Using Secondary Knowledge to Support Decision Tree Classification of Retrospective Clinical Data*

Dympna O'Sullivan[1], William Elazmeh[3], Szymon Wilk[1], Ken Farion[4], Stan Matwin[1,2], Wojtek Michalowski[1], and Morvarid Sehatkar[1]

[1] University of Ottawa, Ottawa, Canada
{dympna,wilk,wojtek}@telfer.uottawa.ca, mseha092@site.uottawa.ca
[2] Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland
stan@site.uottawa.ca
[3] University of Bristol, Bristol, United Kingdom
elazmah@cs.bris.ac.uk
[4] Faculty of Medicine, University of Ottawa, Ottawa, Canada
farion@cheo.on.ca

Abstract. Retrospective clinical data presents many challenges for data mining and machine learning. The transcription of patient records from paper charts and subsequent manipulation of data often results in high volumes of noise as well as a loss of other important information. In addition, such datasets often fail to represent expert medical knowledge and reasoning in any explicit manner. In this research we describe applying data mining methods to retrospective clinical data to build a prediction model for asthma exacerbation severity for pediatric patients in the emergency department. Difficulties in building such a model forced us to investigate alternative strategies for analyzing and processing retrospective data. This paper describes this process together with an approach to mining retrospective clinical data by incorporating formalized external expert knowledge (*secondary knowledge sources*) into the classification task. This knowledge is used to partition the data into a number of coherent sets, where each set is explicitly described in terms of the secondary knowledge source. Instances from each set are then classified in a manner appropriate for the characteristics of the particular set. We present our methodology and outline a set of experiential results that demonstrate some advantages and some limitations of our approach.

1 Introduction

In his book [1], Motulsky submits "*the human brain excels at finding patterns and relationships ...*". Scientists have long exhibited an aptitude to learn and

* The support of the Natural Sciences and Engineering Research Council of Canada, the Canadian Institutes of Health Research and the Ontario Centres of Excellence is gratefully acknowledged.

Z.W. Raś, S. Tsumoto, and D. Zighed (Eds.): MCD 2007, LNAI 4944, pp. 238–251, 2008.

generalize from observations leading them to develop refined methods for detecting patterns and identifying coherent conjectures drawn from experience. Since their early days, intelligent computer systems have inspired scientists with their promising potential of supporting such research in medical domains [2]. However, medical data features many difficult domain-specific characteristics and complex properties [3]. Incompleteness (missing data), incorrectness (noise), sparseness (non-representative values), and inexactness (inappropriate parameter selection) make up the short list of challenges faced by any machine learning technique applied in the medical domain [4]. A comprehensive overview of these and other challenges is presented in [5], where medical data is described as often being heterogeneous in source as well as in structure, and that the pervasiveness of missing values for technical and/or social reasons can create problems for automatic methods for classification and prediction. Furthermore, translating physicians' interpretations based on years of clinical experience to formal models represents a serious and complex challenge.

An important requirement of medical problem solving or decision support applications is interpretability for domain users [6]. Such a stipulation dramatically reduces the choice of machine learning models that can be applied to medical problem solving to those that can offer systematic justification and explanation of the prediction process. Such models include classifiers that estimate probabilities (probabilistic), classifiers that identify training examples similar to a test example (case-based), classifiers that produce rules that can be applied to a given test example (rule-based), and classifiers that describe decisions based on a selected set of attributes (tree-based). In this work we have chosen to focus our prediction efforts on tree-based classifiers. Decision tree classification models are especially useful in medical applications as a result of their simple interpretation but also as they are represented in the form typically used for describing clinical algorithms and practice guidelines. As such a tree-based classification model can easily be represented in a comprehensible and transparent format if required, without the need for computer implementation.

In this work, the clinical prediction task is centered on the domain of emergency pediatric asthma where the goal is to develop a classification model that can provide an early prediction of the severity of a child's asthma exacerbation. Asthma is the most common chronic disease in children (10% of Canadian population), and asthma exacerbations are one of the most common reasons for children to be brought to the emergency department [7]. The provision of computer-based decision support to emergency physicians treating asthma patients has been shown to increase the overall effectiveness of health care delivered in emergency departments [8,9]. For a patient suffering from an asthma exacerbation, early identification of severity (*mild*, *moderate*, or *severe*) is a crucial part of the management and treatment process. Patients with a *mild* attack are usually discharged following a brief course of treatment (less than 4 hours) and resolution of symptoms, patients with a *moderate* attack receive more aggressive treatment over an extended observation in the emergency department (up to 12 hours), and patients with a *severe* attack receive maximal therapy before ultimately being transferred to an in-patient hospital bed for ongoing treatment (after about 16 hours in the emergency department).

