19th International Congress on Modelling and Simulation, Perth, Australia, 12–16 December 2011
http://mssanz.org.au/modsim2011

Stroke prediction in a sample of HIV/AIDS patients: Logistic regression, Bayesian networks or a combination of both?

J. Gutierrez [a], **and C. Yoo** [b]

[a] *Department of Neurology, New York Presbyterian Hospital, Columbia University*

[b] Department of Statistics, Robert Stempel College of public Health and Social Work

Email: drjosegc@hotmail.com

Abstract: Background: Cardiovascular disease is an increasingly frequent diagnosis in patient with HIV/AIDS. Predictive models of stroke in these populations can helps planning targeted strategies to reduced morbidity and mortality in HIV/AIDS populations.

Methods: In a hospital-based sample, we explored associations with stroke. Three different statistical models were used: multivariate logistic regression (LR), Bayesian networks (BN) and a combination of both. Goodness of fit was evaluated with the area under the curve and reliability was tested with three-cross validation.

Results: One hundred and nine patients with HIV/AIDS were included in the analysis. The mean age was 46.87 (SD± 11.6), 56.8% were men, 77% were black, 79% came from low income areas (less than $40,000 year median income). Stroke was associated with high CD4 count, lack of insurance and the presence of cardiovascular risk factors. Using as inputs the presence of cardiovascular risk factors, race, insurance status, HAART administration, CD4 count and medical comorbidities in a multivariate LR, 76 % of the stroke cases were accurately predicted. Using BN analysis, 75% of stroke cases were accurately predicted. Using a combination of LR and BN, only 72% of the cases were correctly classified, although this model had the best discriminant function as evidenced by good positive and negative predictive value.

Conclusions: Stroke is associated with cardiovascular risk factors, lack of medical insurance and higher CD4 counts. Using a combination of LR and graphical probabilistic models improved the discriminant function of the predictive model. Longitudinal, population based studies are needed to confirm these associations.

Keywords: *Stroke, hiv, logistic regression, Bayesian networks, discriminant models*

1. BACKGROUND

Despite significant improvement in the survival of patient diagnosed with AIDS, morbidity and mortality in this population remains a priority for public health specialist. The administration of highly active antiretroviral therapy (HAART) has rendered the HIV infection a chronic process that predisposes to new forms of disease in patients who host the virus.[1, 2] Among the most important implications of the prolonged survival time is cardiovascular disease occurrence, including stroke. [3-5]

Modeling healthcare outcome is of paramount importance for researcher and public health specialist. The use of different statistical models allows researchers to identify important association that later can help planning strategies to tackle the specific problem. Determining the accuracy of the predictive models of stroke in HIV patients is thus an important goal to achieve. There are multiples ways to model the outcomes to be studied. In the case of binary data, the most commonly used is logistic regression (LR). However, there is growing interest in graphical probabilistic model that can analysis important interaction in complex datasets with multiple predictors. [6, 7] Traditionally, the interaction are tested based on medical expertise, however, this poses the problem of increase computational work and sometimes missing important relationship not visible in a pairwise comparison.[8] An alternative approach is to use hybrid model that can help us determining what interaction need to be tested. [7] The addition of interactions using graphical logistic models has been attempted in larger datasets with relative good results. [6] In this paper we show the reliability of the hybrid model in smaller dataset.

We analyzed different predictive models in a small sample of patients with HIV/AIDS with and without stroke to determine the accuracy of the stroke prediction by using the same inputs with different discriminant models.

2. METHODS

Cases with HIV/AIDS-related consultations were selected from the neurology consultation log used for sign-out among residents from July 1st 2009 to June 30th 2010 at Jackson Memorial Hospital. Cases with multiple encounters during the year were counted as one case unless a different reason for consult was established in the same patient. Total of 129 cases were collected but only 109 were kept for the final analysis. Demographic, socioeconomic and clinical variables were obtained from each patient. Three different models were created. The first was logistic regression using the variables mentioned below. The second model consisted of Bayesian networks only. The second model included the six variables used for logistic regression plus interactions identified in the Bayesian network.

