Association for Information Systems AIS Electronic Library (AISeL)

ECIS 2010 Proceedings

European Conference on Information Systems (ECIS)

2010

An Early Warning System for Hospital Acquired Pneumonia

Franziska Oroszi *Friedrich-Schiller-Universität Jena,* Franziska.Oroszi@med.uni-jena.de

Johannes Ruhland Friedrich-Schiller-Universität Jena, J.Ruhland@wiwi.uni-jena.de

Follow this and additional works at: http://aisel.aisnet.org/ecis2010

Recommended Citation

Oroszi, Franziska and Ruhland, Johannes, "An Early Warning System for Hospital Acquired Pneumonia" (2010). ECIS 2010 Proceedings. 93. http://aisel.aisnet.org/ecis2010/93

This material is brought to you by the European Conference on Information Systems (ECIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2010 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.



AN EARLY WARNING SYSTEM FOR HOSPITAL ACQUIRED PNEUMONIA

Journal:	18th European Conference on Information Systems
Manuscript ID:	ECIS2010-0167.R1
Submission Type:	Research Paper
Keyword:	Data mining, Health informatics/health information systems/medical IS, Decision and activity analysis, User involvement/participation



AN EARLY WARNING SYSTEM FOR HOSPITAL ACQUIRED PNEUMONIA

Oroszi, Franziska, Universitätsklinikum Jena, Bachstraße 18, 07745 Jena, Germany, Franziska.Oroszi@med.uni-jena.de

Ruhland, Johannes, Friedrich-Schiller-Universität Jena, Carl-Zeiß-Str. 3, 07743 Jena, Germany, J.Ruhland@wiwi.uni-jena.de

Abstract

Pneumonia is a dangerous, often fatal secondary disease acquired by patients during their stay at Intensive Care Units. ICU patients have scores of data collected on a real time basis. Based on two years of data for a large ICU, we develop an early warning system for the onset of pneumonia that is based on Alternating Decision Trees for supervised learning, Sequential Pattern Mining, and the stacking paradigm to combine the two. Mainly due to decreased stay, the system will save \in 180000 in this hospital alone while at the same time increasing the quality and consistent standard of health care. The ultimate system relies on a rather small numeric data base alone and is thus amenable to integration in a treatment protocol and a newly conceived ICU workflow system.

Keywords: Medical Data – Pneumonia – chaining techniques – ADTrees – sequential pattern mining - treatment protocol

1 INTRODUCTION

The following paper is based on an interdisciplinary project between researchers from the department of information systems of Friedrich Schiller University (FSU), Jena, an intensive care unit (ICU) and the hospital pharmacy of the same institution. Objective of the project was the application of data mining techniques to the ICU database. Focus was set on detecting pneumonia, with emphasis on early identification of the developing illness. Beyond its immediate medical utility, the goal was also to explore the potential of secondary data usage in a project with medical as well as economic value. Unlike conventional medical studies which are typically hypothesis driven and based on the analysis of experiments, this project was data driven. However, to ensure acceptance on behalf of the physicians and later integration into a workflow system, it was required to develop a comprehensive and comprehensible model. As a consequence, black box approaches of data analysis like support vector machines or neural networks were immediately rejected. As organizational framework we followed the cross industry standard process model for data mining (CRISP-DM) to structure the project (Chapman et al. 2000).

Pneumonia is an inflammatory illness of the lung. Nosocomial pneumonia is defined as a lung infection occurring after the hospital admission (Robert-Koch-Institut, 2000). Nosocomial pneumonia is among the most frequent infections in the intensive care unit (primarily due to excessive use of ventilation) and shows the highest mortality rate (Reinhart et al., 2006). Patients who developed pneumonia had a prolonged length of stay¹ in the intensive care unit and as a result increased costs of hospitalization (Safdar et al., 2005 and Heyland et al., 1999). An adequate, timely and sufficiently high dosed initial antibiotic treatment is crucial for a successful therapy and reduces the length of stay for pneumonia patients (Welte, 2006 and Dupont et al., 2000 and Battleman et al., 2002). This measure cannot be applied on a purely preventive basis as the risks of severe side effects with already seriously ill patients and of immunization of germs through too widespread application must be taken into account. To exacerbate the problem, the disease will fully develop very quickly, i.e. within days after infection and tends to be overlooked or underestimated in the typically ultra-hectic environment of an ICU.

