MEDINFO 2007
K. Kuhn et al. (Eds)
IOS Press, 2007

Comparing Decision Support Methodologies for Identifying Asthma Exacerbations

Judith W Dexheimer[a], Laura E Brown[a], Jeffrey Leegon[b], Dominik Aronsky[a,c]

[a] *Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA*
[b] *Vanderbilt Center for Better Health, Vanderbilt University, Nashville, TN, USA*
[c] *Department of Emergency Medicine, Vanderbilt University, Nashville, TN, USA*

Abstract

Objective: To apply and compare common machine learning techniques with an expert-built Bayesian Network to determine eligibility for asthma guidelines in pediatric emergency department patients.

Population: All patients 2-18 years of age presenting to a pediatric emergency department during a 2-month study period.

Methods: We created an artificial neural network, a support vector machine, a Gaussian process, and a learned Bayesian network to compare each method's ability to detect patients eligible for asthma guidelines. Our outcome measures included the area under the receiver operating characteristic curves, sensitivity, specificity, predictive values, and likelihood ratios.

Results: The data were randomly split into a training set (n=3017) and test set (n=1006) for analysis. The systems performed equally well. The area under the receiver operating characteristic curve was 0.959 for the expert-built Bayesian network, 0.962 for the automatically constructed Bayesian network, 0.956 for the Gaussian Process, and 0.937 for the artificial neural network.

Discussion: All four evaluated machine learning methods achieved high accuracy. The automatically created Bayesian network performed similarly to the expert-built network. These methods could be applied to create a real-time detection system for identifying asthma patients.

Keywords:

asthma, decision support techniques, decision support systems, evaluation

Introduction

Asthma is one of the most common pediatric illnesses. It causes an estimated 14 million missed school days and 14.5 million missed workdays yearly in the United States [1]. Asthma exacerbations account for more than 1.8 million emergency department (ED) visits annually [1].

Adherence with guidelines has been shown to improve the clinical care patients receive [2, 3]. A common barrier to guideline initiation is determining eligible patients [4]. Nurses in addition to their normal workload are often charged with identifying eligible patients, a task that is frequently forgotten and leads to lower guideline adherence. Automatically detecting patients could help improve guideline adherence. Ideally, an electronic system that requires no additional data entry would be used to detect guideline eligible patients.

The goal of our study was to evaluate several machine learning techniques using a verified asthma dataset and to compare the techniques with an existing expert-built Bayesian Network (BN) used to identify asthma patients with asthma exacerbations in a pediatric ED [5]. We compared the expert-built BN with a BN automatically learned from data, a Support Vector Machine (SVM), an Artificial Neural Network (ANN), and a Gaussian Process (GP).

Background and methods

Various computerized methods have been developed for identifying asthma patients [6, 7]. These studies have focused on clinical data - such as peak flow, clinic notes or discharge summaries, and computer-based questionnaires [6]. Although some of the studies used ED data, they were not stand-alone systems or integrated into the clinical workflow or with existing information technology infrastructures.

Setting

The study data were collected at Vanderbilt Children's Hospital ED, a 29-bed pediatric facility with more than 40,000 visits annually. The ED information system infrastructure includes an electronic whiteboard, an electronic triage system, an electronic medical record, and a computerized provider order-entry system in the ED. The study was approved by the institution's IRB.

Population

All patients presenting to the pediatric ED during the 2-month study period were screened for inclusion. Patients were included if they were 2-18 years of age and seen in the pediatric ED. Patients were excluded if they did not have a coded chief-complaint, were not treated in the ED (such as left without being examined) or had no final diagnosis for their visit in the paper or electronic patient record.

Design objectives

Our objective was to apply and compare machine learning techniques to an expert-built Bayesian Network. The result of each technique could be developed into a real-time system for detecting asthma patients presenting to a pediatric ED. We adapted the design objectives of a prior BN study to fit our additional techniques [5].

Data sources

We used the database of the previous study [5] which included 4,023 patient encounters. The randomized splits for training and testing of the original BN were preserved. The data included commonly available variables from the electronic medical record, electronic triage, and electronic whiteboard such as 141 ICD-9 coded chief complaints, a past medical history of asthma, billing codes, and asthma medications including beta-agonists, steroids and others. Chief complaints and asthma medication were identified using searches of free-text. Text matching has been shown to be an accurate method to determine the past medical history of asthma [8]. The past medical history of asthma was determined by searching the billing records for prior asthma diagnosis codes (ICD-9 493.*). The patient's chief complaint, acuity level, age, respiratory rate, and oxygen saturation came from the electronic triage. All of these variables are regularly captured and stored in the hospital's information systems. All data elements were available through the computerized hospital information systems, and no additional data collection was necessary.

