Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 133–139
ISBN 978-83-60810-22-4

Data Mining Research Trends in Computerized Patient Records

Payam Homayounfar
Wrocław University of Economics,
ul. Komandorska 118/120
53-345 Wrocław, Poland
Email: p.homayounfar@gmail.com

Mieczyslaw L. Owoc
Wrocław University of Economics,
ul. Komandorska 118/120
53-345 Wrocław, Poland
Email: mieczyslaw.owoc@ue.wroc.pl

***Abstract*—Over the last decades has the research on Data Mining made a great progress. Also the Computerized Patient Records (CPR) as part of Hospital Information Systems have improved in terms of usability, content coverage, and diffusion rate. The number of Health Care Organizations using the CPR is growing. Causally determined is the need for techniques and models to provide solutions for decision making based on the data stored from different sources in CPR. This paper provides an overview on the current research trends and shows the impact on the medical domain with the CPR.**

I. Introduction

INTERNATIONALLY the Computerized Patient Records (CPR) became growingly more important for health care institutions. The number of institutions changing over from the paper based patient files to the CPR are increasing. This evolutionary development will increase with the establishment of Data Mining (DM) and the associated techniques and applications. Before DM the CPR has been known as centralized data storage for patient data with limited the possibilities to analyze, to process, and to use the data for other questions except for some simple cases. This is the reason for using the word CPR instead of Electronic Patient Records which represents the limited analyzability of the patients data in the past.

The amount of medical and patient oriented data stored in CPR has grown strong progressive. The CPR contains medical data, laboratory data, and images from different modalities and organizational data from different sources with the purpose of patients care. DM is the key technology to evaluate, interpret and link information of the large amount of data. DM improves the value of CPR to support the process of decision-making and medical diagnosis [1].

In the context of medical data DM uses algorithms, tools, lifecycles of knowledge, and formalizations to extract patterns, information and knowledge from data stored in the CPR. DM transforms transactional data in the CPR from tacit knowledge into explicit knowledge [2]. In this context it is important to mention the link of DM to Knowledge Management (KM). KM is the system and managerial approach to the gathering, management, use, analysis, sharing, and discovery of knowledge [3]. KM deals with eliciting, representing, and storing explicit data. DM is a subfield of KM and is used as part of the knowledge discovery process. KM and DM have the same fundamental issues and must be combined in the decision making process. Especially in medical applications the interaction and integration of DM and KM is essential [4]-[6].

DM is becoming an area of great interest for clinical practice and research as medical decisions must always be supported by arguments, and the explanation of decisions and predictions should be mandatory for an effective deployment of DM models. DM and KM are the most important technologies for enabling Evidence Based Medicine, which proposes strategies to apply evidence gained from scientific studies for the care of individual patients [7].

There are some scientific research projects with the purpose of merging clinical and research objectives like the I2B2 project at the Harvard University [8].

DM is the essential part of futures CPR. The objective of this paper is to give an overview of current research trends of DM in CPR. Firstly the Computerized Patient Record with its characteristics is described. DM is especially in CRP valuable and essentially necessary. Also the basic tasks of DM will be described in the paper. The following DM models of sophisticated machine learning models will be focused with their impacts on CPR in a separate section: Symbolic Learning vs. Rule induction, Case Based Reasoning, Natural Language Processing, Artificial Neuronal Networks, Bayesian Networks, and Fuzzy Models.

Not in scope of this paper is the view on aspects of technical systems, special tools for the DM techniques, description of underlying methodologies, legal issues, data privacy, and data security.

II. Characteristics of Computerized Patient Records

The CPR is a collection of data in a database or repository, which is managed by application programs. It is a key part of hospital information systems. All relevant data and the investigations and interventions for one patient in one health care institute are collected in a structured manner and without redundancy in the electronic patient record. The data is stored on digital media and are always available electronically. The difference to data files in other segments like economical data of a customer is based on the complexity of medical data of patients [1], [4], [9].

The electronic patient file contains data from different areas [10]-[13] like personal details, billing information, case or medical history, clinical test results, diagnoses from different specialists, therapy information, digital pictures from various modalities, pictorial archiving of historical pathologic findings, important treatment data and results of control tests, representation of specific content for the single specialists (Anesthesiology, Radiology, Pathology, Cardiology, Endocrinology, Pharmacology, Odontology, Accident Medicine, …), and nursing measures.

