GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

# A NEURAL NETWORK MODEL FOR FLIGHTS DELAY: CLASSIFICATION AND PREDICTION

**Ramadan A. ZeinEldin**
Associate Prof. Institute of Statistical Studies and Research (ISSR), Cairo University, Egypt
rzeanedean@yahoo.com

**ABSTRACT**

In the area of flights delay, most of the research done concentrate on developing flight schedules without studying the real reasons for flights delay. The flights delay causes great loss in money and in travelers for the airline companies. In this study, a feed forward multi-layer perceptron (MLP) with two hidden layers type of artificial neural network (ANN) is built to determine the most air flight delay causes, predict the flights delay time and prioritize the factors that affect the flights delay. Most of the delay causes found in this study need good management to minimize the delay time.

**Keywords:** Flights delay, ANN, MLP, Classification, Prediction.

## I. INTRODUCTION

Image processing is a vast area today.

Artificial Neural Networks (ANNs) have been used since the 1980s for complex modeling and various recognition, prediction and filtering tasks. Their ability to learn and evolve has made them attractive to many different fields of research and innovation. ANN is considered as an interesting methodological perspective, especially where the interpretation of complex multivariate systems involves both quantitative and qualitative variables and there is no a priori deterministic interpretation model [1]. Artificial neural network is considered as one of the modern mathematical-computational methods which are used to solve un-anticipated dynamic problems in developed behavioral systems during a time period [2]. ANN can learn to recognize patterns from data and solve the real complex problems. Flights delays cause a loss for airlines companies in profit and number of customers. There are many reasons for flight delay some of them are caused by passengers, others are caused by schedule or the airport such as immigrations, customs, health and others. Reducing flight delay in the air transportation system has become more urgent in recent years as air travel demand has escalated [3] and because On-time performance of airlines schedule is key factor in maintaining satisfaction of current customers and attracting new ones. Also, maintaining economical operations during irregular conditions is essential to achieve expected revenues [4].

A model based on ANN is developed to the flights delayed to know which delay causes have more effect on delaying the air flights and to predict the expected time of delay. Also, the proposed ANN model prioritize the factors that affect the flights delay. The rest of the paper is organized as follows. An overview about ANN is presented in section two, the flight delay problem and a review are presented in section three. The developed ANN model is described in section four. Finally, section five is assigned to the conclusions and future research.

## II. Artificial Neural Networks (ANN)

Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the connections between elements largely determine the network function. Neural network can be trained to perform a particular function by adjusting the values of the connections (weights) between n elements so that a particular input leads to a specific target output [5]. The ANN consists of nodes that are connected to each other and exist in several different layers, resulting in it being often referred to as a Multi Layered Perceptron (MLP) network. These layers are the input layer, the hidden layer, and the output layer. Each of these layers has a design specific amount of individual nodes in them. An individual node works much like its biological counterpart the neuron. It receives input from a multitude of different weighted input connections, sums these inputs and then produces an output that serves as input for other nodes. The output is generally normalized to be between -1 and 1 and typically a sigmoid function of the type can be used [6, 7]. Neural networks can be used to learn to predict future events based on the patterns that have been observed in the historical training data; learn to classify unseen data into pre-defined groups based on characteristics observed in the training data; learn to cluster the training data into natural groups based on the similarity of characteristics in the training data. The areas and applications of using ANN are the economic field, political field, tourism field, environmental area, and informational technology field, marketing, finance, manufacturing, strategic management, airline security control, investment management, risk control and prediction of stock price index [2, 7, 8].

Some researchers used ANN in different real applications such as Khaze et al. [2] used ANN to predict participation in presidential elections in Iran. Wang et al. [9] used particle swarm optimization algorithm to adjust the parameters (weights, initial inputs of the context nodes and self-feedback coefficient). Das et al. [10] used particle swarm optimization to train ANN for solving the problem of channel equalization. Gallo et al. [1] applied an unsupervised ANN for the classification of a series of olive farms according to the communication strategies and market. Ciupan [11] used ANN to determine the optimum a supply batch in inventory management. Abas [12] proposed adaptive competitive learning neural network algorithm to integrate the unsupervised learning and the optimization of the competitive neural networks. Kondo et al. [13] proposed a hybrid multi-layered neural network group method to the computer aided image diagnosis of liver cancer. Yang et al. [14] applied three types of neural networks to model or approximate the landscape of known functions. Zhang [15] reviewed several important issues and recent developments of neural networks for classification problems. Larasati et al. [16] built a model to predict overall customer satisfaction in a student-operated restaurant.