Dataset

The dataset collected consisted of 11 variables: 3 variables (age, oxygen saturation, and respiratory rate) have continuous values, 5 variables are ordinal (acuity, billing codes, and the 3 medication variables), 1 variable (chief complaint) is categorical, and 2 variables (history of asthma and prediction variable of asthma eligibility) are binary. The dataset was discretized following the values given in Table 1 for use in constructing the expert BN.

BN learning algorithms, including the one used in this study, generally require complete datasets, i.e., no missing values. The frequency of missing data in several variables (shown in table 1) limited the choice of methods to handle the missing values (e.g., removal of cases, removal of variables, imputation). The "non-asthma" column includes all patients who were not diagnosed with asthma (n=3,638) and the "asthma" column includes patients diagnosed with asthma (n=385). The missing elements were encoded as an additional value a variable may take. The dataset was randomly split into a training set (n=3017) and testing set (n=1006) with equal proportions of asthma cases.

The ANN, SVM, and GP methods may take continuous values, therefore an alternative dataset was constructed for their use from the original data (i.e. prior to discretization). For this dataset missing data for the continuous variables were imputed using a k-nearest neighbors imputation method. For the medication variables, a missing value was treated as having no history of that medication found in the medical record. Missing acuity values were treated as an additional "0" value. All ordinal and categorical variables were encoded using 1-of-m encoding. Finally, the dataset was scaled to [-1,1] for the SVMs and [0,1] for the ANN technique. The dataset was scaled using Gaussian normalization for the GP.

Table 1 – Missing values

Variable	Values	Non-asthma (%)	Asthma (%)
History of Asthma	Present, Absent	0.0	0.0
Billing Codes	Number: 0, 1, >1	0.0	0.0
Meds: -agonists	Number: 0, 1, >1	51.3	45.2
Meds: Steroids	Number: 0, 1, >1	51.3	45.2
Meds: Other	Number: 0, 1, >1	51.3	45.2
Chief Complaint	141 unique	0.0	0.0
Acuity Level	ESI level: 1-5	1.21	0.26
Age Category	2-3, 4-6, 7-11, 12-18	0.0	0.0
Respiratory Rate	<20, 20-24, 25-29, 30-34, 35-39, >40	4.15	3.38
Oxygen Saturation	<91, 91-92, 93-94, 95-96, 97-98, >98	3.13	0.26

Reference standard

The reference standard for an asthma diagnosis was a free-text diagnosis of "asthma exacerbation," "status asthmaticus," "wheezing," or "reactive airway disease" [5]. A board-certified internal medicine physician determined the asthma diagnosis through manual chart review. Electronic and paper charts were searched for a diagnosis for every ED visit during the study period. Patients without a discharge diagnosis were not included in the study.

Bayesian network

A BN is formalism consisting of a directed acyclic graph with nodes representing variables and a joint probability distribution over the variables. BNs can be created by hand using expert knowledge. The network parameters for the BN are set by the training set and predictions are made on the testing set. BNs are advantageous in that the prediction inference algorithms handle missing data which is prevalent in clinical systems, and they allow an investigator to choose an optimal detection threshold balancing sensitivity and specificity. BNs have been used for disease detection and diagnosis [9].

Max-Min Hill-Climbing

As an alternative to creating the structure of a BN by hand using expert knowledge, a BN was constructed automatically from the data using machine learning techniques. Many algorithms exist for learning BNs; the MMHC algorithm has been shown to outperform on average a number of other prototypical and state-of-the-art BN learning algorithms in an extensive empirical evaluation and was therefore chosen for use in the analysis [10]. MMHC learns the structure of a BN given a dataset, after which the

network parameters are then estimated directly from the data. The MMHC algorithm is available in the Causal Explorer library[1] [11]. The training dataset was used to learn the network structure and estimate the parameters. The Norsys Netica™ API probabilistic inference algorithm was used to predict asthma in the test dataset; the same that was used in the expert BN. This test dataset was used to calculate an AUC for the method.

Support Vector Machine

SVM's, when used in a binary classification problem such as in this study, construct a maximal margin separating hyperplane to discriminate between the two classes of data [12]. SVMs make use of a kernel function to map the input data to a new space typically of much higher dimensions, where an optimization procedure is run to find the linear separating hyperplane. Several different kernel functions are often considered for classification tasks; the full polynomial and radial-basis kernel functions (RBF) were both used in this task. The full polynomial kernel takes two parameters: the degree d of the function, and C, a cost parameter. The RBF kernel also takes two parameters: the cost parameter C, and which determines the width of the function. SVMs have been used to classify clinical data and in clinical data analysis [13, 14].