This paper discusses challenges, issues, and difficulties we face in developing a prediction model for early asthma exacerbation severity using retrospective clinical data. Preliminary analysis of the data without preprocessing resulted in unacceptably low classification accuracy. These results forced a rethink of common methodologies for mining retrospective clinical data. Although not particularly complex, this data set is characterized by a fair amount of missing values such that standard methods of feature extraction and classifier tuning fail to produce acceptable performance. Furthermore, clinically-based "classifiers", such as PRAM (section 3.1) cannot be applied due to the type of data being collected. We employ such a clinical classifier as an external method to evaluate the data which leads to the identification of sets where PRAM can or cannot be readily employed We argue that such partitioning will ultimately improve the classification. Our investigations led us to develop a methodology for classification that involves identification and formalization of expert medical knowledge specific to the clinical domain. This knowledge is referred to as a secondary knowledge source and its incorporation allowed us to exploit implicit domain knowledge in the data for more fine-grained data analysis and processing. This paper demonstrates the usefulness of secondary knowledge to partition medical data and ultimately to to reduce its complexity. Our experimental evaluation demonstrates that with such partitioning a decision tree classifier is capable of overcoming some but not all complexities posed by this dataset. An added benefit is the ability to capture other regularities that should be in asthma data according to PRAM, thus in a sense, we "extend" its interpretation.

This paper is organized as follows. In Section 2 we describe the retrospectively collected asthma data used in this analysis. Section 3 outlines a methodology for identifying, formalizing and applying secondary knowledge sources with the purpose of harnessing and exploiting implicit domain knowledge. An experimental evaluation of this approach is outlined in Section 4, where our results display that the approach can be applied with some degree of success. We conclude with a discussion in Section 5.

2 Retrospective Clinical Dataset

The dataset used in this study was developed as part of a retrospective chart study conducted in 2004 at the Children's Hospital of Eastern Ontario (CHEO), Ottawa, Canada. The study includes patients who visited the hospital's emergency department from 2001 to 2003 for treatment of an asthma exacerbation. To illustrate the underlying structure of the data, we present the workflow by which asthma patients are processed in the emergency department (Figure 1). The workflow shows that a patient is evaluated multiple times by multiple caregivers at variable time intervals. This information is documented on the patient chart with varying degrees of completeness. Furthermore, some aspects of evaluation are objective and therefore reliable measures of the patient's status, however

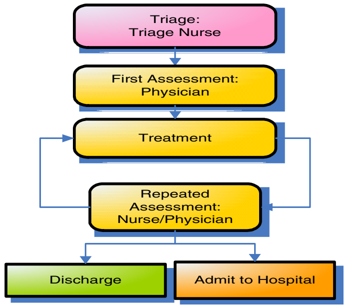

Fig. 1. Asthma Assessment Workflow in the Emergency Department at CHEO

other aspects can be quite subjective and less reliably correlated with the patient's status. In preparing the final dataset, patient information was divided into three subcategories for each record; historical and environmental information, information collected during the triage assessment and information collected at a reassessment approximately 2 hours after triage. The final dataset consisted of 362 records and each record was reviewed by a physician and assigned to one of two classes (*mild* or *moderate/severe*) using predefined criteria related to the duration and extent of treatment required, the final disposition (i.e., discharged or admitted to hospital), and the possible need for additional visits for ongoing symptoms. In this way, the assigned severity group was used as a gold standard for creation and evaluation of a prediction model.

The dynamic nature of asthma exacerbations and the collection of assessments over time would lend itself naturally to a temporal representation for analysis of data. However, inconsistencies in data recording meant it was not possible to incorporate a temporal aspect into the analysis. Further difficulties presented by the data were a significant number of missing values (for some attributes up to 98%), incorrectness, sparseness, and noise due to the variability with which information was recorded, and inexactness due to inappropriate parameter selections as well as the problem of "values as attributes" often encountered in medical data.

