Poster Number

697

11th Conference on Retroviruses and Opportunistic Infections
February 8-11, 2004
San Francisc, CA, USA

Accuracy of Neural Network Models in Predicting HIV Treatment Response from Genotype May Depend on Diversity As Well As Size of Data Sets

B.A. Larder[1], D Wang[1], A Revell[2], R Harrigan[3], J Montaner[3] and C. Lane[4].

1: The HIV Resistance Response Database Initiative (RDI), Cambridge, UK; 2: RDI Ltd, London, UK; 3: The BC Centre for Excellence in HIV/AIDS Vancouver, Canada; 4: National Institute of Allergy and Infectious Diseases (NIAID), Bethesda, MD, USA.

RDI Ltd
14 Union Square
London N1 7DH
UK
Tel: +44 207 226 7314
Email: Info@hivrdi.org

Introduction

A major cause of HIV treatment failure continues to be viral drug resistance. In response, resistance testing is in widespread use as part of standard HIV clinical practice. The two main approaches to resistance testing are genotyping and phenotyping. Whilst both have been shown to help in the selection of effective combination treatment regimens following the development of drug resistance, it is generally accepted that their utility is limited by difficulties in interpretation and application to clinical reality.

Genotyping identifies point mutations in the relevant regions of the viral genome, which then have to be interpreted in terms of their potential impact on response to different therapeutic options. With over 200 individual mutations known to affect drug susceptibility and with more than 20 approved antiretroviral drugs in common use, this represents a formidable challenge. The most common approach is to apply sets of rules that relate individual genetic changes to changes in the phenotypic sensitivity of the virus to individual drugs. There are many such sets of rules in existence, employed by different organisations and comparative studies have demonstrated that different interpretation systems can produce somewhat different results. The virtual phenotype approach matches genotypes directly to existing phenotypes stored in a database. This overcomes some of these interpretation difficulties but still does not resolve the difficulty of relating phenotype to clinical response. For example, how much of a change in phenotypic susceptibility is clinically significant? How can changes in phenotypic susceptibility of individual drugs in the laboratory be related to combinations of drugs in the clinic. It is generally agreed that the development of an accurate and reliable method to predict virological response to combination therapy directly from genotype would be a major advantage.

The RDI is an international collaboration of scientists and clinicians whose aim it is to develop a new approach to the interpretation of HIV genotype, relating genotypic mutation data *directly* to virological response to combination therapy using artificial intelligence in conjunction with a substantial clinical database. The technique currently used by the RDI to model the relationship between genotype and treatment response is artificial neural networks (ANN). This method imitates the way a human brain works by creating connections between processing elements, the computer equivalent of neurons. The organization and weights of the different connections determine the output. ANN are good pattern recognition engines and robust classifiers, with the ability to generalize in making decisions about imprecise input data.

Methods

The RDI's current methodology is to train ANN using a large dataset of input variables (baseline viral load, mutations in the protease and reverse transcriptase coding regions of the HIV-1 genome and treatment details) and output (drug response) data. Three layer (one hidden) ANN are used with full interconnectivity between input nodes and hidden nodes and from hidden nodes to the output nodes. The models are trained using back-propagation. The models are then tested on an independent set of input data and the output predictions of virological response to therapy compared with the actual virological response data as a measure of the model's accuracy.

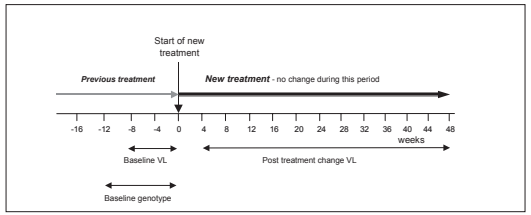
In this study, 228 'treatment change episodes' (TCEs) were obtained from a clinical cohort managed under the auspices of the US National Institute of Allergy and Infectious Diseases (NIAID). TCEs were classified as episodes with a genotype ≤12 weeks and a viral load measure (VL) ≤ 8 weeks before a change to the patient's antiretroviral treatment regimen, and follow-up VL within 4-40 weeks. Four ANN models were developed. In each case the 228 TCEs were partitioned into 171 TCE's which were used to train the ANN and 57 TCE's which were kept back for testing. Results were compared to those obtained with 13 ANN models developed using the NIAID data plus data from a clinical cohort managed by the BC Centre for Excellence in HIV/AIDS (BC), yielding 747 TCE's. The data were finally added to data from multiple sources in the RDI database, yielding 1,581 TCEs and 10 ANN models trained, and tested. All the training and testing sub-datasets were independent, randomly partitioned and normalised.

Methods (continued)

There were 65 input variables for the models: 49 different primary resistance mutations, identified as occurring in ≥1% of the dataset; 14 different antiretroviral drugs involved in ≥1% of the TCEs (both mutations and drugs were used as binary input variables); baseline viral load as log₁₀ copies/ml was used as a single continuous input variable; time to follow-up in days was used as a single continuous input variable and viral load at follow up was the single continuous output variable.

The ANN models were evaluated in terms of the percentage correct prediction of the viral load trajectory at follow-up for the test data and the correlation between predicted and actual absolute follow-up viral load and change from baseline viral load for the test data.