Data mining techniques have been occasionally employed in the context of diagnosis of pulmonary diseases. Xuanyang et al. (2005) have applied methods of image recognition and classification trees to identify SARS patients from X-ray imagery. Hripcsak et al. (2007) used text mining methods on narrative clinical databases to reliably diagnose pneumonia. Our approach is different from both papers in that it relies on the continuous stream of numeric data that online monitoring in an ICU provides and puts large emphasis on timely reaction.

This paper is structured as follows: In section 2 we describe the data set, which was used in the project. The relevant methods are shortly explained in chapter 3. Experimental results are presented in section 4. A prototypical workflow system called Imagina has been implemented at the ICU under study. We will sketch integration of our results in section 5. Finally, we give some remarks about lessons learned in a broader context.

2 DATA SET

The intensive care unit at FSU has a capacity of 78 beds and is one of the largest university ICUs in Germany. Data are collected in a patient data management system. Figure 1 gives a schematic overview of the data.

¹ Safdar et al. (2005) report an average stay increase of 6.1 days and increased costs of 10,019\$, Heyland et al. (1999) a longer stay of 6.55 days in the average resulting in 14,600 DM additional costs per case.



Figure 1 Systematic overview of the intensive care unit data

The data are extremely sparse which is caused by a considerable number of short stays (less than one day on ICU) and the specific tests and treatments given to each patient leading to diverse and almost patient-specific parameter sets. In line with the goal of identifying a widespread secondary disease, only data routinely collected were considered in this study. Large parts of the data are time dependent with different time granularities. For instance body temperature is measured every half an hour, tracheal secretions three times a day. Other values are only analyzed on demand. Cleansing, transformation and aggregation of the data had to observe these restrictions. Automatically measured data underwent an extra processing step to ensure reliable values. Even in the ICU environment with its high quality requirements and work ethic, data errors and inconsistencies remained an issue.

To deploy methods of supervised learning on the database, cases with and without pneumonia had to be unambiguously identified. The Clinical pulmonary infections score (CPIS) is considered the Gold Standard by pulmologists (Calandra et al., 2005). Pneumonia is confirmed when CPIS sums up to ≥ 6 and a new or progressive radiographic infiltrate (Figure 2) is observed. The first day where this condition holds will in the following be named *CPIS reaction day* and is considered the operationalization of the onset of pneumonia.

CPIS, clinical pulmonary infection score The CPIS scoring system grades each of the six features on a scale from 0 to 2, as follows: tracheal secretions: 0 = rare, 1 = abundant, 2 = purulent; radiographic infiltrates: 0 = absent, 1 = patchy or diffuse, 2 = localized; fever (°C): 0 = \geq 36.5 and \leq 38.4, 1 = >38.4 and \leq 38.9, 2 = \geq 38.9 or <36; leukocytosis (/mm3): 0 = \geq 4000 and \leq 11,000, 1 = <4000 or >11,000, 2 = >4000 or <11,000 and \geq 500 band forms; PaO2/FIO2: 0 = >240 or acute respiratory distress syndrome (ARDS), 2 = \leq 240 and no ARDS; microbiology: 0 = negative; 2 = positive.

Figure 2 Clinical pulmonary infection score.

As not all components of the CPIS could be extracted from the data base and had in part to be added manually, the calculation was a time consuming step and also restricted the number of cases. Finally, all pneumonia patient data from two years (2004 - 2005) were analyzed. In both years, they were 11,726 cases in total at the ICU. Out of this were 1,160 cases considered relevant for pneumonia. CPIS could be calculated for 764 patients on 8,868 ICU days, the rest must be excluded due to data situation. For 364 of those 764 patients pneumonia could be verified. It might surprise that pneumonia was confirmed only in half of the cases, but it must be kept in mind that a fast treatment is crucial in

case of pneumonia. So, medical practice displays strong preference for false alert over missed pneumonia.

