AAPS PharmSciTech **2003; 4 (2) Article 26 (http://www.pharmscitech.org).**

The Use of Artificial Neural Networks for the Selection of the Most Appropriate Formulation and Processing Variables in Order to Predict the In Vitro Dissolution of Sustained Release Minitablets

Submitted: February 19, 2003; Accepted: May 8, 2003

Michael M. Leane,[1,2] Iain Cumming,[2] and Owen I. Corrigan[1]

[1]Department of Pharmaceutics and Pharmaceutical Technology, School of Pharmacy, Trinity College, Dublin 2, Ireland

[2]Elan Pharmaceutical Technologies, Biotechnology Building, Trinity College, Dublin 2, Ireland

ABSTRACT

The objective of this work was to apply artificial neural networks (ANNs) to examine the relative importance of various factors, both formulation and process, governing the in-vitro dissolution from enteric-coated sustained release (SR) minitablets. Input feature selection (IFS) algorithms were used in order to give an estimate of the relative importance of the various formulation and processing variables in determining minitablet dissolution rate. Both forward and backward stepwise algorithms were used as well as genetic algorithms. Networks were subsequently trained using the back propagation algorithm in order to check whether or not the IFS process had correctly located any unimportant inputs. IFS gave consistent rankings for the importance of the various formulation and processing variables in determining the release of drug from minitablets. Consistent ranking was achieved for both indices of the release process; ie, the time taken for release to commence through the enteric coat (T_{lag}) and that for the drug to diffuse through the SR matrix of the minitablet into the dissolution medium (T_{90-10}). In the case of the T_{lag} phase, the main coating parameters, along with the original batch blend size and the blend time with lubricant, were found to have most influence. By contrast, with the T_{90-10} phase, the amounts of matrix forming polymer and direct compression filler were most important. In the subsequent training of the ANNs, removal of inputs regarded as less important led to improved network performance. ANNs were capable of ranking the relative importance of the various formulations and processing variables that influenced the release rate of the drug from minitablets. This could be done for all main stages of the release process. Subsequent training of the ANN verified that removal of less relevant inputs from the training process led to an improved performance from the ANN.

KEYWORDS: artificial neural network, formulation development, sustained release, minitablet, input feature selection

Corresponding Author: Owen I. Corrigan, Department of Pharmaceutics and Pharmaceutical Technology, School of Pharmacy, Trinity College, Dublin 2, Ireland. Phone: +353-1-6082782; Fax: +353-1-6082783; Email: occorrign@tcd.ie

INTRODUCTION

Artificial neural networks (ANNs) are computer systems developed to mimic the operations of the human brain by mathematically modeling its neurophysiological structure (ie, its nerve cells and the network of interconnections between them). In an ANN, the nerve cells are replaced by computational units called neurons and the strengths of the interconnections are represented by weights.[1] This unique arrangement can therefore attempt to simulate some of the neurological processing ability of the biological brain such as learning and drawing conclusions from experience.[2]

Each neuron takes 1 or more inputs and creates an output, which may be passed on to another neuron. The method by which the neurons are organized is termed the "network architecture." In an ANN, the neurons are usually organized in layers. In feed-forward networks, there is always 1 input and 1 output layer with 1 or more hidden layers.

The number of neurons in the input and output layers is automatically determined by the number of input and

output variables in the problem being considered. Therefore, the number of neurons to be incorporated into the hidden layer is a key decision. If a hidden layer has too few neurons, the network will lack the power it needs to classify patterns in the data. If a hidden layer has too many neurons, existing patterns will merely be memorized and the network will be unable to generalize (ie, spot patterns in new data).[3]

ANNs are trained using preexisting data. At the beginning of the training process, the connections between the neurons are set to random weight values. During the training process, the input and output data from the training data subset are fed into the network. The difference between the actual output and the training output values is then calculated. The difference is an error value, which is decreased using a training algorithm during the training process by modifying the values of the weights at each neuron. These modifications bring the output of the network closer to the desired output. Once trained, the network can hopefully be used to predict accurate output values for new input data.

ANNs have many advantages over conventional statistical techniques. They give good results when the response variable is highly nonlinear. In addition, historic or literature data can be used for training.[4] Other advantages are that ANNs can make use of incomplete data and that no a priori knowledge of the underlying statistical nature of the problem is required.[5]

Applications in Formulation Development

The current work deals specifically with the area of formulation development and the use of ANNs to examine the relative importance of formulation and processing variables in determining the dissolution profile of a sustained release (SR) dosage form. This is a challenging problem since complex nonlinear relationships exist between the independent formulation and processing variables and the dependent drug release variable. Such complex relationships include the levels of the active drug and the various excipients, possible interactions between the drug and excipients, and possible interactions or synergies between the excipient and numerous processing factors.

The nonlinear processing ability and unique structure of ANNs means that they have significant potential in dealing with such problems and several research groups have already utilized ANNs in related areas, such as the following:

- the modeling of in-vitro release of drugs from SR hydrophilic matrix capsules[4]
- the modeling of in-vitro release of drugs from SR hydrophilic matrix tablets[2, 6-8]
- analysis of fluidized bed granulation processing[9,10]
- evaluation of direct compression IR tablet formulations[11]
- modeling the formulation and processing of tablets[12]

All of these studies have used ANNs as a predictive tool (ie, to train an ANN to try to predict a specific output parameter, for example, tablet dissolution, from a set of input variables such as formulation variables) and have demonstrated that ANNs are capable of a comparable performance to traditional statistical techniques used for this purpose.