The complexity is not only based on the broad range and variety of data in the CPR, but also evolved from the input from many different sources of the data. Different sources of technical systems as well as the multiple groups inserting and extracting data in the CPR give a good impression of the complexity of CPR data.

Another characteristic for CPR is the huge amount of medical data. Pictures made with different modalities like Computer Tomography (CT) or Magnetic Resonance Tomography (MRT) as well as the measurement of laboratory and pathological reports produce a very large and storage intensive amount of data for each single patient. Having this in mind it is understandable that data in CPR reach many Giga-Bytes. The trend is rising as the sophisticated technical possibilities are growing and going much and creating more data.

The CPR provides a multidisciplinary information exchange for communication between the different stakeholders in health care institutions like medical doctors, physicians, laboratories, nurses, other members of the health care team and the administration and controlling staff [14].

The purpose of the CPR is divided into following prioritized areas [11]:

1. Patient care: The CPR provides the documented basis for planning care and treatment.
2. Communication: All stakeholders have the same information on a patient. This allows the communication with each other with the same actual data basis on a patient.
3. Legal documentation: Documentation of the treatment as well as the legal forms signed by the patient.
4. Billing and reimbursement: Coded treatment with e.g. International Classification of Diseases (ICD-10) or Diagnosis-related groups (DRG)
5. Research
6. Quality Management

The characteristics of DM in CPR and the uniqueness of medical data are the challenge in this field [4] are the volume and complexity of medical data, the importance of physicians interpretation, the sensitivity and specificity analysis, poor mathematical characterization of medical data, the canonical form, and ethical, legal as well as social issues (Data ownership, privacy and security)

III. Data Mining in Context with Knowledge Management

Data Mining extracts patterns, explicit knowledge and information from data. The objective of data mining in this context is to support the medical doctor and the health care institution in decision making.

Data mining analyzes data and extracts models that allow the interpretation and transformation of the raw data in the CPR into knowledge. This is the entry point for knowledge management to create tacit knowledge. KM is the system and managerial approach to the gathering, management, use, analysis, sharing, and discovery of knowledge [3], [15]. KM deals with eliciting, representing, and storing explicit data. DM is a subfield of KM and is used as part of the knowledge discovery process. KM and DM have the same fundamental issues and must be combined in the decision making process. Especially in medical applications the interaction and integration of DM and KM is essential [4].

Data analysis in medicine depends more than in other areas on medical background knowledge. Further approaches, such as association and classification rules, joining the declarative nature of rules, and the availability of learning mechanisms are a great potential for effectively merging DM and KM [16].

Over the past years have many ontologies been developed. Ontologies have an important role in DM to facilitate knowledge sharing between different sources. An ontology is a specification of conceptualization. It describes the concepts and relationships that can exist and formalizes the terminology in a domain [17], [18].

IV. Data Mining Research in Computerized Patient Records

The shown characteristics of medical and organizational data in CPR lead to the special problem of analyzing and linking data from different sources and qualities together. Especially in the CPR are many information hidden and are important to be revealed. The hidden information in the raw data are also caused by the complexity of the medical domain in CPR. It is easy to lose the track of a disease if different medical doctors make a diagnosis in their own domain and do not compare their findings with each other. Data mining can bridge the important gap and bring together the essence of the information.

Data mining and knowledge creation is more than a set of techniques for data analysis, it is the key for extracting information out of the mentioned data. Without data mining the storage of the data in the CPR would be not necessary as this makes the difference to the patients files based on paper.

Data mining techniques build a group of heterogeneous tools and techniques to different purposes along the process to create knowledge. There are descriptive and predictive models. The descriptive models identify similar patterns in the analyzed date by using classification, association rules and visualization. Predictive models use classification, regression and time series analysis to show the impact of a treatment to a patient based on the data of the past. Another way to categorize most of the data mining techniques distinguishes them into model building and clustering techniques. Model Building seeks to create a predictive model related to a specific question. Depending upon the techniques chosen, a model may be either opaque (results are clear, but the functions are unclear) or transparent (complete knowledge about the model at any prediction). Clustering attempts to segment a population into one or more groups that have (as far as we are concerned) similar characteristics and are therefore expected to behave in a similar manner.

