

Book Title	Encyclopedia of Machine Learning	
Chapter Number	00040	
Book CopyRight - Year	2010	
Title	Medicine: Applications of Machine Learning	
Author	Particle	
	Given Name	Katharina
	Family Name	Morik
	Suffix	Prof. Dr.
	Email	katharina.morik@tu-dortmund.de
Affiliation	Division	Department of Computer Science
	Organization	Technische Universitat Dortmund
	Street	Baroperstr.301
	Postcode	44227
	City	Dortmund
	State	
	Country	Germany

M

Medicine: Applications of Machine Learning

KATHARINA MORIK

Department of Computer Science,
Technische Universität Dortmund,
Dortmund, Germany

Motivation

Health care has been an important issue in computer science since the 1960s. In addition to databases storing patient records, library resources (e.g., PubMed, a service of the U.S. National Library of Medicine that includes over 16 million citations from journals for biomedical articles back to the 1950s), administrative and financial systems, more sophisticated support of health care has been the aim of artificial intelligence (AI) from the very beginning on. Starting with expert systems which abstract laboratory findings and other vital parameters of a patient before they heuristically classify the patient into one of the modeled diagnoses (Shortliffe, 1976), knowledge acquisition was discovered to be the bottleneck of systems for the automatic medical diagnosis. Machine learning came into play as a means of knowledge acquisition. Learning rules for (medical) expert systems focused on the heuristic classification step within expert systems. Given conveniently abstracted measurements of the patient's state, the classification was learned in terms of rules or [▶decision trees](#). Since the early days, the use of machine learning for health care progressed in two ways:

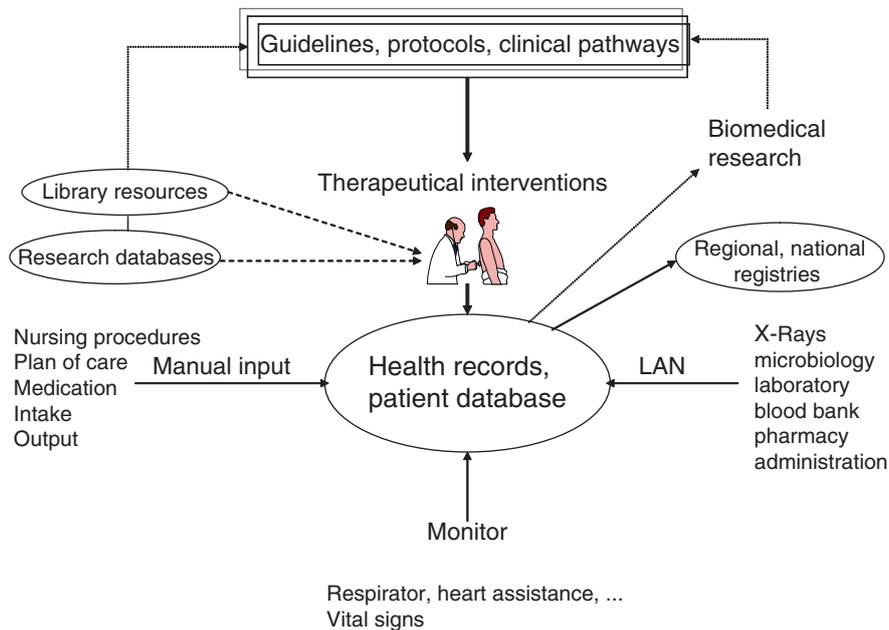
- The abstraction of measurements of a patient's vital parameters is a learning task in its own right. Diverse kinds of data are to be handled: laboratory data, online measurements at the bedside, x-rays or other imaging data, genetic data,... Machine learning is confronted with a *diversity of representations* for the examples.

- Diagnosis is just one task in which physicians are to be supported. There are many more tasks which machine learning can ease. In intensive care, the addressee of the learning results can be a machine, e.g., the respirator. Financing health care and planning the medical resources (e.g., for a predicted epidemic) are yet another important issue. Machine learning is placed in a *diversity of medical tasks*.

