

# Machine Learning for Medical Diagnosis: History, State of the Art and Perspective

*Igor Kononenko*

*University of Ljubljana*

*Faculty of Computer and Information Science*

*Tržaška 25, 1001 Ljubljana, Slovenia*

*tel: +386-1-4768390, fax: +386-1-4264647*

*e-mail: igor.kononenko@fri.uni-lj.si*

## Abstract

The paper provides an overview of the development of intelligent data analysis in medicine from a machine learning perspective: a historical view, a state of the art view and a view on some future trends in this subfield of applied artificial intelligence. The paper is not intended to provide a comprehensive overview but rather describes some subareas and directions which from my personal point of view seem to be important for applying machine learning in medical diagnosis. In the historical overview I emphasize the naive Bayesian classifier, neural networks and decision trees. I present a comparison of some state of the art systems, representatives from each branch of machine learning, when applied to several medical diagnostic tasks. The future trends are illustrated by two case studies. The first describes a recently developed method for dealing with reliability of decisions of classifiers, which seems to be promising for intelligent data analysis in medicine. The second describes an approach to using machine learning in order to verify some unexplained phenomena from complementary medicine, which is not (yet) approved by the orthodox medical community but could in the future play an important role in overall medical diagnosis and treatment.

## 1 Introduction

Artificial intelligence is a part of computer science that tries to make computers more intelligent. One of the basic requirements for any intelligent behaviour is learning. Most of the researchers today agree that there is no intelligence without learning. Therefore, machine learning (Shavlik and Dietterich, 1990; Michie et al., 1994; Mitchell, 1997; Michalski et al., 1998) is one of major branches of artificial intelligence and, indeed, it is one of the most rapidly developing subfields of AI research.

Machine learning algorithms were from the very beginning designed and used to analyse medical data sets. Today machine learning provides several indispensable tools for intelligent data analysis. Especially in the last few years, the digital revolution provided relatively inexpensive and available

means to collect and store the data. Modern hospitals are well equipped with monitoring and other data collection devices, and data is gathered and shared in large information systems. Machine learning technology is currently well suited for analyzing medical data, and in particular there is a lot of work done in medical diagnosis in small specialized diagnostic problems.

Data about correct diagnoses are often available in the form of medical records in specialized hospitals or their departments. All that has to be done is to input the patient records with known correct diagnosis into a computer program to run a learning algorithm. This is of course an oversimplification, but in principle, the medical diagnostic knowledge can be automatically derived from the description of cases solved in the past. The derived classifier can then be used either to assist the physician when diagnosing new patients in order to improve the diagnostic speed, accuracy and/or reliability, or to train students or physicians non-specialists to diagnose patients in a special diagnostic problem.

The aim of this paper is to provide an overview of the development of the intelligent data analysis in medicine from a machine learning perspective: a historical view, a state of the art view and a view on some future trends in this subfield of applied artificial intelligence, which are, respectively described in the following three sections. None of the three sections is intended to provide a comprehensive overview but rather describe some subareas and directions which from my personal point of view seem to be important for medical diagnosis. In the historical overview I emphasize the naive Bayesian classifier, neural networks and decision trees. Section 3 presents a comparison of some state of the art systems, one or two representatives from each branch of machine learning, when applied to several medical diagnostic tasks. The future trends are illustrated by two case studies. Section 4.1 describes a recently developed method for dealing with reliability of decisions of classifiers, which seems to be promising for intelligent data analysis in medicine. Section 4.2 describes an approach to using machine learning in order to verify some unexplained phenomena from complementary medicine, which is not (yet) approved by the orthodox medical community but could in the future play an important role in overall medical diagnosis and treatment.

## 2 Historical overview

As soon as electronic computers came into use in the fifties and sixties, the algorithms were developed that enabled modeling and analysing large sets of data. From the very beginning three major branches of machine learning emerged. Classical work in symbolic learning is described by Hunt et al. (1966), in statistical methods by Nilsson (1965) and in neural networks by Rosenblatt (1962). Through the years all three branches developed advanced methods (Michie et al., 1994): statistical or pattern recognition methods, such as the  $k$ -nearest neighbours, discriminant analysis, and Bayesian classifiers, inductive learning of symbolic rules, such as top-down induction of decision trees, decision rules and induction of logic programs, and artificial neural networks, such as the multilayered feedforward neural network with backpropagation learning, the Kohonen's self-organizing network and the Hopfield's associative memory.

## 2.1 The naive Bayesian classifier

I limit the historical overview of statistical methods to the naive Bayesian classifier. From the very beginning I was very interested in it. The algorithm is extremely simple but very powerful, and later I discovered that it can provide also comprehensive explanations which was confirmed in long discussions with physicians.

I was fascinated with its efficiency and ability to outperform most advanced and sophisticated algorithms in many medical and also non-medical diagnostic problems. For example, when compared with six algorithms, described in Section 3, the naive Bayesian classifier outperformed all the algorithms on five out of eight medical diagnostic problems (Kononenko et al., 1998). Another example is a hard problem in mechanical engineering, called mesh design. In one study, sophisticated inductive logic programming algorithms achieved modest classification accuracy between 12 and 29% (Lavrač and Džeroski, 1994; Pompe and Kononenko, 1997) while the naive Bayesian classifier achieved 35%. The naive Bayesian classifier became for me a benchmark algorithm that in any medical domain has to be tried before any other advanced method. Other researcher had similar experience. For example, Spiegelhalter et al. (1993) were for several man-months developing an expert system based on Bayesian belief networks for diagnosing the heart disease for new-born babies. The final classification accuracy of the system was 65.5%. When they tried the naive Bayesian classifier, they obtained 67.3%.

The theoretical basis for the successful applications of the naive Bayesian classifier (also called simple Bayes) and its variants was developed by Good (1950; 1964). We demonstrated the efficiency of this approach in medical diagnosis and other applications (Kononenko et al., 1984; Cestnik et al., 1987). But only in the early nineties the issue of the transparency (in terms of the sum of information gains in favor or against a given decision) of this approach was also addressed and shown successful in the applications in medical diagnosis (Kononenko, 1989; 1993). This issue is addressed in more detail in Section 3.4 and illustrated in Table 2.

Lately, various variants and extensions of the naive Bayesian classifier have been developed. Cestnik (1990) developed the m-estimate of probabilities that significantly improved the performance of the naive Bayesian classifiers in several medical problems. Kononenko (1991) developed a semi-naive Bayesian classifier that goes beyond the “naivety” and detects dependencies between attributes. The advantage of fuzzy discretization of continuous attributes within the naive Bayesian classifier is described in (Kononenko, 1992). Langley (1993) developed a system that uses the naive Bayesian classifier in the nodes of the decision tree. Pazzani (1997) developed another method for explicit searching of dependencies between attributes in the naive Bayesian classifier. The transparency of the naive Bayesian classifier can be further improved with the appropriate visualization (Kohavi et al., 1997).

## 2.2 Neural networks

After Rosenblatt (1962) developed a basic delta learning rule for single layered perceptrons, Minsky and Papert (1969) proved that this rule cannot solve nonlinear problems. Only few scientists continued with research of neural networks. The field gained a prominent impulse with the seminal works of Hopfield (1982; 1984) on associative neural networks and even more with the publication of the

backpropagation learning rule for multilayered feedforward neural networks (Rumelhart et al., 1986). This learning rule and its variant enabled the use of neural networks in many hard medical diagnostic tasks. However, neural networks were typically used as black-box classifiers lacking the transparency of generated knowledge and lacking the ability to explain the decisions. Lately, many advanced variants of neural network algorithms were developed and some do provide for the transparency of decisions (Haykin, 1994).

