

Using Neural Nets for Decision Support in Prescription and Outcome Prediction in Anticoagulation Drug Therapy

S. Byrne*, P. Cunningham**, A. Barry***, I. Graham****, T. Delaney***, O.I. Corrigan*.

*Department of Pharmaceutics,

**Department of Computer Science,
Trinity College, Dublin 2.

***Pharmacy Department,

****Cardiology Department,
The Adelaide and Meath Hospitals
incorporating the National Children's Hospital
Tallaght, Dublin 24.

Abstract

In this paper we consider the use of Artificial Neural Networks (ANNs) in decision support in anticoagulation drug therapy. In this problem domain ANNs can be used to learn the prescribing behaviour of expert physicians or alternatively to learn the outcomes associated with such decisions. Both these possibilities are evaluated and we show how, through the prediction of outcomes that the prescribing practice of physicians may be improved – the ANN can learn to predict the outcome resulting from prescribing levels better than expert physicians. In evaluating the use of ANNs on these regression problems the possibility of *ensembling* these nets to improve performance is also considered. We show that ensembles of networks do not improve over individual networks, presumably because of the lack of diversity in the outputs of the ANNs.

1. Introduction

Artificial Neural Networks (ANNs) have the ability to discern complex and latent patterns in the data presented to them. This attribute of ANN makes them powerful tools for modelling and predictive purposes. Consequently, neural network technology is finding increasing uses in medicine as medical decision support systems. Maclin *et al.* reported the use of ANN to aid in the diagnosis of renal and hepatic carcinomas.^{1,2} Several other clinical applications have also been reported e.g. in the treatment of cardiac diseases³ and to predict patient gentamicin and cyclosporin peak and trough blood concentrations.^{4,5}

In this paper we present some results of a study on the use of ANNs in anticoagulant drug therapy. In such a situation there are various possible rôles for Machine Learning (ML) in decision support. The ML system may learn the decision-making behaviour of an expert as shown by the arrow at A in Figure 1. In a financial loan approval system this would involve learning some of the skills of the underwriter; in our system this involves learning the prescribing patterns of the clinician. In this scenario the learning system may achieve the competence of the domain expert but it will never surpass the competence of the human expert. Thus, for this scenario to be successful it requires skilled domain experts from which to learn.

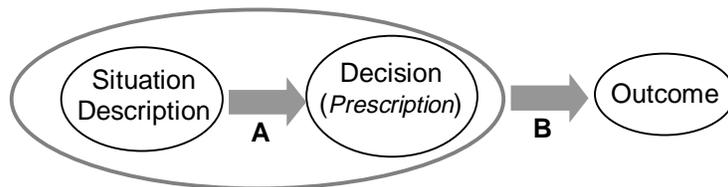


Figure 1. Rôles for Machine Learning systems in decision support.

In many ways, the more exciting prospect is for the ML system to learn likely outcomes associated with specific scenarios (option B in Figure 1). In the loan approval scenario this would involve learning the kinds of loans that turn into bad debts; or in our scenario, the medical outcomes associated with prescribing decisions. In this paper we evaluate both of these approaches in the anticoagulant therapy domain.

Anticoagulant drug therapy inhibits or delays coagulation of the blood. The decision to anticoagulate a patient is made on an individual basis, taking into account both the potential benefits and risks. Warfarin has a very narrow therapeutic index and therefore, for safe rational anticoagulant therapy, an understanding of the interrelationships, which influence an individual's response to the drug, is essential. Several studies have shown that satisfactory clinical control of patients on anticoagulant therapy was maintained in approximately half of the patients at any given time.^{6,7} In order to improve the quality of patient care, we consider that there is a role for the use of ANNs to help in the prediction of warfarin dosage adjustment. We present results on the use of single ANNs and ensembles of ANNs in learning appropriate prescribing levels and also results on predicting outcomes associated with such decisions. The ensembling of ANNs is a *committee of experts* idea where the aggregated results of several networks is better than the results of a single network. The methods and results are presented in the next sections and the paper concludes with an assessment of the results and a discussion on issues of bias in the predictions of the ANNs.