Technically speaking, identifying reaction days is classification task with a highly unbalanced class distribution. To handle this situation, we split the whole data set to obtain an artificially balanced training set and a test set reflecting the real class distributions (Weiss, 2001). Cases were randomly chosen for training and for test set. The training set consists of 299 pneumonia and 300 no-pneumonia patients. As for the test set we defined a 5% portion of pneumonia patients, a relation which meet the natural situation at UKJ and other studies (Bercault & Boulain, 2001; Alp et al., 2004; Vahle, 2004). There were 65 pneumonia and 1,250 no-pneumonia patients in the test set.

Following the data collection process, the variables had to be selected. Laboratory data, medical scores and diagnoses – routinely collected data – were tested for relevance. Additional requirement was a measurement in at least 30 percent of the cases. Attributes were submitted to a pre-selection step. Only attributes that showed a minimum predictive power for pneumonia prediction (operationalized as a significance threshold in oneway ANOVA) were considered eligible for further data mining. Finally, 23 variables were chosen for further consideration.

For data mining purposes, the values at the CPIS reaction day and on the three days before were collected. Older values were excluded from a medical point of view as pneumonia incubation time is below 3 days. Values measured after the CPIS reaction day were not considered due to the goal of identifying the onset and not the development of pneumonia. As a side product of this study, a modified CPIS score was also developed that is highly correlated with the original but can be calculated on the basis of automated data alone. Furthermore, as all time series data for pneumonia patients are measured relative to the reaction day, we had to find an equivalent of the CPIS reaction day for the no-pneumonia patients in the sampling. We analyzed on which day of stay at ICU pneumonia had developed and mirrored this distribution on no-pneumonia patients. On this cohort this day was called the reference day.

3 METHODS

Methods employed were selected for their predictive power as reported in the literature, their ease of communication to physicians and their potential of eventual manual intervention. We will only sketch the ideas here and refer to the literature for a detailed description of the algorithms. Regarding the spate of methods proposed, some arbitrariness in method selection does of course remain.

3.1 Alternating Decision Trees

Alternating Decision Trees (ADTree) combine decision trees with a user-selectable number of boosting steps to generate decision rules. Compared to other decision trees like C4.5 or CART, ADTrees often result in more compact tree structures and are easier to interpret (Freund & Mason, 1999). ADTrees use the boosting algorithm AdaBoost (Freund & Schapire, 1999) for learning. The final tree consists of alternating layers of decision and prediction nodes.

In Figure 3 the same classification rule is represented by a decision tree and by an alternating decision tree. In addition to the well known decision nodes (rectangles), ADTrees contain prediction nodes (ellipses). To classify an instance; all paths from the root to leaves reachable for the current instance are considered. When a path reaches a decision node, it continues with the branch that corresponds to the decision outcome, as in a standard decision tree. On the other hand, when a path reaches a prediction node, all offspring nodes are considered. Unlike decision trees, the classification is not from the label of a leaf; but from collecting and summing up all scores from prediction nodes along any path visited. In the two-class case, class prediction is taken from the sign of the total score (see Fig. 3b). Hence, the contribution of each decision node may be understood in isolation, and summing the individual contribution yields the final prediction and classification (Liu et al., 2005).



Figure 3 Decision Tree (a) versus Alternating Decision Tree (b) [13] A case with a=4 and b=2 collects 0.5-0.7+0.4 = 0.2 points and is classified as "positive score"-class.

3.2 Sequential Pattern Mining

Sequential Pattern Mining (SPM) as an extension of Association Rule Mining was originally proposed by Agrawal and Srikant (Agrawal & Srikant, 1995). It describes the search for frequent subsequences within a given set of sequences, where a predefined support threshold has to be observed.

SPM-algorithms have a lot of attractive features: they can not only process metric variables but also nominal and categorical values. They can handle many variables simultaneously with acceptable complexity and speed as well as differently scaled attributes, and work stable in the presence of missing values. Due to these properties, SPM-algorithms are of high interest for applications in medical area, where the problems cited are routinely encountered.