At the very beginning I was very enthusiastic with neural networks. When I read papers by Hopfield (1982; 1984) and Rumelhart et al. (1986) for the first time in my life I had a feeling that I understand how neurons in the brain can do useful computations. The early inspiration led to my Ph.D. thesis on Bayesian neural networks (Kononenko, 1989a) but later my research interest moved back to symbolic learning.

### 2.3 Symbolic learning

Probably the most promising area for medical data analysis was from the very beginning the symbolic learning of decision trees and decision rules. Hunt et al. (1966) used their Concept Learning System (CLS) for building decision trees in medical diagnosis and prognosis. They state (p. 170):

*“In medicine fairly large files of records may be obtained in the course of routine hospital administration or from a special survey. Such records are often examined in order to plan an intensive, and perhaps expensive, specialized investigation. A drawback to this research strategy is that it is difficult to organize large files of records to reveal complex interactions in a manner that can be understood by the human investigator. Some help can be obtained by using computer oriented techniques of information retrieval, such as program to print selected two- and three-way tables plotting one variable against another. The investigator still must nominate the variable in which he is interested, since such programs have no way of discovering interesting patterns on their own. A CLS program, on the other hand, is designed to do precisely this.”*

Generating decision trees and decision rules became an active research area after Quinlan (1979) developed the famous Iterative Dichotomizer 3 (ID3) algorithm and Michalski and Chilausky (1980) successfully applied the system AQ in a plant disease diagnostic task. Bratko and Mulec (1980) applied ID3 to a hard diagnostic problem in oncology and later various descendants of ID3 were developed and successfully applied to various medical diagnostic problems. For example, our system Assistant (Kononenko et al., 1984; Cestnik et al., 1987) was applied to various problems in oncology (localization of primary tumor, prognosing the recurrence of breast cancer, lymphography), urology (lower urinary tract dysfunctions), and the prognosis of survival in hepatitis. Independently of ID3, Breiman et al. (1984) developed the system CART and applied it to several diagnostic and prognostic tasks in cardiology and oncology.

A very incomplete and only illustrative list of applications of machine learning in medical diagnosis in the eighties include applications in oncology (Elomaa & Holsti, 1989), liver pathology (Lesmo et al., 1982), diagnosis of thyroid diseases (Horn et al., 1985; Hojker et al., 1988; Quinlan et al., 1987), rheumatology (Kononenko et al., 1988; Karalič & Pirnat, 1990; Kern et al. 1990), diagnosing

craniostenosis syndrome (Baim, 1988), dermatoglyptic diagnosis (Chan & Wong, 1989), cardiology (Bratko et al., 1989; Clark & Boswell, 1991; Catlett, 1991), neuropsychology (Muggleton, 1990), gynaecology (Nunez, 1990), and perinatology (Kern et al., 1990).

In nineties the Relief algorithm and its successors were developed (Kira and Rendell, 1992a;b; Kononenko, 1984; Robnik-Šikonja and Kononenko, 1997) that enabled the estimation of the quality of each attribute in the context of other attributes. This amazing algorithm not only significantly improved the applicability of the induction of decision trees and similar algorithms but also improved the transparency of decision trees. The structure of generated trees was more human-like, which was confirmed in several diagnostic tasks (Kukar et al., 1996; Kononenko et al., 1998).

### 3 State of the art

In this section we give a description of specific requirements that any machine learning system has to satisfy in order to be used in the development of applications in medical diagnosis. Several learning algorithms are briefly described. We compared the performance of all the algorithms on several medical diagnostic and prognostic problems and their appropriateness for applications in medical diagnosis is discussed.

#### 3.1 Specific requirements for machine learning systems

For a machine learning (ML) system to be useful in solving medical diagnostic tasks, the following features are desired: good performance, the ability to appropriately deal with missing data and with noisy data (errors in data), the transparency of diagnostic knowledge, the ability to explain decisions, and the ability of the algorithm to reduce the number of tests necessary to obtain reliable diagnosis.

In this section we first discuss these requirements. Then we overview a comparison study (Kononenko et al., 1998) of seven representative machine learning algorithms to illustrate more concretely the points made.

**Good performance:** The algorithm has to be able to extract significant information from the available data. The diagnostic accuracy on new cases has to be as high as possible. Typically, most of the algorithms perform at least as well as the physicians and often the classification accuracy of machine classifiers is better than that of physicians when using the same description of the patients. Therefore, if there is a possibility to measure the accuracy of physicians, their performance can be used as the lower bound on the required accuracy of a ML system in the given problem.

In the majority of learning problems, various approaches typically achieve similar performance in terms of the classification accuracy although in some cases some algorithms may perform significantly better than the others (Michie et al., 1994). Therefore, apriori almost none of the algorithms can be excluded with respect to the performance criterion. Rather, several learning approaches should be tested on the available data and the one or few with best estimated performance should be considered for the development of the application.

**Dealing with missing data:** In medical diagnosis very often the description of patients in patient

records lacks certain data. ML algorithms have to be able to appropriately deal with such incomplete descriptions of patients.

**Dealing with noisy data:** Medical data typically suffer from uncertainty and errors. Therefore machine learning algorithms appropriate for medical applications have to have effective means for handling noisy data.

**Transparency of diagnostic knowledge:** The generated knowledge and the explanation of decisions should be transparent to the physician. She should be able to analyse and understand the generated knowledge. Ideally, the automatically generated knowledge will provide to the physician a novel point of view on the given problem, and may reveal new interrelations and regularities that physicians did not see before in an explicit form.

**Explanation ability:** The system must be able to explain decisions when diagnosing new patients. When faced with an unexpected solution to a new problem, the physician shall require further explanation, otherwise she will not seriously consider the system's suggestions. The only possibility for physicians to accept a "black box" classifier is in the situation where such a classifier outperforms by a very large margin all other classifiers including the physicians themselves in terms of the classification accuracy. However, such situation is typically highly improbable.

**Reduction of the number of tests:** In medical practice, the collection of patient data is often expensive, time consuming, and harmful for the patients. Therefore, it is desirable to have a classifier that is able to reliably diagnose with a small amount of data about the patients. This can be verified by providing all candidate algorithms with a limited amount of data. However, the process of determining the right subset of data may be time consuming as it is essentially a combinatorial problem. Some ML systems are themselves able to select an appropriate subset of attributes, i.e., the selection is done during the learning process and may be more appropriate than others that lack this facility.

### 3.2 Brief description of some state-of-the-art algorithms

In this subsection we briefly describe seven representative algorithms from symbolic learning, statistical learning and neural networks: three decision tree builders (Assistant-R, Assistant-I, and LFC), two variants of the Bayesian classifiers (the naive and the semi-naive Bayesian classifier), a state of the art neural network which uses the backpropagation learning with weight elimination, and the  $k$ -nearest neighbors algorithm.