The urgent need for sophisticated support of health care follows from reports which estimate up to 100,000 deaths in the USA each year due to medical error (Kohn, Corrigan, & Donaldson, 2000).

Structure of the Problem

The overall picture of the medical procedures shows the kinds of data and how they are entered into the database of health records (a synonym is "patient database.") A monitoring system is given in intensive care units, which acquires time series from minute measurements. The observations at the bedside are entered manually into the system. The information from the hospital is entered via a local area network. The physician accesses information from libraries and research databases (dashed lines). Libraries, research databases, and biomedical research also influence the development of guidelines, protocols, and clinical pathways (dotted lines). Guidelines are rather abstract. Protocols of certain actions are integrated to become a clinical pathway which is a plan of both diagnostic and therapeutical actions for a typical patient with a specific diagnosis. The bold arrow shows the intended high-quality therapy. Guidelines and protocols promote evidence-based practices, reduce inter-clinician practice variations and support decision-making in patient care while constraining the costs of care. Computerized protocols can be generated based on guidelines. They have been proved useful in improving the quality



and consistency of healthcare but the protocol development process is time-consuming (Ten Teije, Lucas, & Miksch, 2006). This is where machine learning offers support. Usually, ontologies (e.g., in description logic) or other knowledge-based techniques (in medicine-specific formats like the Arden Syntax, GuideLine Interchange Format (GLIF), PROforma, Asbru, and EON) are used to support the development of protocols (de Clercq, Blomb, Korstenb, & Hasman, 2004). By contrast, machine learning induces the current practices and their outcome from the health records (Smith, Doctor, Meyer, Kalet & Philips, 2009). To reflect such use of Machine Learning, the bold arrows of the picture would need to be turned the other way around, protocols are learned from the data or evaluated based on the data. All (reversed) arrows mark possible applications of machine learning.

Diversity of Representations

The overall health record of a patient includes several types of data, not all of them are digital.

- Laboratory data consist of attributes almost always with numerical values, sometimes with discrete ordinal values, sometimes just binary values like “positive,” “negative.”

- Plain text states anamneses, diagnosis, and observations. From the text, key words can be transformed into attributes for machine learning.
- Online measurements at the bedside are time series. They are analyzed in order to find level changes or trends (Gather, Schettlinger, & Fried, 2006) and alarm the physician (Sieben & Gather, 2007). In order to exploit the time series for further learning tasks, they often are abstracted (e.g., Bellazzi, Larizza, Magni, & Bellazi (2002)). Recently, online measurements from body sensors have raised attention in the context of monitoring patients at home (Amft & Tröster, 2008).
- Sequences can also be considered time series, but the measurements are not equidistant and not restricted to numerical values. Examples are data gathered at doctors’ visits and long-term patient observations.
- X-rays or other imaging data (e.g., ultrasound imaging or more novel molecular imaging techniques like positron emission tomography, magnetic resonance imaging, or computer tomography) cannot be analyzed directly by machine learning algorithms. They require the extraction of features. It has been shown that the adequate extraction of features is more important than the selection of the best suited learning algorithm (Mavroforakis, Georgiou, Dimitropoulos, Cavouras, & Theodoridis, 2006).

The number of extracted features can become quite large. For instance, from 1,619 images of skin lesion, each 752×582 pixels, 107 features were extracted in order to detect melanoma using diverse learning algorithms (Dreiseitl et al., 2001). Hence, feature selection is also an important task in medical applications (Lucaces, Taboada, Albaiceta, Domingues, Enriques & Bahamonde, 2009; Withayachumnankul et al., 2006). Often, different techniques are applied to gather data for the detection of the same disease. For instance, glaucoma detection uses standard automated perimetry or scanning laser or Heidelberg Retina Tomograph or stratus optical coherence tomography. It is not yet clear how important the choice among measurement types (devices) is with respect to feature extraction and machine learning.

