Poster Number

157

XV International HIV Drug Resistance Workshop
13-17 June 2006
Sitges, Spain

Accurate prediction of virologic response to HAART using three computational modelling techniques

B.A. Larder[1], D. Wang[1], F. De Wolf[2], J. Lange[3], A. Revell[1], S. Wegner[4], J. Montaner[5], R. Harrigan[5], J.A. Metcalf[6], and H. C. Lane[6]

1: RDI, London, UK; 2: Netherlands HIV Monitoring Foundation, Amsterdam, the Netherlands; 3: Academic Medical Centre of the University of Amsterdam, the Netherlands; 4: US Military HIV Research Program, Rockville, USA; 5: BC Centre for Excellence in HIV/AIDS, Vancouver, Canada; 6: National Institute of Allergy and Infectious Diseases (NIAID), Bethesda, MD, USA.

hiv resistance response database initiative
A global non-profit initiative against HIV drug resistance

www.hivrdi.org
email: info@hivrdi.org

Introduction

- Virologic response to different highly active antiretroviral therapy (HAART) combinations is highly complex and dependent on many factors.
- Accurate quantitative prediction of response to optimize HAART selection is likely to require sophisticated modelling.
- The RDI has demonstrated that artificial neural networks (ANN) can predict response to combination therapy from genotype, viral load, CD4 count and treatment history.
- Performance of ANN has been more accurate for patients from clinics represented in the training datasets than for patients from unfamiliar clinics.
- Alternative high level modelling methods such as Random Forests (RF) and Support Vector Machines (SVM) are resistant to over-fitting and may be more generalisable that ANN.
- Here we compare the performance of ANN with that of RF and SVM using independent tests sets from the RDI database and unfamiliar clinics.

Methods

- 'Committees' of 10 ANN, RF and SVM models were trained to predict virologic response (ΔVL) using 76 input variables (55 resistance mutations in RT and protease, antiretroviral drugs, baseline viral load, baseline CD4 count, treatment history and time to follow-up) from 1,154 treatment change episodes (TCEs) from the RDI database (from a large number of clinical sources).
- These models were then tested using the input data from two independent tests sets:
  1. 50 TCEs involving HAART (≥3 full-dose drugs) selected at random from the RDI database (different patients from the training set)
  2. 50 TCEs invoplving HAART from clinics in the Netherlands (ATHENA database) without data in the training set.
- The committee average predictions (the predictions of ΔVL for each TCE averaged across all ten models in a committee) were then compared with the actual ΔVL values in terms of:
  1. The correlation between predicted and actual ΔVL values
  2. The mean absolute difference between the predicted and actual ΔVL values across the 50 test TCEs.
- The outputs of the three methods were then combined for each test TCE in three alternative ways:
  1. The mean of the three committees' predictions
  2. Combined in proportion to the inverse of the variance of the scores (differences from actual ΔVL values)
  3. Combined in proportion to the correlations between predicted and actual ΔVL values
- The combined predictions were also compared with actual ΔVL values.

Results

RDI test set

- Correlations between predicted and actual ΔVL for the RDI test set gave r^2 values of 0.69, 0.71 and 0.61 for ANN, RF and SVM respectively (Figure 1).
- The mean absolute differences between predicted and actual ΔVL were 0.54, 0.60 and 0.61 for ANN, RF and SVM respectively
- Combining the committee average predictions of the three methods produced an r^2 value of 0.73 and a mean absolute difference value of 0.55, regardless of the method of combination.

ATHENA test set

- The correlations between predicted and actual ΔVL gave r^2 values of 0.46, 0.47 and 0.48 for ANN, RF and SVM respectively (Figure 2).
- The mean absolute difference scores were 0.71, 0.69 and 0.73.
- The combined outputs produced r^2 values of 0.52-0.53 and absolute difference scores of 0.62-0.63.
- There were no statistically significant differences between the three modelling methods and the combined output for either test set (ANOVA).
- Predictions for the RDI test set were significantly more accurate than for the ATHENA test set with all three methods ($p<0.01$).