**Assistant-R:** is a reimplementaion of the Assistant learning system for top down induction of decision trees (Cestnik et al., 1987). The main difference between Assistant and its reimplementaion Assistant-R is that ReliefF is used for attribute selection (Kononenko, 1994). ReliefF is an extended version of RELIEF, developed by Kira and Rendell (1992a;b), which is a non-myopic heuristic measure that is able to estimate the quality of attributes even if there are strong conditional dependencies between attributes. In addition, wherever appropriate, instead of the relative frequency, Assistant-R

uses the  $m$ -estimate of probabilities, which was shown to often improve the performance of machine learning algorithms (Cestnik, 1990).

**Assistant-I:** A variant of Assistant-R that instead of ReliefF uses information gain for the selection criterion, as the original Assistant does.

**LFC:** Ragavan et al. (1993; Ragavan and Rendell, 1993) use limited lookahead in their LFC (Lookahead Feature Construction) algorithm for top down induction of decision trees to detect significant conditional dependencies between attributes for constructive induction. LFC generates binary decision trees. At each node, the algorithm constructs new binary attributes from the original attributes, using logical operators (conjunction, disjunction, and negation). From the constructed binary attributes, the best attribute is selected and the process is recursively repeated on two subsets of training instances, corresponding to two values of the selected attribute.

**Naive Bayesian Classifier:** A classifier that uses the naive Bayesian formula to calculate the probability of each class  $C$  given the values  $V_i$  of all the attributes for an instance to be classified, assuming the conditional independence of the attributes given the class:

$$P(C|V_1..V_n) = P(C) \prod_i \frac{P(C|V_i)}{P(C)} \quad (1)$$

A new instance is classified into the class with maximal calculated probability. The  $m$ -estimate of probabilities makes the naive Bayesian classifier more robust (Cestnik, 1990).

**Semi-naive Bayesian Classifier:** Kononenko (1991) developed an extension of the naive Bayesian classifier that explicitly searches for dependencies between the values of different attributes. If such dependency is discovered between two values  $V_i$  and  $V_j$  of two different attributes then they are not considered as conditionally independent. Accordingly the term

$$\frac{P(C|V_i)}{P(C)} \times \frac{P(C|V_j)}{P(C)}$$

in Equation (1) is replaced with

$$\frac{P(C|V_i, V_j)}{P(C)}$$

For such replacement a reliable approximation of the conditional probability  $P(C|V_i, V_j)$  is required. Therefore, the algorithm trades-off between the non-naivety and the reliability of approximations of probabilities.

**Backpropagation with weight elimination:** The multilayered feedforward artificial neural network is a hierarchical network consisting of two or more fully interconnected layers of processing units - neurons. The task of the learning algorithm is to determine the appropriate weights on the interconnections between neurons. Backpropagation of error in multilayered feedforward neural network (Rumelhart et al., 1986) is a well known learning algorithm and also the most popular among algorithms for training artificial neural networks. Well known problems with backpropagation are the

Table 1: The appropriateness of various algorithms for medical diagnosis.

classifier	performance	transparency	explanations	reduction	miss. data handling
Assistant-R	good	very good	good	good	acceptable
Assistant-I	good	very good	good	good	acceptable
LFC	good	good	good	good	acceptable
naive Bayes	very good	good	very good	no	very good
semi-naive Bayes	very good	good	very good	no	very good
backpropagation	very good	poor	poor	no	acceptable
$k$ -NN	very good	poor	acceptable	no	acceptable

selection of the appropriate topology of the network and overfitting the training data. An extension of the basic algorithm that uses the weight elimination technique (Weigand et al., 1990) addresses both problems. The idea is to start with too many hidden neurons and to introduce into the criterion function a term that penalizes large weights on the connections between neurons. With such criterion function the algorithm, during training, eliminates an appropriate number of weights and neurons in order to obtain the appropriate generalization on the training data.

**$k$ -NN:** The  $k$ -nearest neighbor algorithm. For a given new instance the algorithm searches for  $k$  nearest training instances and classifies the new instance into the most frequent class of these  $k$  instances.

### 3.3 An overview of comparison of algorithms on medical problems

We compared the performance of the algorithms on eight medical data sets (Kononenko et al, 1998). In the following we discuss how various algorithms fit the requirements. Table 1 summarizes the comparison of algorithms with respect to the appropriateness for developing applications in medical diagnostic and prognostic problems.

Among the compared algorithms only decision tree builders are able to select the appropriate subset of attributes. With respect to the criterion of reduction of the number of tests, these algorithms have clear advantage over other algorithms.

With respect to the performance criterion the algorithms are more similar. The best performance was achieved by the naive and semi-naive Bayesian classifiers. In medical data sets, attributes are typically relatively conditionally independent given the class. Physicians try to define conditionally independent attributes. Humans tend to think linearly and independent attributes make the diagnostic process easier. Therefore, it is not surprising that the Bayesian classifiers show clear advantage on medical data sets. It is interesting that the performance of the  $k$ -NN algorithm is also good in these domains.

With respect to the transparency and the explanation ability criteria there are great differences between the algorithms:

***k*-nearest neighbours:** As *k*-NN does no generalization, the transparency of knowledge representation is poor. However, to explain the decision of the algorithm, a predefined number (*k*) of nearest neighbours from training set is shown. This approach is analogous to the approach used by domain experts who make decisions on the basis of previously known similar cases. Such explanation ability is assessed by physicians as acceptable.

**Naive and semi-naive Bayes:** Here, knowledge representation consists of a table of conditional probabilities which seems to be of interest to physicians. Therefore such knowledge representation is assessed as good. On the other hand, the decisions of Bayesian classifiers can be naturally interpreted as the sum of information gains (Kononenko, 1993). The amount of information necessary to find out that an instance belongs to class *C*, is given by:

$$-\log_2 P(C|V_1, \dots, V_n) = -\log_2 P(C) - \sum_i (-\log_2 P(C) + \log_2 P(C|V_i)) \quad (2)$$

Therefore, the decisions of the Bayesian classifiers can be explained with the sum of information gains from all attributes in favor or against the given class. In the case of the semi-naive Bayesian classifier, the process is exactly the same, except when the tuples of joined attribute/value pairs occur. In this case, instead of simple attribute values, the joined values are used.

Such information gains can be listed in a table to sum up the evidence for/against the decision. Table 2 provides a typical explanation of one decision (Kukar et al, 1996). Each attribute has an associated strength, which is interpreted as the amount of information in bits provided by that attribute. It can be in favor or against the classifier's decision. One of the main advantages of such explanation is that it uses all available attributes. Such explanation was found by physicians as very good and they feel that Bayesian classifiers solve the task in a similar way they diagnose. Namely, they also sum up the evidence for/against a given diagnosis.

**Backpropagation neural networks** have non-transparent knowledge representation and in general cannot easily explain their decisions. This is due to the large number of real-valued weights which all influence the result. In some cases it is possible to extract symbolic rules from the trained neural network. However, the rules tend to be large and relatively complex. Craven and Shavlik (1993) compare rules extracted from a neural network with rules produced by Quinlan's (1993) C4.5 system. The rules for a *NetTalk* data set extracted from a neural network have on the average over 30 antecedents per rule compared to 2 antecedens for C4.5. Such rules are too complicated and hardly offer a useful explanation to a domain expert.