In this study, an alternative use of ANNs in formulation development was explored. Because of their unique abilities in spotting patterns in data, ANNs can be used to rank which of the various formulation and processing variables are most critical in influencing the output parameter of interest, in this case the in vitro release of drug from the sustained release minitablets. Knowledge of critical variables in determining tablet dissolution rate would be of considerable benefit to the formulator.

In addition, such an investigation would also identify relatively unimportant variables. Removal of such variables would be of benefit particularly when training ANNs for predictive purposes because the more inputs an ANN has the more cases it needs in order to be successfully trained—the so-called "curse of dimensionality".[13] Therefore, if less relevant inputs can be eliminated, the ANN has a much better chance of successful generalization, even if some information is lost. This approach is known as "input feature selection" and is described in more detail below.

This study utilized a data set containing 125 cases of the drug CEL50 development project,[14] all of which contained complete formulation and processing details and corresponding dissolution data. This relatively large data set should give greater reliability to the results obtained.

Input Feature Selection

In many formulation problems, a wide range of input variables are available that can be used to train a neural network, but it is hard to define which of them are most relevant, or indeed are useful at all. The situation is further confused when there are interdependencies or correlations between some of the input variables,

which means that any of a number of subsets might be adequate.[15]

Input Feature Selection (IFS) covers a variety of techniques that seek to identify input variables that do not contribute significantly to network performance, so that they can be removed.

However, the difficulty of IFS should not be underestimated. It is literally an exponential search problem in which the number of feature subsets to be considered is 2^n, where n is the number of possible features. For 19 features, there are over half a million possible feature subsets. Therefore, techniques combining mathematical algorithms and neural networks are used. These can be either stepwise algorithms that progressively add or remove variables or a genetic algorithm. These algorithms may discover subsets of inputs that are not discovered by other techniques.

Forward stepwise feature selection starts by finding the input variable that, by itself, best predicts the output variable. It then looks for a second variable, which most improves the model when added to the first. This process is continued until either all variables have been selected or no further improvement is seen. Backwards stepwise feature selection takes the opposite approach. It starts with a model including all variables and then discards one at a time. At each stage backwards stepwise feature selection finds the variable that least degrades the model when it is removed.

An alternative to the 2 feature selection methods outlined above is the genetic algorithm. This mathematical algorithm is loosely based on the Darwinian theory of evolution. To start, the algorithm selects random populations of inputs. It then uses a process similar to natural selection to select better inputs, which are combined or "bred" together to form a new population. Over successive generations, increasingly better populations are produced until eventually the optimum set of inputs is found. Such algorithms are especially efficient at locating interdependencies between variables.

Each of the algorithms displays messages showing how they are progressing. With the stepwise selection methods, these messages serve to give a ranking of the importance of the variables.

Once the appropriate input variables have been selected by the algorithms, a generalized regression neural network (GRNN) is then used to test this new training set. These types of networks are used because they normally train very rapidly; they are able to model nonlinear functions quite accurately; and they are very sensitive to the presence of irrelevant input variables, which is an advantage when trying to decide what variables are required.

Although irrelevant input variables will lead to deterioration in network performance, they may do so by only a tiny amount, making it extremely tricky for the algorithm to locate all of them. Also, as stated earlier, it is usually better to use fewer input variables even if this causes a slight increase in the error, as the generalization power of the network may then be improved. Therefore, a unit penalty factor can be set. Such a penalty is multiplied by the number of selected inputs and added to the error, thus favoring smaller networks. With a penalty of zero, only obviously redundant inputs will be discarded. Higher penalty values will favor smaller networks and will usually improve performance. However, if the factor is set too large, the number of variables becomes more important than the quality of the network, and this will eventually lead to the algorithm screening out all inputs.

The Minitablet Development Project

The data used in the current work were generated in the course of a project aimed at developing extended release, enteric-coated dosage forms containing the equivalent of 40 mg of a drug CEL50. Initial blending of the materials was carried out in a Y-cone blender (Patterson-Kelley, East Stroudsburg, PA). The minitablets were manufactured using a direct compression tableting process. Small-scale and larger batches were compressed on a single station tablet press (Horn Noack, Romaco Inc, Pompton Plains, NJ) or a Fette P2100 rotary tablet press (Fette, Schwarzenbek, Germany) with multitipped tooling (x8), respectively. The Opadry coatings were applied using an aqueous system with the enteric coatings being applied from solvent-based coating suspensions. Small-scale and larger batches were coated in a Freund Hi-Coater (Vector Corp, Cedar Rapids, IA) or an Accela-Cota (Manesty, Knowsley, Merseyside, UK) tablet coater, respectively. A brief description of the ranges of the relevant formulation and processing variables is given in **Table 1**.

