Prediction of Early-Stage Chronic Kidney Disease in an HIV-Positive Population

Omolola I. Ogunyemi[a,b], Chizobam Ani[b], Francis Yemofio[b,c], Wilbert Jordan[b,c], Keith Norris[b]

[a]Center for Biomedical Informatics, [b]Charles Drew University of Medicine and Science, Los Angeles, CA; [c]OASIS Clinic, Los Angeles, CA

Abstract

We assessed the feasibility of predicting an HIV-positive individual's two-year risk of developing chronic kidney disease (CKD) and examined the accuracy of several predictive models based on Bayesian networks, logistic regression, artificial neural networks and support vector machines. Our preliminary study used anonymized data from 92 patients currently seen at a specialty clinic for HIV-positive patients in Los Angeles (the OASIS Clinic). When the predictive models were trained and evaluated on the dataset of 92 patients using leave-one-out cross validation, the best predictive model was a Bayesian network model, with an area under the ROC curve of 0.693. The study suggests that computerized prediction of a CKD-free, HIV-positive patient's two-year risk of developing CKD is feasible. Risk prediction models developed for this purpose could assist clinicians in preventing or delaying the onset of CKD in HIV-positive patients.

Keywords:

HIV-associated nephropathy, Machine learning

Introduction

Kidney disease in the United States affects roughly 27 million adults.[1] Kidney function has been estimated to be abnormal in as many as 30% of individuals with HIV, and the co-occurrence of kidney dysfunction confers a three-fold additional risk of death for HIV-positive individuals.[2] In spite of knowledge of the high risk and incidence of CKD among HIV-positive individuals, and evidence that CKD can be prevented or progression delayed, the quality of CKD-preventive care in many settings continues to be inadequate.

A premise of our study is that providing physicians with computerized decision support tools for CKD prediction in HIV-positive individuals may enhance preventive renal care. As a first step towards this goal, we seek to address the difficult problem of predicting the two-year risk of developing CKD at its earliest stages, in patients who are currently CKD-free.

Methods

To assess the feasibility of predicting an HIV-positive individual's risk of developing CKD within two years, we developed Bayesian network, neural network, support vector machine, and logistic regression models on a small sample of retrospective cases from the OASIS clinic. CKD for this preliminary assessment was defined to include the earliest stages (i.e., eGFR < 90 mL/min/1.73 m^2 or > 90 mL/min/1.73 m^2 with evidence of proteinuria for three or more months). Patients selected for the study did not have CKD upon their initial visit to the clinic but they developed CKD within two years of treatment for HIV.

After an examination of the literature for published CKD risk factors, medical record data on 30 identified risk factors were obtained for classification from a subset of 92 OASIS patients. We performed feature subset selection using backward elimination to reduce the set of risk factors for two-year prediction of CKD and learned Bayesian network (BN), artificial neural network (ANN), support vector machine (SVM) and logistic regression (LR) models using Weka.**[**3**]**

Results

After feature subset selection, 5 variables were implicated in CKD in our HIV-positive population: hypertension, CD4 count, non-steroidal anti-inflammatory drug use, smoking, and HIV RNA count. Table 1 gives the risk prediction accuracy.

Table 1- Accuracy of predictive models

	BN	**ANN**	**SVM**	**LR**
Correctly classified	72	65	63	70
Incorrectly classified	20	27	29	22
Accuracy	78.3%	70.7%	68.5%	76.1%
Sensitivity	46.4%	32.1%	28.6%	39.3%
Specificity	92.2%	87.5%	85.9%	92.2%
AUC	0.693	0.598	0.573	0.657
Cohen's kappa statistic	0.429	0.221	0.163	0.358
Brier score	0.201	0.237	0.338	0.207

Conclusion

The study suggests that computerized prediction of a CKD-free, HIV-positive patient's two-year risk of developing CKD is feasible and shows that Bayesian networks gave the best predictions for our patient sample.

References

[1] Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, et al. Prevalence of chronic kidney disease in the United States. JAMA. 2007 Nov 7;298(17):2038-47.

[2] Olatinwo T, Hewitt RG, Venuto RC. Human Immunodeficiency Virus-Associated Nephropathy: A Primary Care Perspective. 2004:333-6.

[3] Witten I, Frank E. Data Mining: Practical Machine Learning Tools and Techniques. 2nd edition ed: Morgan Kaufmann 2005.