**Decision trees (Assistant-I and Assistant-R):** can be used without the computer and are fairly easy to understand. Positions of attributes in the tree, especially the top ones, often directly correspond to the domain expert's knowledge. However, in order to produce general rules, these methods use pruning which drastically reduces the tree sizes. Correspondingly, the paths from the root to the leaves are shorter, containing only few, although most informative attributes. In many cases the physicians feel that such a tree describes very poorly the diagnoses and is therefore not sufficiently informative (Pirnat et al., 1989). However, as mentioned earlier, the structure of generated trees by Assistant-R is more human-like, which was confirmed in several diagnostic tasks (Kukar et al., 1996;

Table 2: Semi-naive Bayes: an explanation of a decision in the femoral neck fracture recovery problem.

Decision = No complications (correct)

Attribute value	For decision (bit)	Against decision (bit)
Age = 70 - 80	0.07	
Sex = Female		-0.19
Mobility before injury = Fully mobile	0.04	
State of health before injury = Other	0.52	
Mechanism of injury = Simple fall		-0.08
Additional injuries = None	0.00	
Time between injury and operation > 10 days	0.42	
Fracture classification acc. to Garden = Garden III		-0.30
Fracture classification acc. to Pauwels = Pauwels III		-0.14
Transfusion = Yes	0.07	
Antibiotic profilaxis = Yes		-0.32
Hospital rehabilitation = Yes	0.05	
General complications = None		-0.00
<b>Combination:</b>	0.21	
Time between injury and examination < 6 hours AND Hospitalization time between 4 and 5 weeks		
<b>Combination:</b>	0.63	
Therapy = Artroplastic AND Anticoagulant therapy = Yes		

Kononenko et al., 1998).

**Lookahead feature construction (LFC)** also generates decision trees. However, in each node a potentially complex logical expression is used instead of a simple attribute value. The generated trees can therefore be smaller. The expressions may represent valid concepts from the domain. However, on the lower levels of the tree the expressions are often very specific and typically meaningless. Due to complex logical expressions in nodes, the number of attributes used to classify an instance can be greater than in usual decision trees.

## 4 Future trends - two case studies

There are many directions in which future development of machine learning in medical diagnosis may take place. Some may rely on new trends in computer technology or technology of medical equipment, however, probably more important is going to be the development of new machine learning algorithms and the philosophy of medical diagnosis. We do not want to speculate all possible trends. Instead we describe two case studies that illustrate the new trends in the development of machine learning algorithms and how machine learning methodology can support a possible change of philosophy of medical diagnosis.

The first case study describes a recently developed method for dealing with reliability of decisions of classifiers, which seems to be promising for intelligent data analysis in medicine. The second describes an approach to using machine learning in order to verify some unexplained phenomena from complementary medicine, which is not (yet) approved by the orthodox medical community but could in the future play an important role in overall medical diagnosis and treatment.

### 4.1 Reliability of single prediction

#### 4.1.1 Adding new instance to a learning set

When we apply a certain machine learning method we usually estimate the overall reliability of the method, typically in terms of the classification accuracy, information score (Kononenko and Bratko, 1991) or misclassification cost (Kukar et al., 1999). However, what we are really interested in when using the method to solve a given problem, is the reliability of that method on this particular problem. This is also important when we use several classifiers and combine their decisions (Kukar et al., 1996). In such a case we have to weigh the contribution of each classifier to the final decision. The weights should be case dependent, i.e., we have to be able to estimate the reliability of each method on the given case.

A simple idea can be used for that purpose: the decision of a classifier is reliable on the given case when the decision (prediction, class, diagnosis) is not sensitive to adding this case, labeled with this or any other decision (diagnosis), to the learning set. We can verify the reliability simply by labeling the new case in turn with all possible decisions and by adding it to the learning set and rerunning the learning algorithm. If the decision does not vary a lot, we assume that the classifier is quite reliable. On the other hand, if the decisions are sensitive to adding a new case to the learning set, the final decision is not reliable.

Kukar (2001) in his PhD thesis has elaborated this basic idea much further. He developed several metrics for measuring distances between classifications which are then used to measure the variation of classification. He compared several different reliability estimations and empirically showed that a metric based on scalar product of classification vectors performs best when combined with posttest probability. The experimental results on 15 domains confirm that the estimation of the reliability of single prediction provides useful information that can be used to improve the overall applicability of classifiers.

The same idea was used for weighted combination of answers of several classifiers. This approach improves the classification accuracy of a single classifier and considerably improves the robustness of the combined classifier with respect to noisy, random, and default classifiers.

The same idea was used also for problems with non-uniform misclassification costs. Cost sensitive reliability estimations were used for cost-sensitive combination of different classifiers that do not need to be cost-sensitive by themselves. Experimental results show significant decrease of overall misclassification costs (Kukar, 2001). We illustrate the usefulness of the approach on the problem of diagnosing the ischaemic heart disease.

#### **4.1.2 Application in the ischaemic heart disease diagnosis**

Ischaemic heart disease is one of the world's most important causes of mortality, so any improvements and rationalization of diagnostic procedures are very useful. The four diagnostic levels consist of the evaluation of signs and symptoms of the disease and ECG (electrocardiogram) at rest, sequential ECG testing during the controlled exercise, myocardial scintigraphy and finally coronary angiography. The diagnostic process is stepwise and the results are interpreted sequentially, i.e., the next step is necessary only if the results of the former are inconclusive. Because of the possible suggestibility, the results of each step are interpreted separately and only the results of the highest step are valid.

On the other hand, machine learning methods may be able to objectively interpret all available results for the same patient and in this way increase the diagnostic accuracy of each step. The performance of different diagnostic methods is usually described as classification accuracy, sensitivity, specificity, ROC curve, and posttest probability. We shall discuss only the latter, the other performance criteria are discussed in (Kukar et al., 1999).

In our study we used a dataset of 327 patients with performed clinical and laboratory examinations, exercise ECG, myocardial scintigraphy and coronary angiography. In 229 cases the disease was angiographically confirmed and in 98 cases it was excluded. The patients were selected from the population of the approximately 4000 patients who were examined at the Nuclear Medicine Department of University Clinical Center in Ljubljana, Slovenia in years 1991-1994. For the purpose of our study we selected only the patients with complete diagnostic procedures (all four steps).

The positive and the negative diagnosis of the ischaemic heart disease are defined to be reliable if the probability of presence or absence of the disease, respectively, is greater than 0.90 (Diamond and Forster, 1979). For that purpose the tabulated pretest probabilities and the results of various diagnostic steps together with the sensitivity and specificity are used in order to calculate the posttest probabilities (Pollock, 1983).

The standard procedure of the lookup table can be replaced by machine learning algorithms. Kukar

*Table 3:* Results of various classifiers in the ischaemic heart disease diagnosis (Kukar, 2001). The percentage of reliably diagnosed cases together with the amount of wrongly classified cases is given both for the positive and negative cases.

(a) Stepwise calculation of posttest probabilities.

(b) Using all attributes at once to calculate posttest probabilities.