- Tissue and blood: In vitro “data” also belong to health records. Immediately after biopsy or surgery, the tissue is transferred to the pathology department. After the pathologist has taken the sample needed for proper diagnosis, a representative tissue sample will be snap frozen and stored in liquid nitrogen or at -80°C . Also blood cells are stored in a blood bank. From the specimen, the RNA is extracted and the so-called microarrays of gene expressions are developed and then scaled. The huge prognostic value of gene expression in patients with breast cancer has been shown by van't Veer et al. (2002). Genome research aims at revealing the impact of gene regulation and protein expression-regulation (taking into account the regulation of protein synthesis, protein ubiquitination, and post-translational modification) on, e.g., cancer diagnosis and response to therapies. Machine learning, particularly clustering, frequent itemset mining, and classification have been applied successfully (see ►[learning from gene expression microarray data](#)).

In addition to patient records, there are knowledge bases describing particular diseases or computerized protocols for particular therapies.

Medical Tasks

Diagnosis and Medication

Diagnosis is primarily a classification task. Given the description of the patient's state and a set of diseases, the learning algorithm outputs the classification

into one of the classes. If physicians want to inspect the learned classifier, logic-based algorithms are preferred. Decision trees and the conceptual clustering algorithm AQ were used to diagnose breast cancer from nine abstracted descriptions like tumor size: 0 – 4, 5 – 9, ..., 50 – 54, 55 – 59 (Cestnik, Kononenko, & Bratko, 1987; Michalski, Mozetic, Hong, & Lavrac, 1986).

►[Bayesian methods](#) were used to classify, e.g., diseases of the lymph node. Based on the examination of the extracted tissue, a pathologist enters the description. The Bayesian network (BN) outputs not only just one diagnosis, but the conditional probabilities for the diseases (Heckerman, 1990). In particular, diagnosis for rather vague symptoms such as abdominal pain or lower back pain is well supported by BNs (McNaught, Clifford, Vaughn, Foggs, & Foy, 2001). BNs are capable of incorporating given expert knowledge as priors. In order to combine textbook knowledge with empirical data, electronic literature was transformed into priors for BN structures. Then, from health records, the BN was learned as a model of ovarian tumors (Antal, Fannes, Timmerman, Moreau, & De Moor, 2004).

►[Inductive logic programming](#) (ILP) also allows to take into account background knowledge. This was used for an enhanced learning of medical diagnostic rules (Lavrac, Dzeroski, Prinat, & Krizman, 1993). The identification of glaucomatous eyes was effectively learned by ILP (Mizoguchi, Ohwada, Daidoji, & Shirato, 1997). One advantage of ILP is that the learned logic clauses can easily be integrated into a knowledge-based system and, hence, become operational for clinical practice.

Since some tests which deliver information about the patient's state can be costly – both, financially and in terms of a risk for the patient – ►[cost-sensitive learning](#) may be applied.

Since the error of classifying a patient as ill where he or she is not (false positives) is less harmful than classifying a patient as healthy where he or she is not (false negatives), the evaluation of the learning result most often is used in a biased way. The evaluation can be summarized in Table 1.

Precision is the proportion $\frac{A}{A+B}$, and *recall* is the proportion $\frac{A}{A+C}$. *Sensitivity* is synonymous to recall. In medical applications, sensitivity is balanced with respect to

Medicine: Applications of Machine Learning. Table 1**Evaluation measures**

	True +	False –
Predicted +	A	B
Predicted –	C	D

specificity being the proportion $\frac{B}{B+D}$ (synonym *false positives rate*). The analysis of the Receiver Operator Characteristic (ROC) allows to evaluate learning according to sensitivity and specificity (see ►ROC analysis).

