Journal of Forecasting
J. Forecast. **17**, 481–495 (1998)

How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation

MONICA ADYA[1]* AND FRED COLLOPY[2]
[1]*University of Maryland at Baltimore County, USA*
[2]*Case Western Reserve University, USA*

ABSTRACT

Despite increasing applications of artificial neural networks (NNs) to forecasting over the past decade, opinions regarding their contribution are mixed. Evaluating research in this area has been difficult, due to lack of clear criteria. We identified eleven guidelines that could be used in evaluating this literature. Using these, we examined applications of NNs to business forecasting and prediction. We located 48 studies done between 1988 and 1994. For each, we evaluated how effectively the proposed technique was compared with alternatives (effectiveness of validation) and how well the technique was implemented (effectiveness of implementation). We found that eleven of the studies were both effectively validated and implemented. Another eleven studies were effectively validated and produced positive results, even though there were some problems with respect to the quality of their NN implementations. Of these 22 studies, 18 supported the potential of NNs for forecasting and prediction. © 1998 John Wiley & Sons, Ltd.

KEY WORDS artificial intelligence; machine learning; validation

INTRODUCTION

An artificial neural network (NN) is a computational structure modelled loosely on biological processes. NNs explore many competing hypotheses simultaneously using a massively parallel network composed of non-linear relatively computational elements interconnected by links with variable weights. It is this interconnected set of weights that contains the knowledge generated by the NN. NNs have been successfully used for low-level cognitive tasks such as speech recognition and character recognition. They are being explored for decision support and knowledge induction (Shocken and Ariav, 1994; Dutta, Shekhar and Wong, 1994; Yoon, Guimaraes, and Swales 1994).

In general, NN models are specified by network topology, node characteristics, and training or learning rules. NNs are composed of a large number of simple processing units, each interacting

* Correspondence to: Monica Adya, Department of Information Systems, University of Maryland at Baltimore County, Baltimore, MD 21250, USA, E-mail: adya@umbc.edu

CCC 0277–6693/98/050481–15$17.50

with others via excitatory or inhibitory connections. Distributed representation over a large number of units, together with interconnectedness among processing units, provides a fault tolerance. Learning is achieved through a rule that adapts connection weights in response to input patterns. Alterations in the weights associated with the connections permits adaptability to new situations (Ralston and Reilly, 1993). Lippmann (1987) surveys the wide variety of topologies that are used to implement NNs.

Over the past decade, increasing research efforts have been directed at applying NNs to business situations. Despite this, opinions about the value of these technique have been mixed. Some consider them effective for unstructured decision-making tasks (e.g. Dutta *et al.*, 1994); other researchers have expressed reservations about their potential, suggesting that stronger empirical evidence is necessary (e.g. Chatfield, 1993).

The structure of this paper is as follows. First, we explain how studies were selected. Then we describe the criteria that we used to evaluate them. Next, we discuss our findings when we applied these criteria to the studies. Finally, we make some recommendations for improving research in this area.

HOW STUDIES WERE SELECTED

We were interested in the extent to which studies in NN research have contributed to improvements in the accuracy of forecasts and predictions in business. We searched three computer databases (the *Social Science Citation Index*, and the *Science Citation Index*, and *ABI Inform*) and the proceedings of the IEEE/INNS Joint International Conferences. Our search yielded a wide range of forecasting and prediction-oriented applications, from weather forecasting to predicting stock prices. For this evaluation we eliminated studies related to weather, biological processes, purely mathematical series, and other non-business applications. We identified additional studies through citations. This process yielded a total of 46 studies. We subsequently surveyed primary authors of these studies to determine if our interpretation of their work was accurate and to locate any other studies that should be included in this review. Twelve (26%) of the authors responded and two identified one additional study each. These two were included in the review. The current review, therefore, includes 48 studies between 1988 and 1994 that used NNs for business forecasts and predictions.

CRITERIA USED TO EVALUATE THE STUDIES

In evaluating the studies, we were interested in answering two questions. First, did the study appropriately evaluate the predictive capabilities of the proposed network? Second, did the study implement the NN in such a way that it stood a reasonable chance of performing well? We call these effectiveness of validation and effectiveness of implementation respectively.

Effectiveness of validation

There is a well-established tradition in forecasting research of comparing techniques on the basis of empirical results. If a new approach is to be taken seriously, it must be evaluated in terms of alternatives that are or could be used. If such a comparison is not conducted it is difficult to argue that the study has taught us much about the value of NNs for forecasting. In fairness to the researchers conducting the studies, it should be noted that this is not always their objective. Sometimes they are using the forecasting or prediction case as a vehicle to explore the dynamics of a particular technique or domain. (For instance, Piramuthu, Shaw and Gentry, 1994, proposed the use of a modified backpropagation algorithm and tested it in the domain of loan evaluations.) Still, our purpose here is to answer the question, what do these techniques contribute to our understandings and abilities as *forecasters*?

To evaluate the effectiveness of validation, we applied the three guidelines described in Collopy, Adya and Armstrong(1994).