In each model, we have included the following variables:

- *Stroke:* Focal neurological deficit that is irreversible and has a vascular etiology as the most likely cause of the symptoms. Confirmation was required with either Brain MRI or brain CT.
- *Low CD4 count (Low CD4):* Less than 200 lymphocytes CD4+ per dl.
- *Insurance status (Ins status):* Insured or Uninsured, recorded from medical records.
- *Race:* self-defined, black or non-black (Hispanic or non-Hispanic white)
- *Cardiovascular comorbidities:* Presence of at least one modifiable traditional Cardiovascular Risk Factor (CVRF): Hypertension, diabetes, dyslipidemia and smoking were considered modifiable traditional cardiovascular risk factor.
- *HAART:* The appearance in the medical record of any antiretroviral therapy was considered as evidence of administration of HAART.
- *Medical comorbidities (Med comorb):* The presence of at least one medical comorbidity, excluding psychiatric diagnoses and cardiovascular risk factors mentioned above.

Bayesian networks:

A Bayesian Network is a directed acyclic graph in which each node represents a variable and each arc represents probabilistic influence. In Bayesian Networks, each arc is interpreted as a direct influence between a parent node (variable) and a child node, relative to the other nodes in the network. For example, in figure 1, cardiovascular risk factors (CVRF) are related to low CD4 count only through stroke. In the same way, HAART and medical comorbidities are only related through insurance status, otherwise they would be independent. A causal network consists of a structure (Figure 1) and a set of probabilities that parameterize

that structure (Figure 2). In general, for each variable, there is a conditional probability of that variable given the states of its direct causes. Thus, the probability associated with *low CD4* is P (Low CD4 | race, stroke). That is, we give the probability distribution over the values *of low CD4* conditioned on each of the possible values of stroke and race. For variables that have no direct causes in the network, a prior probability is specified. The Markov condition gives the conditional independence relationships that are specified by a Bayesian Network: *A node is independent of its non-descendants* (*e.g., non-effects*) *given its parents* (*i.e., its direct influences*).

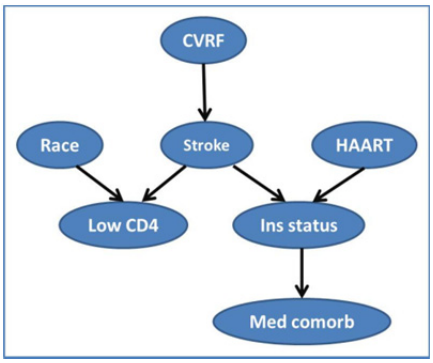

Figure 1: Bayesian Network for stroke outcomes

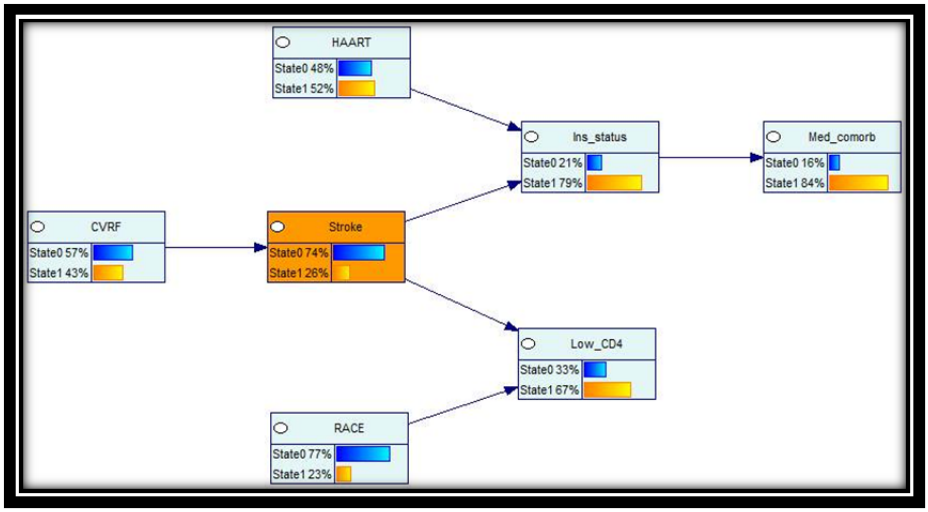

Figure 2: Parameterized Bayesian network

Bayesian networks model (BNonly model):