If not the understandability but only sensitivity and specificity are important, numerical learning algorithms are used to classify the patient's data. In particular, if the patient's state is described by numerical features, no discretization is necessary for numerical learners as is needed for the logic-based ones. Multi-layer perceptrons (see ►Neural Networks), ►support vector machines (SVM), ►mixtures of Gaussians, and mixture of generalized Gaussian classifiers were trained on the numerical data of 189 normal eyes and 156 glaucomatous eyes (Goldbaum et al., 2002). The numerical description of the visual field is given by standard automated threshold perimetry. The medical standard procedure to interpret the visual field is to derive global indices. The authors compared performance of the classifiers with these global indices, using the area under the ROC curve. Two human experts were judged against the machine classifiers and the global indices by plotting their sensitivity–specificity pairs. The mixture of Gaussian had the greatest area under the ROC curve of the machine classifiers, and human experts were not better at classifying visual fields than the machine classifiers or the global indices.

Other approaches to glaucoma detection use different features describing the patient's state (Zangwill et al., 2004) or other numerical learners, e.g., logistic regression (Huang, Chen, & Hung, 2006). For testing the learning from numerical attributes, the UCI Machine Learning Repository offers the arrhythmia database. The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups. About 279 attributes are given, 206 of them being numerical ones.

As has been shown in an application to intensive care, medication can be transformed into a set of classification tasks (Morik, Imhoff, Brockhausen, Joachims, & Gather, 2000). Given measurements of eight vital signs, a decision is made for each of six drugs, whether to increase or to decrease it. This gives a set of classification tasks, which the ►SVM learned. Depending on the drug, the accuracy ranged from 81.3% with 2.5 standard error to 86.9% with 7 standard error. Additionally, on 41 cases, the SVM decision was compared with an expert's decisions when confronted with the same data. In 32 cases the expert chose the same direction of change as did the learned decision function. In 34 cases the learned decision was equal to the actual therapy. Another set of classification tasks were to decide every minute whether to increase, decrease, or leave the doses as it is. Again, each of these classifiers was learned by the SVM. From 1,319 examples decision functions were learned and tested on 473 examples. For training, an unbalanced cost function was used. The SVM cost factor for error was chosen according to $\frac{C_+}{C_-} = \frac{\text{number of negative examples}}{\text{number of positive examples}}$. The results again differed depending on the drug. For adrenaline, 79% of the test cases were equally handled by the physician and the decision function. For adrenaline as well as for dobutamine, only in 1.5% of the test cases the learned rule recommended the opposite direction of change. Again, a blind test with an expert showed that the learned recommendations' deviation from actual therapy was comparable to that of the human expert. Combining the two sets of classifications, for each minute and each patient, the support vector machine's decision function outputs a recommendation of treatment (Morik, Imhoff, Brockhausen, Joachims, & Gather, 2000).

Prognosis and Quality of Care Assessment

Prognosis or outcome prediction is important for the evaluation of the quality of care provided. The standard statistical models use only a small set of covariates and a score variable, which indicates the severity of the illness. Machine learning may also rely on the aggregated score features, but is in addition capable of handling the features underlying the scores. Given health records of patients including the therapy, machine learning is to predict the outcome of care, e.g., classifies into mortal or surviving cases. The prediction of breast cancer survivability has been tested on a very large database

comparing three learning methods (Delen, Walker, & Kadam, 2004). The results indicated that decision trees (here: C5) result in the best predictor with 93.6% accuracy on the holdout sample (this prediction accuracy is better than any reported in the literature), artificial neural networks came out to be the second with 91.2% accuracy, and the logistic regression models came out to be the worst of the three with 89.2% accuracy.

Prediction of survival is a hard task for patients with serious stroke, because there is a long-term risk after the stay at the hospital. The scoring schemes (e.g., the Glasgow coma scale and the Ranking score) are not sufficient for predicting the outcome. In a data situation where 29 attributes (or features) were given for only 327 patient records, BNs were learned and compared with a handmade causal network. The results were encouraging – as soon as more electronic health records become available, the BNs will become closer to medical knowledge. Moreover, the discovery of relations on the basis of empirical data may enhance medical knowledge (Wu, Lucas, Kerr, & Dijkhuisen, 2001).

Carcinogenesis prediction was performed using ILP methods. As has become usual with cancer diagnosis and prognosis, there is a close link with microbiology (Srinivasan, Muggleton, King, & Sternberg, 1994)(see Learning from gene expression microarray data).

