masters, 1993).

18. The noise in the data can be a result of either a systematic or random error. With a systematic error, all inputs are affected by the same amount of fixed noise. The random error, however, affects data measurements at different and random noise levels. The next sections describe the procedure implemented for investigating the impact of both types of error on model performance.

Systematic Error

19. To introduce a systematic error across all variables, the input variables of the validation-test data set were multiplied by a fixed amount of noise and presented to the model. Table 2(a) and Figure 1(b) show the results obtained using fixed noise levels between -50% to +50%.

21. These results are potentially relevant from a practical perspective in the context of the detection technology being used. Different detector technologies may have variable systematic errors which in turn can have a different effect on model performance. The results presented in Table 2(a) can therefore be used to evaluate the impact of the systematic error resulting from using different detection technologies on the performance of the ANN incident detection model.

22. Examination of the results in Table 2(a) reveal that the incident detection performance of the model remains robust for systematic error levels between +1% and +5%. The results also show that the model's performance is more robust for the decreased systematic noise levels than for the increased noise levels, especially for the range between 10-20 percent where the performance as a result of the systematic noise decrease remained fairly constant. The negative impact of the systematic increase in noise levels on the performance of the model, however, was more substantial than that for the systematic decrease in the noise levels, as shown in Figure 1(b).

Random Noise

23. To introduce a random error in the data, the validation-test data set of forty incidents were contaminated with a random amount of noise across all input variables. The random noise was uniformly distributed over the range (-0.01 INto (+0.01 $\times$ nl), where (nl) is the noise level value. Therefore, a noise level of 10 would introduce a uniformly distributed random noise of +/-0.10 across the inputs (the inputs are scaled between 0 and 1). The model is then tested on the validation-test data set and the PEC area calculated. The impact of noise levels on model performance can be determined by repeating the procedure using different noise levels (Masters, 1993). These results, presented in Table 2(b) and Figure 1(c), show that the model's performance remains robust for noise levels below 10% but starts to deteriorate as noise levels increase beyond that. The results also suggest that larger random errors, eg. more than 20%, can have adverse effects on the performance of the model. These results also have practical relevance and can be used to evaluate the model's performance as a result of random traffic fluctuations.

CONCLUSIONS

24. This paper examined how the trained ANN incident detection model responds to changes in one or more of the input parameters. The information obtained from this analysis provides an insight into the inputs that can be eliminated without allowing the model's performance to deteriorate. This issue has many practical implications and can be used to determine the impact of faulty detectors on incident detection performance. Based on the results obtained from the sensitivity analysis by input clamping technique, it was found that:

- Compared to the initial model with all inputs free, clamping the upstream and downstream speed inputs (representing a situation where the model is implemented on a facility that only uses single loop detectors), the performance of the model deteriorated by about 24.5%. Training the ANN model without the upstream and downstream speeds, however, results in only a deterioration of 0.18% in model performance. This issue is clearly an area worthy of more investigation in future research efforts. The results from the input clamping technique clearly showed that failure to provide speed data at a station (which can arise from a detector failure or communications malfunction) can result in a significant deterioration of model performance within that section. In this event, measures could be implemented such that speed data from the immediate upstream station are provided until the detector problem is fixed. It is also recommended that automated procedures be implemented for identifying detector failures at the traffic control centres.

25. Based on the results obtained from 'the backward elimination method', it was found that:

- The overall best incident detection performance is obtained with models that are trained using all the six original variables, ie. speed, flow and occupancy at both the upstream and downstream stations.

- Eliminating any single input from the original model and retraining the model does not result in a significant deterioration of model performance. However, eliminating groups of input variables (especially those including the upstream speed input) and retraining the resulting model was found to affect the model's performance substantially. Eliminating the upstream and downstream speeds from the original ANN model and retraining the model without these inputs resulted in only a 0.18% deterioration in performance compared to the original ANN model that was trained using these inputs.