(c) Using all attributes at once to evaluate the reliability of classification of single new cases.

classifier	positive cases		negative cases	
	reliable (%)	errors (%)	reliable (%)	errors (%)
physicians	73	3	46	8
semi-naive Bayes (a)	79	5	46	3
Assistant-I (a)	79	5	49	8
neural network (a)	78	4	49	8
semi-naive Bayes (b)	90	7	81	11
Assistant-I (b)	87	8	77	6
neural network (b)	86	5	66	9
naive Bayes (c)	<b>89</b>	<b>5</b>	<b>83</b>	<b>1</b>
semi-naive Bayes (c)	<b>91</b>	<b>6</b>	<b>79</b>	<b>2</b>
Assistant-I (c)	77	18	55	18
Assistant-R (c)	<b>81</b>	<b>5</b>	<b>77</b>	<b>2</b>
$k$ -NN (c)	64	12	80	12
neural network (c)	81	11	72	11

and Grošelj (1999) showed that for the stepwise calculation of posttest probabilities machine learning algorithms are able to improve the number of reliably classified positive and negative cases for 6%, which is an important improvement (Table 3 (a)). When we allow the machine learning algorithm to deal with all attributes at once the improvement is even higher, however, this result is not useful, as the number of incorrectly classified negative cases also increases (Table 3 (b)). On the other hand, Kukar (2001) has shown that if machine learning algorithms use the estimation of the reliability of a single prediction the results can be significantly better (Table 3 (c)). The naive and the semi naive Bayes and Assistant-R achieved excellent results. Compared to physicians the naive Bayesian classifier improves the number of reliably classified positive cases for 17% and the number of reliably classified negative cases for 37%!

## 4.2 Machine learning in complementary medicine

### 4.2.1 Kirlian effect - a scientific tool for studying subtle energies

The history of the so called Kirlian effect, also known as the Gas Discharge Visualization (GDV) technique (a wider term that includes also some other techniques is bioelectrography), goes back to 1777 when G.C. Lihtenberg in Germany recorded electrographs of sliding discharge in dust created by static electricity and electric sparks. Later various researches contributed to the development of the technique (Korotkov, 1998b): Nikola Tesla in the USA, J.J. Narkiewich-Jodko in Russia, Pratt and Schlemmer in Prague until the Russian technician Semyon D. Kirlian together with his wife Valentina noticed that through the interaction of electric currents and photograph plates, imprints of living organisms developed on film. In 1970 hundreds of enthusiasts started to reproduce Kirlian photos and the research was until 1995 limited to using a photo-paper technique. In 1995 a new approach, based on CCD Video techniques, and computer processing of data was developed by Korotkov (1998a;b) and his team in St. Petersburg, Russia. Their instrument Crown-TV can be routinely used which opens practical possibilities to study the effects of GDV.

The basic idea of GDV is to create an electromagnetic field using a high voltage and high frequency generator. After a threshold voltage is exceeded the ionization of gas around the studied object takes place and as a side effect the quanta of light – photons are emitted. So the discharge can be fixed optically by a photo, photo sensor or TV-camera. Various parameters influence the ionization process (Korotkov, 1998b): gas properties (gas type, pressure, gas content), voltage parameters (amplitude, frequency, impulse waveform), electrode parameters (configuration, distance, dust and moisture, macro and micro defects, electromagnetic field configuration) and studied object parameters (common impedance, physical fields, skin galvanic response, etc.). So the Kirlian effect is the result of mechanical, chemical, and electromagnetic processes, and field interactions. Gas discharge acts as means of enhancing and visualization of super-weak processes.

Due to the large number of parameters that influence the Kirlian effect it is very difficult or impossible to control them all, so in the development of discharge there is always an element of vagueness or stochastic. This is one of the reasons why the technique has not yet been widely accepted in practice as results did not have a high reproducibility. All explanations of the Kirlian effect apprehended fluorescence as the emanation of a biological object. Due to the low reproducibility, in academic circles there was a widely spread opinion that all observed phenomena are nothing else but fluctuation

of the crown discharge without any connection to the studied object. With modern technology, the reproducibility became sufficient to enable serious scientific studies.

Besides studying non-living objects, such as water and various liquids (Korotkov, 1998b), minerals, the most widely studied are living organisms: plants (leaves, seeds, etc. (Korotkov and Kouznetsov, 1997; Korotkov, 1998b)), animals (Krashenuk et al., 1998), and of course humans. For humans, most widely recorded are coronas of fingers (Kraweck, 1994; Korotkov, 1998b), and GDV records of blood excerpts (Voeikov, 1998). Principal among these are studies of the psycho-physiological state and energy of a human, diagnosis (Gurvits and Korotkov, 1998), reactions to some medicines, reactions to various substances, food (Kraweck, 1994), dental treatment (Lee, 1998), alternative healing treatment, such as acupuncture, 'bioenergy', homeopathy, various relaxation and massage techniques (Korotkov, 1998b), GEM therapy, applied kineziology and flower essence treatment (Hein, 1999), leech therapy, etc., and even studying the GDV images after death (Korotkov, 1998a). There are many studies currently going on all over the world and there is no doubt that the human subtle energy field, as visualized using the GDV technique, is highly correlated to the human's psycho-physiological state, and can be used for diagnostics, prognostics, therapy selection, and controlling the effects of the therapy.

#### **4.2.2 Verifying the map of organs**

Korotkov's team has developed a computer program that generates the corona of the whole human body from coronas of all ten fingertips. The program is based on a map, known from traditional Chinese medicine and described in Mandel's book (1986). This map defines regions (sectors) of each finger's corona to be related with a specific organ or organ system in the body. For example, the corona of the left little finger contains sectors that correspond to the coronary vessels, heart, kidney, respiratory system, small intestine, and ileum. Korotkov (1998b) and his team slightly modified Mandel's map.

For the orthodox medicine this map is meaningless, there is no physiological evidence for the connection of fingertips with different organs. Besides, the Kirlian camera is considered to provide only noisy pictures that are not related to human state of health.

In order to verify the map and the hypothesis that Kirlian camera provides useful information, we performed several experiments (Kononenko et al., 1999a;b; Bevk et al., 2000). In the following we briefly describe one such experiment.

We recorded all ten fingertips of 105 persons that also filled in a questionnaire where they described their health problems. We wanted to distinguish persons, that in the questionnaire had answered that they had no health problem, from persons who had problems with the throat (majority class contained 52.4% of cases). The cases were described by 75 numeric attributes that correspond to areas of sectors of coronas according to the map. We used the C4.5 learning system (Quinlan, 1993) and the result of 10-fold cross validation was 14.5% of error. This indicates that coronas in fact contain useful information for diagnosis.

Even more interesting was the structure of trees. For the root of the tree the algorithm selected most of the times, out of 75 attributes describing sectors of fingers, a sector that corresponds to the throat. There are two such sectors out of 75 sectors (the probability that this could happen by chance

for one tree is  $2/75 = 0.027$ ). The other two most important attributes corresponded to jaw and kidney. Jaw is, also by orthodox medicine, related to throat while kidney is by traditional Chinese medicine directly connected with the throat.

This result and several similar studies (Kononenko et al., 1999a; b; Bevk et al., 2000) indicate that the map of organs makes sense and that it would be beneficial for medicine to study this phenomena and eventually discover the underlying principles.

### 4.2.3 Overview of other studies

We use Kirlian camera to indirectly record subtle bioelectromagnetic field of living organisms, mostly humans. The obtained images are then described with a set of numerical parameters that serve as an input to statistical and ML algorithms. The subtle energies are not recognized by current orthodox scientific community and the aim of our studies is to verify “knowledge” of many practitioners in complementary medicine, who claim that living organisms besides physical body contain also non-measurable subtle levels, such as emotional and mental body.