3 Secondary Knowledge Sources

Evidence-based medicine is a recent movement that has gained prominence in current clinical practice as a methodology for supporting clinical decision making. The practice of evidence based medicine involves integrating individual clinical expertise with the best available external clinical evidence from systematic research [10]. Individual clinical expertise refers to the proficiency and judgment that individual clinicians acquire through clinical experience and external clinical evidence describes clinically relevant research usually evaluated using randomized control trials. In practice evidence based medicine is applied in a number of ways, including, through the use of clinical practice guidelines, specialty-specific literature and clinical scoring systems.

In this research, we utilize external clinical evidence to support the classification task. The incorporation of a secondary knowledge source into classification leads us to define a three step approach to mining retrospective clinical data. In the first step relevant medical evidence is identified, for example in the form of a clinical practice guideline for the particular clinical domain. The second step is to formalize the medical evidence so it can be applied to available data. The third step involves developing a framework that makes use of the evidence to support the automatic classification task. The advantage of integrating such knowledge is that it allows for more effective and natural organization of information along existing and important data characteristics. As such secondary knowledge can be viewed as a proxy for an expert built classifier and may be incorporated to improve the predictive accuracy of the automatic classification task.

3.1 Secondary Knowledge Sources for Pediatric Asthma

The secondary knowledge source identified as relevant for our retrospective asthma data is the Preschool Respiratory Assessment Measure (PRAM) asthma index [11]. PRAM provides a discriminative index of asthma severity for preschool children. It is based on five clinical attributes commonly recorded for pediatric asthma patients, *suprasternal indrawing, scalene retractions, wheezing, air entry* and *oxygen saturation*. PRAM is based on a 12 point scale (see Table 1) and is calculated using scores of 0, 1, 2, and 3. These scores are assigned to attributes depending on the presence or absence of values as well as observed increasing or decreasing values of attributes. PRAM has been clinically validated as a reliable and responsive measure of the severity of airway obstruction. A patient with a PRAM score of 4 or less is considered to have a *mild* exacerbation,

Table 1. PRAM Scoring System

Signs	0	1	2	3
Suprasternal indrawing	absent		present	
Scalene retractions	absent		present	
Wheezing	absent	expiratory	inspiratory and expiratory	Audible without stethoscope /absent with no air entry
Air entry	normal	decreased bases	widespread decrease	absent/ minimal
Oxygen saturation	≥95%	92-95%	<92%	

a score between 5 and 8 corresponds to a *moderate* exacerbation, and a score of 9 or higher corresponds to a *severe* exacerbation.

In order to identify if the PRAM scoring system was appropriate secondary knowledge, the retrospective asthma dataset was analyzed for the presence of PRAM attributes. It was found that four of the five PRAM attributes were present in our data and values for these attributes may be collected twice for each record, once at triage and again at reassessment. The next step of our approach was to formalize the secondary knowledge source so that it could be applied to the classification task. This process is described in the next subsection.

3.2 Formalizing Secondary Knowledge Sources for Classification

The formalization of the secondary knowledge source involved determining a mapping from the set of attributes outlined by PRAM to a subset of attributes from the retrospective asthma data and an associated assignment of scores for attribute values. This was necessary as not all attributes required to calculate the PRAM score were present in the retrospective asthma data, and for some other attributes a 1:1 mapping did not exist. Specifically, the retrospective data did not contain an attribute corresponding to "Suprasternal Indrawing", and "Wheezing" was captured using two attributes in the retrospective data, inspiratory wheezing and expiratory wheezing. Also, the PRAM scoring system describes "Air Entry" using four values (normal, decreased bases, widespread decrease and absent/minimal), whereas our data defined air entry as either good (i.e., normal) or reduced. Therefore the formalized mapping was developed in conjunction with a domain expert (emergency physician), and the rules devised

Table 2. Mapping PRAM attributes and scores

Attribute(s)	Value(s)	Score
Oxygen Saturation	Greater than 95%	0
Oxygen Saturation	Greater than 92% and Less than 95%	1
Oxygen Saturation	Greater than 88% and Less than 92%	2
Oxygen Saturation	Less than 88%	3
Air Entry	Good	0
Air Entry (class = mild)	Reduced	1
Air Entry (class = other)	Reduced	3
Retractions AND Air Entry	Absent AND Good	0
Retractions AND Air Entry	Absent AND Reduced	1
Retractions AND Air Entry	Absent AND "Missing"	2
Retractions	Present	2
Expiratory AND Inspiratory Wheeze	Absent AND Absent	0
Expiratory AND Inspiratory Wheeze	Present AND Absent	1
Expiratory AND Inspiratory Wheeze	Present AND Present	2
Expiratory AND Inspiratory Wheeze	Absent AND Present	Undefined

for mapping attributes used by the PRAM system to attributes in our data and their corresponding score assignments are shown in Table 2.