Comparisons with well-accepted models

Forecasts from a proposed model should perform at least as well as some well-accepted reference models. For example, if a proposed model does not produce forecasts that are at least as accurate as those from a naive extrapolation (random walk), it cannot really be argued that the modelling process contributes knowledge about the trend.

Use of ex ante validations

Comparison of forecasts should be based on *ex ante* (out-of-sample) performance. In other words, the sample used to test the predictive capabilities of a model must be different from the samples used to develop and train the model. This matches the conditions found in real-world tasks, where one must produce predictions about an unknown future or a case for which the results are not available.

Use of a reasonable sample of forecasts

The size of the validation samples should be adequate to allow inferences to be drawn. We examined the size of the validation samples used in the classification and time series studies separately. Most of the classification studies used 40 or more cases to validate. Time series studies typically used larger samples. Most of them used 75 or more forecasts in their validations.

Effectiveness of implementation

For studies that have effectively validated the NN we asked a second question: How well was the proposed architecture implemented? While a study that suffers from poor validation is not of much use in assessing the applicability of the technique to forecasting situations, one that suffers from poor implementation might still have some value. If a method performs comparatively well, even when it has not benefited from the best possible implementation, there is reason to be encouraged that it will be a contender when it has.

In determining the effectiveness with which a NN had been developed and tested, we used the guidelines for evaluating network performance suggested by Refenes (1995). Our implementation of some of the criteria (particularly that regarding stability of an implementation), varies from that of Refenes (1995).

- *Convergence*: Convergence is concerned with the problem of whether the learning procedure is capable of learning the classification defined in a data set. In evaluating this criterion, therefore, we were interested in the in-sample performance of the proposed network since it determines the network's convergence capability and sets a benchmark for assessing the generalizabilty, i.e. *ex ante* performance, of the network. If a study does not report in-sample performance on the network, we suggest caution in acceptance of its *ex ante* results.

- *Generalization*: Generalization measures the ability of NNs to recognize patterns outside the training sample. The accuracy rates achieved during the learning phase typically define the bounds for generalization. If performance on a new sample is similar to that in the convergence phase, the NN is considered to have learned well.
- *Stability*: Stability is the consistency of results, during the validation phase, with different samples of data. This criterion, then, evaluates whether the NN configuration determined during the learning phase and the results of the generalization phase are consistent across different samples of test data. Studies could demonstrate stability either through use of iterative resampling from the same data set or by using multiple samples for training and validation.

The criteria are sufficiently general to be applicable to any NN architecture or learning mechanism. Furthermore, they represent a distillation of the literature's best practice. The fact that a study failed to meet the criteria is not necessarily an indictment of that study. If we wish to use empirical studies to make a case for or against the applicability of NNs to forecasting or prediction, though, we must be able to determine which represent good implementations for that purpose.

In summary then, studies were classified as being of three types. Those that are well implemented and well validated are of interest whatever their outcome. They can be used either to argue that NNs are useful in forecasting or that they are not, depending upon outcome. These would seem to be the most valuable studies. The second type are studies which have been well validated, even though their implementation might have suffered in some respects. These are important when the technique they propose does well despite the limitations of the implementation. They can be used to argue that NNs are applicable and to establish a lower bound on their performance. Finally, there are studies that are of little interest, from the point of view of telling us about the applicability of neural nets to forecasting and prediction. Some of these have little value because their validation suffers. Others are effectively validated but produce null or negative results. Since it is not possible to determine whether these negative results are because the technique is not applicable or the result of implementation difficulties, the studies have little value as forecasting studies.

RESULTS

Twenty-seven of the studies were effectively validated. Appendix A reports our assessment of the validation effectiveness of each of the 48 studies. Eleven of the studies met the criteria for both implementation and validation effectiveness. Of the remaining 37 studies, 16 were effectively validated but had some problems with implementation. Eleven of these reported NN performance that was better than comparative models. Twenty-two (46%) studies, then, produced results that are relevant to evaluating the applicability of neural networks to forecasting and prediction problems. Table I provides a summary.

Five studies that met the criteria for effective validation but failed to meet those for effective implementation produced negative or mixed results. The most common problem with these studies was their failure to report in-sample performance of the NN, making it difficult to assess the appropriateness of the NN configuration implemented. It also makes it difficult to evaluate

Table I. Relationship of effectiveness to outcomes (number of studies)

	NN better	NN worse or inconclusive	Not compared
Problems with validations	11	3	7
Problems only with implementation	**11**	5	0
No problems either criteria	**8**	**3**	0

Studies in bold contribute to forecasting knowledge.

the generalizability of the NN since there is no benchmark for comparison. Consequently, the results of these studies must be viewed with some reservation. Of the 48 studies, 27 were effectively validated. Appendix B contains the evaluation of the implementations for each of these.

Effectively validated and implemented

Of the eleven studies that met the criteria for both implementation and validation effectiveness, eight were implemented in classification domains such as bankruptcy prediction. The remaining three studied time-series forecasting.