We have performed several studies in which we analysed the influence of various parameters on plant and human bioelectromagnetic field:

**Apple skin:** We recorded the coronas of apple peels that were cut off from apples in a standard way. We used four sorts of apples of two different ages. We succeeded by the means of ML to extract useful information for distinguishing apples of different sorts and of different ages (the achieved classification accuracy was low but significantly higher than random classification). We were unable to extract any information to distinguish sun/shadow sides of apples (Kononenko et al., 1999b).

**Grapes:** The aim of the study was to verify whether the Kirlian camera could be used to describe grapevines and if the berry bioelectromagnetic field is influenced by disease. With the Kirlian camera we recorded coronas of grape berries. We tested this method on eight grapevine cultivars, performing different tests using ML algorithms. The results show that the coronas of grapevine berries contain significant information about the cultivars and their sanitary status (Kononenko et al., 2000b).

**Menstrual cycle:** For the preliminary study we recorded coronas of all ten fingertips of 13 female students in four weeks, one recording per week. Each recording was classified into one of four menstrual phases. The results of the analysis indicate that the coronas seem to be correlated with menstrual phases and that sectors of organs make sense. Out of 225 numerical parameters we automatically extracted 15 most important parameters. Fourteen of those parameters corresponded to sectors of three fingers which by Chinese medicine are directly connected with organs that are by official medicine affected by/responsible for the menstrual cycle (Kononenko et al., 1999b).

**T-shirts:** We wanted to evaluate the effect of different T-shirts on the human bioelectromagnetic field: color T-shirts developed by physicist Dr. Tom Chalko from University of Melbourne, ‘healing’ T-shirt developed by Vitalis from Slovenia, and an ordinary black and ordinary white T-shirt.

We measured 5 groups of people (with a control group). The analysis confirmed that black and white T-shirts have no significant influence on the coronas, while Vitalis and color T-shirts do have positive influence - they improve in time the coronas of humans fingers by means of larger area and lower fragmentation (Kononenko et al., 1999b).

**Glass 2000:** Vili Poznik from Celje, Slovenia uses orgon technology to encode information into a glass which affects in some way the water with which you fill the glass (this is of course nonsense for orthodox science). We recorded coronas of 34 persons before and after drinking the tap water from an ordinary glass and from a Glass 2000, coded by Vili Poznik. The results show that there was a significant improvement of coronas (larger area and lower fragmentation) when drinking water from Glass 2000 while the effect of drinking from ordinary glass was insignificant (Kononenko et al., 2000b).

**The art of living:** We performed three studies in order to verify the effects of The Art of Living Programme (exercises in communication, relaxation and breathing) on its participants. The results showed significant improvement of coronas (larger area) for participants of a 2-hour meeting and of a 6 day seminar compared to control groups which had no significant differences (Trampuž et al., 2000).

**Mobile telephones:** We recorded coronas of all ten fingertips of five groups of persons that were carrying the mobile telephone above their heart for a period of one hour under different conditions: without any protection, with two different energetic protections (which are nonsense for orthodox science), with placebo (fake) protection and a control group (without mobile telephones). Results indicate that mobile telephones negatively affect the human BEM field, that energetic protections work well while the placebo protection does not work (Kononenko et al., 2000a).

**Energetic diagnosis:** We recorded coronas of all ten fingertips of 110 persons for whose the extrasense healer provided the energetic diagnosis. We used machine learning to interpret the GDV coronas in order to verify three hypothesis: (a) the GDV images contain useful information about the patient, (b) the map of organs on coronas of 10 fingers does make sense, and (c) the extrasense healer is able to see by himself (with his natural senses) the energetic disorders in the human body. The results support all three hypotheses (Bevk et al., 2000).

## 5 Discussion

The historical development of machine learning and its applications in medical diagnosis shows that from simple and straightforward to use algorithms, systems and methodology have emerged that enable advanced and sophisticated data analysis. In the future, intelligent data analysis will play even a more important role, due to the huge amount of information produced and stored by modern technology. Current machine learning algorithms provide tools that can significantly help medical practitioners to reveal interesting relationships in their data.

Our experiments show that in medical domains various classifiers perform roughly the same. So one of the important factors when choosing which classifier to apply is its explanation ability. Our experiments show that the physicians prefer explanations as provided by the Bayesian classifiers and

decision tree classifiers: Assistant-R and LFC. However, instead of selecting a single best classifier, it seems that the best solution is to use all of them and combine their decision when solving new problems. The physicians found that the combination of classifiers was the appropriate way of improving the reliability and comprehensibility of diagnostic systems. The combination should be done in an appropriate way and the reliability of each classifier on the given new case should be taken into account, as the results of Kukar (2001) clearly demonstrate.

Regarding the future role of machine learning in medical diagnosis, our views are as follows:

- Machine learning technology has not been accepted in the practice of medical diagnosis to an extent that the clearly demonstrated technical possibilities indicate. However, it is hard to expect that this disproportion between the technical possibilities and practical exploitation will remain for very much longer.
- Among the reasons for slow acceptance perhaps the most reasonable one is that the introduction of machine learning technology will further increase the abundance of tools and instrumentation available to physicians. Any new tool has the undesirable side effect of further increasing the complexity of the physician's work which is already sufficiently complicated. Therefore machine learning technology will have to be integrated into the existing instrumentation that makes its use as simple and natural as possible.
- Machine learning based diagnostic programs will be used as any other instrument available to physicians: as just another source of possibly useful information that helps to improve diagnostic accuracy. The final responsibility and judgement whether to accept or reject this information will, as usual, remain with the physician.
- Complementary medicine is becoming in recent years more and more important, which can be seen also by the amount of money people spend on various complementary medicine treatments. Physicians are becoming aware of the efficiency and the benefits of complementary medicine and they need verification procedures in order to acknowledge the benefits and issue licences for the use of complementary approaches. Machine learning can play an important role in this process in particular due to the transparency of data analysis.

## Acknowledgements

Special thanks to Ivan Bratko, Matjaž Kukar, and Nada Lavrač for long term joint work on projects related to intelligent data analysis in medicine. Experiments with the Kirlian camera were done with the invaluable help and support from Matjaž Bevk, Zoran Bosnić, Tom Chalko, Minnie Hein, my wife Irena, Milan Mladženić, Barbara Novak, Petar Papuga, Vili Poznik, Bor Prihavec, Marko Robnik-Šikonja, Aleksander Sadikov, Danijel Skočaj, Slobodan Stanojević, Tatjana Zrimec, and many others. I thank Nada Lavrač and Elpida Keravnou for their corrections that significantly improved the paper. This research was supported by the Slovenian Ministry of Science and Technology.