## III. Flight Delay review

Flights delays are common events in the airline industry; about 25% of all flights in the US reported delays in the first two months of 2011[17]. Flight can be delayed by many reasons some of these reasons are related to the management of airports, some are related to the passengers and their awareness and some unavoidable factors such as weather. There are a number of uncertain factors that could result in flight delays into the gate, and in two adjacent flights delayed flights ahead will be spread to different levels of rear flight plans to arrive time [18]. Some researchers mentioned that flights are delayed because of the assignment and the pilots. On average, about 42% of the delayed flights is delayed because their assigned aircrafts are not timely ready. Furthermore, pilots and FAs (Federal Aviation Administration (FAA)) contribute to 18% and 12% as a reason of the projected delayed flights, respectively. Flight schedules are often subjected to numerous sources of irregularity. In particular, weather accounts for nearly 75% of system delays. The total flight-minute delays generally increases with the increase in number of flights in the issued ground delay program [4]. Different causes of flight delays may have different effects on airline schedule reliability and the effects of flight delays resulting from the same delay cause may not be the same in all cases [19]. So the target of this study is to focus on the real delay causes that are not mentioned at the previous study. Flight delays have a negative impact on an airline's market share and profitability. Flight delay causes loss in money and in number of customers. Of the delay costs analyzed by Austrian Airlines, only 22% can be attributed directly to the effect of delays; 24% stem from permanent loss of passenger loyalty and 54% stem from induced knock-on delays in aircraft rotation schedules [19]. United Airlines estimates that it saves approximately $1.6 million by using a flight delay projection model during the first quarter of 2004 [4].

Most of researchers handle the flight delay problem as a scheduling problem without considering the real reasons of these delays. So the researchers developed optimization approaches and heuristics and metaheuristic approaches for flight scheduling problem without concentrating on the real delay reasons. Yan and Tang [20] developed a heuristic approach to make airport gate assignments that are sensitive to stochastic flight delays. Abdelghany et al. [4] presented a flight delay projection model. Cao and Fang [3] proposed genetic-simulated annealing approach to flight departure delays. Maharjan and Matis [21] presented a binary integer program formulation for the flight gate reassignment problem to minimize the walking distance of passengers between disrupted flights. Muller and Santana [22] used simulation to analyze the flight delay and their costs in the Sao Paulo terminal area in Brazil also, Zheng et al. [18] used simulation to analyze quantitatively and systematically the chaos of an aftereffect delay spread time series. Lubbe and Victor [23] used a simple quantitative methodology to measures the cost of flight delays to corporations. Zou and Hansen [24] proposed an assessment framework considering the interplay among passenger demand, air fare, flight frequency, aircraft size, and flight delay. Gao et al. [25] built two nonlinear mixed integer programming models in aviation production engineering based on flight priority levels. Jungai and Hongjun [26] presented a simulated annealing algorithm for optimizing arrival flight delays to minimize delay cost. Livet et al. [27] presented a quadratic optimization technique to design a robust flight control system. Ferrer et al. [17] analyzed the effects of flight delays on passengers' future purchasing behavior. Ding and Li [28] proposed a forecasting model to solve the problem of flight delay. Britto et al. [29] studied the impact of flight delays on passenger demand, airfares and consumer and producer welfare. Baumgarten et al. [30] estimated the impact of hubbing behavior on delays. Fan et al. [31] used directional distance function to evaluate the technical efficiency of twenty major Chinese airports from 2006 to 2009 of desirable and undesirable output such as flight delays.

## IV. The Proposed ANN Model for Flight Delay

The multilayer perceptron (MLP) type of ANN is used. The MLP is a feed-forward, supervised learning network with two hidden layers. Each layer consists of units. The input network is in the size of features which have educational samples and enters to the input layer, simultaneously. These inputs are transported through input layer and then become weighted. Then, they are transported to the second layer of pseudo-neuron units which known as hidden layer. The output of hidden layer can be entered to another hidden layer as an input of output layer. The weighted output or hidden layers are used as input of output layer in which they can reach to the ideal weight and network learning by educational data series and provide classification and anticipation operation for tuples and test samples [32, 2]. Larasati et al. [16] showed that the best neural network model to use is the MLP neural network model with two hidden layers. The MLP network is a function of one or more predictors or inputs or independent variables that minimizes the prediction error of one or more target variables (also called outputs). Predictors and targets can be a mix of categorical and scale variables.