There are two main approaches of SPM-algorithms – with or without candidate generation. Candidate generation algorithms like GSP (Srikant & Agrawal, 1996) or SPADE (Zaki, 2001) start with counting the support of single items to find frequent patterns. Only items that fulfill a defined minimum support are relevant for further testing and are considered as candidates for longer patterns. They are using the apriori-principle for reducing the number of possible candidates. Prefix-Span (Pei et al., 2001) and Free-Span (Han et al., 2000) represent sequential pattern algorithms without candidate generation. They also start by counting the support of single items. In a next step the frequent single items are sorted in descending order. Those frequent single items are used as prefixes and separate the whole set into subsets which are recursively searched for frequent patterns. Hence, this is a divide-and-conquer-approach.

For our project we have chosen the SPAM algorithm (Ayres et al., 2002), a representative of candidate generation algorithms, which is a very fast algorithm and offers a complete download and documentation. SPAM uses a bitmap representation of the data. This allows an efficient support counting, significant bitmap compression and fast join-operations to generate new candidates for subsequences. In Figure 4, the initial transformation of given sequences is displayed.



Figure 4 Tree SPAM algorithm - transformation into a bitmap

SPAM uses a lexicographic sequence tree (Figure 5) for finding sequential patterns. By traversing the tree, candidate sequences are generated, followed by support counting to test the frequency.



Figure 5 The Lexicographic Sequence

SPAM differentiates sequence-extension steps (S-Steps) and itemset-extension steps (I-Step). In a sequence-extension step a new item will be attached at the end of the parent sequence, so the appended transaction will occur later than all other items of the parent sequence. In case of an itemset-extension, the new transaction will take place at the same time as the last itemset of the parent sequence.

Both extensions are illustrated in Figure 5. Processing an I-Step is a simple join operation (Figure 6). For a sequence extension a transformation is necessary first. The transformation assures that items are added at a later point in time. (Figure 7)







Figure 7 SPAM - S-Step processing

Using bitmaps for sequential pattern mining results in very fast operations. On small data sets the costs for building the bitmaps can outweigh the benefits of fast support counting. For large data set and longer sequences, SPAM has empirically proven its strengths (Yang & Kitsuregawa, 2005).

3.3 Chaining of classifiers

As far as "false positives" and "false negatives" accuracy is concerned, both classifiers developed have their specific strengths and weaknesses. Combining the two through a slight modification of the *stacked generalization* ("stacking") techniques proposed by Wolpert (1992) yields a satisfactory overall result. The details will be reported in the next section.

Stacking is a meta-learning algorithm that merges the output of different classifiers for a final classification. It is hoped that a combination of weak classifiers can lead to a stronger classifier. However, there is no deep theory and no generally accepted recommendation as to how carry out this combination. Interpretation and analysis is also said to be difficult (Wolpert (1992)), but has not been found a problem in this application.

4 EXPERIMENTAL RESULTS

The results of applying ADTrees are shown in Table 1. As there is no automatic criterion when to stop building up the tree; we evaluated different numbers of boosting rounds on a test set. We also considered different standard measures for prediction quality. In Table 1 can be seen that the best overall classification (77.7 %) was reached after just one boosting round. But there is a poor recognition of pneumonia (36.8%). The good overall result is a consequence of the predominance of no-pneumonia patients; this unbalanced class distribution as already mentioned.

Number Boosting-Rounds	Overall classification	Pneumonia (True Positive class=1)	No-Pneumonia (True Positive class=0)	Precision Pneumonia	Area Under ROC Curve (AUC)
1	77.7 %	0.368	0.798	0.085	0.5547
2	70.4 %	0.692	0.705	0.107	0.7041
3	71.7 %	0.677	0.719	0.109	0.6677
4	63.0 %	0.815	0.621	0.099	0.7399
5	71.8 %	0.769	0.715	0.121	0.7573
6	65.4 %	0.738	0.65	0.097	0.7439
7	74.7 %	0.723	0.748	0.127	0.7804
8	73.8 %	0.723	0.739	0.124	0.7862
9	72.2 %	0.754	0.72	0.121	0.7928
10	70.9 %	0.785	0.705	0.119	0.7818

Table 1 Experimental results using ADTree

The best detection of pneumonia patients was achieved after 4 boosting rounds (81.5%). But the corresponding no-pneumonia recognition was not satisfying (62.1%). The best precision for pneumonia in combination with one of the best AUC values was obtained after 7 rounds. Here, the pneumonia and no-pneumonia detection were both mediocre.