Eudragit L and Eudragit S coatings dissolve at different rates. The Eudragit L polymer dissolves at pH values greater than 6.0, while the Eudragit S polymer dissolves at pH values greater than 7.0. Therefore, the Eudragit S-coated minitablets are intended to present their dose at more distal regions of the intestine. When the coatings become permeable and dissolve, they allow progressive ingress of water into the SR cores and release of the active ingredient in a controlled manner. Release from the SR cores is by erosion or by diffusion

Table 1. Properties of Minitablets

CEL50 content	16.8%
Methocel content	10%-20%
Klucel content	5%-10%
Aerosil content	1%
Magnesium stearate content	1%-3%
Avicel content	49.2%-66.2%
Target weight gain after Eudragit L coating	16%
Target weight gain after Eudragit S coating	10%
Target weight gain after Opadry coating	2.5%
Spray rate of coating solution	3-420 g/min
Original blend size of the batch	1.5-150 kg
Batch size	1.5-150 kg
Blend time with lubricant	5-20 minutes
Tablet press speed	30 000-70 000 revs/hr
Filomatic speed setting	1-6
Tablet punch size	3.8 mm, round

through the resulting viscous layer. The combination of the components results in a product that in vivo exhibits an extended duration of action suitable for once-a-day dosing.

MATERIALS AND METHODS

Drug Release Data

Dissolution data (125 profiles) obtained from SR minitablets, both coated and uncoated, were gathered for analysis. The original outputs for the neural network were the dissolution profiles for these SR tablets. A USP apparatus type 2 containing 900 mL of 0.015 M citrate/phosphate buffer, pH 6.8, was used for tablet dissolution testing. The complete dissolution of both enteric coatings (Eudragit L and S) required that the medium pH was greater than pH 6.5. Sodium lauryl sulphate (SLS), 0.5%, was introduced to achieve sink conditions for the CEL50.

The profiles were normalized between 0.0% and 100%. The times taken for 10% of the dosage form and 90% of the dosage form to be released were calculated from the profiles. These were converted to 2 outputs for use in the neural network:

1. Time taken for 10% of the drug to be released (T_{lag}) was taken to be an indication of the lag time before release began (ie, the extent of the delaying effect of the enteric coating).
2. Time taken for 90% of the drug to be released less the time taken for 10% of the drug to be released (T_{90-10}) was taken to be an indication of the time taken for the drug to be released from the SR matrix once the outer coat had dissolved.

Similar simple dissolution parameters, such as the time to 50% drug dissolution have been used in other studies.[2,4]

The 2 separate phases in the dissolution profile (lag time followed by SR) are illustrated in **Figure 1**.

Software

The Trajan Neural Network Simulator, version 4, (Washington, Tyne and Wear, UK) was the software package used in this study. This is a Windows-based package, which supports numerous types of neural networks along with the fastest state-of-the-art training algorithms. The package includes other algorithms that carry out a variety of tasks such as pre- and post-processing of data, input feature selection, network design, and selection. The software also gives extensive statistical feedback on each network.[15]

Input variables

The number of inputs used for the network was 19 (see **Tables 2**, **3**, and **4**). The inputs included 11 variables related to the composition of the formulation and 8 variables related to the processing conditions: percentage Cel50 in the formulation, percentages of methocel K15M, methocel K100M, methocel K100LV, klucel, aerosil, magnesium stearate, avicel PH101, percentage weight gain after Eudragit L or Eudragit S coating, percentage weight gain after Opadry coating, spray rate of the coating solution (g/min), blend size of the original tablet batch (kg), batch size at that stage of processing (kg), time (minutes) the formulation was blended with the lubricant (magnesium stearate), press speed (revolutions per hour), Filomatic speed setting (setting on dial that governs how fast the blend was fed down from the hopper into the tablet press), tablet hardness (newtons), and tablet weight (mg).

Input Feature Selection

The input feature selection algorithms in Trajan were used to select the most important inputs in the CEL50

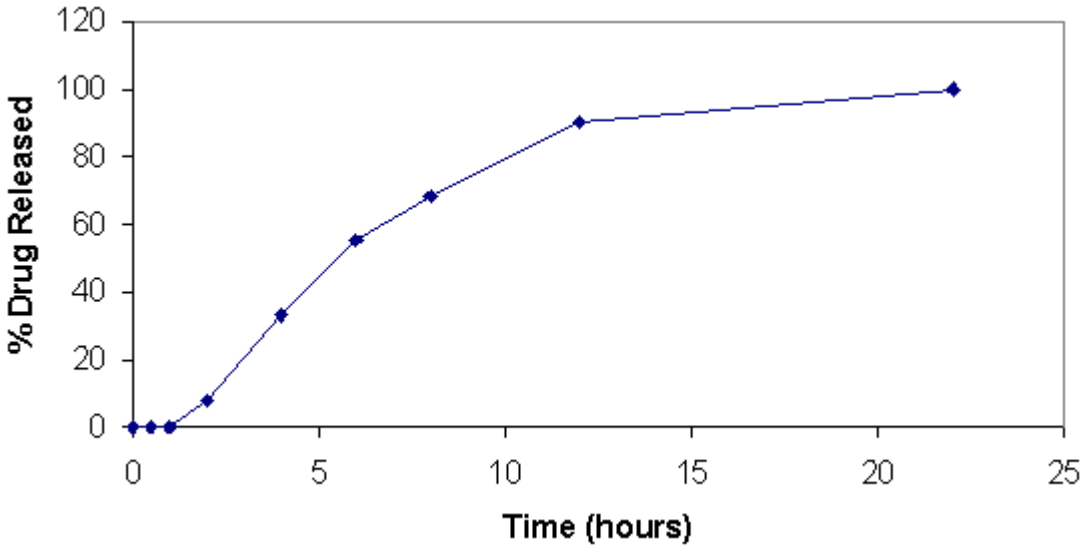

Figure 1. Dissolution profile for minitablet batch PD14859: a sigmoidal profile comprising an initial lag phase (T_{lag}) followed by a relatively linear release phase (T_{90-10}).