Two of the eight classification studies satisfied all of the effectiveness criteria yet failed to support their hypotheses that NNs would produce superior predictions. Gorr, Nagin and Szczypula (1994) compared linear regression, stepwise polynomial regression, and a three-layer NN with a linear decision rule used by an admissions committee for predicting student GPAs in a professional school. In a study of bankruptcy classification, Udo (1993) reported that NNs performed as well as, or only slightly better than, multiple regression although this conclusion was not confirmed by statistical tests.

Wilson and Sharda (1994) and Tam and Kiang (1990, 1992) developed NNs for bankruptcy classification. Wilson and Sharda (1994) reported that although NNs performed better than discriminant analysis, the differences were not always significant. The authors trained and tested the network using three sample compositions: 50% each of bankrupt and non-bankrupt firms, 80% of non-bankrupt and 20% of bankrupt firms, and 90% of non-bankrupt and 10% of bankrupt firms. Each such sample was tested on a 50/50, 80/20, and 90/10 training set yielding a total of nine comparisons. The NN outperformed discriminant analysis on all but one sample combination for which performance of the methods was not statistically different.

Tam and Kiang (1990, 1992) compared the performance of NNs with multiple alternatives: regression, discriminant analysis, logistic, *k* Nearest Neighbour, and ID3. They reported that the NNs outperformed all comparative methods when data from one year prior to bankruptcy was used to train the network. In instances where data for two years before bankruptcy was used to train, discriminant analysis outperformed NNs. In both instances, a NN with one hidden layer outperformed a linear network with no hidden layers.

In a similar domain, Salchenberger, Cinar and Lash (1992) and Coats and Fant (1992) used NNs to classify a financial institution as failed or not. Salchenberger *et al*. (1992) compared the performance of NNs with logit models. The network performed better than logit models in most instances where the training and testing sample had equal representation of failed or non-failed institutions. The NN outperformed logit models in a diluted sample where about 18% of the sample was comprised of failed institutions' data. Coats and Fant (1993) used the Cascade Correlation algorithm for predicting financial distress. Comparative assessments were made with discriminant analysis. The NN outperformed discriminant analysis on samples with large percentages of distressed firms, but failed to do so on those with a more equal mix of distressed and non-distressed firms.

Refenes, Azema-Barac and Zapranis (1993) tested NNs in the domain of stock ranking. Comparisons with multiple regression indicated that the proposed network gave better fitness on the test data over multiple regression by an order of magnitude. The network outperformed regression on the validation sample by an average of 36%.

Three of the eleven effective studies compared the performance of alternative models in the prediction of time series. Of these, one indicated mixed results in this comparison of neural networks with alternative techniques. Ho, Hsu and Young (1992) tested a proposed algorithm, the Adaptive Learning Algorithm (ALA), in the domain of short-term load forecasting. The ALA automatically adapts the momentum of the training process as a function of the error. Performance of the network was compared to that of a rule-based system and to the judgmental forecasts of the operator. Although the network performed slightly better than the rule-based system and the operator, the Mean Absolute Errors (MAEs) were not very different for the three approaches and no tests were performed to determine if the results were significantly better with the NN.

Foster, Collopy and Ungar (1992) compared the performance of linear regression and combining with that of NNs in the prediction of 181 annual and 203 quarterly time series from the M-Competition (Makridakis *et al.*, 1982). They used one network to make direct predictions (network combining). The authors reported that while the direct network performed significantly worse than the comparative methods, network combining significantly outperformed both regression and simple combining. Interestingly, the networks became more conservative as the forecast horizon increased or as the data became more noisy. This reflects the approach that an expert might take with such data.

Connor, Martin and Atlas (1994) compared the performance of various NN configurations in the prediction of time series. They compared performance of recurrent and feedforward nets for power load forecasting. The recurrent net outperformed the traditional feedforward net while successfully modelling the domain with more parsimony than the competing architecture.

Effectively validated with positive results despite implementation issues

Eleven additional studies that were effectively validated reported NN performance that was better than comparative models. Dutta *et al.* (1994) used simulated data, corporate bond rating, and product purchase frequency as test beds for their implementation of a NN. NNs performed better than multiple regression on the simulated data, despite a training advantage for the regressions. In the prediction of bond rating, NNs consistently outperformed regression, while only one configuration outperformed regression in the purchase frequency domain.

Lee and Jhee (1994) used a NN for ARMA model identification with Extended Sample Autocorrelation Function (ESACF). The NN demonstrated superior classification accuracy on simulated data. The NN was then tested on data from three prior studies where the models were identified using traditional approaches. The authors report that the NN correctly identified the model for US GNP, Consumer Price Index, and caffeine data.

Other studies in the domain of prediction included those by Fletcher and Goss (1993), DeSilets *et al.* (1992), and Kimoto *et al.* (1990). Fletcher and Goss (1993) developed NNs for bankruptcy classification and compared their NN with a logit model. The NN outperformed logit models, having a lower prediction error and less variance. DeSilets *et al.* (1992) compared the performance of regression models with NNs in the prediction of salinity in Chesapeake Bay. Results indicated that NNs performed effectively as compared to regression models.

