MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

DRUG DEVELOPMENT AND INDUSTRIAL PHARMACY®
Vol. 29, No. 3, pp. 349–355, 2003

RESEARCH PAPER

Comparison of Neural Network and Multiple Linear Regression as Dissolution Predictors[#]

Pradeep M. Sathe[1,*] and Jurgen Venitz[2]

[1]U.S. Food and Drug Administration, Rockville, Maryland, USA
[2]Virginia Commonwealth University, Richmond, Virginia, USA

ABSTRACT

The predictive performance of an artificial neural network (NN) was compared with the first-order multiple linear regression (MLR) using mean dissolution data of 28 diltiazem immediate release tablet formulations. The performance was evaluated using ''Weibull'' function parameters alpha and beta. Weibull parameters were used as dissolution markers of the eight principal, mainly compositional, variables. The parameters were obtained by fitting the Weibull function to the mean ($n = 12$) dissolution profiles of 28 diltiazem hydrochloride tablet formulations. The generated set of 28 pairs of Weibull function parameters was evaluated for internal and external predictability using both the MLR and the artificial NN. A three-layered 8-5-2 feedforward NN was found to be an adequate descriptor of the dissolution data. Internal predictions were based on the data of 24 products. External predictions used the 24 product data to test four products not used in the training phase. The predictive performances of the two techniques were evaluated using bias (mean prediction error; MPE) and precision (mean absolute error; MAE). The study results suggested that, for the studied data set, NN is a superior internal and external predictor to MLR.

	Internal (alpha, beta)		External (alpha, beta)	
	MPE (%)	MAE (%)	MPE (%)	MAE (%)
MLR	0, 0	12.5, 5.7	11.2, −7.1	18.1, 7.1
NN	0.3, 0.1	3.3, 1.7	−0.7, −0.4	3.3, 0.9

[#]This paper represents the personal opinions of the authors and does not necessarily represent the views or policies of the U.S. Food and Drug Administration.

*Correspondence: Pradeep M. Sathe, U.S. Food and Drug Administration, Office of Generic Drugs, Division of Bioequivalence, 7500 Standish Place, Rockville, MD 20855, USA; E-mail: sathe@cder.fda.gov.

DOI: 10.1081/DDC-120018209
0363-9045 (Print); 1520-5762 (Online)

www.dekker.com

The artificial NN predicted order of the formulation composition variables, influencing the dissolution parameters as follows: hydrogenated oil > microcrystallinecellulose > ethyl cellulose > eudragit > hydroxypropylcellulose > coat > hydroxypropylmethylcellulose > Speed.

Key Words: Neural network; Multiple linear regression; Dissolution; Prediction.

INTRODUCTION

Multiple linear regression (MLR) is frequently used to assess the impact of compositional variables on the formulation release characteristics. Multiple linear regression simultaneously relates a variety of input variables to the output response yielding slopes and intercepts. Using this relationship, the desired output response is predicted corresponding to selected input levels.

Artificial intelligence[1–3] (AI) is a branch of computer science devoted to programming computers to carry out tasks that, if carried out by humans, would require intelligence. Pattern recognition is a prerequisite to most AI tasks. An AI program extracts significant patterns from the problem situation and uses these as clues to the solution. ''Artificial neural network'' (NN) is an application of AI, which uses pattern recognition. Artificial NNs are simulations of collections of model biological neurons. They do not model the actual biology, chemistry, or physics of a real neuron, but model several aspects of information. By combining and recognizing patterns, NNs associate, emulate, analyze, and predict.

In the past, NNs have been used for computing pharmaceutical product formulations.[4] Hussain et al.[5] studied NNs for predicting human pharmacokinetic parameters from animal data. More recently, Peh et al.[6] used NNs to predict the dissolution profiles of matrix-controlled release theophylline pellets containing different ratios of microcrystalline cellulose and glyceryl monostearate. The current research involved the use of NNs for studying drug release characteristics of the marketed immediate release diltiazem hydrochloride tablets. No assumption of the release characteristics was made.