For sequential pattern mining a different approach was used - the data set was split in two groups: pneumonia and no-pneumonia patients. The SPAM algorithm was applied on both groups separately to find frequent sequential patterns. Score values and laboratory values were discretized into alternative numbers of bins to improve results. SPAM was able to find a lot of rules up to a maximum support of 40% and it proved its claim to fastness. Nevertheless, none of the rules could be found just in one of the groups, displaying discriminating power. Although there was a different support of the sequential patterns, there could not be identified a singular and strong pattern for the separation task. No pattern could be found which spans more than 2 hospital days (Figure 8).

Glasgow Coma Score: $t_{.1}=15$; $t_0 = 15$

Pneumonia

	1 / 0	
	No-Pneumonia	30,6 %
	Pneumonia	12,1 %
Urea: t	$t_{1}=5; t_{0}=5$	
	No-Pneumonia	29,5 %
	Pneumonia	19,5 %
Lactate	: t ₀ = appr. 2	
	No-Pneumonia	34,8 %

Figure 8 Examples for rules found with SPAM algorithm

The idea for a chaining of different data mining techniques was driven by the results of the above mentioned algorithms. The good identification of the pneumonia patients with ADTrees should be retained. The prediction of no-pneumonia patients should be improved by using sequential patterns (Figure 9).

49,5 %



Figure 9 Classification with intelligent chaining of data mining techniques

We have chosen the ADTree with the best result in predicting pneumonia. The confusion matrix of the complete, chained model is shown in Table 2.

	Actual no- pneumonia	Actual pneumonia	
Classified as no-pneumonia	1181	96	1277
Classified as pneumonia	12	53	65
	1193	149	

Table 2 Confusion matrix for chained model from Fig. 9

The quality measures achieved are the following:

Overall classification:	92.0	(1234 of 1342 correctly classified)
Pneumonia (TP=0):	81.5%	(53 of 65 correctly classified)
No-pneumonia (TP=1):	92.5%	(1181 of 1277 correctly classified)
Precision pneumonia:	35.6%	(53 of 149 correctly classified)
Precision no-pneumonia:	99.9 %	(1181 of 1193 correctly classified)

Figure 10 displays the used ADTree for the final model. The tree shows that prior diagnoses of lung illnesses have the most influence. Furthermore, a medical score indicating the decubitus risk is relevant as well as the oxygenation index, both values measured on the previous day.



Figure 10 Resulting ADTree after 4 boosting rounds

After the data have passed the ADTree those cases, which have been classified as no-pneumonia patients by the tree, will be checked by a SPAM rule. The best SPAM Rule found is a very simple one – if an inflammatory parameter is exceeding a certain threshold (*Figure 11*). Although this SPAM rule did not show good results applied as stand-alone algorithm, it worked well after the data set was preprocessed by the ADTree.

CRP: t₀ = approx. 250

(C-reactive proteine; an inflammatory parameter)

No-Pneumonia	14,6 %
Pneumonia	25 %

Figure 11	Selected	SPAM-H	Rule for	chaining	model
-----------	----------	--------	----------	----------	-------

All of the chosen input variables were considered as plausible by the physicians involved in the project. Furthermore, all of the measures are easy to collect and a fast prediction is delivered. As a

next step we calculated the economic impact of this predictor. In preparation of the project we had identified 99 patients whose onset of pneumonia was not detected in time. Due to this delay in starting antibiotic treatment it caused in average 5 days more length of stay at ICU, leading to 445,500 Euro extra costs. Running the developed model, we could have identified 40 patients out of these 99 patients. This corresponds to 180,000 Euro and a possible cost decreasing of 40% because a longer hospitalization could have been avoided.

The classifier scores badly along the "precision" criterion. It will thus create a lot of false positives, i.e. false alarms. Facing the inevitable tradeoff between first and second order statistical error, we accepted this situation, as the detection of pneumonia is considered to be more important from both the medical as well as the economic point of view. In a previous stage of the project, we had found out that costs of antibiotic treatment are much less than those of extended ICU stay. It must also be seen that many of our "false positives" do indeed develop respiratory diseases of a lighter calibre and do need treatment. They just fall short of the CPIS threshold of 6 that defines pneumonia in the clinical sense.