The following list shows the commonly used techniques of data mining for knowledge discovery [19]-[22]:

1. Summarization: The relevant Data from the CPR have to be generalized and abstracted. The result of this step is the set of task-relevant data.
2. Classification: Classification is the process of assigning data items to one or many predefined classes. The classification model contains a set of classification rules. These rules are also used for future data. It derives a function or model, which determines the class of a model based on its attributes [22]. For the CPR is the definition of the rules a complex task as the rules have to cover a deep understanding of medical and economical knowledge. In the medical diagnosis, the classification is the most critical data mining technique. An example for classification is the definition of a group of patients with high blood pressure and the assignment of equivalent patients to the group. Classification is in the most important task of DM in CPR [20]. Classification is a DK task for which also other Artificial Intelligence approaches like neural networks and decision trees are often used.
3. Association: Search the records and finding association patterns by using defined rules. For example search for a set of symptoms of diseases of patient, hat also occur to patients with other diseases. Automated systems are filtering association rules based on findings from medical transaction databases: Association rules are ranked for medical knowledge by using formal ontologies.
4. Clustering: The clustering identifies classes or groups for a set of objects. The clustering maximizes the similarities of objects assigned to a class. This is based on the criteria defined on the attributes of the objects. After the decision and assignment process of an object to the cluster, the objects are labeled with the corresponding cluster.
5. Trend analysis or time series analysis: The Trend is the result of comparing time related data over a period of time, e.g. blood pressure over a period of one month. The objects are snapshots of entities with certain values that can change over the time.
6. Forecasting: This is the prediction of the value for an object based on the data from the past.
7. Visualization techniques: They help to discover patterns in medical data sets as a starting point. Afterwards other data mining techniques have to be used to determine the details of the patterns.

V. Setting the Trend of DM with a Selection of Techniques and Models

The selection of DM techniques and models shows the trend of DM in CPR with a brief review of the key concepts.

Since the beginning of DM it was the aim to automize DM techniques and models and to reduce the participation of human actions to a minimum. The beginning of this chapter describes Machine Learning, before the machine learning methods are described.

Machine Learning algorithms can be divided into supervised and unsupervised learning algorithms. In supervised learning, training examples exist of input-output pair patterns. Learning algorithms try to predict output values based of new examples, based on their input values. In unsupervised there are only input patterns without an explicit output available. Here is the aim of the algorithms to use input values to discover meaningful associations or patterns [3].

A. Probabilistic and Statistical Models

Probabilistic and statistical analysis techniques and models have a strong theoretical foundation in DM research. Assigned to the statistical techniques are regression analysis, discriminant analysis, time series analysis, principal component analysis, and multi-dimensional scaling. Because of their maturity those models are often used as benchmarks for comparison with newer machine learning techniques [3].

An advanced and popular probabilistic model for CPR is the Bayesian Model. It was originated in pattern recognition research and frequently used to classify different objects into predefined classes based on a set of features. The model stores the probability of each class, each feature, and each feature given each class, based on the training data. New instances will be classified according to the existing probabilities. There are many variations of the Bayesian Model.

An important and popular machine learning technique is the Support Vector Machines (SVM). It is based on statistical learning theory, which aims to find a hyperplane to best separate two or many classes. The applied model has shown encouraging results as it has the performance among other learning methods in document classification [3].

B. Symbolic Learning

Symbolic learning is implemented by applying algorithms that attempt to induce general concept descriptions that describe different classes of training example [3]. Many algorithms have been developed using algorithms to identify patterns that are useful in generating a concept description. Given a set of objects, symbolic learning created a decision tree that classifies all given objects correctly. At each step, the algorithm finds the attribute that best divides the objects into the different classes by minimizing the uncertainty. This way it is possible to create complete treatment plans in CPR [23]

C. Case Based Reasoning

CBR is a problem solving paradigm that utilizes the specific knowledge of previously experienced situations or cases. It consists in retrieving past cases that are similar to the current one and in reusing solutions which were used successfully in the past, the current case can be retained. CBR provides a solution for solving new problems and understanding unfamiliar situations.

In medicine, CBR can be seen as a suitable instrument to build decision support tools able to use tacit knowledge [24].

The algorithms find similarities of cases by using tacit knowledge. The data of experiences are compared to similar solutions that were successful in past cases [25].