Research Objectives

The research involved the use of NNs to study drug release performance of the marketed formulations. Drug product dissolution was characterized by the empirical ''Weibull'' model. Research objectives were to:

1. Develop an adequate NN to characterize the diltiazem tablet dissolution of 28 products.
2. Compare the performance of a first-order MLR to the NN developed in step 1. Comparisons were made using the predictions of in vitro dissolution from the formulation variables (based on both the internal and external predictions).

EXPERIMENTAL

Composition data from 24 diltiazem hydrochloride, immediate release tablets were pooled. Tablet strengths were 30 mg, 60 mg, 90 mg, and 120 mg. They came from six different manufacturers ($6 * 4 = 24$ compositions). One manufacturer had dissolution results using a different rotation speed on the four strengths. The total evaluated products were, therefore, $24 + 4 = 28$. Mean dissolution profiles (mean of 12 units) corresponding to each product were studied. The dissolution data had been generated by the formulators as per the U.S. Pharmacopeia 23 method.

The predictive performance of NN and the first-order MLR was evaluated using mean dissolution profiles. Instead of analyzing the simplistic percentage dissolved vs. time plots, the profiles (of all 28 products) were characterized by Weibull function parameters. This gave us a better handle on the dissolution rate. Both the internal ($n = 24$) and external ($n = 4$) predictions were studied with respect to bias and precision.[6] The evaluation was carried out through the following steps:

1. Weibull function ($X_t = X_{\text{inf}}(1 - e * *((\text{alpha}) * ((t) * *(\text{beta}))))$ was fitted to mean ($n = 12$) diltiazem hydrochloride tablet dissolution data using the Scientist® (version 2.01, Micromath, SLC, UT, USA) program.[7]
2. Weibull parameters (alpha, beta) were generated for all 28 products.

3. Relationship of compositional variables and rotation speed to Weibull parameters was studied.
4. Different NN schemes were developed and trained to select the adequate network, which described the behavior of Weibull parameters. Multiple linear regression was conducted describing the relationship between all eight variables and Weibull parameters.
5. Adequate NN and MLR performances were compared using 24 products (internal prediction).
6. Using the NNs and MLR from step 4, four additional products with similar compositions, but different rotation speed, were tested to predict Weibull parameter estimates (external prediction).
7. Influence of each variable on the Weibull parameters was assessed using both NN and MLR.

NN Settings

Neural network analysis was conducted using Neuralyst®, version 1.3 (Cheshire Engineering Corp.)[8] All NNs were fully interconnected perceptrons with a sigmoid threshold function. More than 50,000 epochs (iterations) were run for each of the three-layered feedforward network schemes, using the following training parameters:

Learning Rate: 1, Momentum: 0.9, Input Noise: 0, Training Tolerance: 0.1, Testing Tolerance: 0.3.

Learning rate determined the magnitude of correction term applied to adjust each neuron's weights while training. Larger values resulted in quicker training. Momentum determined the "lifetime" of a correction term as the training progressed. Values closer to 1 caused the NN to retain more of the impact of previous corrections to the current corrections. Input noise provided a slight random variation to each input value for every input epoch. Training tolerance defined the percentage error allowed in comparing the NN output to the target value to be scored as "right" during the training process. Testing tolerance was similar to the training tolerance. It was applied to NN outputs and target values of the test data. Neural network output and test target values were scored as "right" if they were within the testing tolerance limits.

RESULTS AND DISCUSSION

Scatterplots of compositional variables vs. Weibull parameters yielded seven ingredients, influencing the dissolution. They were coat (qualitative), ethyl cellulose, Eudragit, hydrogenated oil, hydroxypropylcellulose, hydroxypropylmethylcellulose, and microcrystalline cellulose. Along with the rotation speed, this resulted in eight main variables influencing the drug product dissolution. The relationships between different variables and Weibull parameters are given in Fig. 1A,B. The output Weibull parameters, alpha and beta, reflected the extent and shape of dissolution.

To select an adequate network, different NN schemes were evaluated based on the above 8 input and 2 output variables. An 8-5-2 NN scheme was judged as adequate, based on model selection statistics. The optimal NN scheme is shown in Fig. 2. Neural network selection criteria and relevant statistics are given in Table 1. The internal prediction comparison is given in Table 2, and the external prediction comparison is given in Table 3.