The SVM classifiers were implemented using LibSVM [15]. The choice of parameters for the classifiers was optimized with stratified 10-fold nested cross validation over the training dataset using empirical area under the receiver operating characteristic curve as a performance measure. The values for each parameter was selected from the following ranges: $C - \{10^{-6}, 10^{-4}, 10^{-2}, 1, 10^{2}, 10^{4}\}$, $d - \{1, 2, 3, 4, 6, 8\}$, and $- \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\}$. The best classifier was selected and trained using the entire training dataset, and evaluated on the test data set.

Artificial Neural Network

ANNs are modeled after the brain's interconnecting neurons [16]. They use a non-linear approach to create statistical models. ANNs can be used to discover patterns in a dataset, and have been applied to identify asthma patients using responses to a questionnaire [6]. "Learning" occurs through adjusting the connection weights between nodes, finding a result, and then re-adjusting the connection weights.

Our network was developed using the Netlab [16]. A nested 10-fold cross-validation was employed to select the network architecture. The networks were trained and tested using gradient descent with adaptive learning back-propagation with mean squared error as the fitness measure. The best network was selected and trained on the entire training set. An independent test set was then used to calculate a receiver operating characteristic curve (ROC) and asthma probabilities.

Gaussian Process

A GP applies Bayesian techniques to an ANN to create a probabilistic structure that can be used to calculate probabilities. Bayesian methods have been applied to ANNs by placing prior probabilities on the weights of the network. Using Bayesian methods with ANNs elevates the need for a monitoring data set and allows parameters to be determined on the network being trained [17]. GPs are an extension to BNs, but they place prior probabilities on the function.

We used the Gaussian Processes for Machine Learning (GPML) toolbox in Matlab [17, 18] to develop a GP for asthma prediction. The hyperparameters were optimized using the supplied function in the GPML toolbox. We applied a Laplace's Approximation for a binary Gaussian process and selected the commonly used squared exponential covariance function. The hyperparameters are related to the squared exponential covariance function to determine the amplitude and shift of the function in space. The hyperparameter lengths are associated with each variable and determine how much the variable is can vary in its dimension.

Analysis

Performance was evaluated using ROC curves [19]. The ROC curve measures overall test performance and is obtained by plotting sensitivity versus 1-specificity. The area under the ROC curve (AUC) was the primary outcome measure [20] for all techniques with the exception of the SVM. We determined standard operational characteristics for each method including sensitivity, specificity, predictive value, and likelihood. For probabilistic systems, sensitivity was varied from 80%, 85%, 90%, and 95% to determine standard operational characteristics. To compare the methods with the expert-built BN, sensitivity was fixed at 90% as in the previous paper [5].

Results

There were 4,115 patient visits during the study period. Of these, 92 visits were excluded, and 385 (9.6%) had a final diagnosis of asthma. The patient demographics were reported in the previous study. The Bayesian network structure displayed in the left hand side of figure 1 was constructed manually using expert knowledge. The right hand side of figure 1 displays the network structure learned from the data using MMHC.

1 Causal Explorer containing MMHC can be downloaded at http://www.dsl-lab.org.

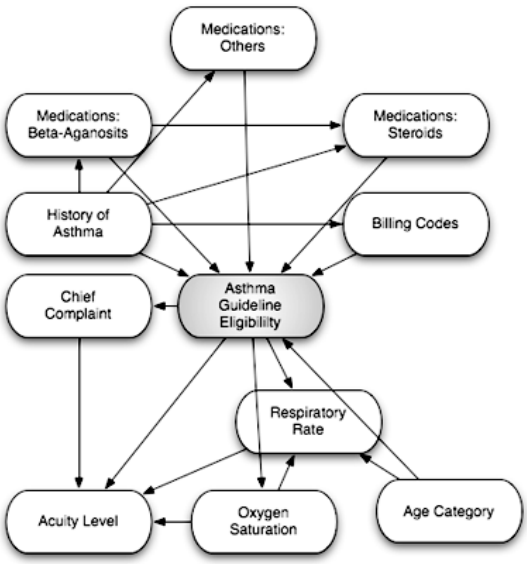
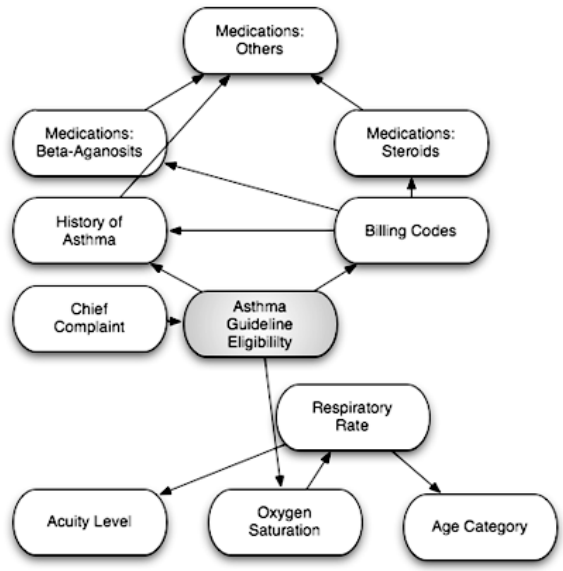