A binary code was used for all variables in the model. The graphical representation of our dataset was obtained using Banjo.[9] The consensus graph was kept as the best fitting network for our dataset after analyzing it with three independent three hours analyses in a Pentium (R) dual-core 3.20 GHz machine. The three analyses produce highly consistent results (Figure 1)

With the Bayesian network obtained from banjo, we used GeNIe (Graphical Network Interface) to parameterize the Bayesian network. [10] We used all the samples (109 cases) to learn parameters given the Bayesian network identified by banjo (Figure 2). Then, with the learned parameters, we calculated the conditional probability of *stroke* given the states of other input variables, such as , *low CD4, Ins status, race, CVRF, HAART* and *med comorb.*

Logistic regression without interactions (LRonly model):

Six variables showed some association with *stroke: ins status, race, low CD4, HAART, med comorb and CVRF*. To compare LRonly model with other models, we have fit the model with the six inputs using the entire sample (109 cases). Using the result model, we have estimated P(*stroke* | *Race, Ins Status, Low CD4, HAART, CVRF, Med Comorb*) using formula (1):

$$\frac{1}{1+e^{-p}} \qquad (1)$$

Where p=0.868·*Race*-0.887·*Ins Status*-1.617·*Low CD4*-0.602·*HAART*+1.232·*CVFR*-0.305·*Med Comorb*

We used PASW statistics, version 18.0 to learn this model. We used the above logit function and calculated the probability of stroke given the inputs plus interactions among them.

Logistic regression with interactions from Bayesian networks (Hybrid model):

In additional to learning the best Bayesian network that fits the data, banjo reports influence scores. An influence score can be either positive or negative and it describes the magnitude that a parent node can have on a child node. The influence scores obtained between modeled variables from the banjo was determined as evidence of the top scoring interactions (See table 1). Additionally, by analyzing the Bayesian network, we identified interactions that could be obtained by following the Markov rule of conditional probability. For example, from Figure 1, *Low CD4* and *Ins Status* are independent given *Stroke*, (a.k.a., diverging – note that arcs are diverging from *Stroke* – structure) however, *HAART* and *Stroke* are dependent given *Ins Status* (a.k.a., converging – note that arcs are converging into *Ins Status* – structure).[11] Similarly, *Race* and *med comorb* are independent from each other because of diverging connections between them. Additional to the interactions from influence scores, we explore two more interactions: *low CD4* count by *ins status* (dependent through *stroke*) and *low CD4* by *med comorb* (dependent through stroke and insurance status). The maximum likelihood for each given combination of the six input variables plus one, two or three interactions was obtained. The best fitting model was decided calculating the p value for the statistical power using chi-square distribution (table 2). [12]The best fitting model was used to calculate the prediction of stroke. PASW statistics, version 18.0 was used for the calculated group membership.

Table 1. Influence score obtained with Banjo analytic software

Influence score	value
Stroke -> Low CD4	-.3765
HAART -> Insurance Status	.3304
Ins Status -> Med comorbidities	.3098
CVRF -> Stroke	.2362
Stroke -> Insurance Status	-.2001
Race -> low CD4	.0000

TABLE 2: Maximum likelihood of the models tested for the best fitting combination of interactions.