26.Finally, the ANN model was examined to determine how much noise at the input layer can be tolerated by the network. This is a very useful indicator of the power of the network to detect incidents with noisy or corrupted data. Based on the results obtained from this part of the analysis, it was found that:

- The incident detection performance of the model remains robust for systematic error levels between +1% and +5%. These results, however, are based on an identical systematic error that was introduced across all input variables. From a practical perspective, it is possible that different input variables could be affected by non-identical systematic errors. This is clearly an area where further research work is also needed. The results from this analysis can be used to evaluate the impact of the systematic error resulting from using different detection technologies on the performance of the ANN incident detection model.

- When corrupted by random noise, the model's performance remains robust for noise levels below 10%. The results also suggest, that larger random errors, eg. more than 20%, can have adverse effects on the performance of the model. These results can be used to evaluate the impact of random traffic fluctuations on the performance of the ANN incident detection model.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Victoria State Road Authority (VicRoads) and in particular the generous assistance of Mr. Francis Sin and Mr. Anthony Snell.

REFERENCES

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

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

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

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

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

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

ANN Model's Input						Incident Detection Performance Decision Threshold=0.5			Validation-Test Data Set	
Upstream			Downstream			DR (%)	FAR (%)	MTTD (Sec)	PEC Area	% decrease in PEC Area
Speed US	Flow UF	Occ UO	Speed DS	Flow DF	Occ DO					
✓	✓	✓	✓	✓	✓	95.0	2.1	135	9963[a]	----
✓	✓	✓	✓	✓	✗	92.5	2.2	138	9963	0.00
✓	✗	✓	✓	✓	✓	100.0	7.3	94	9922	0.41
✓	✓	✗	✓	✓	✓	92.5	1.5	158	9814	1.49
✓	✓	✓	✓	✗	✓	92.5	2.1	140	9806	1.58
✓	✓	✓	✗	✓	✓	62.5	1.1	142	9012	9.54
✗	✓	✓	✗	✓	✓	42.5	8.45	117	7527	24.5
✗	✓	✓	✓	✓	✓	100.0	91.6	24	6722	32.5

✓: Free input ✗: Clamped input

[a]: Represents the initial model with all inputs free

Table 1(a) : Impact of input clamp (input set to zero) on model performance-validation-test data set

ANN Model	Inputs Eliminated	Incident Detection Performance Decision Threshold = 0.5			PEC Area	% Decrease in PEC Area[a]
		Detection Rate (%)	False Alarm Rate (%)	Mean Time-to-detect (Second)		
Initial[b]	None	95.0	2.10	135	9963	---
21	US,DS	95.0	2.33	141	9945	0.18
14	UO,DS,DF	97.5	2.73	125	9939	0.24
11	UO,DO	100.0	2.65	126	9937	0.26
3	UO	97.5	2.71	122	9933	0.30
2	UF	97.5	2.52	128	9931	0.32
4	DS	95.0	2.18	130	9931	0.32
9	UO,DS	95.0	2.03	135	9931	0.32
8	UF,UO	97.5	2.75	131	9930	0.33
6	DO	92.5	2.06	137	9928	0.35
5	DF	100.0	3.78	118	9924	0.39
18	UO,DS,DF,DO	95.0	2.21	133	9917	0.46
17	UF,UO,DS,DO	82.5	1.11	153	9914	0.49
15	UO,DS,DO	92.5	2.37	135	9904	0.59
1	US	97.5	7.83	110	9880	0.83
10	UO,DF	95.0	1.65	147	9815	1.49
13	UF,UO,DS	95.0	2.12	132	9801	1.62
7[c]	US,UO	85.0	30.9	52	8571	13.97
16[c]	US,UO,DS,DO	92.5	50.3	52	7206	27.67
20[c]	US,UO,DS,DF,DO	100.0	67.2	47	6640	33.35
19[c]	US,UF,UO,DS,DO	95.0	67.4	30	6421	35.55
12[c]	US,UO,DS	92.5	65.3	37	6406	35.70

[a] : Compared to the initial model in the first row of the table

[b] : Represents the original model with all six inputs being used

[c] : Represents models which exhibited higher variations in incident detection performance

Table 1(b): Summary of the impact of input elimination on incident detection performance