2. Method

The Adelaide and Meath Hospitals in Dublin, incorporating the National Children's Hospital was identified in previous work as the first hospital in the Rep. of Ireland with a pharmacist involved in it's anticoagulant clinic.⁸ Ethics committee approval for the study was obtained and a researcher attended the outpatient department of the Tallaght Hospital twice weekly when the anticoagulant clinic was in operation between October 1998 and July 1999. 163 patients were approached and asked to take part in the study. 105 patients gave written consent and by doing so, agreed to undergo a structured interview with the researcher every time they attended the clinic and to have additional liver function tests (LFTs) carried out during the study period.

After an initial literature review 25 parameters were measured, these were age, weight, height, gender, marital status, reason for anticoagulation, target INR (International Normalised Ratio) range, measured INR, previous and subsequent dose of warfarin, duration of therapy, number of hours since last warfarin dose, number of adverse drug reactions, concurrent prescribed and over the counter medication, LFT results, smoking habits, alcohol consumption and compliance to prescribed medication.

As outlined in the introduction, work to date has concentrated on training neural networks to learn a patient's dose of warfarin prescribed by clinicians and also on predicting a patient's INR reading given a particular dose. The INR reading is a international standardised method of reporting a patient's prothrombin time (the time it takes for the patient's blood to clot). Initial work using an in-house neural network package concentrated on developing optimum

networks for the above output parameters i.e. the number of hidden units, the learning rates and momentum variables were adjusted until the error of the networks was minimised. The training algorithm used on each occasion was the backpropagation algorithm.

3. Results and Discussion

Two patients decided to drop out of the study, leaving a total of 103 patients in the study population. Data was collected from the 103 patients on 656 occasions, mean=6.5 (S.D.±2.7). This corresponds to 49.89 patient treatment years (PTY) of anticoagulation therapy. The population had a mean age of 57.2yrs, mean height of 171.5cm and a mean weight of 80.1Kg. The main indication for anticoagulation was atrial fibrillation 40.8%(n=42). Overall, patients spent 21.56 PTY outside their desired INR ranges and for 14.44 (67.0%) of these PTY, the patients were under anticoagulated. This difficulty in control indicated a rôle for ML based decision support systems.

3.1. Learning Prescription Practice

The data collected was firstly used to train a neural network with a patient's dose as the desired output variable – the objective being to learn the practice of expert clinicians in this area. The input layer to the network had 23 nodes and the output layer had 1 node. The network had one hidden layer and the number of nodes within this layer was varied, along with the learning and momentum rates, to produce a model that best fitted validation data. The neural network structure that gave the best results had 1 node in the hidden layer and learning and momentum rates equal to 0.1. Using 10-fold cross validation on the 656 data samples available we obtained an estimation of generalisation error of 6.6%. That is, the network should be able to prescribe doses to within 6.6% of that of an expert clinician.

In addition an ensemble of 5 neural nets was built and tested on 101 data points held back from the training. The output from the ensemble was produced by simply averaging the outputs from the individual nets in the ensemble for each test point. This produced an average error of 6.24% on the 101 test data points. (This figure is not directly comparable to the 6.6% figure which is derived from cross validation.) In a regression task such as this the error reduction due to the ensemble is expressed as:

$$E = \overline{E} - \overline{A} \quad (1)$$

where E is the overall error of the ensemble, \overline{E} is the average generalisation error of the ensemble components and \overline{A} is the ensemble diversity (i.e. variance in individual predictions). In this situation there is little scope for the ensemble to reduce the error because it is already small and more importantly there is little diversity (ambiguity) among the ensemble members (see Cunningham & Carney, 2000 for more details on this).⁹

Subsequently it was decided to design and train another neural network with a patient's dose as the desired output variable, but in this case the LFT data was to be omitted from the input variables (LFT data would not normally be available to the clinicians). Therefore the number of input parameters was reduced to 17 and similarly the neural network structure that gave the best error measurements had 1 node in the hidden layer and learning and momentum rates of 0.1. A 10-fold cross validation test was performed to produce an estimate of error and a figure of 6.1% was obtained. This result is surprising because it indicates that LFT data does not inform the prescribing process. Again, an identical ensemble exercise as above was conducted and produced an error of 4.3% on the 101 test samples. Since this scenario with the LFT data excluded is the more realistic the results are shown in more detail in Figure 2. It

can be seen from this graph that the predictions across all 101 cases are good and the average error of 4.3% does not hide any individual large errors.