Model 0 No interactions Df=6	Model 1 One interaction Df=7	Model 2 Two interaction Df=8	Model 3 Three interactions Df=9	Model 4 All interactions Df=10
a)96.603	a)86.270 HAART*Insurance	a)82.420 HAART*Insurance CD4*med_co	a) 81.151 HAART*Insurance CD4*med_co CD4*Insurance	a)80.224 HAART*Insurance CD4*med_co CD4*Insurance Insurance*med_co
	b)95.209 CD4*med_co	b)84.343 HAART*Insurance CD4*Insurance	b) 81.718 HAART*Insurance CD4*med_co Insurance*med_co	
	c)96.410 Insurance*med_co	c)86.269 HAART*Insurance Insurance*med_co	c)84.308 HAART*Ins status Insurance*med_co CD4*Ins status	
	d)96.600 CD4*Insurance	d)94.891 Insurance*med_co CD4*med_co	b)94.715 Insurance*med_co CD4*med_co CD4*Insurance	
		e)94.959 CD4*Insurance CD4*med_co		
		f)96.410 Insurance*med_co CD4*Insurance		
P value for comparisons only models in row "a" (the SMALLEST -2log from each Model) M0-M4=0.003, M4-M3=0.335, M3-M2=0.260, M4-M2=0.333, M2-M0=<0.001, M2-M1= 0.049				

Evaluation of the Models

The area under the receiver operating curve (AUROC) was used to evaluate the three models, i.e., BNonly, LRonly, and Hybrid. Also we have evaluated the models using three-fold cross validation of the data. In the three-fold cross validation, we report average of AUROC for all three test data additional to the average positive predictive value and the average negative predicted value.

3. RESULTS

A total of 129 cases were identified. Only 109 patients were included in the analysis. 20 cases were excluded from the analysis due to lack of information or repeated visits with the same complain The mean age was 46.87 (SD± 11.6), 56.8% were men, 77% were black, 79% came from low income areas (less than 40,000 year median income). Half of the patients were receiving HAART (52%) although 67% had CD4 count less than 200. Forty three percent had at least one cardiovascular comorbidity, and 84% had at least one non-cardiovascular comorbidity. Twenty one percent of the patients were uninsured, and 26% had a stroke.

In univariate LR analysis, having insurance and low CD4 count was less often associated with stroke (OR 0.25, 95% CI 0.09-0.67 and OR 0.21, 95%CI 0.08-0.52, respectively). The presence of CVRF was associated with stroke (OR 3.7, 95% CI 1.5-9.2). These same three associations remained unchanged after adjusting for age, gender, race, and HAART administration, psychiatric and medical comorbidities.

The Bayesian network obtained from the dataset is shown in Figure 1. It can be seen that there is an association between *CRVF, Stroke and Low CD4* count. Also *Low CD4* and *ins status* are conditionally independent given stroke. Note that *stroke* and *Race* become dependent given *Low CD4* and also s*troke* and *HAART* become dependent given insurance status. Using this Bayesian network with updated probabilities on *CVRF, Race, Low CD4, ins status, HAART and med comorb* (Figure 2) yielded an AUROC of 0.750±0.063 (Table 3). The multivariate logistic regression model using the same inputs as Bayesian networks yielded an AUROC of 0.760±0.086. Different interactions were explored and the maximum likelihood helped us determining the best fitting model. The chosen model included the same six variables used in Bayesian networks and multivariate logistic regression plus interactions, i.e., *HAART by Ins status and low CD4 by med comorb*. *Low CD4 by med comorb* was obtained by evaluating the conditional relationship shown in the Bayesian network that would otherwise have been missed in traditional pairwise comparisons. Adding interactions to simple logistic regression improved the model accuracy (Table 2).

Comparing the three AUROCs, the evaluated models performed relatively similar in terms of point estimates and variance. However, it was the hybrid model combining Bayesian input and logistic regression that provided the best positive and negative predictive value compared with the other two models. Of particular interest, the reduction of the sample size did not affect the predictive capability of the hybrid model as demonstrated by the stable AUROC.