Table 1: Summary of modelling results

	RDI test set		ATHENA test set	
	r^2	Absolute difference	r^2	Absolute difference
ANN	0.69	0.54	0.46	0.71
RF	0.71	0.60	0.47	0.69
SVM	0.61	0.61	0.48	0.73
Combined (mean)	0.73	0.55	0.52	0.63
Combined (1/variance)	0.73	0.55	0.52	0.62
Combined (r^2)	0.73	0.55	0.53	0.62

- Examination of the performance of the individual models within each of the three committees revealed that the correlations between predicted and actual ΔVL values were significantly lower for the ANN models (r^2=0.32-0.55) compared with RF (0.59-0.75) and SVM (0.30-0.72) (RDI test set, $p<0.05$). The same pattern was observed with absolute diferences between predcited and actual ΔVL values and for the ATHENA data.
- The RDI and ATHENA test sets were comparable in terms of the degree of similarity of their data to the training dataset, as assessed by the frequency of individual drugs, drug combinations, mutations and mutation patterns,

Figure 1: Correlations with actual ΔVL: RDI test set

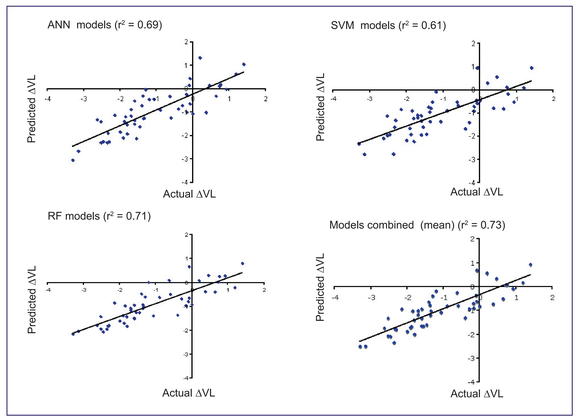
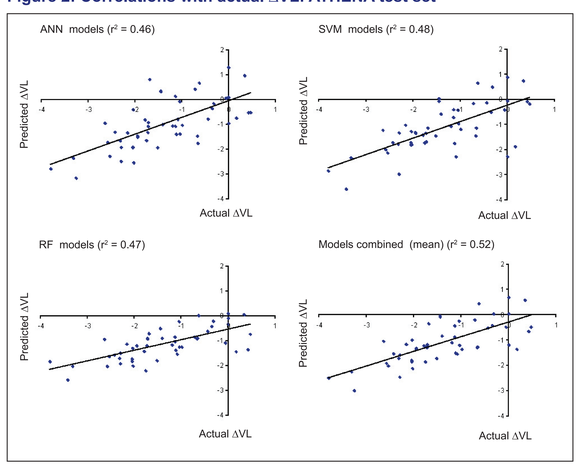


Figure 2: Correlations with actual ΔVL: ATHENA test set



Discussion

- The committee average performance of the three modelling methodologies was comparable accounting for 61-71% of the variance in virological response (RDI test set) and 46-48% (ATHENA test set).
- Individual RF and SVM models achieved more accurate predictions than ANN models and the predictions of RF models were the most consistent.
- The committee average methodology improved the performance of ANN much more than RF or SVM, equalising the accuracy of prediction.
- The superior accuracy of the models with the RDI test set *vs* ATHENA cannot readily be explained in terms of the similarity between data sets.

Conclusions

- All three methods can yield fairly accurate predictions of virological response to HAART and their combination may provide a modest improvement in accuracy over any one method.
- As with previous studies, the predictions for 'familiar' clinics were somewhat more accurate than for 'unfamiliar' clinics but the reason for this remains unclear.
- There was no evidence for RF or SVM being more generalisable than ANN.
- The accuracy of these models, trained using relatively small datasets, was encouraging.
- Further investigation of alternative methods and the potential combination of their predictions is warranted.

Acknowledgments

The RDI thanks all the centres around the world, and their patients for providing the data used in this study, including:

- BC Centre for Excellence in HIV/AIDS, Vancouver, BC, Canada
- Community Programs for Clinical Research on AIDS (CPCRA).
- Fundacion IrsiCaixa, Badelona, Spain
- Hospital Clinic of Barcelona, Spain
- ICONA cohort, Italy
- Italian ARCA database, University of Siena, Siena, Italy
- Italian MASTER Cohort (coordinated by the University of Brescia, Italy)
- National Centre in HIV Epidemiology and Clinical Research, Sydney, Australia
- National Institute of Allergy and Infectious Disease, Bethesda, MD, USA
- Netherlands HIV Monitoring Foundation, Amsterdam, The Netherlands
- PhenGen study group (via the University of Brescia, Italy)
- Ramon y Cajal Hospital, Madrid, Spain
- Royal Free Hospital, London, UK
- US Military HIV Research Program, Rockville, MD, USA

This project has been funded in part with Federal Funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1-CO-12400.