Kimoto *et al*. (1990) predicted the buying and selling time for stocks in the Tokyo Stock Exchange. Their system, consisting of multiple NNs, was compared to multiple regression. Correlation coefficients with the actual stock movements showed a higher coefficient for the NNs than for regression. In the same domain, Yoon *et al*. (1993) compared the performance of NNs with discriminant analysis for prediction of stock price performance. Although the study did not perform cross-validations, results indicated that NNs performed significantly better than discriminant analysis in classifying the performance of stocks.

In the domain of time series forecasting, Chen, Yu and Moghaddamjo (1992) used a NN for electric load forecasting. The NN provided better forecasts than ARIMA models. It also adapted better to changes, indicating robustness. Park *et al*. (1991) also developed a NN for the domain of electric load forecasting and compared its performance with the approach used by the electric plant. Their NN outperformed the traditional approach significantly. Tang, de Almeida and Fishcwick (1991) tested the performance of NNs in the prediction of domestic and foreign car sales and of airline passenger data. They reported that the NN performed better than Box–Jenkins for long-term (12- and 24-month) forecasts, and as well as Box–Jenkins for short-term (1- and 6-month) forecasts.

Further evaluation of backpropagation implementations

Of the 48 studies, 44 (88%) used error backpropagation as their learning algorithm. It is well established in the literature that this approach can suffer from three potential problems. First, there is no single configuration that is adequate for all domains or even within a single domain. The topology must, therefore, be determined through a process of trial and error. Second, such NNs are susceptible to problems with local minima (Grossberg 1988). Finally, they are prone to overfitting. Refenes (1995) suggests five control parameters that can be used to guide the effective design of a NN. We examined the 27 studies that met our effectiveness of validation criteria with respect to their approach to these controls:

- *Network architecture*: Several variables such as the number of hidden layers and nodes, weight interconnections, and bottom-up or top-down design can determine the most effective NN architecture for a problem. We considered whether a study had done sensitivity analyses with the number of layers and nodes in the architecture. Evaluating the other features of network architecture proves difficult given the level of disclosure typical of these studies.
- *Gradient descent*: Manipulation of learning rate during training has been shown to lead to more effective gradient decent into the error surface.
- *Cross-validation*: To prevent overfitting, Refenes (1995) recommends that cross-validation be performed during learning. This facilitates the termination of learning and controls over-fitting.
- *Cross function*: While we identified the cost functions used, we did not attempt to evaluate their relative merits, as the literature on this remains inconclusive.
- *Transformation function*: All the studies that reported them used sigmoid functions.

Of the 27 studies that were effectively validated, 18 (67%) did sensitivity analyses to determine the most appropriate network architecture. In general, most found the use of a single hidden layer effective for the problem being solved. However, there was little consensus regarding the number of nodes that should be included in the hidden layer, suggesting a need for further empirical research on this. Eleven (41%) studies attempted to control the gradient descent by implementing dynamic changes to the learning rate. Once again, further empirical work needs to be done before an appropriate range of learning rate adjustments can be suggested. Interestingly, only three of the 27 studies attempted to control the potential problem of overfitting that can arise during learning by using cross-validations (Refenes, *et al.*, 1993; Fletcher and Goss, 1993; Kimoto *et al.*, 1990). This is a disappointing finding particularly in light of the fact that backpropagation NNs are known to be seriously prone to overfitting. Eighteen (67%) of the 26 studies reported the use of the sigmoid activation function. The remaining nine did not report the particular transformation function. These study features are summarized in Appendix C.

CONCLUSIONS

Of the 48 studies we evaluated, only eleven met all of our criteria for effectiveness of validation and implementation. Of the remaining 38, 17 presented effective validations but suffered with respect to implementation. Eleven of these reported positive results despite implementation problems. Altogether then, of the 48 studies, 22 contributed to our knowledge regarding the applicability of NNs to forecasting and prediction. Nineteen (86%) of these produced results that were favourable, three produced results that were not.

Two conclusions emerge, then, from our evaluation of NN implementations in forecasting and prediction. First, NNs, when they are effectively implemented and validated, show potential for forecasting and prediction. Second, a significant portion of the NN research in forecasting and prediction lacks validity. Over half of the studies suffered from validation and/or implementation problems which rendered their results suspect. We recommend, therefore, that future research efforts in this area attend more explicitly to validity.

Until the value of NNs for forecasting is established, comparisons must be made between NN techniques and alternative methods. The alternatives used for comparison should be simple and well-accepted. The forecasting literature expresses a preference for simpler models unless a strong case has been made for complexity (Collopy, Adya and Armstrong, 1994). Moreover, research findings indicate that relatively simple extrapolation models are robust (Armstrong, 1984). Comparisons should be based on out-of-sample performance. Finally, to be convincing a substantial sample of forecasts must be generated and compared.