Table 3: SUMMARY TABLE FOR STROKE PREDICTION

	LR	BN	LR+BN
	AUC		
Total sample	0.760±0.086	0.750±0.063	0.729±0.065
PPV	81.8 %	78.9 %	85.1 %
NPV	79.3 %	86.7 %	86.7 %
	3-FOLD CROSS VALIDATION		
Sample A	0.711±0.093	0.767±0.087	0.740±0.096
PPV	75.0 %	95.0 %	86.7 %
NPV	78.6 %	76.1 %	84.3 %
Sample B	0.729±0.117	0.757±0.123	0.732 ±0.128
PPV	68.5 %	82.3 %	78.9 %
NPV	88 %	68.2 %	84.6 %
Sample C	0.739±0.075	0.724±0.139	0.752 ±0.153
PPV	72.4 %	79.5 %	87.2 %
NPV	80.3 %	87.3 %	82.6 %

4. DISCUSSION

Our results suggest that being insured, having high CD4 count and the presence of cardiovascular risk factors are statistically associated with the presence of stroke. These associations have been widely discussed in multiple governmental reports using population based samples. [2, 4, 13-16] Being uninsured is related to deficient medical follow up, unrecognized diagnosis and treatment of cardiovascular risk factors than can lead to stroke. [1, 17, 18] The fact that in our sample having insurance was associated with stroke is probably due to the effect of having insurance leads to the administration of HAART and higher CD4. It is widely accepted that the prevalence of cardiovascular risk factors is the major determinant in the occurrence of cardiovascular events, in this case, stroke. [19-26] In our sample, the presence of CVRF was the second most important determinant for stroke after low CD4. Of interest in our analysis, higher CD4 count was associated with stroke. There is no a clear explanation for this finding, however, it is likely that patients with higher CD4 count are truly taken HAART. The administration of HAART has been associated with the occurrence of stroke mainly through accelerated atherosclerosis and dyslipidemia. [27, 28] We could not find an association between HAART and stroke in our sample using LR, but we found dependency of HAART and Stroke given

Insurance status. This could be explained by error in the data collection since compliance was not taken into account.

Modeling probability of binary data can be accomplished with different statistical methods. In our dataset, we tried to demonstrate that combining BN with LR will be better than any one model alone in the prediction of stroke. In this case, the presence of stroke in HIV/AIDS patients was attempted using six covariates. We evaluated the three different statistical models: logistic regression with no interactions, Bayesian networks and a combination of logistic regression with interaction inputted from Bayesian network inferred conditional probability. The three models seem to perform relatively well in terms of AUROC, but the hybrid model has the best combination of positive and negative predicted value.

The approach to probability of outcome using binary data with combining different models has been attempted by a few authors.[6, 7, 29-36] In general, it has been shown that using hybrid model can improve the accuracy of the prediction. Large dataset seem to be better analyzed by combination of models. [36] Stojadinovic et al. showed in dynamic data that using graphical probabilities models can show interaction not usually seen in traditional regression methods. [6] The 10-fold cross validation of their dataset show robustness of method. Although not comparable, the 10-fold cross validation could have improved the prediction since it uses a large proportion of the sample to calculate the parameters. Huttenhower et al. compared the expert-estimated conditional probabilities and showed that Bayesian network input outperformed the expert-base knowledge, although this effect diminishes as the amount of available data decreases. The authors suggest that overall, Bayesian learning provides a consistent benefit in data integration, but its performance and the impact of heterogeneous data sources must be interpreted from the perspective of individual functional categories.[34] Gevaert et al. showed that the integration of structure into predictive model results in improved AUROC compared to logistic regression without structure information. [33]

There are important limitations to our study. The sample used for analysis reflects a high risk population as evidence by the low income, the high prevalence of uninsured individuals and multiple comorbidities. If the associations found hold true for other type of AIDS patient with better health profile remains to be seen. The cross-sectional nature of the analysis precludes further causal inferences. However, we may use Causal Bayesian Networks [37, 38] together with latent variable models [39, 40]to further search for causal relationships among the modeled variables in this paper.

In conclusion, the presence of stroke in our high-risk sample is associated with lack of medical insurance, higher CD4 count and the presence of cardiovascular risk factors. Additionally, the integration of graphical probabilistic designs into logistic regression can improve the model discriminant capacity even in small samples. Longitudinal, population-based samples are needed to confirm our results.