Noise (%)	Incident Detection Performance Decision Threshold = 0.5			PEC Area	% Change in PEC Area[a]
	Detection Rate (%)	False Alarm Rate (%)	Mean Time-to-detect (Second)		
0	95.0	2.10	135	9963	---
1	92.5	2.01	138	9962	-0.01
2	95.0	2.02	136	9962	-0.01
3	92.5	1.96	140	9963	0.00
4	92.5	1.92	140	9966	0.03
5	92.5	1.32	147	9968	0.05
10	92.5	1.60	144	9723	-2.41
20	92.5	1.30	151	9600	-3.64
30	90.0	1.09	162	9596	-3.68
40	90.0	0.96	173	9594	-3.70
50	87.5	0.86	173	9496	-4.69
-1	95.0	2.18	125	9966	0.03
-2	95.0	2.17	135	9965	0.02
-3	95.0	2.23	134	9962	-0.01
-4	95.0	2.38	129	9959	-0.04
-5	97.5	2.58	125	9959	-0.04
-10	100.0	3.72	113	9962	-0.01
-20	100.0	13.7	76	9951	-0.12
-30	100.0	53.5	38	9819	-1.45
-40	100.0	97.6	21	9614	-3.50
-50	100.0	99.9	20	9467	-4.98

[a]: Compared to the original model in the first row of the table

Table 2(a): Impact of increased/decreased systematic noise on model performance

Noise (%)	Incident Detection Performance Decision Threshold = 0.5			PEC Area	% Change in PEC Area[a]
	Detection Rate (%)	False Alarm Rate (%)	Mean Time-to-detect (Second)		
0	95.0	2.10	135	9963	---
1	95.0	2.06	135	9963	0.00
2	95.0	2.07	135	9965	0.02
3	97.5	2.16	134	9969	0.06
4	97.5	2.21	133	9968	0.05
5	100.0	2.20	135	9970	0.07
10	100.0	2.93	123	9967	0.04
20	97.5	8.40	105	9922	-0.41
30	100.0	15.80	101	9806	-1.58
40	100.0	18.62	74	9661	-3.03
50	100.0	22.27	73	9421	-5.44
60	100.0	23.38	68	9379	-5.86
70	100.0	23.68	67	9285	-6.81
80	100.0	24.16	82	9268	-6.98
90	100.0	24.70	73	9141	-8.25
100	100.0	24.40	68	9203	-7.63