3.3 Building a Classifier by a Secondary Knowledge Source

The final step of our approach was to use secondary knowledge to build a model for predicting asthma severity. In the retrospective asthma data a decision (class label) is recorded for each patient along with clinical and historical information. This class is the final outcome for the patient as recorded in the patient chart (not the result of the assessment at the 2-hour point) and indicates whether the patient has suffered a *mild* or *moderate/severe* exacerbation. Using the attributes, values and associated scores mapped from the PRAM scoring system we calculated a PRAM score for each patient in the dataset. This score had possible values between 0 and 12, where a score of less than 5 indicated a *mild* exacerbation and a score of greater than 5 indicated a *moderate/severe* exacerbation. (In our data the *moderate* and *severe* categories outlined by PRAM are collapsed into one group, *moderate/severe*). The score is then compared with the class label for each record in the dataset and the set of patients who comply with the PRAM scoring system are identified. The assignment of PRAM scores allows for the dataset to be partitioned into instances for which all PRAM attributes were present and thus a complete and correct PRAM score could be calculated and instances for which only a partial or no PRAM score could be calculated due to the absence of values for the PRAM attributes. PRAM attributes may be collected at two stages in the asthma workflow (triage and reassessment), however analysis of our data demonstrated that such attributes were more likely to be collected at reassessment (there were many missing values for triage attributes) and as such the dataset was partitioned using the larger set of reassessed values. This resulted in a dataset with 147 instances for which the PRAM score was complete and correct, 206 instances where only a partial or no PRAM score could be calculated due to missing values and 9 instances for which the score calculated by PRAM and the class label completely disagreed. These 9 cases were considered outliers and deleted from the dataset for evaluation.

4 Experimental Evaluation

4.1 Experimental Design

Our evaluation reports results from a number of experiments involving the retrospective asthma dataset where each experiment involved building a decision tree using the J48 decision tree classifier in Weka[12] to classify data. The first experiment involved building a classifier on the entire dataset prior to any application of secondary knowledge. These results serve as a baseline for classifier performance upon which to evaluate all subsequent results. In the next experiment secondary knowledge in the form of the PRAM scoring system was applied to partition the dataset into two sets, one containing PRAM complete and correct instances and one containing PRAM partial or incomplete instances. The purpose of this experiment is to demonstrate that the incorporation of secondary knowledge into the classification tasks allows for enhanced representation of data which results in reducing the complexity of the retrospective clinical dataset for classification. In the final experiment we applied feature selection to the complete original data set twice, once using automatic feature selection (available in Weka), and a second time by manually selecting combinations of expertly selected attributes and removing the remaining attributes. The function of this experiment was to show that neither automatic or expert feature selection can identify and reduce complexities in the data as efficiently as a classifier that incorporates secondary or expert medical knowledge, selects important features and partitions data into sets of similar characteristics.

For each experiment we report classifier performance in terms of percentages of Sensitivity (Sens) and Specificity (Spec), Predictive Accuracy (Acc) and Area Under the Curve (AUC) on the positive class. Sensitivity (the true positive rate) measures how often the classifier finds a set of positive examples. For instance in this research we consider *moderate/severe* to be the critical/positive class, therefore the sensitivity of *moderate/severe* measures how often the classifier correctly identifies patients suffering *moderate/severe* asthma exacerbations. Specificity (1 - false positive rate) measures how often what the classifier finds, is indeed what it was looking for. Therefore the specificity of the positive class (*moderate/severe*) measures how often what the classifier predicts is indeed a patient with a *moderate/severe* asthma exacerbation. Analyzing the trade-off between sensitivity and specificity is common in medical domains and is analogous to Receiver Operating Characteristics (ROC) analysis [13,14] used in machine learning [15].