Figure 1 - Expert-built Bayesian Network



MMHC Bayesian Network (structure learned form data)

The AUC for the expert BN was 0.959 (95% CI: 0.933 – 0.977). The MMHC created network AUC was 0.962 (95% CI: 0.935 – 0.980). The ANN AUC with 160 hidden nodes and 160 hidden layers was 0.936 (95% CI: 0.902 – 0.961), and the AUC for the GP was 0.956 (95% CI: 0.923 – 0.976). The ROC curves for the original BN, the BN learned with MMHC, the ANN, and the GP are shown in figure 2. The SVM produces binary predictions with the threshold implicit in the SVM formalism therefore, an SVM ROC curve is not included in this portion of the analysis.

Table 2 shows operational characteristics of the methods with the fixed 90% sensitivity (for methods where the threshold can be varied). The results are reported directly on the predictions made by the SVM since sensitivity cannot be varied. The final parameters selected and used by the SVM was the full polynomial kernel function with a degree of 1 and cost parameter of 1. Operational characteristics for multiple sensitivities were calculated, as in the previous study, for the MMHC network are shown in table 3 for comparison.

Table 2 - Operational characteristics

	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	PLR	NLR
BN	90.0	88.3	44.7	98.9	7.69	0.11
MMHC	90.0	90.1	49.2	98.9	9.16	0.10
ANN	90.0	84.4	38.0	98.8	5.81	0.11
GP	90.0	90.3	49.7	98.9	9.37	0.10
SVM	71.9	98.7	85.2	97.1	54.5	0.29

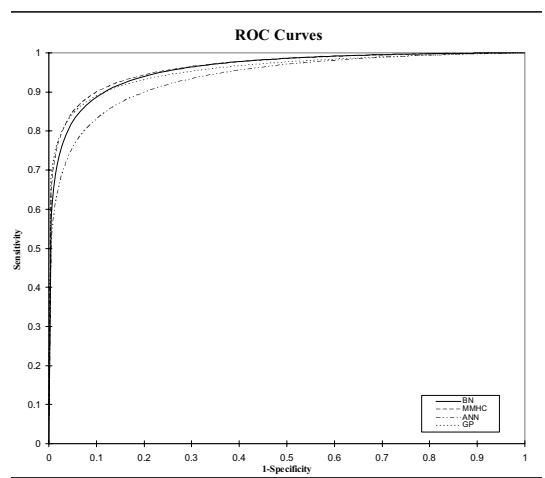


Figure 2 – ROC Curves

The test set had 96 asthma cases and 910 non-asthma cases. At 90% sensitivity, the expert-built BN had 86 true positive predictions and 105 false positive predictions, 805 true negative predictions and 10 false negative predictions. The MMHC network also had 87 true positive predictions and 90 false negative predictions, 820 true negative predictions and 9 false negative predictions. The ANN had 87 true positive predictions and 142 false positive predictions, 768 true negative predictions and 9 false negative predictions. The GP had 87 true positive predictions and 88 false positive predictions, 822 true negative predictions and 9 false negative predictions. At the resulting 72% sensitivity, the SVM had 69 true positive predictions and 12 false positive predictions, 898 true negative predictions and 27 false negative predictions.