If we compare the average error of this ensemble with its component nets we see that it represents a very slight improvement. This improvement is shown in Figure 3 where the range of errors for the component nets is also shown. Clearly the advantage of the ensemble is not material in reducing the error; it is significantly better than the worst component nets but not better than the best net in the ensemble.

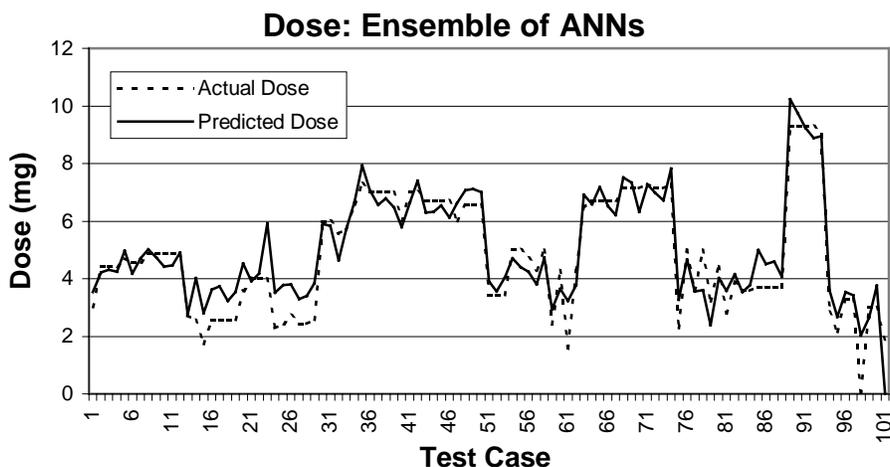


Figure 2. Results of an Ensemble of 5 ANNs in determining dose for 101 test cases.

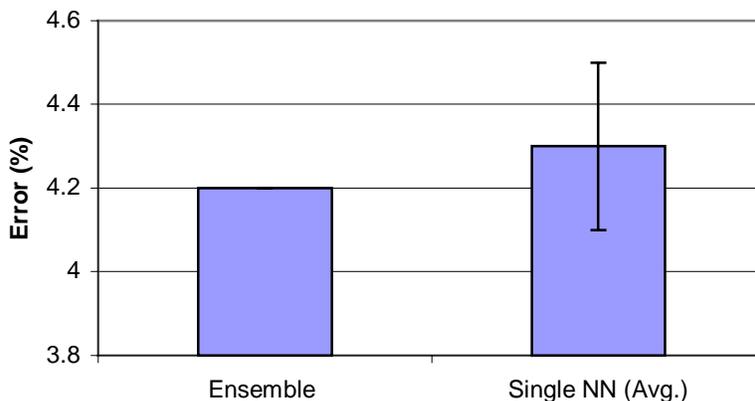


Figure 3. Comparison of error of ensemble of 5 ANNs with that of the component networks.

3.2. Predicting Outcomes

In order to illustrate ANNs flexibility it was decided to design and train a network using a patient’s INR measurement as the desired output variable. The input layer had 22 nodes and the output layer had 1 node. The neural network structure that gave the minimum error values had 4 nodes in the hidden layer and learning and momentum rates equal to 0.1. This more elaborate network structure suggests that this process is more complex than the prescribing process described above. 10-fold cross validation indicated a generalisation error of 8.6% and again an ensemble of 5 networks did not produce an appreciable error reduction on its component networks (8.8% for the ensemble and 8.9% on average for the component

networks). Given the higher error here it might be expected that the ensemble would have a greater benefit however it appears that there is little diversity in the component networks. This in turn suggests that the amount of training data available covers the problem domain well (see Cunningham, Carney & Jacob, 2000 for related discussion on this issue).¹⁰

It is important to assess the quality of this prediction of INR outcome because it offers the potential to improve the current prescribing practice. With this objective the target INR that the physician had in mind was also recorded in the data gathering process. For the given dose this is the INR that the physician expects to achieve, so we can compare this directly to the INR prediction produced by the ANN.