5 WORKFLOW INTEGRATION

Patient care at normal hospital wards can be supported by workflow systems that in their software architecture closely mirror their industrial counterparts. While their implementation typically meets resistance initially, they have proven to be successful tools in both curbing cost and increasing patient care quality (for up-to-date reviews see e.g. Damiani et al. (2010), Chen, Dolan, Lin (2004) and Kuperman (2007) for an application to medication-related issues). Operation at an ICU is substantially different, as almost no standard workflows exist and timeliness of reaction becomes a prime requirement.

The hectic, often fire drill-like operations call for a system that more closely follows the groupware paradigm and is based upon "alerts" that are delivered to doctors and nurses typically over wireless intercom or via SMS. In current systems, alerts are generated

- at fixed points in time
- in response to a treatment plan individually prescribed or set as ward standard (so-called treatment protocol)
- whenever a single bedside data item monitored in real time exceed thresholds (triggers)

In order to make this system more flexible and integrate more sophisticated methods, the Imagina System (Zimmerer, 2007) has been developed and put to prototypical implementation at the ICU considered. The architecture is given in Figure 12 and follows the "Model-View-Controller" paradigm of software design.

All scheduling, module calls and eventually alerts are carried out by the kernel; data to all modules is provided by a database that is fed in real-time from bedside systems and stores them as time series. Master data on patients and stay are added from administrative systems through converters. The database thus serves as a data warehouse for all the modules. Individual modules (alias: monitors) at present do little more than monitor thresholds. They will decide

- if to request an alert through the kernel
- if and when to request the load of other surveillance modules (a respiratory module may, e.g., decide to request immediate load of a cardiac module plus reload of itself after five minutes)

Results are written to the database and are available for processing through other modules. Typically, they will in due course be informed of the user reaction.

It is obvious that our pneunomia early warning system can be seamlessly integrated into this concept. The respective module will realize the system from Section 4, and Figure 9 but will implement the same interface as the current modules. New versions can be implemented even when the system is running without the user even taken notice. The system as a whole shows a consistent user interface as

all user interaction is done through the kernel (or its specialized sub-modules). The Imagina-System thus provides a platform to implement further extensions.



Figure 12 Architecture of IMAGINA

6 LESSONS LEARNED

While the development of the pneumonia detection model needed quite a lot of resources, the daily use of an implemented pneumonia detecting system will be very easy and fast. In our scenario the support system will calculate a prognosis of pneumonia every 12 hours and for every patient. No additional bedside parameters are necessary to collect as just standard measures are considered by the support system.

In a positive case of pneumonia there will be an alert either at the patient's bed or on a mobile device of a doctor. He can now decide whether to start treatment, to trigger further testing or to ignore the alert.

The decision support system fulfills the requirements of a comprehensive solution. All utilized variables made sense from the medical point of view. Careful and extensive preparation of the data was absolutely indispensable; the promise of a fully automated data analysis operating directly on a "dirty" database did not deliver. On the other hand, the project has proven the potential of data mining techniques to reap significant benefit from medical data collected in a totally different context. The Imagina architecture allows to flexibly phase in this module into an ICU groupware system.

References

Agrawal, R. & Srikant, R.(1995). Mining Sequential Patterns. : IEEE Computer Society.

- Ayres, J., Flannick, J., Gehrke, J. & Yiu, T.(2002). Sequential PAttern Mining using a bitmap representation. New York, NY, USA: ACM.
- Battleman, D.; Callahan, M.; Thaler, H. T.(2002). Rapid Antibiotic Delivery and Appropriate Antibiotic Selection Reduce Length of Hospital Stay of Patients with Community-Acquired Pneumonia. In: Archives of Internal Medicine, Vol. 162, pp. 682-688.
- Calandra, T. & Cohen, J.(2005). The International Sepsis Forum Consensus Conference on Definitions of Infection in the Intensive Care Unit. Critical Care Medicine, 33 (7), pp. 1538-1548.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. & Wirth, R.(2000). CRISP-DM 1.0. CRISP-DM consortium.