To build the proposed model, we got data of delayed flights for all the year of 2012 and selected three months (January, June and November). There are about sixty flights delay reasons; each of these reasons take a code and description. Table 1 shows the most repeated delays, No. of repeat, total delayed time (min) for each delay and delay description during the selected three months.

**Table 1: The most repeated delays**

| Delay code | No. of repeat | Total Delayed Time (min) | delay code description |
|---|---|---|---|
| 12 | 306 | 4121 | Late check-in, congestion in check-in area |
| 16 | 166 | 2549 | Commercial/publicity, passenger convenience, VIP, press, |

| | | | |
|---|---|---|---|
| | | | ground meals and missing personal items |
| 32 | 1221 | 24590 | Loading/unloading, bulky, special load, cabin load, lack of loading staff |
| 36 | 174 | 2521 | Fuelling/defueling, fuel supplier |
| 43 | 560 | 24054 | Non-scheduled maintenance, special checks and/or additional works beyond normal maintenance schedule |
| 47 | 545 | 15203 | Standby aircraft, lack of planned standby aircraft for technical reasons |
| 67 | 137 | 1721 | Cabin crew shortage, sickness, awaiting standby, flight time limitations, crew meals, valid visa, health documents etc. |
| 81 | 181 | 4670 | Air traffic flow management restrictions - ATFM due to ATC en-route demand/capacity, standard demand |
| 85 | 570 | 9732 | Mandatory security |
| 86 | 164 | 2538 | Immigrations, customs, health |
| 87 | 215 | 5296 | A/p facilities parking, stands, ramp congestion, lighting, buildings, gates, etc |
| 89 | 913 | 9220 | Restrictions at a/p of departure with or without ATFM restrictions including air traffic, start-up |
| 93 | 3317 | 109780 | Aircraft rotation, late arrival of aircraft from another flight or previous sector |
| 99 | 828 | 25591 | Undefined - explain in SI section |

Table 2 summarizes the case processing of flight delay where N is the number of delayed flights for the selected three months. A random sample is taken from the entire data set and split into training (70%) and testing samples (30%).

**Table 2: The Case Processing Summary**

| | | N | Percent |
|---|---|---|---|
| Sample | Training | 7098 | 70.1% |
| | Testing | 3025 | 29.9% |
| Valid | | 10123 | 100.0% |

| | | |
|---|---|---|
| Excluded | 72 | |
| Total | 10195 | |

Table 3 shows the network information which includes the input layer, number of units, the output layers, activation function and error function. The input layer (the factors) includes the flight number, the place that the flight took off from, the aircraft type, delay code and number of delays. The output layer includes the delay time and total delay time for each flight. For the activation function we use Hyperbolic Tangent and for the error function we use Sum-of-Squares.

**Table 3: Network Information**

| | | | |
|---|---|---|---|
| Input Layer | Factors | 1 | FlightNo |
| | | 2 | From |
| | | 3 | ACT_Type |
| | | 4 | DelayCode |
| | | 5 | NoOfDelays |
| | Number of Unitsa | | 785 |
| Hidden Layer(s) | Number of Hidden Layers | | 2 |
| | Number of Units in Hidden Layer 1a | | 9 |
| | Number of Units in Hidden Layer 2a | | 7 |
| | Activation Function | | Hyperbolic tangent |
| Output Layer | Dependent Variables | 1 | DelayTime |
| | | 2 | TotDelayTime |
| | Number of Units | | 2 |
| | Rescaling Method for Scale Dependents | | Adjusted Normalized |
| | Activation Function | | Hyperbolic tangent |

| | |
|---|---|
| Error Function | Sum of Squares |

a. Excluding the bias unit

**Table 4 shows the model summary which includes the training sum of square error, stopping rule training time and testing sum of square error.**

**Table 4: Model Summary**

| | | | |
|---|---|---|---|
| Training | Sum of Squares Error | | 78.432 |
| | Average Overall Relative Error | | .996 |
| | Relative Error for Scale Dependents | DelayTime | 1.016 |
| | | TotDelayTime | .980 |
| | Stopping Rule Used | | 1 consecutive step(s) with no decrease in errora |
| | Training Time | | 00:00:23.572 |
| Testing | Sum of Squares Error | | 21.145 |
| | Average Overall Relative Error | | .992 |
| | Relative Error for Scale Dependents | DelayTime | 1.029 |
| | | TotDelayTime | .969 |

a. Error computations are based on the testing sample.