The descriptive performance of the network was assessed by accuracy (mean prediction error, MPE) and precision[9] [root mean square error (RMS) and mean absolute error (MAE)]. The model goodness-of-fit was evaluated through the correlation coefficient r between the observed and predicted values for alpha and beta. The study of different NN schemes (Table 1) indicated that the 8-5-2 scheme yielded the least RMS error. Accuracy as reflected by the MPE was better for both alpha and beta parameters compared with the 8-4-2 scheme. The r values for both schemes were comparable. The internal predictions based on 24 sets of data indicated that the NN analysis (8-5-2 scheme) improved the precision and goodness-of-fit compared with the MLR (Table 2). The standard errors of predictions for NN were MPEAlpha = 4.5%, MPEBeta = 0.4%, MAEAlpha = 0.5%, and MAEBeta = 0.3%. For MLR, they were MPEAlpha = 3.2%, MPEBeta = 1.3%, MAEAlpha = 1.9%, and MAEBeta = 0.6%. The external predictions of four new data sets, which were based on the training conducted on the 24 data sets, also indicated superior NN performance compared with MLR. The standard error of predictions were for MPEAlpha = 2.0%, MPEBeta = 0.5%, MAEAlpha = 0.9%, MAEBeta = 0.3% for NN vs. MPEAlpha = 10.9%, MPEBeta = 2.5%, MAEAlpha = 7.1%, and MAEBeta = 2.5% for MLR.