Table 2. Input Feature Selection Algorithm Results for the T_{lag} Output

	Forward Stepwise	**Backward Stepwise**	**Genetic Algorithm**
Inputs rejected at unit weight penalty of zero	% Aerosil Filomatic speed Tablet hardness Tablet weight	% Cel50 % Methocel K15M % Methocel K100LV % Aerosil Filomatic speed Tablet weight	% Cel50 % Methocel K15M % Methocel K100LV % Aerosil Filomatic speed Tablet weight
Inputs rejected at unit weight penalty of 0.001	% Cel50 % Methocel K15M % Methocel K100LV % Eudragit L coating Batch size Tablet press speed	% Eudragit L coating Tablet press speed Tablet hardness	% Eudragit L coating Tablet press speed Tablet hardness
Inputs rejected at unit weight penalty of 0.002	% Methocel K100M % Magnesium stearate	% Methocel K100M % Magnesium stearate % Avicel Batch size	% Methocel K100M % Klucel % Magnesium stearate % Avicel Blend time with magnesium stearate
Inputs rejected at unit weight penalty of 0.003	% Klucel % Avicel Blend time with magnesium stearate	% Klucel Blend time with magnesium stearate	Batch size
Inputs rejected at unit weight penalty of 0.004	None	None	None
Inputs rejected at unit weight penalty of 0.005	Blend size of the original batch	None	None
Inputs remaining at unit weight penalty of 0.005	% Eudragit S coating % Opadry coating Spray rate of coating solution	% Eudragit S coating % Opadry coating Spray rate of coating solution Blend size of the original batch	% Eudragit S coating % Opadry coating Spray rate of coating solution Blend size of the original batch

data set. The forward selection, backward selection, and genetic algorithms were all used. Each algorithm was run 10 times. An input was deemed to be useful if it was selected as important in more than half the cases (ie, 6 or more times). The unit weight penalty was set in the range 0 to 0.005 (increasing in increments of 0.001).

An additional analysis was done by observing the order in which the inputs were added during the running of the forward selection algorithm with zero unit weight penalty. The first input selected was taken to be the most important and so on.

A similar analysis was carried out on the backward selection algorithm with a high unit weight penalty (ie, one where all or nearly all of the input would ultimately be rejected). The first input to be discarded was regarded as being the least important and so on.

A percentage of Aerosil was included in the inputs as a negative control as its level did not change in any of the formulations included. If the algorithms function correctly, they should select the percentage of Aerosil as unimportant and rank it as the least important input.

The above experiments were first carried out for the T_{lag} output and then repeated for the T_{90-10} output.

Training the ANN

ANNs were trained using a 5-fold cross-validation model to produce estimations of generalization error (ie, the 125 cases in the Cel50 data set were randomly divided into 5 subsets of 25 cases that were used for verification on models trained with the remaining 100 cases). The verification subsets were included to ensure that the ANN was generalizing and not merely memorizing the data, so called "over-learning." Cross-validation was used as it gives a more reliable estimate of generalization performance compared with merely using a single verification set. Training was stopped when the error of the verification subset began to rise (this is an indication that over-learning is occurring).

Initially, all formulation and processing variables in the Cel50 data set were used for training along with one output (either the T_{lag} output or the T_{90-10} output). The performance of the ANN was assessed by measuring the error (root mean squared) separately for both the training and verification subsets. All data were analyzed using one-way analysis of variance (ANOVA). Where a significant difference was found between groups, Tukey's post-hoc test was performed to identify the location of this difference using GraphPad Prism, version 3.00, software (GraphPad Software Inc, San Diego, CA).

Then, inputs were systematically removed to see if removal of these less relevant inputs (as determined by IFS above) could lead to improved performance. The network architecture used in this study was the feed-forward multilayer perceptron (MLP). This type of network has been extensively described in the literature and has been found to be very efficient in dealing with regression problems like this.[2,16] Only one hidden layer was used in all cases and the numbers of hidden neurons in that layer were kept to a minimum. This was done to minimize the possibility of over-learning. The ANN was trained using the back propagation algorithm with a sigmoidal activation function for both the hidden and output layers.[1,2,10] This algorithm has the advantage that because of its slower convergence and randomized order of presentation of cases, it is less inclined to over-learn.

The following standard algorithm parameters were maintained throughout: learning rate, 0.1; momentum, 0.3.

RESULTS AND DISCUSSION

Input Feature Selection

In general, the 3 different methods for input feature selection showed broad similarities in their assessment of which inputs were important/unimportant in determining dissolution profiles (**Tables 2** and **3**). The orders in which the forward and backward stepwise algorithms added/subtracted the inputs confirmed the same rank order of importance of variables (**Table 4**).

For the T_{lag}, the indicator of the extent of the lag time before dissolution began, coating parameters such as percentage Eudragit S and Opadry coating applied and the spray rate of the coating solution were among the most important inputs highlighted (**Table 2**). This would be expected because the most important factor delaying the release from the minitablets is the enteric coating. What is less obvious is that the blend size of the original batch was also selected as important. This selection reflects the influence of scale-up on the final dissolution characteristics of the product.