Researchers have been hopeful about the potential for NNs in business applications. We evaluated 48 empirical studies that applied NN approaches to business forecasting and prediction problems. About 48% of the studies failed to effectively test the validity of the proposed NNs. Of the remaining 26 studies 54% failed to adequately implement the NN technique, so that its failure to outperform the alternatives does not provide much valuable information about the utility of NNs generally. This means that we must base any conclusions about the utility of NNs for forecasting and prediction on only about 46% of the studies done in the area. These 22 studies contain promising results. In 19 (86%) of them, NNs outperformed alternative approaches. In eight studies where comparisons were made, NNs performed less well than alternatives. But in five of these there were issues related to the quality of the NN implementation. This calls for some reservation in interpreting their results. A further caution remains that the bias against publication of null and negative results may mean that successful applications are over-represented in the published literature.

APPENDIX A: VALIDITY OF STUDIES

Study	Comparison with alternative methods	*Ex ante* validation	Adequate sample
Classification studies			
Chu and Widjaja (1994)		●	∞
Dasgupta *et al*. (1994)	Discriminant analysis Logistic regression	●	●
Dutta *et al*. (1994)	Regression models Configurations	●	●
Gorr *et al*. (1994)	Multiple and Stepwise regression, decision rule	●	●
Lee and Jhee (1994)	Previously identified models	●	●
Piramuthu *et al*. (1994)	ID3 NEWQ Probit Configurations	●	
Wilson and Sharda (1994)	Discriminant analysis	●	●
Yoon *et al*. (1994)	Discriminant analysis	●	●
Coats and Fant (1993)	Discriminant analysis	●	●
Fletcher and Goss (1993)	Logit	●	●
Kryzanowski *et al*. (1993)		●	●
Refenes *et al*. (1993)	Multiple regression	∞	∞
Udo (1993)	Multiple regression	●	●
Yoon *et al*. (1993)	Discriminant analysis Configurations	●	●
Coats and Fant (1992)	Discriminant analysis	●	●
DeSilets *et al*. (1992)	Regression	●	●
Hansen *et al*. (1992)	Five Qualitative response models Logit Probit ID3	●	●
Karunanithi and Whitley (1992)	Five Software reliability models	●	
Salchenberger *et al*. (1992)	Logit	●	●
Swales and Yoon (1992)	Discriminant analysis Configurations	●	
Tam and Kiang (1992)	Discriminant Regression Logistic *k* Nearest Neighbour ID3	●	●
Tanigawa and Kamijo (1992)	Experts	●	●
Hoptroff (1991)	Leading indicators	●	
Lee *et al*. (1991)	Results from prior studies	∞	
Tam (1991)	Discriminant Factor-Logistic *k* Nearest Neighbour ID3	●	
Odom and Sharda (1990)	Discriminant analysis	●	
Surkan and Singleton (1990)	Discriminant analysis Configurations	●	

Appendix A continued over page

APPENDIX A: CONTINUED

Study	Comparison with alternative methods	*Ex ante* validation	Adequate sample
Tam and Kiang (1990)	Discriminant analysis Factor Logistic *k* Nearest Neighbour	•	•
Dutta and Shekhar (1988)	Regression Configurations	•	
Time series forecasting			
Coporaletti *et al*. (1994)	Traditional estimation approaches	∞	∞
Connor *et al*. (1994)	Configurations	•	•
Grudnitski and Osburn (1993)		•	•
Hsu *et al*. (1993)	Various NN learning algorithms	•	•
Peng *et al*. (1993)	Box–Jenkins	•	∞
Baba and Kozaki (1992)		•	
Bacha and Meyer (1992)	Configurations	•	
Caire *et al*. (1992)	ARIMA	•	•
Chakraborty *et al*. (1992)	Moving Average approach of Tiao and Tsay (1989)	•	
Chen *et al*. (1992)	ARIMA	•	•
Foster *et al*. (1992)	Linear regression, Combining A	•	•
Ho *et al*. (1992)	Configurations	•	•
Tang *et al*. (1991)	Box–Jenkins	•	•
Srinivasan *et al*. (1991)	Exponential smoothing Winter's linear method Two-parameter MA model Multiple regression Simple Reg. and Box–Jenkins	•	
Kimoto *et al*. (1990)	Multiple regression	•	•
Park *et al*. (1991)	Approach used by plant	•	•
Sharda and Patil (1990)	Box–Jenkins	•	•
Wolpert & Miall (1990)		•	•
White (1988)		•	•