The classic CBR uses a cycle with the steps of retrieve, reuse, revise, and retain. An example for CBR in using CPR is if a medical doctor wants to decide whether or not to prescribe a special medication for a patient or not. With CBR the decision would include the medical history of the patient and all patients with similar patterns, their physical states, emotional states, observed behaviors, cognitive status, as well as safety concerns. Each case in a CBR application aims to support the decision making process and therefore contains specific information of these factors for an individual patient [25]. The knowledge base is the collection of those cases in a library or the case base. The case base is organized to facilitate retrieval of the most similar, or most useful experiences when a new case arises. The past solutions would be the starting point for solving the new case. This is the first step of CBR, retrieve. Reuse is the step where adjustments are necessary to fit the case to the new situation. The result of the reuse step is a suggested solution. In case of further necessary adjustments is the revise step necessary. Here is the output the tested and repaired case with a confirmed solution. The last step in the CBR cycle enables the system to grow and to learn from the acquired experiences. According to [25] CBR is particularly useful and applicable in health science and the CPR because of established histories for health care professionals, many publications that can be easily encoded in cases, reasoning from many existing examples, extensive data stores are available in CPR, and cases can complement general treatment guidelines to support personalized medical care for individual patients

The formalization of medical facts and their relations results in a high complexity, but there are many examples and cases of diseases, treatments and outcomes. Furthermore, many diseases are not well enough understood in medicine to have universal applicable treatments

The main selection of CBR systems in health care in their order of chronological appearance are SHRINK: Aid with psychiatric diagnosis and treatment [26], MNAOMIA: For diagnosing and treating psychiatric eating disorders [25], PROTOS: Diagnosis of audiological and hearing disorders [25], [27], CASEY: Diagnosis of heart failure patients by comparing them to earlier cardiac patients with known diagnoses [25], ICONS: Therapy planning system that recommends antibiotic therapy for patients with bacterial infections in the intensive care unit (ICU) [28], KASIMIR: Breast cancer decision support system that also takes missing data and the threshold effect into account [29], and HR3Modul: Decision support system for diagnosing stress related disorders, including signal measures like breathing and heart rate expressed as physiological time series [30].

An important trend of CBR for CPR is the integration of multimedia case representation. This allows using CBR for medical image interpretation for comparing pictures of different modalities of different patient [31]. CBR is also useful for including other factors in the decision process like the co-occurrence of multiple diseases, time series features, overlapping diagnostic categories, the need to abstract features from time series representing temporal history, sensor signals, and continuous data.

D. Natural Language Processing

The content of CPR include a rich source of data and are often the major bottleneck for the deployment of effective clinical applications because textual information is difficult to access by computerized processes. Natural Language Processing (NLP) systems are automated methods containing some linguistic knowledge that aim to improve the management of information in text [32]. NLP allows the extraction of information and knowledge from medical notes, discharge summaries, and narrative patients reports. Current efforts on the construction of automated systems for filtering rules learned from medical transaction databases is an important area for CPR. The availability of a formal ontology allow the ranking of association rules by clarifying what are the rules confirming available medical knowledge, what are surprising, and which are to be filtered out. Currently, NLP systems in clinical environments process CPR to index and categorize reports, extract, structure, and codify clinical information in the reports to make them usable for other computerized applications, generate text to produce patient profiles and summaries, and improve interfaces to health care systems.

The challenges of NLP in the clinical domain described by [32] are the performance of the application, the availability of clinical text and confidentiality, the evaluation and sharing information across institutions, the Expressiveness as language can describe the same medical concept in many different ways, the heterogeneous formats, as there are no standards for writing a report in CPR, the abbreviated text in medical reports often omit information that can be interfered by health care employees based on their individual knowledge, the interpretation of clinical information as evident part of a medical report, as often further information are necessary to associate findings with potential diagnoses, and rare events can make it difficult to enable enough training examples for stabilization of NLP

NLP is based on in advance prepared formalization of the knowledge. NLP can be useful for ontology development, it can be used as a component in an ontology-driven information system and an NLP application can be enhanced with ontology.

There are different approaches to NLP in the CPR. Most approaches are using a combination of syntactic and semantic linguistic knowledge as well as heuristic domain knowledge [32]. Some use manually developed rules, and others are more statistically oriented. The NLP extraction process has two phases; first the report analyzer processes the report in order to identify segments and to handle irregularities. In the second phase the text analyzer as information extraction component uses linguistic knowledge associated with syntactic and semantic features. Also a conceptual model of the domain is used to structure and encode the clinical information [32]. The output is stored for subsequent use in clinical databases. NLP systems have different components that can vary from case to case. The main components according to [32] are morphological analysis: Process to break up original words into canonical forms, lexical look up: Words or phrases are matched against a lexicon to determine their syntactic and semantic properties, syntactic analysis: Determination of the structure of a sentence to establish relationships of the words in a sentence, semantic analysis Process to show the clinical relevance of words and phrases, and encoding: Process to map the clinically relevant terms to established concepts. This is important to achieve widespread use of the structured information.