Prognosis need not be restricted to mortality rates. In general, it is a means of quality assessment of clinical treatments. For instance, hemodialysis services have been assessed through temporal data mining by Bellazzi et al. (2002).

Finding subgroups of patients with devious reactions to a therapy might lead to a better understanding of a certain medical process (Atzmueller, Baumeister, Hensing, Richter, & Puppe, 2005). While the before mentioned study aims at an enhanced expert – system interaction, a Dutch study aims at a more precise modeling of prognoses (Abu-Hanna & Lucas, 2001). In an extensive study for eight different hospitals and 7,803 patients, two different models were combined: one for determining the subgroups and the other for building a model for each subgroup. For the prognoses of patients in an intensive care unit, subgroups have been detected using decision trees. The decision tree was trained to classify patients into the survival class and the mortality class on the basis of the nonaggregated features underlying the illness score. The leaves of the tree become

subgroups. These are then used for training a logistic regression model of mortality based on the aggregated features.

Verification and Validation

Verification is the process of testing a model against a specification. In medicine, this often means to check clinical practice against expert protocols, or to check an actual diagnosis against one derived from textbook knowledge. Since many logic-based machine learning algorithms consist of a generalization and a specialization step, they can be used for verification. Generalization delivers rules from clinical data which can then be compared with given expert rules (protocols). Specialization is triggered by facts that contradict a learning hypothesis. Hence, using an expert rule as hypothesis, the learning algorithm counts the contradicting clinical cases and specializes the rule. For an early case study on verification and rule enhancement see, e.g., (Morik, Potamias, Moustakis, & Charissis, 1994). A more recent study compares a given clinical protocol for intensive care with actual therapies at another hospital (Scholz, 2002). Decision trees and association rules have been learned in order to inspect and enhance the knowledge base of a web-based teledermatology system (Ou, West, Lazarescu, & Clay, 2007). While verification means to build the system right, validation means to build the right system. The borderline between verification and validation is fuzzy. On the one hand, medical practice is investigated with respect to the guidelines (verification), on the other hand, the guidelines are enhanced on the basis of medical practice (validation).

Moreover, learned models can be verified with respect to expert knowledge and validated with respect to clinical practice. A study on the hemodynamic monitoring of the critically ill integrated machine learning into a knowledge-based approach to evidence-based medicine. A knowledge base on drug effects was verified using patient records. Only 18% of the observations showed vital signs of patients in the opposite direction than predicted by the knowledge base. Then, the knowledge base was used to validate therapeutical interventions proposed by a learned model. Accuracy measures of a model only reflect how well the learning result fits actual behavior of the physician and not how well it fits the “gold standard.” Hence, a proposed intervention should be validated with respect to its effects on

the patient. If the known effects push vital signs in the direction of the desired value range, the recommendation is considered sound, otherwise it is rejected. Using past data, the learned model was found to recommend an intervention with the desired effects in 81% of the cases (Morik, Joachims, Imhoff, Brockhausen, & Rüping, 2002).

Intelligent Search in Medical Literature

Intelligent search in the overwhelming number of research publications supplies the information when it is needed. ILP has been successfully put to use for finding relevant medical documents (Dimec, Dzeroski, Todorovski, & Hristovski, 1999). Also the intelligent search in clinical free-text guidelines is an issue (Moskovitch et al., 2006). The techniques for text categorization can be applied to medical texts in the usual way. If the search engine not only labels the overall document but, in addition, phrases within it, the search could become more focused and also deliver paragraphs instead of complete texts. The biomedical challenge for *named entity recognition* requires the automatic extraction and classification of words referring to DNA, RNA, proteins, cell types, and cell lines from texts (Kim, Ohta, Tsuruoka, Tateisi, & Collier, 2004). Even more difficult is the discovery of medical knowledge from texts (Sanchez & Moreno, 2005).