Figure 1: The Treatment Change Episode (TCE)



Results

ANN models developed using NIAID data

The correlation between the predicted and actual follow-up VL for the four ANN models developed from the NIAID cohort data gave a mean r^2 value of 0.71. The mean correlation between predicted and actual change in viral load was 0.65 and the mean correct trajectory prediction rate was 76% (Table 1). The best performing model was Model 2, with a correlation between predicted and actual VL of 0.79 (ΔVL r^2 = 0.73, Figure 1)

Table 1: Test performance of ANN models developed using NIAID data

	Model 1	Model 2	Model 3	Model 4	Mean
Trajectory	70%	81%	77%	75%	76%
r^2 VL	0.59	0.79	0.71	0.75	0.71
r^2 ΔVL	0.40	0.73	0.71	0.77	0.65

Models developed using NIAID and BC data

The 13 ANN models from the combined NIAID and BC cohorts gave a mean r^2 value for absolute follow up viral load of 0.65 and mean correct trajectory prediction rate of 78% (Table 2). The best model overall was Model 7 with a correlation between predicted and actual follow-up viral load of 0.79 (ΔVL r^2 = 0.73, Figure 3).

Results (continued)

Figure 2: Scatter plot of predicted *vs* actual ΔVL for ANN Model 2 (NIAID data)

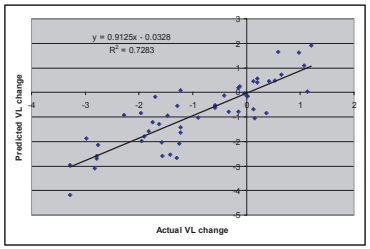


Figure 3: Scatter plot of predicted *vs* actual ΔVL for ANN Model 7 (NIAID + BC data)

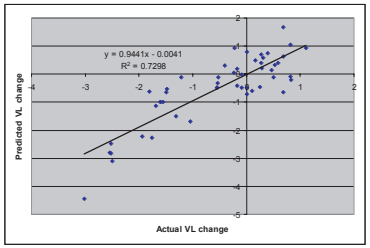
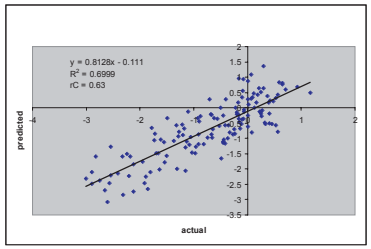


Figure 4: Scatter plot of predicted *vs* actual ΔVL for ANN Model 9 (combined dataset)



Results (continued)

Table 2: Test performance ANN models developed using NIAID + BC data

Model	1	2	3	4	5	6	7	8	9	10	11	12	13	Mean
Trajectory	81%	74%	84%	84%	72%	84%	72%	72%	81%	77%	83%	79%	73%	78%
r^2 VL	0.38	0.76	0.62	0.64	0.66	0.68	0.79	0.64	0.70	0.62	0.53	0.63	0.78	0.65
r^2 ΔVL	0.35	0.76	0.57	0.62	0.60	0.69	0.73	0.59	0.65	0.62	0.60	0.46	0.76	0.62

Models developed using combined dataset

The 10 ANN models from the combined dataset gave a mean r^2 value for absolute follow up viral load and for change in viral load of 0.55. The mean correct trajectory prediction rate was 74% (Table 3). The best performing model overall was Model 9 with a correlation between predicted and actual viral load of 0.74 (ΔVL, r^2 = 0.74, Figure 4).

Table 3: Performance ANN models developed using combined dataset

Model	1	2	3	4	5	6	7	8	9	10	Mean
Trajectory	77%	81%	67%	71%	73%	75%	73%	79%	72%	74%	74%
R^2 VL	0.52	0.50	0.61	0.63	0.52	0.50	0.49	0.46	0.74	0.58	0.55
R^2 ΔVL	0.50	0.50	0.59	0.60	0.51	0.52	0.48	0.49	0.70	0.57	0.55

Paired comparisons using t-tests revealed that the correlation between predicted and actual absolute viral load was significantly greater for the ANN models developed using NIAID data than those using the combined dataset (p=0.019). There were no other significant differences in the performance of the models.

A simple distance measure was calculated for the NIH and the combined datasets as a measure of their diversity. This was performed by calculating the distance from the origin of each TCE in n-dimensional space, where n=the number of input parameters and where each drug used and each mutation found = 1 and the baseline viral load and time to follow-up values are normalised between 0 and 1. This simple measure revealed that the mean distance measure for the NIAID dataset (1.95, SD=0.38) was less than the mean distance measure for the remainder of the combined RDI dataset (2.68, SD=0.84). This suggests that adding data to the NIAID dataset increases the diversity of the dataset and made it more difficult for ANN to model the outcomes.

Conclusions

- ANN models developed using limited data from a single clinical centre can be surprisingly accurate
- Models using larger data sets form multiple centres are not necessarily more accurate
- The accuracy of ANN models may depend not only on the size but the relative diversity of training and test data sets
- Our results underline the potential of ANN models as useful aids to treatment decision making

Acknowledgements

The RDI gratefully acknowledges the support of the BC Centre for Excellence in HIV/AIDS and ICONA for their support as Founding Academic Members

The RDI is also grateful to the following for their additional support:
National Institute of Allergy and Infectious Diseases
Gilead Sciences Inc (Founding Corporate Supporter)
Boehringer Ingelheim Inc
The Hospital Clinic of Barcelona
HIV and Hepatitis.com