In order for NLP to become a main method in CPR it is important to further develop the standardization of report structures in the CPR as well as the standardization of the information model representing clinical information and vocabularies.

E. Artificial Neutral Networks

Artificial Neural Network (ANN) or Neural Networks are computerized paradigms based on mathematical models with strong pattern recognition capabilities [33]. ANN are also called connectionist systems, parallel distributed systems, or adaptive systems, because they are comprised by a series of interconnected processing elements which work parallel in time [33]. ANN aim to build up information structures according to the human nervous system with a representation of neurons and synapses. An ANN is a graph consisting of many nodes connected to each other by weighted links. The knowledge in the ANN is represented by the totality of nodes and weighted links. Learning algorithms or learning rules can be used to adjust the connection weights in networks to predict or classify unknown examples. ANN work in the training mode and after stabilization in the testing mode. The process of ANN starts with a set of random weights and adjusts its weights automatically according to each learning example in an iterative process until the network stabilizes. This mode is called the learning mode, where the weights of the connections can change in order to respond to a present input.

Different types of ANN can solve many problems, like pattern recognition, pattern completion, determining similarities between pattern and data, interpolation of missing and noisy data, and automatic classification [34].

Many different types of ANN have been developed in the last two decades, e.g. the Self-organizing Map of Kohonen and the Hopfield networks described by Chen [3], [35]. Later in this chapter are Neuro Fuzzy Systems described, that also refer to ANN and are separated because of their growing importance. Particularly in the field of medicine and for usage of DM in CPR are ANN valuable as it is possible to build models with a high complexity, e.g. with multilayer feed forward networks. They can be defined as an array of processing elements arranged in layers. Information flows through each element in an input-output manner, where each element receives signals, manipulates them, and sends the signals to other connected elements.

F. Bayesian Networks

A particularly useful method for the CPR is represented by the Bayesian Networks (BN) which is used in different areas of medical applications. The BN represents the conjunction of knowledge representation, automated reasoning, and machine learning. The BN allows to explicitly represent the knowledge available in terms of a directed acyclic graph structure and a collection of conditional probability tables, and to perform probabilistic inference. BN use a directed acyclic graphical model to represent a set of random variables like quantities, latent variables, unknown parameters or hypotheses. Also the conditional relationships and independencies between the random variables are represented in the BN. The graphical structure represents knowledge about an uncertain domain. Each conditional relationship has its own probability function. Learning in BN is performed by intelligent algorithms. Influence diagrams help to generalize BN solve decision problems under uncertainty. In many practical settings the BN is unknown and one needs to learn it from the data. This problem is known as the BN learning problem, which can be stated informally as follows: Given training data and prior information (e.g. expert knowledge, casual relationships), estimate the graph topology (network structure) and the parameters [36]. Graphical models with undirected edges are generally called Markov random fields or Markov networks. These networks provide a simple definition of independence between any two distinct nodes based on the concept of a Markov blanket. Markov networks are popular in fields such as statistical physics and computer vision [36].

Machine learning and system learning for BN is to find the best matching Bayesian network graph with the best data fit for the decision problem.

G. Analytic Learning, Fuzzy Logic, and Neuro Fuzzy Systems

Knowledge is represented in analytical learning as logical rules and the performance of proofs for the rules. Traditional analytic learning systems depend on hard computing rules. As in the reality there is usually no distinction between values and classes, therefore fuzzy systems have been developed. Other concepts aim to avoid imprecise and vague information as they have a negative influence on the computed results. Fuzzy Systems use deliberately this type of information [34]. The result is often a simpler approach with more suitable models that are easier to handle. In the past Fuzzy Logic was not the first choice for DM in CPR because the simplicity did not fit to the complexity of medical and patient oriented data in CRM, but the trend has changed by combining different simple concepts, eg. Neuro Fuzzy Sytems.