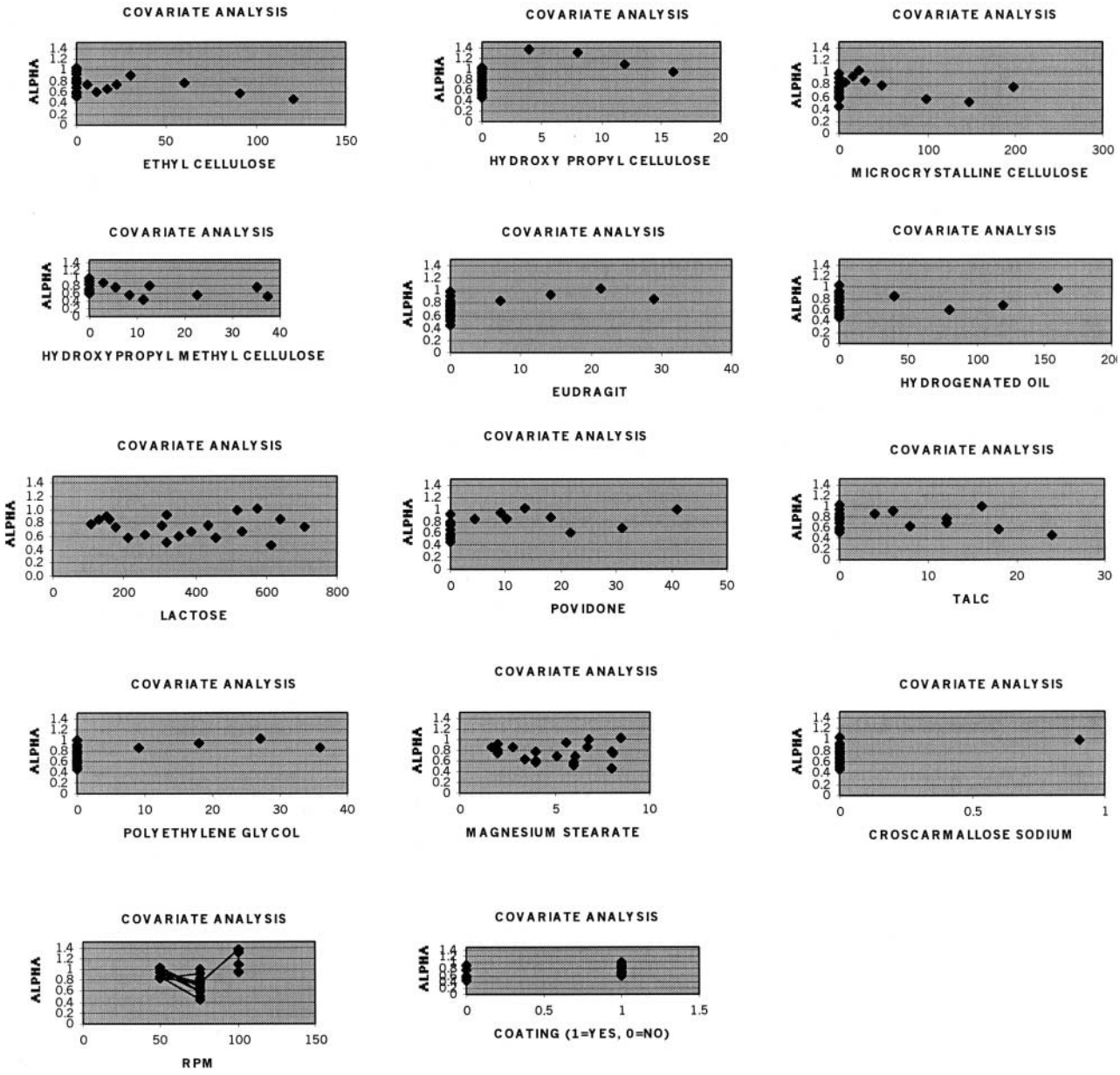

Figure 1A. Alpha values corresponding to different univariates.

To assess the influence of each variable on dissolution markers, the RMS errors of 7-5-2 schemes were studied. One variable was omitted from the analysis at a time to render a 7-5-2 scheme. An NN run was made, and RMS values were studied. The RMS error values indicated that the rotation speed influenced dissolution the least while hydrogenated oil influenced it the most. The influence of different celluloses and Eudragit on dissolution ranked in between these two extremes. The preliminary observation on the rank order of variables influencing alpha and beta taken together was hydrogenated oil > microcrystalline cellulose > ethyl cellulose > Eudragit > hydroxypropylcellulose > coat > hydroxypropylmethylcellulose > speed. The MLR rank order for these variables (absolute values) differed from the NNs, as well as between alpha and beta taken individually. The rank order for alpha was coat > hydroxypropylcellulose > hydroxypropylmethylcellulose > speed > microcrystalline cellulose > ethyl