Inputs regarded as being of moderate importance included the percentages of the main components along with 1 processing parameter and the blend time after the addition of lubricant (magnesium stearate). It has long been known that changes in the duration and

Table 3. Input Feature Selection Algorithm Results for the T_{90-10} Output

	Forward Stepwise	**Backward Stepwise**	**Genetic Algorithm**
Inputs rejected at unit weight penalty of zero	% Methocel K15M % Methocel K100LV % Aerosil Blend time with magnesium stearate Filomatic speed Tablet hardness Tablet weight	% Methocel K15M % Methocel K100LV % Aerosil Tablet hardness Tablet weight	% Methocel K15M % Methocel K100LV % Aerosil Tablet hardness Tablet weight
Inputs rejected at unit weight penalty of 0.001	% Cel50 % Eudragit L coating % Opadry Tablet press speed	% Cel50 % Eudragit L coating % Opadry Blend time with magnesium stearate Tablet press speed Filomatic speed	% Cel50 % Eudragit L coating % Opadry Spray rate of coating solution Blend time with magnesium stearate Tablet press speed Filomatic speed
Inputs rejected at unit weight penalty of 0.002	% Methocel K100M % Eudragit S coating Spray rate of coating solution Batch size	% Eudragit S coating Spray rate of coating solution Batch size	% Methocel K100M % Eudragit S coating Batch size
Inputs rejected at unit weight penalty of 0.003	% Magnesium stearate Blend size	% Methocel K100M % Magnesium stearate Blend size	% Magnesium stearate Blend size
Inputs rejected at unit weight penalty of 0.004	% Klucel	% Klucel % Avicel	None
Inputs rejected at unit weight penalty of 0.005	% Avicel	None	% Klucel % Avicel
Inputs remaining at unit weight penalty of 0.005	None	None	None

mechanism of the lubricant mixing process can lead to changes in the dissolution properties of tablets.[17,18]

Less important variables included most of the processing variables and properties of the finished tablet, such as tablet weight and hardness. The latter 2 are probably unimportant because these factors were tightly controlled throughout the project. A change in dissolution would probably not be seen unless hardness or weight fell below a critical value, and such values were never generated in this data set.

More surprisingly, the percentage of drug and the percentage of Eudragit L coating applied were selected as being relatively unimportant. The differences in importance between the Eudragit L and Eudragit S coating levels on minitablets may have been artificially accentuated due to the poor solubility of the Eudragit S polymer at the pH 6.8 value of the dissolution medium.

The percentages of methocel K15M and K100LV were also regarded as relatively unimportant. This is most likely because these polymers were used in very few formulations, being rejected early in the development program in favor of the methocel K100M grade.

For the T_{90-10} output, the ANN was less definite as to which inputs were most important with no inputs remaining when the unit weight penalty was raised to 0.005 (**Table 3**). This reflects the increased complexity of the process controlling the release of the drug from the hydrophilic matrix compared with the previous (T_{lag}) output.

The most important inputs were selected to be the percentage of klucel (one of the matrix-forming polymers) and the percentage of avicel (an insoluble filler, the largest percentage constituent of the minitablets) (**Table 4**). Next in line came the blend size of the

Table 4. Order of Importance of Variables for the T_{lag} and T_{90-10} Outputs as Determined by the Forward and Backward Stepwise Algorithms

	T_{lag}		T_{90-10}	
	Forward Stepwise Ranking	**Backward Stepwise Ranking**	**Forward Stepwise Ranking**	**Backward Stepwise Ranking**
% Opadry	1	2	7	9
% Eudragit S	2	1	7	6
Spray rate of coating solution	3	3	9	8
Blend size of original tablet batch	4	4	4	3
% Klucel	5	6	2	2
Blend time with magnesium stearate	6	5	14	12
% Avicel	7	7	1	1
Batch size	8	9	6	7
% Magnesium stearate	9	8	5	5
% Methocel K100M	10	10	3	4
Press speed	11	11	10	10
% Eudragit L	12	13	11	11
% Cel50	13	15	12	13
Tablet hardness	14	12	17	16
% Methocel K15M	15	17	13	16
% Methocel K100LV	16	16	14	19
Filomatic speed	17	14	16	14
Tablet weight	18	19	18	18
% Aerosil	19	18	18	15

original batch (effect of scale-up), the percentage of methocel K100M (the other matrix-forming polymer) and the percentage of magnesium stearate (another insoluble constituent). Interestingly, some coating parameters such as percentage Eudragit S coating applied and the spray rate of coating solution applied were still relatively important showing that the enteric coat was still having an influence on the release of the tablet even after that release was well under way.

Less important were the other processing parameters and the percentages of other constituents such as Eudragit L, Opadry, Cel50, aerosil and the 2 methocel polymers K15M and K100LV. In contrast to the T_{lag} output, the blend time with magnesium stearate was ranked as relatively unimportant.

The percentage of aerosil was included in the inputs as a negative control as its level did not change in any of the formulations included. All the algorithms selected it as unimportant and it was ranked at or near the bottom of importance by both the forward and backward stepwise algorithms.