REFERENCES

Ovbiagele B, Nath A. Increasing incidence of ischemic stroke in patients with hiv infection. *Neurology*. 2011;76:444-450

Epidemiology of hiv/aids--united states, 1981-2005. *MMWR Morb Mortal Wkly Rep*. 2006;55:589-592

Connor MD, Lammie GA, Bell JE, Warlow CP, Simmonds P, Brettle RD. Cerebral infarction in adult aids patients: Observations from the edinburgh hiv autopsy cohort. *Stroke*. 2000;31:2117-2126

Dobbs MR, Berger JR. Stroke in hiv infection and aids. *Expert Rev Cardiovasc Ther*. 2009;7:1263-1271

Ortiz G, Koch S, Romano JG, Forteza AM, Rabinstein AA. Mechanisms of ischemic stroke in hiv-infected patients. *Neurology*. 2007;68:1257-1261

Stojadinovic A, Eberhardt J, Brown TS, Hawksworth JS, Gage F, Tadaki DK, Forsberg JA, Davis TA, Potter BK, Dunne JR, Elster EA. Development of a bayesian model to estimate health care outcomes in the severely wounded. *J Multidiscip Healthc*. 2010;3:125-135

Pang BC, Kuralmani V, Joshi R, Hongli Y, Lee KK, Ang BT, Li J, Leong TY, Ng I. Hybrid outcome prediction model for severe traumatic brain injury. *J Neurotrauma*. 2007;24:136-146

Gustafson P, Kazi AM, Levy AR. Extending logistic regression to model diffuse interactions. *Stat Med*. 2005;24:2089-2104

Hartemink AJ. Banjo: Structure learning of static and dynamic bayesian networks. . 2010

Druzdzel MJ. Intelligent decision support systems based on smile. *Software 2.0*. 2005;2: 12-33

Charniak E. Bayesian networks without tears: Making bayesian networks more accessible to the probabilistically unsophisticated. *AI Mag*. 1991;12:50-63

Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: Wiley; 2000.

Disparities in diagnoses of hiv infection between blacks/african americans and other racial/ethnic populations--37 states, 2005-2008. *MMWR Morb Mortal Wkly Rep*. 2011;60:93-98