• Criterion was satisfied
∞ Criterion not reported/unclear

APPENDIX B: IMPLEMENTATION DETAILS OF VALIDATED STUDIES

Study	Learning Algorithm	Convergence	Generalization	Stability	Results
Classification studies					
Wilson and Sharda (1994)	Backpropagation	•	•	•	+
Refenes *et al*. (1993)	Backpropagation	•	•	•	+
Tam and Kiang (1992)	Backpropagation	•	•	•	+
Tam and Kiang (1990)	Backpropagation	•	•	•	+
Coats and Fant (1993)	Cass-Corr	•	•	•	+
Salchenberger *et al*. (1992)	Backpropagation	•	•	•	+
Gorr *et al*. (1994)	Backpropagation	•	•	•	=
Udo (1993)	Backpropagation	•	•	•	=
Dutta *et al*. (1994)	Backpropagation	•	•		+
Coats and Fant (1992)	Backpropagation	•	•		+
Tam (1991)	Backpropagation	•	≠	•	+
Fletcher and Goss (1993)	Backpropagation	∞	≠	•	+
DeSilets *et al*. (1992)	Backpropagation	∞	≠	•	+
Lee and Jhee (1994)	Backpropagation	∞	≠	•	+
Dasgupta *et al*. (1994)	Backpropagation	∞	≠	•	=
Hansen *et al*. (1992)	Backpropagation	∞	≠	•	−
Tanigawa and Kamijo (1992)	Backpropagation	∞	≠		=
Time series forecasting					
Connor *et al*. (1994)	Backpropagation	•	•	•	+
Foster *et al*. (1992)	Backpropagation	•	•	•	+
Ho *et al*. (1992)	Backpropagation	•	•	•	=
Chen *et al*. (1992)	Backpropagation	∞	≠	•	+
Park *et al*. (1991)	Backpropagation	∞	≠	•	+
Kimoto *et al*. (1990)	Backpropagation	∞	≠	•	+
Tang *et al*. (1991)	Backpropagation	∞	≠	•	+
Caire *et al*. (1992)	Backpropagation	•	•		=
Sharda and Patil (1990)	Backpropagation	∞	≠	•	=

• Criterion was satisfied
∞ Criteria not reported/unclear
≠ Interpreted with caution
\+ Positive NN result
= NN same as benchmark
− Negative NN result
Blank cells: criteria not met

APPENDIX C: IMPLEMENTATION DETAILS OF BACKPROPAGATION STUDIES

Study	Network architecture	Gradient Descent	Cross-validation	Cost function	Squashing function
Classification studies					
Refenes *et al*. (1993)	•	•	•	RMSE % Change	Sigmoid
Wilson and Sharda (1994)				∞	∞
Tam and Kiang (1992)	•			∞	Sigmoid
Tam and Kiang (1990)	•			∞	Sigmoid
Salchenberger *et al*. (1992)	•	•		MSE	Sigmoid
Gorr *et al*. (1994)	•	•		MSE	Sigmoid
Udo (1993)	•			∞	∞
Dutta *et al*. (1994)	•			Total Sum of Sq	Sigmoid
Coats and Fant (1992)				∞	∞
Yoon *et al*. (1994)	•	•		MSE	Sigmoid
Tam (1991)	•			∞	Sigmoid
Fletcher and Goss (1993)	•		•	LSE	
DeSilets *et al*. (1992)	•	•		∞	Sigmoid
Lee and Jhee (1994)	•			MSE	Sigmoid
Dasgupta *et al*. (1994)	•	•		∞	Sigmoid
Hansen *et al*. (1992)				Total Sum of Sq	∞
Tanigawa and Kamijo (1992)		•		∞	∞
Time-series forecasting					
Connor *et al*. (1994)				MSE	Sigmoid
Foster *et al*. (1992)	•	•	MSE	Sigmoid	
Ho *et al*. (1992)	•	•		RMSE	Sigmoid
Caire *et al*. (1992)				∞	Sigmoid
Chen *et al*. (1992)	•				Sigmoid
Park *et al*. (1991)	•			∞	∞
Kimoto *et al*. (1990)		•	•	∞	Sigmoid
Tang *et al*., (1991)	•	•		MSE	Sigmoid
Sharda and Patil (1990)				MSE	∞

• Parameter was tested
∞ Parameter not reported/unclear
Blank cells: Parameter not tested

ACKNOWLEDGEMENTS

Many people have commented on previous versions of this paper. We especially wish to thank Scott Armstrong, Miles Kennedy, Raghav Madhavan, Janusz Sczypula, and Betty Vandenbosch.

REFERENCES

Armstrong, J. S., 'Forecasting by extrapolation: conclusions from 25 years of research', *Interfaces*, **14** (1984), 52–66.

Baba, N., and Kozaki, M. 'An intelligent forecasting system of stock price using neural networks', *IEEE/INNS International Joint Conference on Neural Networks*, I 1992, 371–377.

Bacha, H. and Meyer, W. 'A neural network architecture for load forecasting', *IEEE/INNS International Joint Conference on Neural Networks*, Vol. II, 1992, 442–447.

Caire, P., Hatabian, G. and Muller, C., 'Progress in forecasting by neural networks', *IEEE/INNS International Joint Conference on Neural Networks*, Vol. II, 1992, 540–545.

Caporaletti, L. E., Dorsey, R. E., Johnson, J. D. and Powell, W. A., 'A decision support system for in-sample simultaneous equation systems forecasting using artificial neural systems', *Decision Support Systems*, **11** (1994), 481–495.

Chakraborty, K., Mehrotra, K., Mohan, C. K. and Ranka, S., 'Forecasting the behavior of multivariate time series using neural networks', *Neural Networks*, **5** (1992), 961–970.

Chatfield, C., 'Editorial: Neural networks: forecasting breakthrough or passing fad?'. *International Journal of Forecasting*, **9** (1993), 1–3.

Chen, S. T., Yu, D. C. and Moghaddamjo, A. R., 'Weather sensitive short-term load forecasting using nonfully connected artificial neural network', *IEEE Transactions on Power Systems*, **7** (1992), 3, 1098–1105.