In addition we report accuracy and AUC where accuracy is the rate at which the classifier classifies patients (in both classes) correctly while AUC represents the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [16]. Hence AUC measures the classifier's ability to discriminate the positive class from the negative class. In our experiments, we aim to analyze decision tree performance by measuring its ability to discriminate each positive patient with a *moderate/severe* asthma exacerbation. For a given classifier and positive class, an ROC curve [15,16] plots the true positive rate against the rate of false positives produced by the classifier on the test data. The points along the ROC curve are produced by varying the classification threshold from most positive classification to most negative classification and the AUC of a classifier is the area under the ROC curve [16]. For this reason as well as the relatively small size of the dataset we evaluate the classifier for each patient in the dataset using leave-one-out cross-validation.

4.2 Classifying the Entire Dataset

The first experiment involved building a decision tree on the original dataset of 362 instances. The results of this experiment are shown in the first row of Table 3 and demonstrate that the retrospective clinical dataset is complex and that good classification accuracy is difficult to achieve without performing some degree of data preprocessing. We include these results as a baseline by which to measure subsequent classifier performance.

4.3 Classifying PRAM and Non-PRAM Sets

In this experiment the dataset was partitioned by applying the formalized mapping from PRAM scoring system to attributes from the retrospective asthma dataset. This resulted in the dataset being partitioned into those that were PRAM complete and correct (PRAM set) and those that were either PRAM partial or incomplete (non-PRAM set). A decision tree was built for each set and the results are shown in the last two rows of Table 3. Also included for reference purposes are the results for the entire dataset in the first row.

Table 3. Decision Trees built on PRAM and non-PRAM sets

Set	Size	Sens	Spec	Acc	AUC
Entire	362	73	63	69	69
PRAM Set	**147**	**93**	**96**	**95**	**98**
Non-PRAM Set	**206**	**89**	**53**	**74**	**77**

From Table 3 we observe that splitting the datset into different sets based on formalized secondary knowledge increases classification accuracy of the PRAM set. For the non-PRAM set classification improves in terms of Sensitivity, Accuracy and AUC. In particular sensitively on the PRAM set increases by 20% from the baseline. In addition a large gain in AUC from the baseline reflects the increased probability that a positive example is ranked higher than a negative example. In fact, when the decision tree is supplemented with secondary knowledge (the PRAM set) we gain an increase in AUC, and when secondary knowledge cannot be so easily applied (non-PRAM set), the performance only improves marginally on that of the baseline. These results demonstrate that the incorporation of formalized secondary knowledge sources can help with classification in such domains "by exposing" the concept to be learned by the decision tree and thus reducing the overall complexity of the dataset by exploiting domain knowledge implicitly present in the data. The results represent an overall improvement on previous research into classification of clinical data with tree-based classifiers [8,9].

However from the results in Table 3 we also observe a decrease of 10% in specificity between the Non-PRAM set and the baseline. This performance is inadequate in terms of achieving a balance between high sensitivity and high specificity. We note however that in terms of the problem domain high sensitivity and low specificity on the positive class translates to the fact that the classifier is very accurate in identifying *moderate/severe* patients and recommending they are kept for an extended time in the emergency department, however at the same time the classifier is over conservative in recommending that *mild* patients are

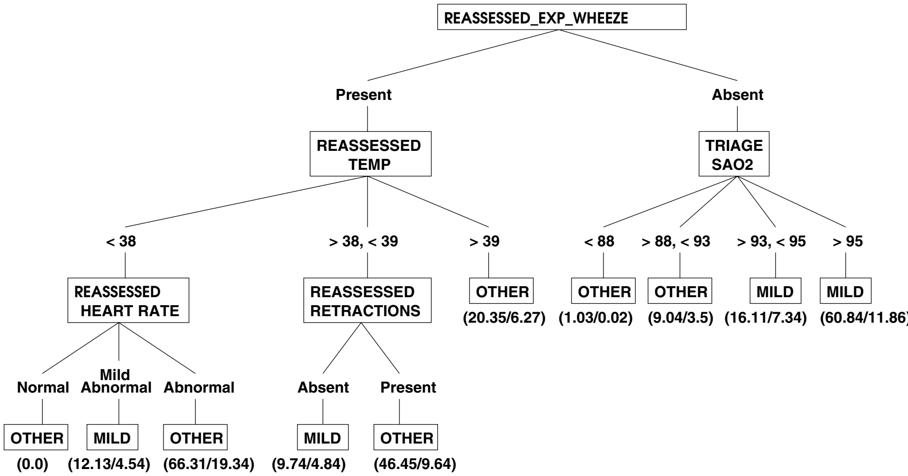

Fig. 2. The resulting decision tree for the classifier trained on the entire data