Figure 4 below shows the mean physicians' INR predictions and the neural networks mean INR predictions compared to the actual INR values for a random sample of 101 meetings with the study population. The physicians' INR prediction was on average 1.05 INR units away from the actual INR value and the neural network was 0.75 INR units away from the actual value. Using a paired sample t-test, one can conclude that the network's prediction was significantly closer to the actual patient's INR measurement than that of the physician ($p < 0.00014$). This offers the prospect of developing a tool that will predict the outcome for a given dose more accurately than at present. This would help the physician to refine his/her prescribing to better target the desired outcome.

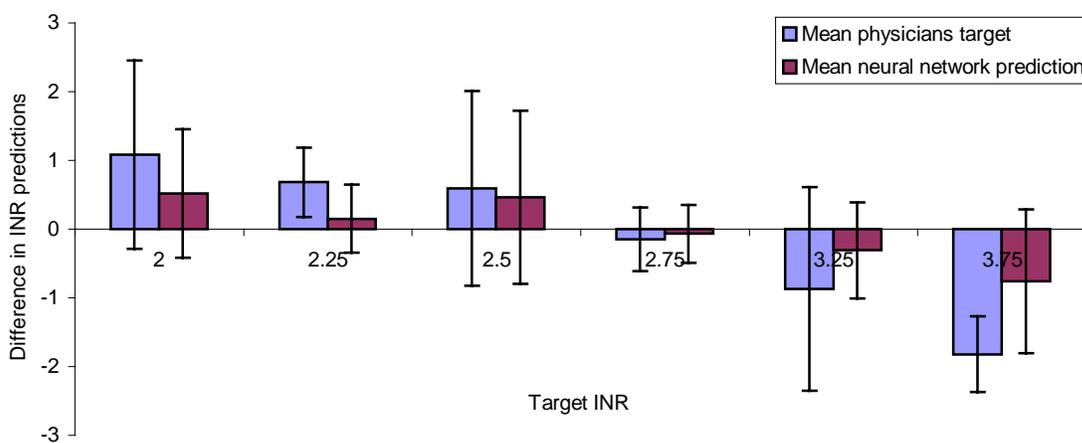


Figure 4. Plot of differences between the physician's INR prediction and the desired INR value and the neural networks INR prediction and the desired INR value.

In Figure 4 the results are divided into intervals according to target INR outcome. The predictions for physician and ANN are most accurate for INR values around 2.75. Above that the ANN and physician underestimate and below that there are overestimations. If we look at the actual distributions of INR in the training data we can see the reason for this (see Figure 5). The distribution of INR values in the training data is skewed with a large number of samples around 2.25 to 2.75. It seems that ANNs trained with this *a priori* distribution are biased toward these values. INR values around 2 are poorly represented in the training data and these outcomes in the test data are overestimated towards the mean of the training data. In turn high INR values are also poorly represented in the training data and in turn test cases in this region are underestimated.