## References

- Baim P.W., A Method for Attribute Selection in Inductive Learning Systems, *IEEE Trans. on PAMI*, Vol.10, No. 6, 1988, pp.888-896.
- Bevk M., Kononenko I., Zrimec T., Relation between energetic diagnoses and GDV images, *Proc. New Science of Consciousness: 3rd Int. Conf. on Cognitive Science*, Ljubljana, October 2000, pp. 54-57.
- Bratko I., Mozetič I., Lavrač N., *KARDIO: A study in deep and qualitative knowledge for expert systems*, Cambridge,MA: MIT Press, 1989.
- Bratko I., Mulec P., An Experiment in Automatic Learning of Diagnostic Rules, *Informatika*, Ljubljana, Vol.4, No.4, 1980, pp. 18-25.
- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. (1984) *Classification and Regression Trees*, Wadsworth International Group.
- Catlett J., On changing continuous attributes into ordered discrete attributes, *Proc. European Working Session on Learning-91*, Porto, March 4-6, 1991, pp. 164-178.
- Cestnik B., Estimating Probabilities: A Crucial Task in Machine Learning, *Proc. European Conf. on Artificial Intelligence*, Stockholm, August, 1990, pp. 147-149.
- Cestnik B., Kononenko I. & Bratko I., ASSISTANT 86 : A knowledge elicitation tool for sophisticated users, in: I.Bratko, N.Lavrač (eds.): *Progress in Machine learning*, Wilmslow: Sigma Press, 1987.
- Chan K.C.C. & Wong A.K.C., Automatic Construction of Expert Systems from Data: A Statistical Approach, *Proc. IJCAI Workshop on Knowledge Discovery in Databases*, Detroit, Michigan, August, 1989, pp.37-48.
- Clark P. & Boswell R., Rule Induction with CN2: Some Recent Improvements, *Proc. European Working Session on Learning-91*, Porto, Portugal, March, 1991, pp.151-163.
- Craven M.W. and Shavlik J.W., Learning symbolic rules using artificial neural networks, *Proc. 10<sup>th</sup> Intern. Conf. on Machine Learning*, Amherst, MA, Morgan Kaufmann, 1993, pp.73-80.
- Diamond G.A. and Forester J.S., Analysis of probability as an aid in the clinical diagnosis of coronary artery disease, *New England J. of Medicine*, 300:1350, 1979.
- Elomaa T., Holsti N., An Experimental Comparison of Inducing Decision Trees and Decision Lists in Noisy Domains, *Proc. 4th European Working Session on Learning*, Montpeiller, Dec. 4-6, 1989, pp.59-69.
- Good I.J., *Probability and the Weighing of Evidence*. London: Charles Griffin, 1950.
- Good I.J., *The Estimation of Probabilities – An Essay on Modern Bayesian Methods*, Cambridge: The MIT Press, 1964.
- Gurvits B. and Korotkov K., A new concept of the early diagnosis of cancer, *Consciousness and Physical Reality*, 1(1)84–89, 1998.
- Haykin S., *Neural Networks: A Comprehensive Foundation*, New York: Macmillan College Publ. Comp, 1994.
- Hein M., Bio-Synergetix: A new paradigm in subtle energy health care, *Proc. Int. Conf. Biology and Cognitive Sciences*, pp. 72-79, Slovenia: Ljubljana, October 12-14, 1999.
- Hojker S., Kononenko I., Jauk A., Fidler V. & Porenta M., Expert System's Development in the Management of Thyroid Diseases, *Proc. European Congress for Nuclear Medicine*, Milano, Sept., 1988.
- Hopfield J.J., Neural networks and physical systems with emergent collective computational abilities. *Proc. National Academy of Sciences* 79:2554–2558, 1982.
- Hopfield J.J., Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. National Academy of Sciences* 81:4586–4590, 1984.
- Horn K.A., Compton P., Lazarus L., Quinlan J.R., An Expert System for the Interpretation of Thyroid Assays in a Clinical Laboratory, *The Australian Computer Journal*, Vol. 17, No. 1, 1985, pp.7-11.

- Hunt E., Martin J & Stone P., *Experiments in Induction*, New York, Academic Press, 1966.
- Karalič A., Pirnat V., Significance Level Based Classification with Multiple Trees, *Informatica*, Ljubljana, Vol.15, No. 1, 1991, pp.54-58.
- Kern J., Deželič G., Težak-Benčič M., Durrigl T., Medical Decision Making Using Inductive Learning Program (in Croatian), *Proc 1st Congress on Yugoslav Medical Informatics*, Beograd, Dec. 6-8, 1990, pp.221-228.
- Kira K. & Rendell L., A practical approach to feature selection, *Proc. Intern. Conf. on Machine Learning* (Aberdeen, July 1992) D.Sleeman & P.Edwards (eds.), Morgan Kaufmann, 1992a, pp.249-256.
- Kira K. & Rendell L., The feature selection problem: traditional methods and new algorithm. *Proc. AAAI'92*, San Jose, CA, July 1992b.
- Kohavi R., Becker B., Sommerfield D., Making sense of simple Bayes, Technical report, Data Mining and Visualization group, SGI Inc., 1997.
- Lavrač N. and Džeroski S., *Inductive Logic Programming*, Ellis Horwood, 1994.
- Kononenko I., Interpretation of neural networks decisions, IASTED Internat. Conf. Expert systems & applications, Zurich, June 26-29 1989, pp.224-227 (also: Proc. ISSEK Workshop, Udine, Sept. 1989).
- Kononenko I. Bayesian Neural Networks, *Biological Cybernetics Journal*, 61:361-370, 1989a.
- Kononenko I., Semi-naive Bayesian classifier, *Proc. European Working Session on Learning-91* (Y.Kodratoff (ed.), Springer-Verlag), Porto, March 4-6 1991, pp.206-219.
- Kononenko I., Naive Bayesian classifier and continuous attributes, *Informatica*, 16(1)1-8, 1992.
- Kononenko I., Inductive and Bayesian learning in medical diagnosis. *Applied Artificial Intelligence*, 7:317-337, 1993.
- Kononenko I., Estimating attributes: Analysis and extensions of RELIEF. *Proc. European Conf. on Machine Learning* (Catania, April 1994), L. De Raedt & F.Bergadano (eds.), Springer Verlag, 1994, pp. 171-182.
- Kononenko I. & Bratko I., Information based evaluation criterion for classifier's performance, *Machine Learning*, Vol.6, No.1, 1991, pp.67-80.
- Kononenko I., Bratko I., Roškar E.: Experiments in automatic learning of medical diagnostic rules, International School for the Synthesis of Expert's Knowledge Workshop, Bled, Slovenia, August, 1984.
- Kononenko I., Jauk A. & Janc T., Induction of Reliable Decision Rules, International School for the Synthesis of Expert's Knowledge Workshop, Udine, Italy, 10-13 Sept., 1988.
- Kononenko I., Bratko I., Kukar M., Application of machine learning to medical diagnosis. In R.S.Michalski, I.Bratko, and M.Kubat (eds.): *Machine Learning, Data Mining and Knowledge Discovery: Methods and Applications*, John Wiley & Sons, 1998.
- Kononenko I., Zrimec T., Prihavec B., Bevk M., Stanojević S., Machine learning and GDV images: Diagnosis and therapy verification, *Proc. Biology and Cognitive Science*, Ljubljana, October 1999a, pp. 84-87.
- Kononenko I., Zrimec T., Sadikov A., Mele K., Milharčič T., Machine learning and GDV images: Current research and results, *Proc. Biology and Cognitive Science*, Ljubljana, October 1999b, pp. 80-83.
- Kononenko I., Bosnič Z., Žgajnar B., The influence of mobile telephones on human bioelectromagnetic field, *Proc. New Science of Consciousness: 3rd Int. Conf. on Cognitive Science*, Ljubljana, October 2000a, pp. 69-72.
- Kononenko I., Zrimec T., Sadikov A., Skočaj D., GDV images: Current research and results, *Proc. New Science of Consciousness: 3rd Int. Conf. on Cognitive Science*, Ljubljana, October 2000b, pp. 65-68.
- Korotkov K., *Light after Life: A scientific Journey into the Spiritual World*, Fair Lawn, USA: Backbone Publ. Comp, 1998a.
- Korotkov K., *Aura and Consciousness: A New Stage of Scientific Understanding*, St.Petersburg, Russia: State Editing & Publishing Unit "Kultura", 1998b.