ANN Training

The number of hidden neurons to be used in the ANN was selected by a preliminary experiment in which networks were trained using a randomly selected verification set of 40 cases and increasing the number of hidden neurons 1 at a time (data not shown). The optimum number of hidden neurons was defined as the network that gave the lowest verification error. This was found to be 4 hidden neurons for the T_{lag} and 5 hidden neurons for the T_{90-10} output. This architecture was used in all subsequent experiments.

When training on the T_{lag} output using all 19 inputs, the ANN performed well giving acceptably low errors for both training and verification subsets (**Table 5**). Significantly different training and verification errors were obtained depending on which subset of cases was used as the verification set. This justified the approach of using a 5-fold cross-validation model rather than just a single verification set.

Table 5. Results of ANN Training for T_{lag} Output*

Subset No.	No. of Inputs	Training Error	Verification Error	No. of Inputs	Training Error	Verification Error
1	19	0.48 ± 0.01	0.24 ± 0.02	19	1.30 ± 0.03	1.13 ± 0.07
1	13	0.48 ± 0.01	0.24 ± 0.02	14	1.34 ± 0.07	0.99 ± 0.04†
1	10	0.42 ± 0.12	0.25 ± 0.02	12	1.54 ± 0.20‡	1.07 ± 0.04
2	19	0.45 ± 0.03	0.23 ± 0.02	19	1.51 ± 0.16	1.99 ± 0.07
2	13	0.46 ± 0.02	0.27 ± 0.04	14	1.32 ± 0.15	1.88 ± 0.14
2	10	0.44 ± 0.13	0.31 ± 0.03‡	12	1.57 ± 0.13	1.98 ± 0.04
3	19	0.32 ± 0.04	0.58 ± 0.04	19	1.71 ± 0.09	1.34 ± 0.06
3	13	0.23 ± 0.04	0.44 ± 0.04†	14	1.64 ± 0.21	1.24 ± 0.03†
3	10	0.24 ± 0.03†	0.34 ± 0.04†	12	1.78 ± 0.05	1.27 ± 0.03†
4	19	0.21 ± 0.02	0.29 ± 0.03	19	1.53 ± 0.34	1.41 ± 0.09
4	13	0.25 ± 0.06	0.28 ± 0.08	14	1.41 ± 0.04	1.18 ± 0.04†
4	10	0.24 ± 0.03	0.22 ± 0.03	12	1.47 ± 0.13	1.12 ± 0.11†
5	19	0.45 ± 0.03	0.54 ± 0.03	19	1.68 ± 0.31	2.59 ± 0.10
5	13	0.23 ± 0.04†	0.39 ± 0.07†	14	0.99 ± 0.28†	2.41 ± 0.08†
5	10	0.27 ± 0.05†	0.44 ± 0.04†	12	1.27 ± 0.35	2.43 ± 0.06†

*ANN indicates artificial neural network. Four hidden neurons were used for the T_{lag} and 5 hidden neurons for the $T_{90\text{-}10}$ outputs. The 6 inputs removed in the 13 input networks for the T_{lag} output were the percentage of Cel50, the percentage of methocel K15M, the percentage of methocel K100LV, the percentage of aerosil, Filomatic speed, and tablet weight. The 9 inputs removed in the 10 input networks for the T_{lag} output were above 6 plus the percentage of Eudragit L coating applied, tablet press speed, and tablet hardness. The 5 inputs removed in the 14 input networks for the $T_{90\text{-}10}$ output were the percentage of methocel K15M, the percentage of methocel K100LV, the percentage of aerosil, tablet hardness, tablet weight. The 7 inputs removed in the 12 input networks for the $T_{90\text{-}10}$ output were above 5 plus blend time with magnesium stearate and Filomatic speed.

† Significantly better than 19 input network $P < .05$

‡ Significantly worse than 19 input network $P < .01$

In order to see if generalization performance could be improved, inputs were removed from consideration. Initially, 6 inputs were eliminated. Those eliminated had been rejected at the unit weight penalty of zero, by both the backward stepwise and genetic algorithms (namely, percentage of Cel50, percentage of methocel K15M, percentage of methocel K100LV, percentage of aerosil, Filomatic speed, and tablet weight). Training was then carried out using the remaining 13 inputs. This training produced error values that were similar to or, in some cases, significantly lower than those obtained by using all 19 inputs, showing that the removal of unimportant inputs had enabled the ANN to generalize better.

In order to see whether or not the elimination of more inputs could lead to a further improvement, another 3 inputs were removed. These were the inputs rejected by the backward stepwise and genetic algorithms at a unit weight penalty of 0.001 (namely, percentage Eudragit L coating, tablet press speed, and tablet hardness). In general however, when training was carried out using the remaining 10 inputs, performance was not significantly improved compared with using 13 inputs. A plot of actual versus estimated values for T_{lag} is shown in **Figure 2**.

The ANN did not perform as well during the training of the $T_{90\text{-}10}$ output using all 19 inputs, reflecting the greater difficulty of this problem with both the training and verification errors being much higher than those obtained for the T_{lag} output (**Table 5**). The greater error is also evident from a comparison of **Figures 2** and **3**. In the latter, the actual values of $T_{90\text{-}10}$ are plotted against the ANN estimates.