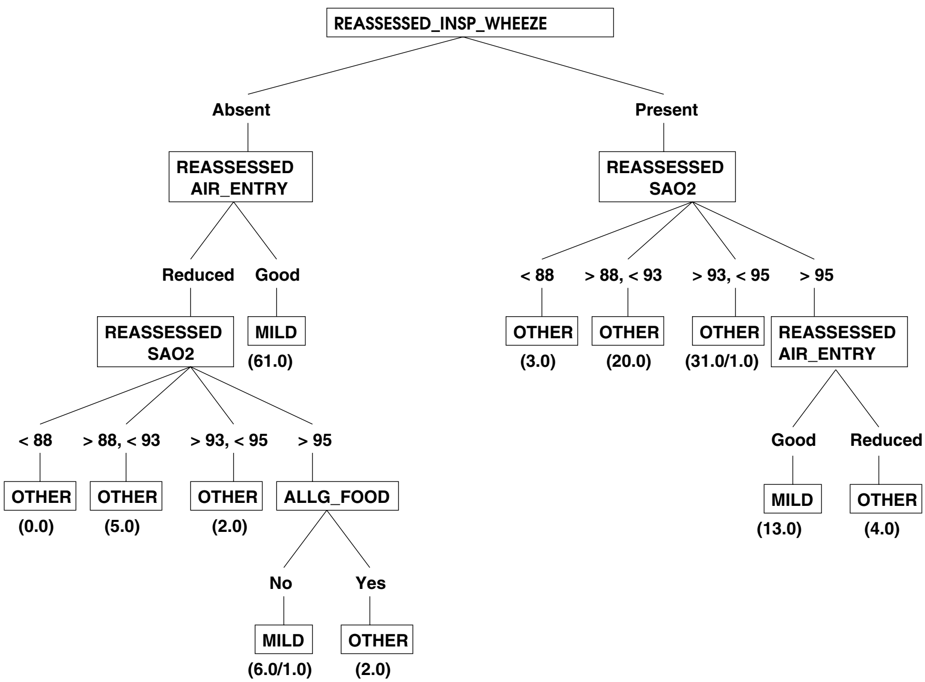

Fig. 3. The resulting decision tree for the classifier trained on PRAM data

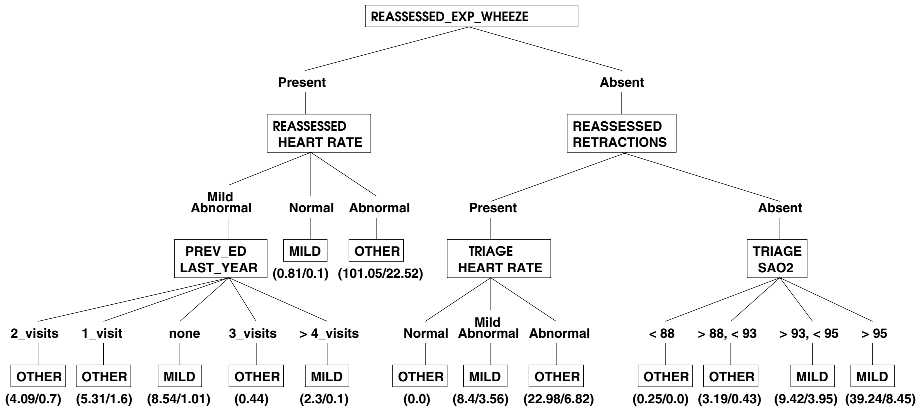

Fig. 4. The resulting decision tree for the classifier trained on Non-PRAM data

kept for longer than usual stays. Such direction of classification a less serious error than one occurring in the opposite situation.

Furthermore, consider figures 2, 3 and 4 which present the resulting decision tree classifier obtained from training with the entire data, PRAM data only, and Non-PRAM data respectively. We clearly see that the classifier trained on the entire data (2) focuses on clinical measures recorded during the `TRIAGE` and the `REASSESSMENT` phase of the workflow, yet, this classifier delivers the worst classification performance. On the other hand, building the other two classifiers (figures 3 and 4) allows for the consideration of other historically relevant attributes, such as `ALLG_FOOD` in figure 3 and `PREV_ED_LAST_YEAR` in figure 4. Despite the difference in all of these classifiers, lower levels of clinically measured Oxygen saturation remains an indication of severe asthma exacerbation, see `REASSESSED_SAO2` and `TRIAGE_SAO2` subtrees in all three figures.