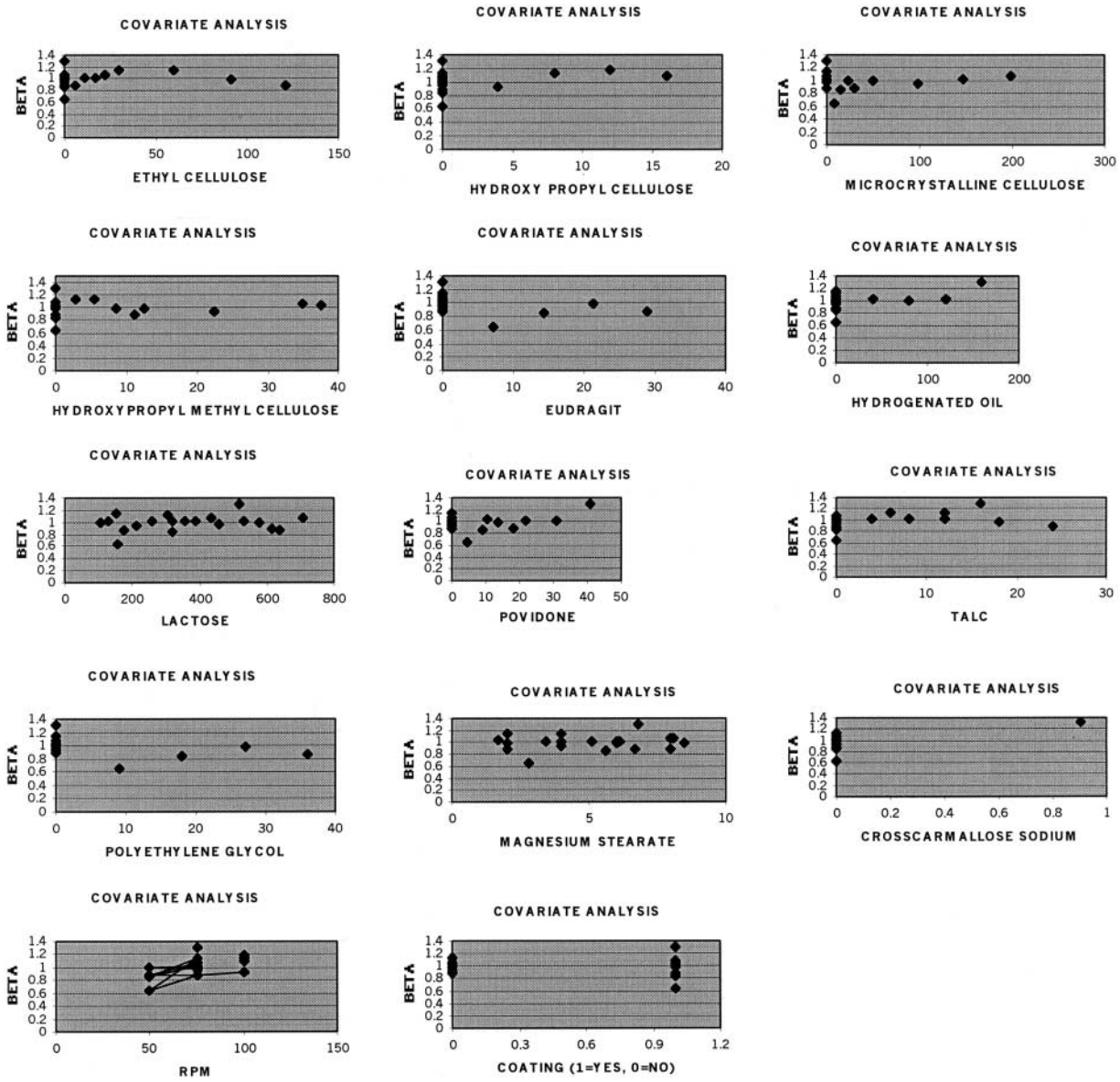

Figure 1B. Beta values corresponding to different univariates.

cellulose > hydrogenated oil > Eudragit. For beta, the rank order was hydroxypropylcellulose > coat > hydroxy- propylmethylcellulose > ethyl cellulose > speed > hydrogenated oil > microcrystallinecellulose > Eudragit. Efforts are currently underway to validate the above rank-order observation.

CONCLUSIONS

The artificial NN was found to be an effective tool for multivariate analysis. For the given data set, the following could be concluded:

1. An 8-5-2 NN scheme adequately characterized the dissolution data.
2. NN was a better internal and external predictor than MLR.
3. Based on a 7-5-2 NN scheme (deleting each variable at a time and running the network), for the given data set, the order of variables influencing the Weibull parameters taken together was as follows: hydrogenated oil > microcrystalline

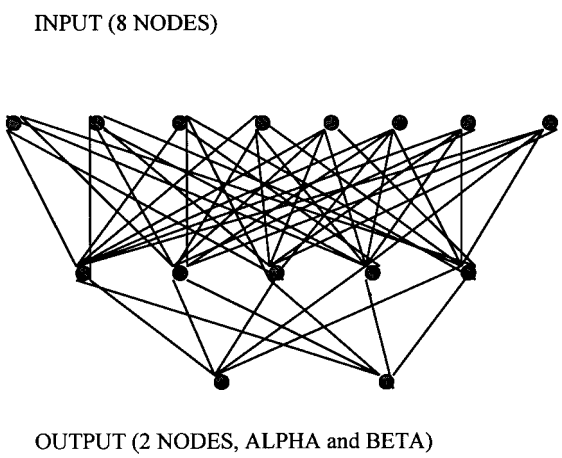

Figure 2. The 8-5-2 feedforward NN.

cellulose > ethyl cellulose > Eudragit > hydroxypropyl cellulose > coat > hydroxypropylmethylcellulose > speed.