Once again, less relevant inputs as detected by IFS were removed to see if network error could be lowered. Initially the 5 inputs, selected as unimportant by all 3 algorithms at a unit weight penalty of zero (ie, percentage of methocel K15M, percentage of methocel K100LV, percentage of aerosil, tablet hardness, and tablet weight) were discarded. When networks were

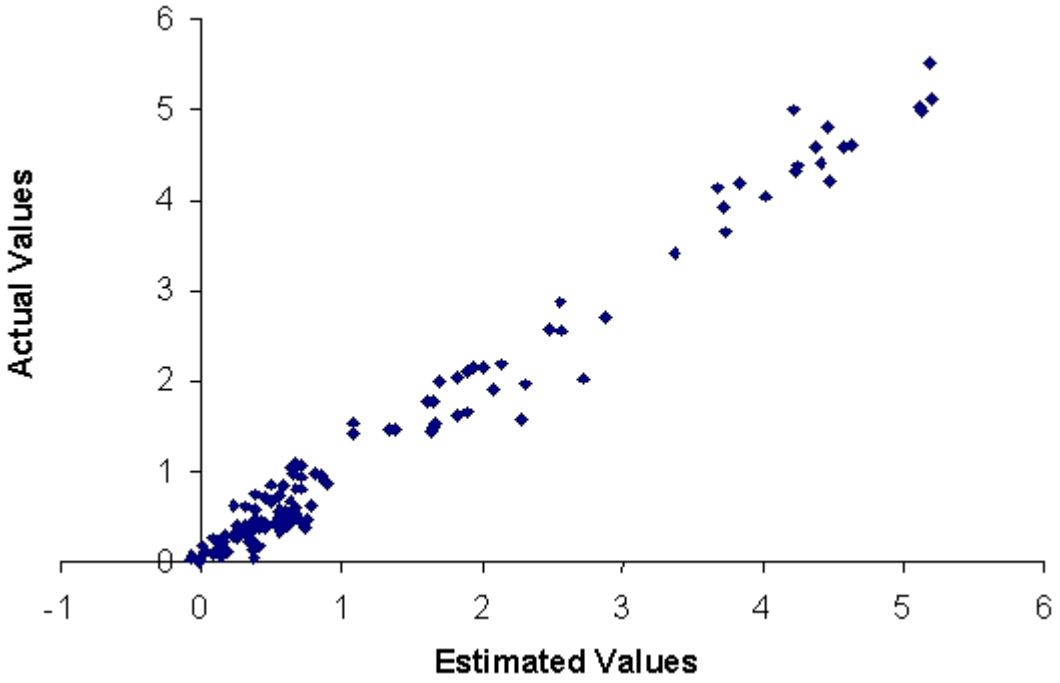

Figure 2. Plot of actual values against estimated values for the T_{lag} output.

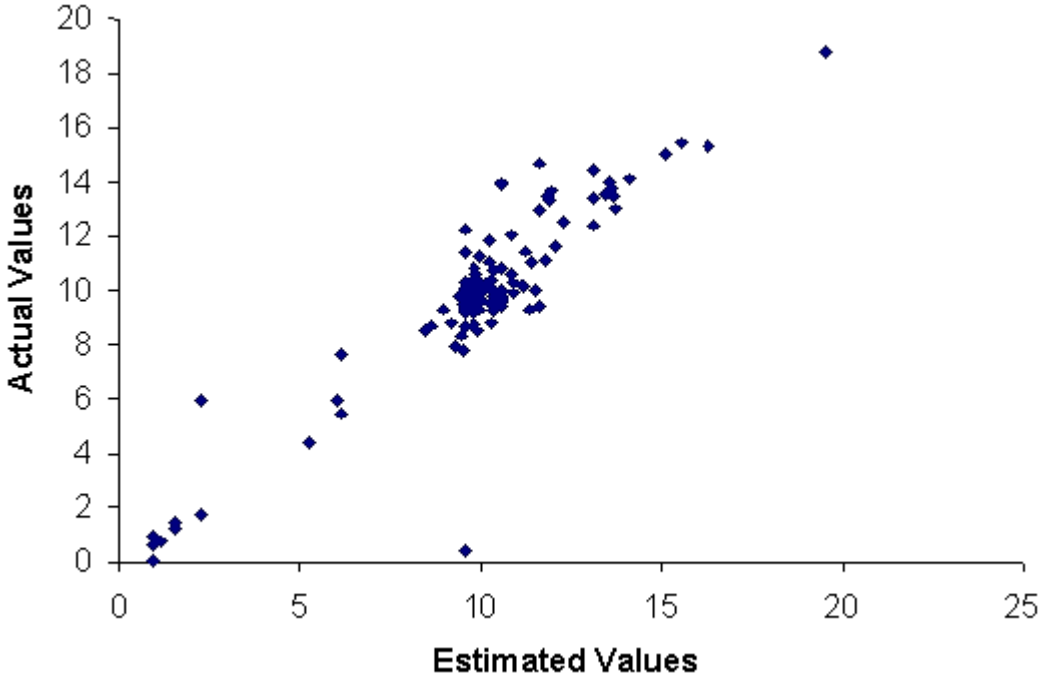

Figure 3. Plot of actual values against estimated values for the T_{90-10} output.

trained with the remaining 14 inputs, the training and verification errors were either similar or, in some cases, significantly lower than when all 19 inputs were used.

An additional 2 inputs were removed to see if further improvement could be obtained. These inputs were the additional pair rejected at the unit weight penalty of zero by the forward stepwise algorithm, namely, blend time with magnesium stearate and Filomatic speed.