Chu, C. H. and Widjaja, D., 'Neural network system for forecasting method selection', *Decision Support Systems*, **12** (1994), 13–24.

Coats, P. K. and Fant, L. F., 'A neural network approach to forecasting financial distress', *The Journal of Business Forecasting*, (1992), Winter, 9–12.

Coats, P. K. and Fant, L. F., 'Recognizing financial distress patterns using a neural network tool', *Financial Management*, **22** (1993), 3, 142–155.

Collopy, F., Adya, M. and Armstrong, J. S., 'Principles for examining predictive validity: the case of information systems spending forecasts', *Information Systems Research*, **5** (1994), 2, 170–179.

Connor, J., Martin, R. D. and Atlas, L. E., 'Recurrent neural networks and robust time series prediction', *IEEE Transactions on Neural Networks*, **5** (1994), 2, 240–254.

Dasgupta, C. G., Dispensa, G. S. and Ghose, S., 'Comparing the predictive performance of a neural network model with some traditional response models', *International Journal of Forecasting*, **10** (1994), 235–244.

DeSilets, L., Golden, B., Wang, Q. and Kumar, R., 'Predicting salinity in Chesapeake Bay using back-propagation', *Computers & Operations Research*, **19** (1992), 3/4, 277–285.

Dutta, S. and Shekhar, S. 'Bond rating: a non-conservative application of neural networks', *IEEE International Conference on Neural Networks*, II, 1988, 443–450.

Dutta, S., Shekhar, S.. and Wong, W. Y., 'Decision support in non-conservative domains: generalization with neural networks', *Decision Support Systems*, **11** (1994), 527–544.

Fletcher, D. and Goss, E., 'Forecasting with neural networks: an application using bankruptcy data', *Information & Management*, **24** (1993), 159–167.

Foster, W. R., Collopy, F. and Ungar, L. H., 'Neural network forecasting of short, noisy time series', *Computers in Chemical Engineering*, **16** (1992), 4, 293–297.

Gorr, W. L., Nagin, D. and Szczypula, J., 'Comparative study of artificial neural network and statistical models for predicting student grade point averages', *International Journal of Forecasting*, **10** (1994), 1, 17–33.

Grossberg, S., *Neural Networks and Natural Intelligence*, Cambridge, MA: The Mitt Press, 1988.

Grudnitski, G. and Osborn, L., 'Forecasting S&P and gold futures prices: an application of neural networks', *The Journal of Futures Markets*, **3** (1993), 6, 631–643.
Hansen, J. V., McDonald, J. B. and Stice, J. D., 'Artificial intelligence and generalized qualitative response models: an empirical test on two audit decision-making domains', *Decision Sciences*, **23** (1992), 708–723.
Hill, T., Marquez, L., O'Connor, M. and Remus, W., 'Artificial neural network models for forecasting and decision making', *International Journal of Forecasting*, **10** (1994), 1, 5–17.
Ho, K. L., Hsu, Y. Y. and Yang, C. C., 'Short term load forecasting using a multilayer neural network with an adaptive learning algorithm', *IEEE Transactions on Power Systems*, **7** (1992), 1, 141–149.
Hoptroff, R. G., Bramson, M. J. and Hall, T. J., 'Forecasting economic turning points with neural nets', *IEEE*, I (1991), 347–352.
Hsu, W., Hsu, L. S. and Tenorio, M. F., 'A ClusNet architecture for prediction', *IEEE International Conference on Neural Networks*, 1993, 329–334.
Karunanithi, N. and Whitley, D., 'Prediction of software reliability using feedforward and recurrent neural nets', *IEEE/INNS International Joint Conference on Neural Networks*, 1992, I-800–I-805.
Kimoto, T., Asakawa, K., Yoda, M. and Takeoka, M., 'Stock market prediction system with modular neural networks', *IEEE/INNS International Joint Conference on Neural Networks*, 1990, I, 1–6.
Kryzanowski, L., Galler, M. and Wright, D. W., 'Using artificial neural networks to pick stocks', *Financial Analysts Journal*, (1993), 21–27.
Lee, J. K. and Jhee, W. C., 'A two-stage neural network approach for ARMA model identification with ESACF', *Decision Support Systems*, **11** (1994), 461–479.
Lee, K. C., Yang, J. S. and Park, S. J., 'Neural network-based time series modelling: ARMA model identification via ESACF approach', *International Joint Conference on Neural Networks*, I, (1991), 43–49.
Lippmann, R. P., 'An introduction to computing with neural nets', *IEEE ASSP Magazine*, (1987), 4–21.
Makridakis, S., Anderson, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. 'The accuracy of extrapolation (time series) methods: results of a forecasting competition', *Journal of Forecasting*, **1** (1982), 111–153.
Odom, M. D. and Sharda, R., 'A neural network model for bankruptcy prediction', *IEEE/INNS International Joint Conference on Neural Networks*, II, 1990, 163–168.
Park, D. C., El-Sharkawi, M. A., Marks II, R. J., Atlas, L. E. and Damborg, M. J., 'Electric load forecasting using an artificial neural network', *IEEE Transactions on Power Systems*, **6** (1991), 2, 442–449.
Peng, T. M., Hubele, N. F. and Karady, G. G., 'An adaptive neural network approach to one-week ahead load forecasting', *IEEE Transactions on Power Systems*, **8** (1993), 3, 1195–1203.
Piramuthu, S., Shaw, M. J. and Gentry, J. A., 'A classification approach using multi-layered neural networks', *Decision Support Systems*, **11** (1994), 509–525.
Ralston, A. and Reilly, E. D. *Encyclopedia of Computer Science*, New York: Van Nostrand Reinhold, 1993.
Refenes, A. N., 'Neural network design considerations', in Refenes, A. N. (ed.), *Neural networks in the Capital Market*, New York: John Wiley, 1995.
Refenes, A. N., Azema-Barac, M. and Zapranis, A. D., 'Stock ranking: neural networks vs multiple linear regression', *IEEE International Conference on Neural Networks*, 1993, 1419–1426.
Salchenberger, L. M., Cinar, E. M. and Lash, N. A., 'Neural networks: a new tool for predicting thrift failures', *Decision Sciences*, **23** (1992), 899–916.
Schocken, S. and Ariav, G., 'Neural networks for decision support: problems and opportunities', *Decision Support Systems,* **11** (1994), 393–414.
Sharda, R. and Patil, R. B., 'Neural networks as forecasting experts: an empirical test', *Proceedings of the International Joint Conference on Neural Networks*, II, Washington, DC, 1990, 491–494.
Srinivasan, D., Liew, A. C. and Chen, J. S. P., 'A novel approach to electrical load forecasting based on neural network', *International Joint Conference on Neural Networks*, 1991.
Swales, G. S., and Yoon, Y., 'Applying artificial neural networks to investment analysis', *Financial Analysts Journal*, (Sept.–Oct. 1992) 78–80.
Surkan, A. J. and Singleton, J. C., 'Neural networks for bond rating improved by multiple hidden layers', *IEEE/INNS International Joint Conference on Neural Networks*, Vol. 2, 1990, 157–162.
Tam, K. Y., 'Neural network models and prediction of bank bankruptcy', *OMEGA: International Journal of Management Science*, **19** (1991), 5, 429–445.