Table 3 - MMHC operational characteristics with fixed sensitivity values

SEN (%)	SPEC (%)	PPV (%)	NPV (%)	PLR	NLR
80	97.1	74.8	97.9	28.1	0.20
85	93.8	59.1	98.3	13.7	0.17
90	90.1	49.2	98.9	9.16	0.10
95	86.3	42.1	99.4	6.90	0.06

SEN: Sensitivity, SPEC: Specificity, PPV: Positive Predictive Value, NPV: Negative Predictive Value, PLR: Positive Likelihood Ratio, NLR: Negative Likelihood Ratio

Discussion

The accuracy of the MMHC discovered BN, the SVM, and the GP were comparable to the expert-built BN, however, the ANN did not perform as well as the expert-built BN.

Sparse data may have caused problems in some of the techniques. In our dataset, asthma prevalence was 9.6% of the cases. With this few cases, the SVM and ANN, without adjustment, may not have been able to properly detect the asthma exacerbations. We did not perform any procedures for handling imbalanced data in this study (e.g., over-sampling, under-sampling, one class SVMs, etc). However, such adjustment may lead to better performance of the SVM and ANN.

The MMHC algorithm depends on tests of conditional independence. Accurate estimates for those tests depend on the number of samples and the domain of the variables involved in the testing. For this dataset the Chief Complaint variable can take one of 141 possible values. With the amount of sample provided to the learning algorithm caution must be taken when considering the network produced.

In summary, we believe these methods could be applied to create a real-time detection system for identifying asthma patients using commonly available clinical data.

Acknowledgments

The first author was supported by NLM T15 007450-03.

References

[1] Allergy & Asthma Advocate. Quarterly patient newsletter of the American Academy of Allergy, Asthma and immunology. 2004

[2] National Heart, Lung, and Blood Institute, National Asthma Education and Prevention Program. Expert Panel Report 2: Guidelines for the diagnosis and management of asthma. 1997

[3] Scribano PV, Lerer T, Kennedy D, Cloutier MM. Provider adherence to a clinical practice guideline for acute asthma in a pediatric emergency department. Acad Emerg Med. 2001;8:1147-1152.

[4] Grimshaw JM, Eccles MP, Walker AE, Thomas RE. Changing physicians' behavior: what works and thoughts on getting more things to work. J Contin Educ Health Prof. 2002;22:237-243.

[5] Sanders DL, Aronsky D. Detecting Asthma Exacerbations in a Pediatric Emergency Department Using a Bayesian Network. Proc AMIA Symp. 2006;:684-688.

[6] Sanders DL, Aronsky D. Biomedical informatics applications for asthma care: a systematic review. J Am Med Inform Assoc. 2006. [E-pub].

[7] Hirsch S, Shapiro JL, Turega MA, Frank TL, Niven RM, Frank PI. Using a Neural Network to Screen a Population for Asthma. Ann Epidemiol. 2001;11:369-376.

[8] Sanders DL, Gregg W, Aronsky D. Identifying asthma exacerbations in a pediatric emergency department: A feasibility study. Int J Med Inform. 2006. [E-pub].

[9] Aronsky D, Haug PJ. Diagnosing community-acquired pneumonia with a Bayesian network. Proc AMIA Symp. 1998;;632-636.

[10] Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. Machine Learning. 2006;65:31-78.

[11] Aliferis CF, Tsamardinos I, Statnikov A, Brown LE. Causal Explorer: A Probabilistic Network Learning Toolkit for Biomedical Discovery. In International Conference on Mathematics and Engineering Techniques in Medicine and Biological Sciences (METMBS). 2003;:371-376.

[12] Vapnik V. Statistical Learning Theory. Wiley-Interscience; 1998

[13] Lagor C, Aronsky D, Fiszman M, Haug PJ. Automatic identification of patients eligible for a pneumonia guideline: comparing the diagnostic accuracy of two decision support models. Medinfo. 2001;10:493-497.

[14] Kote-Jarai Z, Matthews L, Osorio A, Shanley S, Giddings I, Moreews F, Locke I, Evans DG, Eccles D, Williams RD, Girolami M, Campbell C, Eeles R. Accurate prediction of BRCA1 and BRCA2 heterozygous genotype using expression profiling after induced DNA damage. Clin Cancer Res. 2006;12:3896-3901.

[15] Chang CC and Lin CJ. LIBSVM: a library for support vector machines. 2001

[16] Nabney IT. Netlab. Springer; 2004

[17] MacKay DJC. Information Theory, Inference & Learning Algorithms. Cambridge University Press; 2002.

[18] Matlab® v 7.3. http://www.mathworks.com

[19] Hanley JA and McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29-36.

[20] Hand DJ and Till RJ. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning. 2001;45:171-186.

Address for correspondence

Judith W Dexheimer, 2209 Garland Ave, Nashville, TN, 37232-8340, judith.dexheimer@vanderbilt.edu