Vital signs: Hiv testing and diagnosis among adults--united states, 2001-2009. *MMWR Morb Mortal Wkly Rep*. 2010;59:1550-1555
Racial/ethnic disparities among children with diagnoses of perinatal hiv infection - 34 states, 2004-2007. *MMWR Morb Mortal Wkly Rep*. 2010;59:97-101
Karon JM, Rosenberg PS, McQuillan G, Khare M, Gwinn M, Petersen LR. Prevalence of hiv infection in the united states, 1984 to 1992. *JAMA*. 1996;276:126-131
McWilliams JM, Meara E, Zaslavsky AM, Ayanian JZ. Differences in control of cardiovascular disease and diabetes by race, ethnicity, and education: U.S. Trends from 1999 to 2006 and effects of medicare coverage. *Ann Intern Med*. 2009;150:505-515
Oladele CR, Barnett E. Racial/ethnic and social class differences in preventive care practices among persons with diabetes. *BMC Public Health*. 2006;6:259
Bang OY, Saver JL, Liebeskind DS, Lee PH, Sheen SS, Yoon SR, Yun SW, Kim GM, Chung CS, Lee KH, Ovbiagele B. Age-distinct predictors of symptomatic cervicocephalic atherosclerosis. *Cerebrovasc Dis*. 2009;27:13-21
Bang OY, Saver JL, Liebeskind DS, Pineda S, Yun SW, Ovbiagele B. Impact of metabolic syndrome on distribution of cervicocephalic atherosclerosis: Data from a diverse race-ethnic group. *J Neurol Sci*. 2009;284:40-45
Birns J, Morris R, Jarosz J, Markus H, Kalra L. Ethnic differences in the cerebrovascular impact of hypertension. *Cerebrovasc Dis*. 2008;25:408-416
Boden-Albala B, Cammack S, Chong J, Wang C, Wright C, Rundek T, Elkind MS, Paik MC, Sacco RL. Diabetes, fasting glucose levels, and risk of ischemic stroke and vascular events: Findings from the northern manhattan study (nomas). *Diabetes Care*. 2008;31:1132-1137
Brown DW, Giles WH, Greenlund KJ. Blood pressure parameters and risk of fatal stroke, nhanes ii mortality study. *Am J Hypertens*. 2007;20:338-341
Cushman M, Cantrell RA, McClure LA, Howard G, Prineas RJ, Moy CS, Temple EM, Howard VJ. Estimated 10-year stroke risk by region and race in the united states: Geographic and racial differences in stroke risk. *Ann Neurol*. 2008;64:507-513
De Silva DA, Woon FP, Lee MP, Chen CL, Chang HM, Wong MC. Metabolic syndrome is associated with intracranial large artery disease among ethnic chinese patients with stroke. *J Stroke Cerebrovasc Dis*. 2009;18:424-427
Deleu D, Hamad AA, Kamram S, El Siddig A, Al Hail H, Hamdy SM. Ethnic variations in risk factor profile, pattern and recurrence of non-cardioembolic ischemic stroke. *Arch Med Res*. 2006;37:655-662
Treatment. Haart may increase risk of stroke in hiv-infected. *AIDS Policy Law*. 2011;26:1
Voelker R. Stroke increase reported in hiv patients. *JAMA*. 2011;305:552
Stephenson N, Beckmann L, Chang-Claude J. Carcinogen metabolism, cigarette smoking, and breast cancer risk: A bayes model averaging approach. *Epidemiol Perspect Innov*. 2010;7:10
Foltran F, Berchialla P, Giunta F, Malacarne P, Merletti F, Gregori D. Using vlad scores to have a look insight icu performance: Towards a modelling of the errors. *J Eval Clin Pract*. 2010;16:968-975
Nissan A, Protic M, Bilchik A, Eberhardt J, Peoples GE, Stojadinovic A. Predictive model of outcome of targeted nodal assessment in colorectal cancer. *Ann Surg*. 2010;251:265-274
Sakai S, Kobayashi K, Nakamura J, Toyabe S, Akazawa K. Accuracy in the diagnostic prediction of acute appendicitis based on the bayesian network model. *Methods Inf Med*. 2007;46:723-726
Gevaert O, De Smet F, Kirk E, Van Calster B, Bourne T, Van Huffel S, Moreau Y, Timmerman D, De Moor B, Condous G. Predicting the outcome of pregnancies of unknown location: Bayesian networks with expert prior information compared to logistic regression. *Hum Reprod*. 2006;21:1824-1831
Huttenhower C, Troyanskaya OG. Bayesian data integration: A functional perspective. *Comput Syst Bioinformatics Conf*. 2006:341-351
Chakraborty S, Ghosh M, Maiti T, Tewari A. Bayesian neural networks for bivariate binary data: An application to prostate cancer study. *Stat Med*. 2005;24:3645-3662
Lee SM, Abbott P, Johantgen M. Logistic regression and bayesian networks to study outcomes using large data sets. *Nurs Res*. 2005;54:133-138
Cooper GF. Causal discovery from data in the presence of selection bias *Proceedings of the Workshop on Artificial Intelligence and Statistics*. 1995
Meek C. Causal inference and causal explanation with background knowledge. *Proceedings of the Conference on Uncertainty in Artificial Intelligence*. 1995
Yoo C, V. Thorsson, and G.F. Cooper. Discovery of a gene-regulation pathway from a mixture of experimental and observational DNA microarray data. . *Pacific Symposium on Biocomputing*. 2002
Yoo CaC, G. An evaluation of a system that recommends microarray experiments to perform to discover gene-regulation pathways. *Journal of Artificial Intelligence in Medicine*. 2004;31:169-182