- Chen, J.C., Dolan, M., Lin, B. (2004) Improve processes on healthcare: current issues and future trends. Int J Electron Healthc. 1(2):149-64.
- Damiani, G., et al (2010). The effectiveness of computerized clinical guidelines in the process of care: a systematic review. BMC Health Serv Res. Jan 4;2010(1):2.
- Dupont, H.; Mentec, H.; Sollet, J. P., Bleichner, G. (2000). Impact of appropriateness of initial antibiotic therapy on the outcome of ventilator-associated pneumonia. In: Intensive Care Medicine, Vol. 27, pp. 355-362.
- Freund, Y. & Mason, L. (1999). The Alternating Decision Tree Learning Algorithm: Morgan Kaufmann Publishers Inc..
- Freund, Y. & Schapire, R. (1999). A short introduction to boosting. Journal of Japanese Society for Artificial Intelligence, 14 (5), pp. 771-780.
- Han, J., Pei, J., Mortazavi-As, B., Chen, Q., Dayal, U. & Hsu, M. (2000). FreeSpan: frequent patternprojected sequential pattern mining. KDD '00: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 355-359). Boston, Massachusetts, United States: ACM Press.
- Heyland, D. K., Cook, D. J., Griffith, L., Keenan, S. P. & Brun-Buisson, C.(1999). The attributable Morbidity and Mortality of Ventilator-Associated Pneumonia in the Critically III Patient. American Journal of Respiratory and Critical Care Medicin, 159, pp. 1249-1256.
- Hripcsak, G. et al (2007). Using discordance to improve classification in narrative clinical databases: an application to community-acquired pneumonia. Comput Biol Med. 37(3):296-304.
- Kommission für Krankenhaushygiene und Infektionsprävention am Robert Koch-Institut (2000). "Prävention der nosokomialen Pneumonie," Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz, pp. 302 – 309.
- Kuperman, G.J. et al (2007). Medication-related clinical decision support in computerized provider order entry systems: a review. J Am Med Inform Assoc. 14(1):29-40.
- Liu, K., Lin, J., Zhou, X. & Wong, S. (2005). Boosting alternating decision trees modeling of disease trait information. BMC Genetics, 6 (Suppl 1), 132.
- Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U. & Hsu, M. (2001). PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth. Los Alamitos, CA, USA: IEEE Computer Society.
- Quinlan, J. R. (1993). C4.5 programs for machine learning. San Mateo, Calif.: Morgan Kaufmann Publishers.
- Reinhart, K., Brunkhorst, F. M., Bone, H. G., Gerlach, H., Gründling, M., Kreymann, G., Kujath, P., Marggraf, G., Mayer, K. & Meier-Hellmann, A.(2006). Diagnose und Therapie der Sepsis. S2. Der Internist, 47 (4), pp. 356-373.
- Safdar, N., Dezfulian, C., Collard, H. R. & Saint, S.(2005). Clinical and economic consequences of ventilator-associated pneumonia: A systematic review. Critical care medicine, 33 (10), pp. 2184 – 2193.
- Srikant, R. & Agrawal, R. (1996). Mining sequential patterns generalizations and performance improvements. Yorktown Heights, N.Y.: IBM Research Division.
- Weiss, G. M. & Provost, F. (2001). The Effect of Class Distribution on Classifier Learning: An Empirical Study.
- Welte, T.(2006). Die nosokomiale Pneumonie. Intensivmedizin und Notfallmedizin, 43 (4), pp. 301 309.
- Wolpert, D. H.(1992). Stacked Generalization. Neural Networks, 5 (2), pp. 241-259.
- Xuanyang, X., et al. (2005). Computer Aided Detection of SARS Based on Radiographs Data Mining.Conf Proc IEEE Eng Med Biol Soc. 2005;7:7459-62.
- Yang, Z. & Kitsuregawa, M. (2005). LAPIN-SPAM: An Improved Algorithm for Mining Sequential Pattern. Washington, DC, USA: IEEE Computer Society.
- Zaki, M. J. (2001). SPADE: An Efficient Algorithm for Mining Frequent Sequences. Machine Learning, 42 (1), 31 60.
- Zimmerer, D.(2007). Imagina Ein Entscheidungsunterstützungs-system für Intensivstationen Friedrich-Schiller-Universität Jena, Wirtschaftswissenschaftliche Fakultät.