Epidemiology and Outbreak Detection

Understanding the transmission of infectious diseases and forecasting epidemics is an important task, since infections are distributed globally. Statistical approaches to spatio-temporal analysis of scan data are regularly used. There, a grid partitions the map into regions where occurrences of the disease are shown as points. “Hot spot” partitions are those of high density. By contrast, clustering detects hot spot regions depending on the data, hence, the shape of regions is flexible. Taking into account the a priori density of the population, a risk-adjusted nearest neighbor hierarchical clustering discovers “hot spot” regions. Also a risk-adjusted support vector machine with Gaussian kernel has successfully been applied to the problem of detecting regions with infectious disease outbreak. The discovery of hot spot regions can be exploited for predicting virus activity, if an indicator is known which

can easily be observed. For instance, dead crows indicate activity of the West Nile virus. An overview of infectious disease informatics is given by (Zeng, Chen, Lynch, Eidson, & Gotham, 2005).

Machine learning can also contribute to the understanding of the transmission of infectious diseases. A case study on tuberculosis epidemiology uses BNs to identify the distribution of tuberculosis patient attributes. The learning results captured the known statistical relationships. A relational model learned from the database directly using structured statistical models revealed several novel associations (Getoor, Rhee, Koller, & Small, 2004).

Cross References

- ▶ Classification
- ▶ Classifier Systems
- ▶ Cost-Sensitive Learning and the Class Imbalance Problem
- ▶ Decision Trees
- ▶ Feature Selection
- ▶ Inductive Logic Programming
- ▶ ROC Analysis
- ▶ Support Vector Machine
- ▶ Time Series

Recommended Reading

- Abu-Hanna, A., & Lucas, P. J. F. (2001). Prognostic models in medicine: AI and statistical approaches [Editorial]. *Methods of Information in Medicine*, 40(1), 1–5.
- Amft, O., & Tröster, G. (2008). Recognition of dietary events using on-body sensors. *Artificial Intelligence in Medicine*, 42(2), 121–136.
- Antal, P., Fannes, G., Timmerman, D., Moreau, Y., & De Moor, B. (2004). Using literature and data to learn BNs as clinical models of ovarian tumors. *Artificial Intelligence in Medicine*, 30(3), 257–281.
- Atzmueller, M., Baumeister, J., Hensing, A., Richter, E.-J., & Puppe, F. (2005). Subgroup mining for interactive knowledge refinement. In *Artificial intelligence in medicine (AIME)* (pp. 453–462). Berlin/Heidelberg: Springer.
- Bellazzi, R., Larizza, C., Magni, P., & Bellazi, R. (2002). Quality assessment of dialysis services through intelligent data analysis and temporal data mining. In *Workshop at the 15th European conference on AI about intelligent data analysis in medicine and pharmacology* (pp. 3–9). Lyon, France.
- Cestnik, B., Kononenko, I., & Bratko, I. (1987). ASSISTANT 86: A knowledge-elicitation tool for sophisticated users. In I. Bratko, & N. Lavrac (Eds.), *Progress in machine learning* (pp. 31–45). Wilmslow, GB: Sigma Press.
- de Clercq, P. A., Blomb, J. A., Korstenb, H. H., & Hasman, A. (2004). Approaches for creating computer-interpretable guidelines that