[a]: Compared to the original model in the first row of the table

Table 2(b): Impact of random noise levels on model performance

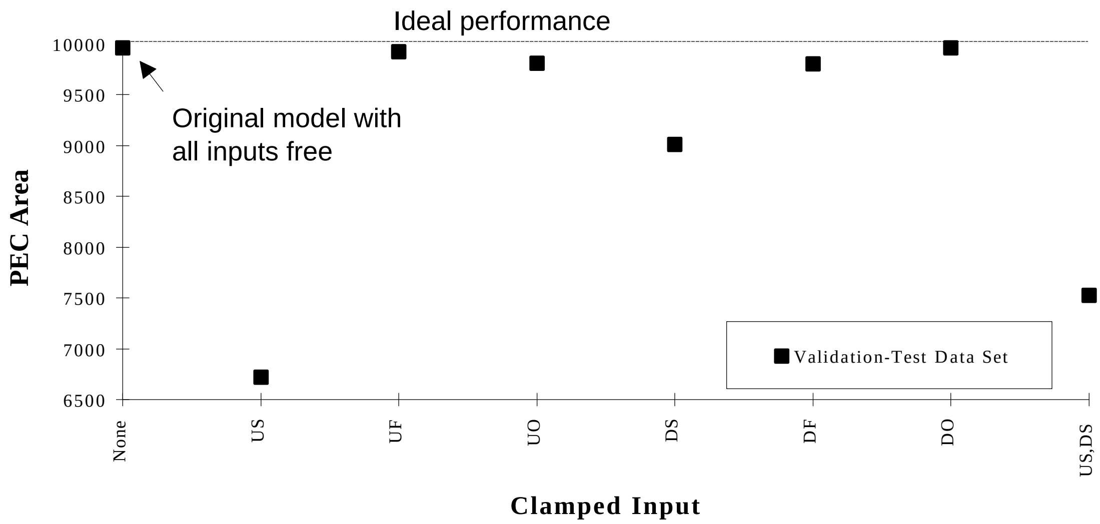

Figure 1(a)- Impact of input clamp (input set to zero) on model performance

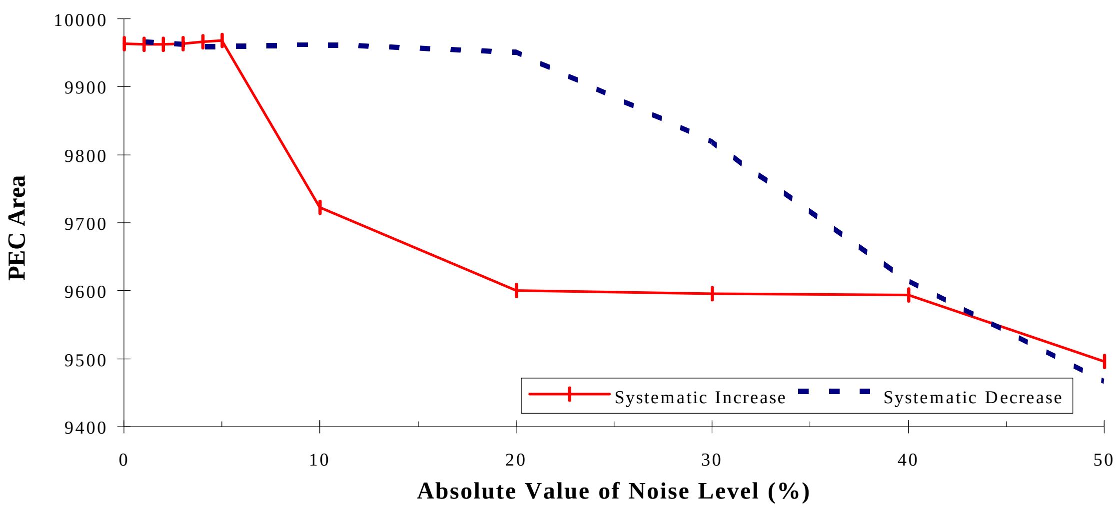

Figure 1(b)- Impact of increased/decreased systematic noise on model performance

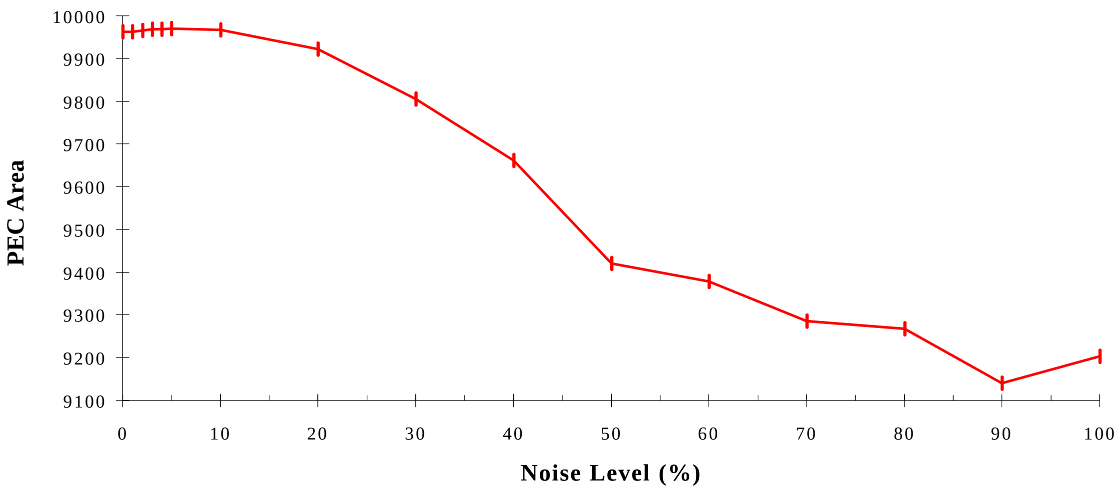

Figure 1(c)- Impact of random noise levels on model performance