Fuzzy logic is an extension of traditional proposition logic. It deals with approximate reasoning by extending the binary membership. In contrast to classical set theory, in which an object or a case either is a member of a given set, fuzzy set theory makes it possible that an object or a case belongs to a set only to a certain degree [34]. Interpretations of membership degrees include similarity, preference and uncertainty. They can state how similar an object or case it to a prototypical one. They can indicate preferences between sub optimal solutions to a clinical problem, or they can model uncertainty in case of an imprecisely described situation or term [34]. For the CPR the set up of a fuzzy system is useful as many medical information are linguistic, vague or imprecisely described because a complete description would be too complex. However, the limitation of the fuzzy system is reached when fuzzy concepts have to be represented by concrete membership degrees, which ensures that the system works as expected. A fuzzy system can be used to solve a problem, if knowledge exists about the solution in the form of if-then rules.

The development of Neuro Fuzzy Systems enhances the Fuzzy Systems where knowledge is represented in an interpretable manner, by the ability of ANN of learning. The advantage of the combination is that a problem can be solved without the need to analyze the problem itself in detail.

For CPR are the hybrid Neuro Fuzzy models interesting, which combine neuronal networks with fuzzy systems in a homogeneous architecture. The architecture can either be interpreted as a special neuronal network with fuzzy parameters, or as a fuzzy system implemented in a parallel distributed way.

H. Evolution Based Models

Evolution based models refer to computer-based methods inspired by biological mechanisms of natural evolution.. Evolution based algorithms have been applied to various optimization problems. They were developed on the basis of genetic principles. A population of potential solutions is initiated in the first step. The iterative process changes the population based on the operations mutation and crossover in different generations. The crossover operation is a high level process that aims at exploitation while mutation is a unary process that aims at exploration [3]. The selection process goes over different generations and selects the best fitting individuals. At the end of the process is the best solution presented. Due to the stochastic and global-search capability this technique is popular in medical informatics research [3].

VI. Conclusion

CPR contain heterogeneous data in heterogeneous information systems and from heterogeneous sources. Not only the information technology improves the complexity of data mining in electronic patient files, but also the business site is very complex in terms of the medical description of doctors who try to describe the disease of a patient. DM is particularly useful in this domain.

The described techniques and methods of DM in CPR prove the fast development of research trends over the last decades. The usage of the research findings and developed new techniques and models will reach the full momentum after the CPR coverage rate has reached a much higher value. Currently, many systems in health care are separated isle solutions with a low integration rate. The benefits of DM research in CPR will be fully unlooked when the data will be interlinked. All methods have shown that the result of the decision proposal is relying on the quality of the data basis. This is obvious in Data Mining and shows the growing importance for Data Mining research and the usage in CPR. Future internet technologies will allow to use Data Mining in the Web over a broad data basis and link the results to existing CPR. This will allow to access knowledge in a today not known dimension and revolutionize the decision making process. 'The current solutions of today build the foundation for the future scenarios.

Most of the described examples of DM techniques and methods related to practical problems in CPR are directed on one single problem, e.g. diagnosis for stress related heart attacks. Future trends will be integrating the different approaches, technologies, methodologies, and constructs into a DM framework of methodologies that link together different approaches. The start is already made with the linkage of ANN with Fuzzy Logic into Neuro Fuzzy Systems.

The challenges of data mining will also remain in future to deal with different scientific areas, to understand the patterns, to deal with complex relationships between attributes, interpolate missing or noisy data, mining very large databases, handle changing data and integrate the data with other data base systems. All these challenges are particularly important for CPR.

References

[1] K.J . Cios and G. W. Moore, "*Medical data mining and knowledge discovery*," Berlin Heidelberg: Springer, 2001, pp. 1-67.

[2] S. S. R. Abidi, "Knowledge management in healthcare towards 'knowledge-driven' decision support services," *Intl J Med Inf*, 63, 5-18.

[3] H. Chen, S. S. Fuller, C. Friedmann, and W. Hersch, "Knowledge management, data mining, and text mining in medical informatics," in *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*, H. Chen, S.S. Fuller, C. Friedmann, and W. Hersch, Eds., New York: Springer Science + Business Media, 2005, pp. 3-22.

[4] K. J. Cios and G. W. Moore, "Uniqueness of medical data mining," in *Artificial Intelligence in Medicine*, Elsevier, 26 (2002), 2002, 1–24, http://www.cs.uwm.edu/~mani/fall05/smi/link/pdf/aimj-medkdd1.pdf (25.5.2011).