Distribution of INR in Training Data

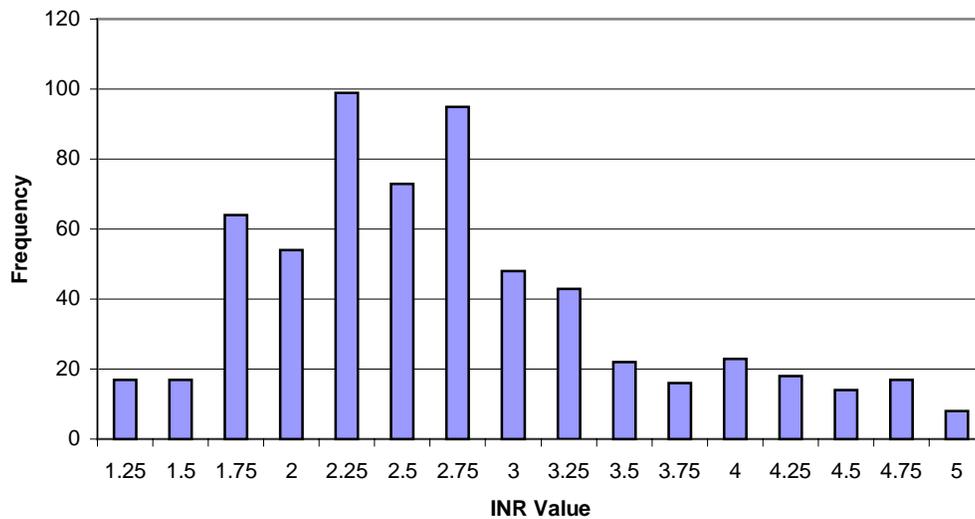


Figure 5. Distribution of INR values in training data.

This is a manifestation of a general problem of skewed distributions in machine learning. This problem is well documented in classification problems (Provost & Fawcett, 2000)¹¹ where it is well known that classifiers are biased away from minority classes. What we see here is a regression situation where the ANN is biased toward commonly occurring values.

4. Conclusions

The evaluation presented here shows that ANNs can be used to learn the prescribing patterns of expert physicians in anticoagulant drug therapy to within an accuracy of 6%. Further, ANNs can be used to predict the outcome of prescribing decisions. The evaluation shows that these predictions of outcome are about 30% more accurate than those of physicians. This offers the possibility of improving the performance of expert physicians by using these accurate predictions of outcome to guide and refine prescribing decisions.

We also evaluated the use of ensembles of ANNs to improve on the error of an individual net on these tasks. No significant improvement was found presumably because of the lack of diversity amongst individual nets. We speculate that this suggests that the 656 training samples used in this study provide good coverage of the problem domain. A good model that does not overfit to training data can be built with just one ANN.

It emerged in the evaluation that the accuracy on predicting the outcome (INR level) was not uniform across the distribution of outcomes. It appears that this is due to the skewed distribution of INR outcomes. An analysis of the errors across the distribution of INR values shows that there is a bias towards the dominant values. Correcting this bias warrants further investigation and is a promising area for future research.

Reference

1. Maclin P.S., Dempsey J., Brooks J. and Rand J., Using neural networks to diagnose cancer, *J Med Syst* 1991, **15** (1), 11-19

2. Maclin P.S., Dempsey J., Using an artificial neural network to diagnose hepatic masses. *J Med Syst* 1992, **16** (5), 215-25
3. Furlong J.W., Dupuy M.E., and Heinsimer J.A., Neural network analysis of serial cardiac enzymes data. A clinical application of artificial machine intelligence, *Am J Clin Pathol*, 1991, **96** (1), 134-41
4. Brier M.E., Neural network gentamicin peak and trough predictions from prior dosing data, *Clin Pharm and Ther*, 1995, **57** (2), 157
5. Brier M.E., Neural network predictions of cyclosporin blood concentrations, *Clin Pharm and Ther*, 1995, **57** (2), 157
6. Duxbury B. McD., Therapeutic control of anticoagulant treatment, *BMJ*, 1982, **284**, 702-704
7. Radley A.S., Hall J., Farrow M. and Carey P.J., Evaluation of Anticoagulant Control in a Pharmacist Operated Anticoagulant Clinic, *J Clin Pathol* 1995,**48**,545-547
8. Byrne S., Corrigan O.I., An overview of pharmaceutical manpower and services in hospitals situated in the republic of Ireland during 1998, *IPJ*, 1999, **77** (7), 220-25
9. Cunningham, P., Carney, J., Diversity versus Quality in Classification Ensembles based on Feature Selection, *accepted for presentation at the 11th European Conference on Machine Learning (ECML 2000)*, Barcelona, Spain, May 2000.
10. Cunningham, P., Carney, J., Jacob, S., Stability Problems with Artificial Neural Networks and the Ensemble Solution, to appear in *AI in Medicine*, 2000, Vol. **20**.
11. Provost, F., T. Fawcett, Robust Classification for Imprecise Environments, to appear in *Machine Learning*, 2000 (available at <http://www.stern.nyu.edu/~fprovost/>).