Comparison of NN and MLR

An Artificial NN can be used for linear and nonlinear systems.[10] It's use allows us to study the covariate interactions. By adjusting weights, multivariate outputs can also be evaluated. Multiple linear regression can be used only with linear systems. It uses slopes and intercepts to ascertain the variable response relationship. The analysis involves evaluation of the multidimensional univariate outputs. The study of covariate interactions is possible,

Table 1. Different network schemes with the corresponding statistics.

NN Configuration				Alpha			Beta		
No.	Scheme	PAR	RMS	MPE	MAE	*r*	MPE	MAE	*r*
1	8-5-2	57	0.0269	0.0028	0.0332	0.9836	0.0010	0.0170	0.9868
2	8-4-2	46	0.0322	0.0037	0.0272	0.9860	0.0150	0.0190	0.9921
3	8-3-2	35	0.0733	−0.0064	0.0365	0.9793	−0.0020	0.0550	0.8637
4	8-2-2	24	0.1029	0.0004	0.0634	0.9422	−0.0040	0.0730	0.7419
	MLR	9		0	0.1186	0.7912	0	0.0750	0.6939

PAR, no. of parameters.

Table 2. Internal predictions statistics of NN and MLR.

	Optimal NN				MLR			
	Alpha	SE (%)	Beta	SE (%)	Alpha	SE (%)	Beta	SE (%)
R	0.9836		0.9868		0.7291		0.8457	
MPE	0.0028	4.5	0.0014	0.4	0.0	3.2	0.0	1.3
MAE	0.0332	0.5	0.0172	0.3	0.1249	1.9	0.0573	0.6

Table 3. External predictions statistics of NN and MLR.

	Actual		NN predicted		MLR predicted	
Product no.	Alpha	Beta	Alpha	Beta	Alpha	Beta
1	1.3080	1.0974	1.2784	1.1013	1.0864	0.9894
2	1.2221	1.1529	1.1828	1.1447	1.0294	0.9827
3	0.9749	1.1137	1.0259	1.1181	0.9723	0.9759
4	0.8729	1.0961	0.8626	1.0786	0.9153	0.9692
MPE SE (%)			2.0	0.5	10.9	2.5
MAE SE (%)			0.9	0.3	7.1	2.5

provided they are specified prospectively. In data analysis, NN appeared to be less biased and more precise, compared with MLR.

REFERENCES

1. Barr Avron; Fagenbaum Edward. *The Handbook of Artificial Intelligence*; William Kaufman Publishers, 1981; 1–31.
2. Graham Neil. *Artificial Intelligence*; Tab Books Publishers: 1979, 1–33.
3. Winston, P.H. *Artificial Intelligence*, 2nd Ed.; Addison-Wesley Publishers, 1984; pp. 1–43.
4. Hussain, A.S.; Yu, X.; Johnson, R.D. Application of neural computing in pharmaceutical product formulation. Pharmaceutical Research **1991**, *8*, 1248–1252.
5. Hussain, A.S.; Johnson, R.D.; Vachhrajani, N.; Ritshel, W.A. Feasibility of developing a neural network for prediction of human pharmacokinetic parameters from animal data. Pharmaceutical Research **1993**, *10*, 466–469.
6. Peh Kok Khiang; Lim Chee Peng; Quek Siow San; Khoh Kean Hock. Use of artificial neural network to predict drug dissolution profiles and evaluation of network performance using similarity factor. Pharmaceutical Research **2000**, *17*, 1384–1388.
7. Micromath^{R} Scientist^{R} for Windows 1995.
8. 'Neuralyst^{C}' Users Guide (Cheshire Corporation), 1994.
9. Sheiner, L.B.; Beal, S.L. Some suggestions for measuring predictive performance. Journal of Pharmacokinetics and Biopharmaceutics **1981**, *9*, 503–512.
10. Takayama, K.; Takahara, J.; Isowa, K.; Nagai, T. Artificial neural network as a novel method to optimize pharmaceutical formulations. Pharmaceutical Research **1999**, *16*, 1–6.