[5] V. Sintchenko, "Information processing in clinical decision making," in *Handbook of research on informatics in healthcare and biomedicine*, INITIAL Lazakidou, Ed., Hersey London: Idea, 2006, pp. 1-8.

[6] K. P. Soman, S. Diwakar, and V. Ajay, "*Insight into data mining theory and practice*," New Delhi: Prentice-Hall, 2006, pp. 1-19.

[7] D. L. Sackett, W. M. Rosenberg, J. A. Gray, R. B. Haynes, and W. S. Richardson,"Evidence based medicine: What it is and what it isn't," *BMJ* 312 (7023), 71-2, 2009.

[8] D. T. Heinze, M. L. Morsch, B. C. Potter, R. E. Jr. Sheffer, "Medical i2b2 NLP smoking challenge: the A-Life system architecture and methodology", *J Am Med Inform Assoc*, 15(1), 40-3, 2008.

[9] S. Bullas and J. Bryant, "Complexity and its impacts for health systems Implementation," in *Information Technology in Health Care 2007*, J.I. Westbrook, E.W. Coiera, J.L. Callen, and J. Aarts, Eds., Amsterdam Lancaster Fairfax: IOS, 2007, pp.37-44.

[10] P. Schmücker, "Elektronic patient file and digital archivein hospitals," in *MEDNET Workbook integrated health care*, U. Eissing, N. Kuhr,and G. Noelle, Eds., ORT UND VERLAG, 2003, pp.103-115.

[11] K. A. Wagner, F. W. Lee, and J. P. Glaser, *"Health Care Information Systems: A Practical Approach for Health Care Management*," San Francisco: Wiley, 2005, pp. 3-42.

[12] O. Galani and A. Nikiforou, "Electronic health record," in *Handbook of research on informatics in healthcare and biomedicine*, INITIAL Lazakidou, Ed., Hersey London: Idea, 2006, pp.1-8.

[13] M.A. Montero and S. Prado, "Electronic health record as a knowledge management tool in the scope of health, " in *Knowledge Management for Health Care Procedures: ECAI 2008 Workshop K4HelP 2008*, D. Riano, Ed., Berlin Heidelberg: Springer, 2008, pp. 152-166.

[14] B. Fong, A. C. M. Fong, and C. K. Li, "*Telemedicine technologies: information technologies in medicine and telehealth*," West Sussex: Wiley, 2011, pp. 67-105.

[15] M. L. Owoc, "Knowledgebases: a management context and development determinants," in Proc. of 2003 Informing Science and Information Technology Education Conference, Pori, 2003, pp. 1193-1199, http://informingscience.org/proceedings/IS2003Proceedings/docs/147Owoc.pdf (20.05.2011).

[16] J. van der Zwaan, E. T. K. Sang, and M. de Rijke, "An experiment in automatic classification of pathological reports," in *Artificial Intelligence in Medicine, AIME 2007 Amsterdam, July 2007, Proceedings*, R. Bellazzi, A. Abu-Hanna, and J. Hunter, Eds., Berlin Heidelberg: Springer, 2007, pp. 207-216.

[17] M. Gruninger and J. Lee, "Onthology: applications and design," *Communication of the ACM,* 45(2), 2002, pp. 39-41.

[18] C. Romero-Tris, D. Riano, and F. Real, "Ontology-based retrospective and prospective diagnosis and medical knowledge personalization," in *Knowledge Representation for Health-Care, ECAI 2010 Workshop KR4HC 2010, Lisbon, Portugal, August 2010*, D. Riano, A. Teije, S. Miksch, and M. Peleg, Eds., Berlin Heidelberg: Springer, 2011, pp. 1-15.

[19] Yo. Wang, D. Niu, and Ya. Wang, "Power load forecasting using data mining and knowledge discovery technology," in *Intelligent Information and Database Systems: Second International Conference, ACIIDS, March 2010*, N.T. Nguyen, M.T. Le, and J. Swiatek, Eds., Berlin Heidelberg New York: Springer, 2010, pp. 319-328.

[20] S. K. Wasan,V. Bhatnagar, H. Kaur, "The impacts of data mining techniques on medical diagnostics," *Data Science Journal*, Volume 5, 19 October 2006, http://www.jstage.jst.go.jp/article/dsj/5/0/119/_pdf (24.05.2011).

[21] S. Tangsripairoj and M. H. Samadzadeh, "A taxonomy of data mining applications supporting software reuse," in *Intelligent System Design and Applications*, A. Abraham, K. Franke, and M. Köppen M, Eds., Berlin Heidelberg: Springer, 2003, pp. 303-311.