- facilitate decision support. *Artificial Intelligence in Medicine*, 31(1), 1–27.
- Delen, D., Walker, G., & Kadam, A. (2004). Predicting breast cancer survivability: A comparison of three data mining methods. *Artificial Intelligence in Medicine*, 34(2), 113–127.
- Dimec, B., Dzeroski, S., Todorovski, L., & Hristovski, D. (1999). WWW search engine for slovenian and english medical documents. In *Proceedings of the 15th international congress for medical informatics* (pp. 547–552). Amsterdam: IOS Press.
- Dreiseitl, S., Ohn-Machado, L., Kittler, H., Vinterbo, S., Billhardt, H., & Binder, M. (2001). A comparison of machine learning methods for the diagnosis of pigmented skin lesions. *Journal of Biomedical Informatics*, 34, 28–36.
- Gather, U., Schettlinger, K., & Fried, R. (2006). Online signal extraction by robust linear regression. *Computational Statistics*, 21(1), 33–51.
- Getoor, L., Rhee, J. T., Koller, D., & Small, P. (2004). Understanding tuberculosis epidemiology using structured statistical models. *Artificial Intelligence in Medicine*, 30(3), 233–256.
- Goldbaum, M. H., Sample, P. A., Chan, K., Williams, J., Lee, T-W., Blumenthal, E., et al. (2002). Comparing machine learning classifiers for diagnosing glaucoma from standard automated perimetry. *Investigative Ophthalmology and Visual Science*, 43, 162–169.
- Heckerman, D. (1990). *Probabilistic similarity networks*. Technical report STAN-CS-1316, Department of Computer Science and Medicine at Stanford.
- Huang, M. L., Chen, H. Y., & Hung, P. T. (2006). Analysis of glaucoma diagnosis with automated classifiers using stratus optical coherence tomography. *Optical Quantum Electronics*, 37, 1239–1249.
- Kim, J. D., Ohta, T., Tsuruoka, Y., Tateisi, Y., & Collier, N. (2004). Introduction to the bio-entity recognition task at JNLPBA. In N. Collier, P. Ruch, & A. Nazarenko, (Eds.), *Proceedings of the international joint workshop on natural language processing in biomedicine and its applications* (pp. 70–76). Morristown, NJ: ACL.
- Kohn, L. T., Corrigan, J. M., & Donaldson, M. (Eds.) (2000). *To err is human - building a safer health system*. Washington, DC: National Academic Press.
- Lavrac, N., Dzeroski, S., Prinat, V., & Krizman, V. (1993). The utility of background knowledge in learning medical diagnostic rules. *Applied Artificial Intelligence*, 7, 273–293.
- Lucaces, O., Taboada, F., Albaiceta, G., Domingues, L. A., Enriques, P., & Bahamonde, A. (2009). Predicting the probability of survival in intensive care unit patients from a small number of variables and training examples. *Artificial Intelligence in Medicine*, 45(1), 63–76.
- Mavroforakis, M., Georgiou, H., Dimitropoulos, N., Cavouras, D., & Theodoridis, S. (2006). Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers. *Artificial Intelligence in Medicine*, 37(2), 145–162.
- McNaught, K., Clifford, S., Vaughn, M., Foggs, A., & Foy, M. (2001). A Bayesian belief network for lower back pain diagnosis. In P. Lucas, L. C. van der Gaag, & A. Abu-Hanna (Eds.), *Bayesian models in medicine - Workshop at AIME*. Caseais, Portugal.
- Michalski, R., Mozetic, I., Hong, J., & Lavrac, N. (1986). The multi-purpose incremental learning system AQ15 and its testing application on three medical domains. In *Proceedings of the 5th national conference on artificial intelligence* (pp. 1041–1045). San Mateo, CA: Morgan Kaufmann.
- Mizoguchi, F., Ohwada, H., Daidoji, M., & Shirato, S. (1997). Using inductive logic programming to learn classification rules that identify glaucomatous eyes. In N. Lavrac, E. Keravnou, & B. Zupan, (Eds.), *Intelligent data analysis in medicine and pharmacology* (pp. 227–242). Norwell, MA: Kluwer.
- Morik, K., Imhoff, M., Brockhausen, P., Joachims, T., & Gather, U. (2000). Knowledge discovery and knowledge validation in intensive care. *Artificial Intelligence in Medicine*, 19(3), 225–249.
- Morik, K., Joachims, T., Imhoff, M., Brockhausen, P., & Rüping, S. (2002). Integrating kernel methods into a knowledge-based approach to evidence-based medicine. In M. Schmitt, H. N. Teodorescu, A. Jain, A. Jain, S. Jain, & L. C. Jain, (Eds.), *Computational intelligence processing in medical diagnosis*, (Vol. 96) Studies in fuzziness and soft computing, (pp. 71–99). New York: Physica-Verlag.
- Morik, K., Potamias, G., Moustakis, V. S., & Charissis, G. (1994). Knowledgeable learning using MOBAL: A medical case study. *Applied Artificial Intelligence*, 8(4), 579–592.
- Moskovitch, R., Cohen-Kashia, S., Drora, U., Levya, I., Maimona, A., & Shahar, Y. (2006). Multiple hierarchical classification of free-text clinical guidelines. *Artificial Intelligence in Medicine*, 37(3), 177–190.
- Ou, M., West, G., Lazarescu, M., & Clay, C. (2007). Dynamic knowledge validation and verification for CBR teledermatology system. *Artificial Intelligence in Medicine*, 39(1), 79–96.
- Sanchez, D., & Moreno, A. (2005). Web mining techniques for automatic discovery of medical knowledge. In *Proceedings of the 10th conference on artificial intelligence in medicine*. Aberdeen, Scotland.
- Scholz, M. (2002). Using real world data for modeling a protocol for ICU monitoring. In P. Lucas, L. Asker, & S. Miksch, (Eds.), *Working notes of the IDAMAP 2002 workshop*, (pp. 85–90). Lyon, France.
- Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C., et al. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nature Medicine*, 8(1), 68–74.
- Shortliffe, E. H. (1976). *Computer based medical consultations: MYCIN*. New York, Amsterdam: Elsevier.
- Sieben, W., & Gather, U. (2007). Classifying alarms in intensive care—analogy to hypothesis testing. In *11th conference on artificial intelligence in medicine (AIME)* (pp. 130–138). Berlin: Springer.
- Smith, W. P., Doctor, J., Meyer, J., Kalet, I. J., & Philips, M. H. (2009). A decision aid for intensity-modulated radiation-therapy plan selection in prostate cancer based on a prognostic Bayesian network and a Markov model. *Artificial Intelligence in Medicine*, 46(2), 119–130.
- Srinivasan, A., Muggleton, S. H., King, R. D., & Sternberg, M. J. E. (1994). Carcinogenesis prediction using inductive logic programming. In B. Zupan, E. Keravnou, & N. Lavrac (Eds.), *Intelligent data analysis in medicine and pharmacology* (pp. 243–260). Norwell, MA: Kluwer.
- Ten Teije, A., Lucas, P., & Miksch, S. (Eds.), (2006). *Workshop on AI techniques in healthcare: Evidence-based guidelines and protocols, held in conjunction with ECAI-2006*. Italy.
- van't Veer, L. J., Dai, H. Y., van de Vijver, M. J., He, Y. D. D., Hart, A. A., Mao, M., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415, 530–536.