An important observation is that most PRAM attributes (data attributes mapped to attributes shown in table 1 on page 242) are being used in all three decision trees. This illustrates the significant relevance of these PRAM attributes to the classification task, thus, providing data-driven evidence to support PRAM. In addition, these decision trees allows us to extend PRAM and present a more fine grained representation of it, including dependencies which could be easily explained and presented to physicians.

4.4 Automatic and Expert-Driven Feature Selection

It is acceptable to state that a reduction in dimensionality of data can reduce the complexity of underlying concepts that it may represent. The purpose of this experiment is to demonstrate that data complexity in this domain requires more than dimensionality reduction to reduce its complexity. We compare results obtained from applying automatic and expert-driven feature selection methods to those obtained by partitioning the data according to PRAM secondary knowledge. Automatic feature selection is based on standard methods used by the data

Table 4. Automatic and Expert Feature Selection

Feature Selection Mode	Mode	Size	Sens	Spec	Acc	AUC
Information Gain	Automatic	362	72	63	68	69
Chi-squared	Automatic	363	72	63	68	69
Combinatorial	Automatic	362	72	65	69	71
Wrapper with Naive Bayes	Automatic	362	71	60	70	77
On All Attributes	Expert	362	72	66	70	73
On Only PRAM Attributes	Expert	362	77	78	70	71

mining community and are available in the Weka software. The expert-driven feature selection methods are based on selecting attributes observed to useful to classification from our repeated experiments and by an expert and those outlined by the PRAM scoring system. 10 methods of feature automatic selection were applied to the dataset where each was used in conjunction with a decision tree for classification. The results for the best four methods are shown in rows 1-4 of Table 4. Comparing the results for automatic feature selection to those for the baseline as outlined in Table 3 we can conclude it is not successful in reducing the complexity of the dataset. In general results do not display any improvement in classification except in the case where a wrapper using a Naive Bayes classifier for optimization is used for feature selection. Here we note an increase in AUC, however this is at the expense of a large decrease in specificity. In applying expert feature selection, we built one classifier using all data records of attributes collected during the reassessment only and another classifier using only the attributes that were mapped from the PRAM scoring system while still using all instances available in the dataset. The results for these two experiments are shown in rows 5-6 of Table 4. Again comparing these results to those outlined for the baseline in Table 3 we observe no significant improvement.

However, by comparing the results from Table 4 to those for classification on the PRAM and non-PRAM sets in Table 3 a number of important conclusions can be drawn. Partitioning data into different sets for classification based on secondary knowledge results in much improved classification that of using either automatic or expert feature selection. Augmenting the developed classifier with external knowledge allows for more effective classification by exploiting underlying domain knowledge in the dataset and by organizing data according to these concepts. Such classification accuracy cannot be captured by a classification model developed on the data alone. The partitioning of data does not reduce the dimensionality of the dataset like traditional methods for classification such as feature selection, however it manages to reduce the complexity of the dataset by using secondary knowledge to identify more coherent sets into which data more naturally fits.

The intention is to use the classification results from the PRAM and non-PRAM sets from Table 3 to implement a prediction model for asthma severity. This can be achieved in a number of ways. One option is to develop a metaclassifer that could learn to direct new instances to either the model built on the

PRAM set or the model built on the non-PRAM set. For such a metaclassifier values of PRAM attributes alone may be sufficient to make the decision or it may be necessary to develop a method by which unseen patients can be related to the sets (PRAM and non-PRAM) we identify in the dataset. Alternatively the predictions from both sets could be combined to perform the prediction task. One option is to use a voting mechanism, another is to build these classifiers in a manner that produce rankings of the severity of the exacerbation. With such a methodology the classifier with the highest ranking provides a better insight into the condition. However, such an approach introduces additional issues in terms of interpretations and calibrations of ranks and probabilities. Such a study remains as part of our future research directions.