[22] M. Kwiatkowska, M. S. Atkins, L. Matthews, N. T. Ayas, and C. F. Ryan, "Knowledge-based induction of clinical rediction rules," in *Data Mining and Medical Knowledge Management: Cases and Applications*, P. Berka, J. Rauch, and D.A. Zighed, Eds., Hershey London: Idea, 2009, pp. 350-375.

[23] S. N. S. Saad, A. M. Razali, A. A. Bakar, and N. R. Suradi, "Developing treatment plan support in outpatient health care delivery with decision trees technique", in *Advanced Data Mining and Applications*, L.Cao, Y. Feng, and J. Zhong, Eds., Berlin Heidelberg: Springer, 2010, pp. 475-482.

[24] R. Schmidt, S. Montani, R. Bellazzi, L. Portinale, and L. Gierl, „Case-based reasoningfor medical knowledge-based systems," Intl J Med Inf 64(2-3), 2001, pp. 355-367.

[25] I. Bichindaritz and C. Marling C, "Case-based reasoning in health science: foundations and research directions," in *Computational Intelligence in Healthcare 4: Advanced Methodologies*, I. Bichindaritz, S. Vaidya, A. Jain, and L. C. Jain, Eds., Berlin Heidelberg: Springer, 2010, pp. 127-157.

[26] J. L. Kolodner and R. M. Kolodner, "Using experience in clinical problem solving: introduction and framewort," *IEEE Transactions on Systems, Man and Cybernetics*, 17(3), 1987, pp. 420-431.

[27] S. Cox, M. Oakes, S. Wermter, and M. Hawthorne, "AudioMine: medical data mining in heterogeneous audiology records. world academy of science", *Engineering and Technology*, 1. 2005, http://www.waset.org/journals/waset/v1/v1-2.pdf (25.5.2011).

[28] R. Schmidt, B. Pollwein, and L. Gierl, "Case-based reasoning for antibiotics therapy advice," in ICCBR 1999. LNCS (LNAI), K.-D. Althoff, R. Bergmann, and L.K. Branting, Eds., vol. 1650, Heidelberg: Springer, 1999, pp. 550-559.

[29] M. d'Aquin, J. Lieber, and A. Napoli,"Adaptation knowledge acquisition: a case study for case-based decision support in oncology," *Computerational Intelligence*, 22(3-4), 2006, pp. 161-176.

[30] P. Funk and N. Xiong, "Case-based reasoning and knowledge discovery in medical applications with time series," *Computerational Intelligence*, 22(3-4), 2006, pp. 238-253.

[31] T. S. Yoo, "3D medical informatics," *in Medical Informatics: Knowledge Management and Data Mining in Biomedicine*, H. Chen, S. S. Fuller, C. Friedmann, and W. Hersch, Eds., New York: Springer Science + Business Media, 2005, pp. 333-355.

[32] C. Friedmann, "Semantic text parsing for patient records," in *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*, H. Chen, S. S. Fuller, C. Friedmann, and W. Hersch W, Eds., New York: Springer Science + Business Media, 2005, pp. 423-448.

[33] M. Sordo, S. Vaidya, and L.C. Jain, "An introduction to computational intelligence in healthcare: New Directions," in *Advanced Computational Intelligence Paradigms in Healthcare*, M. Sordo, S. Vaidya, and L. C. Jain, Eds., 3rd ed., Berlin Heidelberg: Springer, 2010, pp. 1-26.

[34] A. Klose, "Extracting fuzzy classification rules from partially labeled data," in *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 8, Berlin Heidelberg: Springer, 2004, pp. 417 427.

[35] S. Eggers, Z. Huang, H. Chen, L. Yan, C. Larson, A. Rashid, M. Chau, and C. Lin C, "Mapping medical informatics research," in *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*, H. Chen, S.S. Fuller, C. Friedmann, and W. Hersch W, Eds., New York: Springer Science + Business Media, 2005, pp. 35-58.

[36] I. Ben-Gal, "Bayesian networks," in *Encyclopedia of Statistics in Quality & Reliability*, F. Ruggeri, F. Faltin, and R. Kenett, Eds., Wiley, 2008, http://onlinelibrary.wiley.com/doi/10.1002/ 9780470061572. eqr089/full (27.05.2011).