- Withayachumnankul, W., Ferguson, B., Rainsford, T., Findlay, D., Mickan, S. P., & Abbott, D. (2006). T-ray relevant frequencies for osteosarcoma classification. In D. Abbott, Y. S. Kivshar, H. H. Rubinstein-Dunlop, & S.-H. Fan, (Eds.), *Proceedings of SPIE*. Brisbane, Australia.
- Wu, X., Lucas, P., Kerr, S., & Dijkhuisen, R. (2001). Learning bayesian-network topologies in realistic medical domains. In *Intelligent data analysis in medicine and pharmacology*. Medical Data Analysis, (pp. 302–307). Berlin/Heidelberg: Springer.
- Zangwill, L. M., Chan, K., Bowd, C., Hao, J., Lee, T. W., Weinreb, R. N., et al. (2004). Heidelberg retina tomograph measurements of the optic disc and parapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Investigative Ophthalmology and Visual Science*, 45(9), 3144–3151.
- Zeng, D., Chen, H., Lynch, C., Eidson, M., & Gotham, I. (2005). Infectious disease informatics and outbreak detection. In H. Chen, S. Fuller, C. Friedman, & W. Hersh, (Eds.), *Medical informatics: knowledge management and data mining in biomedicine* (pp. 359–395). New York: Springer.