5 Discussion

We have introduced an approach to mining complex retrospective clinical data by incorporating secondary knowledge to supplement the classification task by reducing the complexity of the dataset. The methodology involves identification of a secondary knowledge source suitable for the clinical domain, formalization of the knowledge to analyze and organize data according to the underlying principle of the secondary knowledge, and incorporation of the secondary knowledge into the chosen classification model. In this research we concentrated on classifying information using a decision tree to satisfy the requirement that classification should be easily interpreted by domain users. From our experimental results we draw a number of conclusions. Firstly we have demonstrated that domain knowledge is implicit in the data as the dataset partitions naturally into two sets for classification with the application of a formalized mapping from the PRAM scoring system. This is in spite of the fact that the mapping was inexact; our dataset only contained four of the five attributes outlined by PRAM and some attribute values had slightly different representations. In such a way the application of secondary knowledge reduces the complexity of the dataset by allowing for the exploitation of underlying domain knowledge to supplement data analysis, representation and classification. As outlined, this approach is more successful than traditional methods for reducing data complexity such as feature selection which fail to capture a measure of the expert knowledge implicit in the retrospectively collected data. A further advantage of the approach was demonstrated by the ability of the secondary knowledge to help identify outlier examples in the data. However, the results are still somewhat disappointing in terms of achieving a balance acceptable in medical practice between high sensitivity and high specificity in the non-PRAM set. We believe that a high proportion of missing values in this set is causing difficulties for the classification model. This issue remains an open problem for future research. In other future work we are interested in further investigating attributes used by the PRAM system and to test whether all attributes used by PRAM are necessary for enhanced classification.

References

1. Motulsky, H.: Intuitive Biostatistics. Oxford University Press, New York (1995)
2. Ledley, R.S., Lusted, L.B.: Reasoning foundations of medical diagnosis. Science 130, 9–21 (1959)
3. Mullins, I.M., Siadaty, M.S., Lyman, J., Scully, K., Garrett, C.T., Miller, W.G., Muller, R., Robson, B., Apte, C., Weiss, S., Rigoutsos, I., Platt, D., Cohen, S., Knaus, W.A.: Data mining and clinical data repositories: Insights from a 667,000 patient data set. Computers Biology and Medicine 36(12), 1351–1377 (2006)
4. Magoulas, G.D., Prentza, A.: Machine learning in medical applications. In: Paliouras, G., Karkaletsis, V., Spyropoulos, C.D. (eds.) ACAI 1999. LNCS (LNAI), vol. 2049, pp. 300–307. Springer, Heidelberg (2001)
5. Cios, K.J., Moore, G.W.: Uniqueness of medical data mining. A. I. in medicine 26(1-2), 1–24 (2002)
6. Lavrac, N.: Selected techniques for data mining in medicine. Artificial Intelligence in Medicine 16(1), 3–23 (1999)
7. Lozano, P., Sullivan, S., Smith, D., Weiss, K.: The economic burden of asthma in us children: estimates from the national medical expenditure survey. The Journal of allergy and clinical immunology 104(5), 957–963 (1999)
8. Kerem, E., Tibshirani, R., Canny, G., Bentur, L., Reisman, J., Schuh, S., Stein, R., Levison, H.: Predicting the need for hospitalization in children with acute asthma. Chest 98, 1355–1361 (1990)
9. Lieu, T.A., Quesenberry, C.P., Sorel, M.E., Mendoza, G.R., Leong, A.B.: Computer-based models to identify high-risk children with asthma. American Journal of Respiratory and Critical Care Medicine 157(4), 1173–1180 (1998)
10. Sackett, D., Rosenberg, W., Muir Gray, J., Haynes, R., Richardson, W.: Evidence based medicine: what it is and what it isn't. British Medical Journal (1996)
11. Chalut, D.S., Ducharme, F.M., Davis, G.M.: The preschool respiratory assessment measure (pram): A responsive index of acute asthma severity. Pediatrics 137(6), 762–768 (2000)
12. Witten, I.H., Frank, E.: Data mining: Practical machine learning tools and techniques (2005)
13. Sox, H.C.J., Blatt, M.A., Higgins, M.C., Marton, K.I.: Medical Decision Making, Boston (1998)
14. Faraggi, D., Reiser, B.: Computer-based models to identify high-risk children with asthma. American Journal of Respiratory and Critical Care Medicine 157(4), 1173–1180 (1998)
15. Provost, F., Fawcett, T.: Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. In: The Third International Conference on Knowledge Discovery and Data Mining, pp. 34–48 (1997)
16. Fawcett, T.: Roc graphs: Notes and practical considerations for data mining researchers, Technical Report HPL-2003-4, HP Labs (2003)