This change did not result in much further improvement however. Indeed, some of the training errors were significantly worse. This indicated that inputs necessary for the training of the ANN had been removed.

CONCLUSION

With regard to input feature selection, all 3 algorithms performed well and gave consistent rankings for the importance of inputs. In the case of the T_{lag} output, the main coating parameters were highlighted, as would be expected. Less obviously, the blend size of the original batch was shown to have some influence. Most other processing parameters were shown to be relatively unimportant with the exception of the blend time after the addition of magnesium stearate.

In the case of the T_{90-10} output release rate parameter, the percentage of matrix-forming polymer along with the amount of direct compression base was classed as most important. Coating parameters were still seen as having some influence, and processing parameters were once more classified as being less important. All algorithms were able to distinguish and reject clearly redundant inputs.

The ANN was more successfully able to train networks with low errors when the T_{lag} output was used compared with the very high errors seen with the T_{90-10} output. The factors affecting the release of the drug from the SR matrix (represented by the T_{90-10} output) are evidently more numerous and complex than those governing the lag time from coated tablets (represented by the T_{lag} output).

Removal of the 6 least important inputs as determined by input feature selection for the T_{lag} output led to an overall improvement in both the training and verification errors obtained. This improvement indicates that although removal of less important inputs may lead to the loss of some information, the increased simplicity can lead to better generalization ability by the ANN. When an additional 3 inputs were eliminated, no further significant improvement was seen.

Removal of inputs also changed the performance of the ANN in training for the T_{90-10} output. The discarding of 5 inputs gave a marked overall improvement, whereas removing an additional 2 led to signs of deterioration. Because of the high errors seen with the T_{90-10} output, it is likely that this problem was too complex for the ANN to model accurately. A larger data set may therefore be necessary to successfully model this release phase.

The techniques employed above show promise, particularly if applied to large datasets. They are capable of extracting valuable information from historical formulation data previously considered of little importance. Knowledge of critical formulation variables would be of considerable benefit to the formulation scientist at the development stage.

In addition, data from this project could be combined with data generated from other projects using similar water-soluble drugs and thus extend the scope of these ANNs. Once such large databases are in existence they can serve as a form of "institutional memory" aiding in the selection of suitable formulations and accelerating the development process.

REFERENCES

1. Jha BK, Tambe SS, Kulkarni BD. Estimating diffusion coefficients of a micellar system using an ANN. J Coloid Interface Sci. 1995;170:392-398.

2. Bourquin J, Schmidli H, Van Hoogevest P, Leuenberger H. Basic concepts of ANN modeling in the application to pharmaceutical development. Pharm Dev Technol. 1997;2(2):95-109, 111-121.

3. Erb RJ. Introduction to back propagation NN computing. Pharm Res. 1993;10(2):165-170.

4. Hussain AS, Yu X, Johnson RD. Application of neural computing in pharmaceutical product development. Pharm Res. 1991;8(10):1248-1252.

5. Hussain AS, Johnson RD, Vacharajani NN, Ritschel WA. Feasibility of developing an NN for prediction of human pharmacokinetic parameters from animal data. Pharm Res. 1993;10(3):466-469.

6. Hussain AS, Shivanand P, Johnson RD. Application of neural computing in pharmaceutical product development: computer aided formulation design. Drug Dev Ind Pharm. 1994;20(10):1739-1752.

7. Chen Y, McCall T, Baichwal A, Meyer M. The application of an ANN and pharmacokinetic simulations in the design of CR dosage forms. J Control Release. 1999;59:33-41.

8. Ebube NK, McCall T, Chen Y, Meyer M. Relating formulation variables to in vitro dissolution using an ANN. Pharm Dev Technol. 1997;2(3):225-232.

9. Murtoniemi E, Merkku P, Kinnunen P, Leiviska K, Yliruusi J. Effect of NN topology and training end-point in modeling the fluidised bed granulation process. Int J Pharm. 1994;110:101-108.

10. Murtoniemi E, Yliruusi J, Kinnunen P, Merkku P, Leiviska K. The advantages by the use of NN in modeling the fluidised bed granulation process. Int J Pharm. 1994;108:155-164.

11. Turkoglu M, Ozarslan R, Sakr A. ANN analysis of a direct compression tabletting study. Eur J Pharm Biopharm. 1995;41(5):315-322.

12. Kesavan JG, Peck GE. Pharmaceutical granulation and tablet formulation using NN. Pharm Dev Technol. 1996;1(4):391-404.

13. Bishop CM. Neural Networks for Pattern Recognition. Oxford, England: Clarendon Press; 1995.

14. Ledwidge MT, Corrigan OI. Effects of surface active characteristics and solid state forms on the pH solubility profiles of drug-salt systems. Int J Pharm. 1998;174:187-200.

15. Trajan Software Ltd. Trajan Users Manual. Washington, Tyne and Wear, UK:Trajan Software Ltd; 1999.

16. Rowe RC, Colbourn EA. Applications of neural computing in formulation. Pharmaceutical Visions. 2002;spring:4-8.

17. Van der Watt JG. The effect of particle size of microcrystalline cellulose on tablet properties in mixtures with magnesium stearate. Int J Pharm. 1987;36:51-54.

18. Billany MR, Richards JH. Batch variation of magnesium stearate and its effect on the dissolution rate of salicylic acid from solid dosage forms. Drug Dev Ind Pharm. 1982;8(4):497-511.