**Figures 1 and 2 shows the predicted delay time and total delay time.**

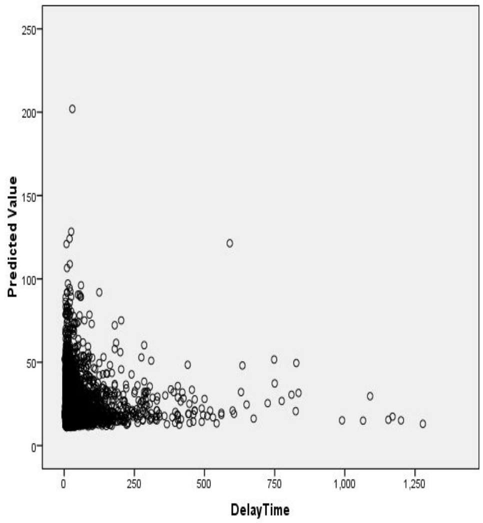


**Figure 1: Predicted delay time**

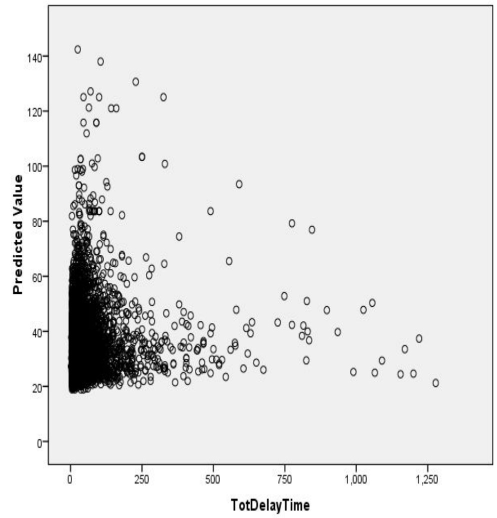


**Figure 2: Predicted total delay time**

**Figures 3 and 4 shows the predicted residual for delay time and total delay.**

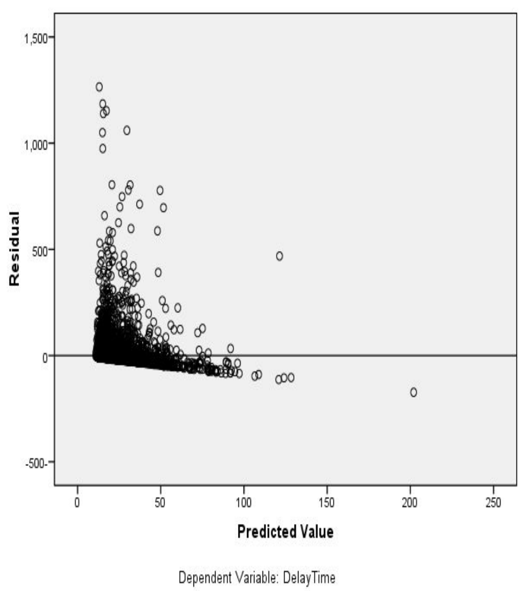


Figure 3: Predicted residual for delay time

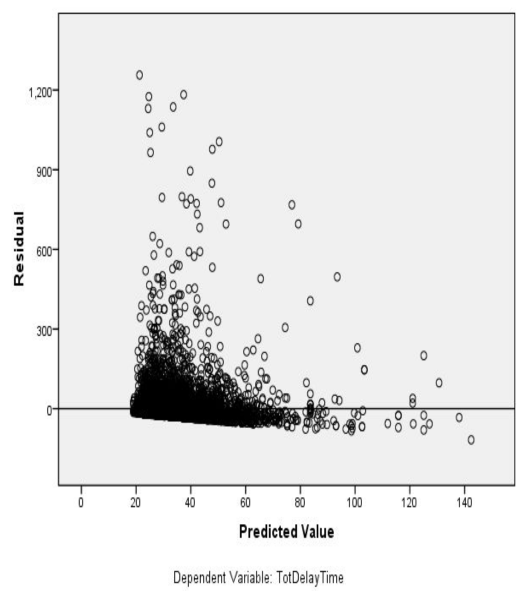


Figure 4: Predicted residual for total delay time

Table 5 shows the Independent Variable Importance where the flight number has about 26% of effect on the flight delay, the airport that the flight took off from and the delay reason have about 24% of effect on the flight delay. The number of delays has about 17% of effect on the flight delay. Finally, the aircraft type has about 10% of effect on the flight delay.

**Table 5: The factors importance**

| | Importance |
|---|---|
| FlightNo | .255 |
| From | .236 |
| ACT_Type | .102 |
| DelayCode | .235 |
| NoOfDelays | .172 |

## V. Conclusion

In this study, an ANN model for flights delays is developed. We considered the factors that are measured which are already registered in the flights system such as aircraft type, departure airport, flight number, delay reason and number of delays. The results showed that the considered factors have an important weight on the flights delays. We think that successful management can reduce the delays flights times. There are other factors that are not registered in the system and they are very important to a future study. From these factors; the travelers and clerks behavior, crew readiness, customs, immigration, security and others. The problem is how to measure these factors, so the fuzzy approach is recommended to deal with these factors.

## *VI.* REFERENCES

*[1] Gallo, C., Contò, F., La Sala, P., Antonazzo, A. P., "A Neural Network Model for Classifying Olive Farms", Procedia Technology, Vol. 8, pp. 593 – 599, 2013.*