Tam, K. Y., 'Neural networks for decision support', *Decision Support Systems*, **11** (1994), 389–392.
Tam, K. Y. and Kiang, M. Y., 'Predicting bank failures: a neural network approach', *Applied Artificial Intelligence*, **4** (1990), 265–282.
Tam, K. Y. and Kiang, M. Y., 'Managerial applications of neural networks: the case of bank failure predictions', *Management Science*, **38** (1992), 7, 926–947.
Tang, Z., de Almeida, C. and Fishcwick, P. A., 'Time series forecasting using neural networks vs. Box–Jenkins methodology', *Simulation*, **57** (1991), 5, 303–310.
Tanigawa, T. and Kamijo, K., 'Stock price pattern matching system', *IEEE/INNS International Joint Conference on Neural Networks*, 1992, II-465–II-471.
Udo, G., 'Neural network performance on the bankruptcy classification problem', *Proceedings of the 15th Annual Conference on Computers and Industrial Engineering*, **25** (1993), 1–4, 377–380.
White, H., 'Economic prediction using neural networks: a case of IBM daily stock returns', *IEEE International Conference on Neural Networks*, II, 1988, 451–458.
Wilson, R. L. and Sharda, R., 'Bankruptcy prediction using neural networks', *Decision Support Systems*, **11** (1994), 545–557.
Wolpert, D. M. and Miall, R. C., 'Detecting chaos with neural networks', *Proceedings of the Royal Society of London*, **242** (1990), 82–86.
Yoon, Y., Swales, G. and Margavio, T. M., 'A comparison of discriminant analysis vs artificial neural networks', *Journal of Operational Research Society*, **44** (1993), 1, 51–60.
Yoon, Y., Guimaraes, T. and Swales, G., 'Integrating artificial neural networks with rule-based expert systems', *Decision Support Systems*, **11** (1994), 497–507.

Authors' biographies:
Monica Adya is an Assistant Professor in Information Systems at the University of Maryland Baltimore County. Her research interests are in Al applications to forecasting, knowledge elicitation and representation, and judgement and decision making.

Fred Collopy is an associate professor of management information systems in the Weatherhead School of Management at Case Western Reserve University. He received his PhD in decision sciences from the Wharton School of the University of Pennsylvania. His research has been published in leading academic journals including *Management Science*, *Information Systems Research*, *Journal of Market Research*, *Journal of Forecasting*, *International Journal of Forecasting*, as well as in practice-oriented publications such as *Interfaces*, and *Chief Executive*.

Authors' addresses:
Monica Adya, Department of Information Systems, University of Maryland at Baltimore County, Baltimore, MD 21250, USA.

Fred Collopy, Management Information & Decision Systems, The Weatherhead School of Management, Case Western Reserve University, Cleveland, OH 44106, USA.