*[2] Khaze, S. R., Masdari, M. and Hojjatkhah, S., "Application Of Artificial Neural Networks In Estimating Participation In Elections", International Journal of Information Technology, Modeling and Computing (IJITMC), Vol.1(3), pp. 23 – 31, August 2013.*

*3- Cao, W., Fang, X., "Airport Flight Departure Delay Model on Improved BN Structure Learning", Procedia, Vol. 33, pp. 597 – 603, 2012.*

*4- Abdelghany, K. F., Shah, S. S., Raina, S., Abdelghany, A. F., "A model for projecting flight delays during irregular operation conditions", Journal of Air Transport Management, Vol.10, pp. 385–394, 2004.*

*[5] Diab, A. M., Elyamany, H. E., Abd Elmoaty, M. A., Shalan, A. H., "Prediction of concrete compressive strength due to long term sulfate attack using neural network", Alexandria Engineering Journal, In Press, 2014.*

*6- Lee, H. M, Huang, T.C., and Chen, C.M., "Learning Efficiency Improvement of Back Propagation Algorithm by*

*Error Saturation Prevention Method", International Joint Conference on Neural Networks 1999 (IJCNN '99), Volume 3, pp. 1737 – 1742, July 1999.*

*7- Hakimpoor, H., Arshad, K. A., Tat, H. H., Khani, N. and Rahmandoust, M., "Artificial Neural Networks'Applications in Management", World Applied Sciences Journal, Vol.14 (7), pp. 1008-1019, 2011.*

*8- Li, E. Y., "Artificial neural networks and their business applications", Information & Management, Vol. 27, pp. 303-313, 1994.*

*9- Wang, X., Ma, L., Wang, B., Wang, T., "A hybrid optimization-based recurrent neural network for real-time data prediction", Neurocomputing, Vol. 120, pp. 547–559, 2013.*

*10- Das, G., Pattnaik, P. K., Padhy, S. K., "Artificial Neural Network trained by Particle Swarm Optimization for non-linear channel equalization", Expert Systems with Applications, Vol. 41, pp. 3491–3496, 2014.*

*11- Ciupan, E., "A Study Regarding the Possibility of Optimizing the Supply Batch using Artificial Neural Networks", Procedia Engineering, Vol. 69, pp. 141 – 149, 2014.*

*12- Abas, A. R., "Adaptive competitive learning neural networks", Egyptian Informatics Journal, Vol. 14, pp. 183–194, 2013.*

*13- Kondo, T., Ueno, J., Takao, S., "Hybrid Multi-layered GMDH-type Neural Network Using Principal Component Regression Analysis and Its Application to Medical Image Diagnosis of Liver Cancer", Procedia Computer Science, Vol. 22, pp. 172 – 181, 2013.*

*14- Yang, S., Ting, T.O., Man, K.L. and Guan, S., "Investigation of Neural Networks for Function Approximation", Procedia Computer Science, Vol. 17, pp. 586 – 594, 2013.*

*15- Zhang, G. P. (2000) "Neural Networks for Classification: A Survey", IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, VOL. 30, NO. 4, pp. 451 – 462, Nov. 2000.*

*16- Larasati, A., DeYong, C., Slevitch, L., "The Application of Neural Network and Logistics Regression Models on Predicting Customer Satisfaction in a Student-Operated Restaurant", Procedia - Social and Behavioral Sciences, Vol. 65, pp. 94 – 99, 2012.*

*17- Ferrer, J., Oliveira, P. R., Parasuraman, A., "The behavioral consequences of repeated flight delays", Journal of Air Transport Management, Vol. 20, pp. 35-38, 2012.*

*18- Zheng, P., Hu, S., Zhang, C., "Chaotic Phenomena and Quantitative Analysis of Aftereffect Delay Spread Time on Flights", Journal of Transportation Systems Engineering and Information Technology, Vol 10(4), pp. 68-72, August 2010.*

*19- Wong, J., Tsai, S., "A survival model for flight delay propagation", Journal of Air Transport Management, Vol. 23, pp. 5-11, 2012.*

*20- Yan, S., Tang, C., "A heuristic approach for airport gate assignments for stochastic flight delays", European Journal of Operational Research, Vol. 180, pp. 547–567, 2007.*

*21- Maharjan, B., Matis, T. I., "An optimization model for gate reassignment in response to flight delays", Journal of Air Transport Management, Vol. 17, pp. 256-261, 2011.*

*22- Muller, C., and Santana, E.S.M., "Analysis of flight-operating costs and delays: The Sao Paulo terminal maneuvering area", Journal of Air Transport Management, Vol. 14, pp. 293–296, 2008.*

*23- Lubbe, B., Victor, C., "Flight delays: Towards measuring the cost to corporations", Journal of Air Transport Management, Vol. 19, pp. 9-12, 2012.*

24- *Zou, B., Hansen, M., "Flight delays, capacity investment and social welfare under air transport supply-demand equilibrium", Transportation Research Part A, Vol. 46, pp. 965–980, 2012.*

25- *Gao, M., Chi, H., Hu, Y., Xu, B. (2012) "Models Responding to Large-area Flight Delays in Aviation Production Engineering", Systems Engineering Procedia, Vol. 5, pp. 68 – 73, 2012.*

26- *Jungai, T., Hongjun, X., "Optimizing Arrival Flight Delay Scheduling Based on Simulated Annealing Algorithm", Physics Procedia, Vol. 33, pp. 348 – 353, 2012.*

27- *Livet, T., Kubica, F. and Magni, J.F., "Robust Flight Control Design With Respect To Delays, Control Efficiencies And Flexible Modes", ControlEng. Practice, Vol. 3(10), pp. 1373-1384, 1995.*

28- *Ding, J., Li, H., "The Forecasting Model of Flight Delay Based On DMT-GMT Model", Physics Procedia, Vol. 33, pp. 395 – 402, 2012.*

29- *Britto, R., Dresner, M., Voltes, A., "The impact of flight delays on passenger demand and societal welfare", Transportation Research Part E, Vol. 48, pp. 460–469, 2012.*

30- *Baumgarten, P., Malina, R., Lange, A., "The impact of hubbing concentration on flight delays within airline networks: An empirical analysis of the US domestic market", Transportation Research Part E: Logistics and Transportation Review, Vol. 66, pp. 103-114, June 2014.*

31- *Fan, L.W., Wu, F., Zhou, P. "Efficiency measurement of Chinese airports with flight delays by directional distance function", Journal of Air Transport Management, Vol. 34, pp. 140-145, January 2014.*

32- *Towell, G. G., and Shavlik, J. W., "Knowledge-Based Artificial Neural Networks", Artificial intelligence, Vol. 70(1), pp. 119-165, 1994.*