- Korotkov K. and Kouznetsov A., The theory of morfogenic synergization of biological objects and the phantom leaf effect, *Proc. 3rd Int. Conf. for Medical and Applied Bio-Electrography*, Helsinki, Finland, April 1997, pp. 55–57.
- Krashenuk A., Krashenuk S., Korotkov K., Buzian N., Lesiovskaya E., Bogaeva N., GDV analysis of hirudotherapy effect to rats, *Proc. Int. Scientific Conf. Kirlionics, White Nights 98*, Federal Technical University SPIFMO, St. Petersburg, Russia, June 1998.
- Kraweck A., *Life's Hidden Forces: A Personal Journey into Kirlian Photography*, Edmonton, Canada: Triune-Being Research Organization Ltd, 1994.
- Kukar M., Estimating the reliability of classifications and cost sensitive combining of different machine learning methods, PhD Thesis (in Slovene), University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia, 2001.
- Kukar M. and Grošelj C., Machine learning in stepwise diagnostic process, *Proc. Joint European Conf. on Artificial Intelligence in Medicine and Medical Decision Making*, pp.315-325, Aalborg, Denmark, 1999.
- Kukar M., Kononenko I., Silvester T., Machine learning in prognostics of the femoral neck fracture recovery, *Artificial intelligence in medicine*, 8:431-451, 1996.
- Kukar M., Kononenko I., Grošelj C., Kralj K., Fettich J., Analysing and improving the diagnosis of ischaemic heart disease with machine learning, *Artificial Intelligence in Medicine*, 16:25-50, 1999.
- Langley P., Induction of recursive Bayesian classifiers, *Proc. European Conf. on Machine Learning*, Vienna, April 1993.
- Lee S.D., The application of kirlian photography in dentistry, *Proc. Int. Scientific Conf. Kirlionics, White Nights 98*, Federal Technical University SPIFMO, St. Petersburg, Russia, June 1998.
- Lesmo L., Saitta L., Torasso P., Learning of Fuzzy Production Rules for Medical Diagnoses, In: Gupta M.M. & Sanchez E.(eds.) *Approximate reasoning in Decision Analysis*, North-Holland, 1982.
- Mandel P., *Energy Emission Analysis*. Synthesis Publ. Comp., 1986.
- Michalski R.S., Chilausky R.L., Learning by being told and learning from examples: An experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis. *Int. Journal of Policy Analysis and Information Systems*, 4:125–161, 1980.
- Michalski R.S., Bratko I., and Kubat M. (eds.): *Machine Learning, Data Mining and Knowledge Discovery: Methods and Applications*, John Wiley & Sons.
- Mitchell T., *Machine Learning*, McGraw Hill, 1997.
- Michie D., Spiegelhalter D.J., Taylor C.C (eds.) *Machine learning, neural and statistical classification*, Ellis Horwood, 1994.
- Minsky M. in Papert S., *Perceptrons*. Cambridge, MA: MIT Press, 1969.
- Muggleton S., *Inductive Acquisition of Expert Knowledge*, Turing Institute Press & Addison-Wesley, 1990.
- Nilsson N., *Learning Machines*, McGraw-Hill, 1965.
- Nunez M., Decision Tree Induction Using Domain Knowledge, In: Wielinga B. et al. (eds.) *Current Trends in Knowledge Acquisition*, Amsterdam: IOS Press, 1990.
- Pazzani M., Searching for dependencies in Bayesian classifiers, *Artificial Intelligence and Statistics IV*, Lecture Notes in Statistics, Springer-Verlag, New York, 1997.
- Pirnat V., Kononenko I., Janc T. & Bratko I., Medical Estimation of Automatically Induced Decision Rules, *Proc. of 2nd Europ. Conf. on Artificial Intelligence in Medicine*, City University, London, August 29-31, 1989, pp.24-36.
- Pollock B.H., Computer assisted interpretation of noninvasive tests for diagnosis of coronary artery disease. *Cardiovasc. Rev. Rep.* 4, pp.367-375, 1983.
- Quinlan, J.R., Discovering rules from large collections of examples. Michie D. (ed.) *Expert Systems in the Microelectronic Age*, Edinburgh University Press, 1979.
- Quinlan J.R., Induction of Decision Trees. *Machine Learning*. Vol. 1, No. 1, 1986, pp. 81-106.

- Quinlan J.R., *C4.5: Programs for Machine Learning*, San Mateo, CA, Morgan Kaufmann, 1993.
- Quinlan R., Compton P., Horn K.A., Lazarus L., Inductive knowledge acquisition: A case study, in: J.R.Quinlan (ed.) *Applications of expert systems*, Turing Institute Press & Addison- Wesley, 1987 (Also: *Proc. 2nd Australian Conf. on Applications of Expert Systems*, Sydney, May 14-16, 1986).
- Pompe U. and Kononenko I., Probabilistic first-order classification, In Lavrač. N. and Džeroski S. (eds.) *Inductive Logic Programming - Proc. 7th Int. Workshop ILP-97*, Springer Verlag, pp. 235-242, 1997.
- Ragavan H. & Rendell L., Lookahead feature construction for learning hard concepts. *Proc. 10th Intern. Conf. on Machine Learning*. (Amherst, MA, June 1993), Morgan Kaufmann, 1993, pp.252-259.
- Ragavan H, Rendell L., Shaw M. & Tessmer A., Learning complex real-world concepts through feature construction. Technical Report UIUC-BI-AI-93-03. The Beckman Institute, University of Illinois, 1993.
- Robnik-Šikonja M. and Kononenko I., An adaptation of Relief for attribute estimation in regression, *Proc. Int. Conf. on Machine Learning ICML-97*, Nashville, July 1997, pp. 296-304.
- Rosenblatt F., *Principles of Neurodynamics*, Washington, DC: Spartan Books, 1962.
- Rumelhart D.E., Hinton G.E. in Williams R.J., Learning internal representations by error propagation. Rumelhart D.E. in McClelland J.L. (eds.) *Parallel Distributed Processing, Vol. 1: Foundations*. Cambridge: MIT Press, 1986.
- Shavlik J.W., Dietterich T.G. (eds.) *Readings in machine learning*, Morgan Kaufmann Publ., 1990.
- Spiegelhalter D.J., Philip Dawid A., Lauritzen S.L. and Cowell R.G., Bayesian analysis in expert systems, *Statistical Science*, 8(3):219-283, 1993.
- Trampuž A., Kononenko I., Rus V., Experiential and biophysical effects of the Art of Living programme, *Int. Journal of Psychology*, 35(3/4)12, 2000.
- Voeikov V.L., Living blood outside an organism, In *Proc. Int. Scientific Conf. Kirlionics, White Nights 98*, Federal Technical University SPIFMO, St. Petersburg, Russia, June 1998.
- Weigand S., Huberman A., and Rumelhart D.E., Predicting the future: a connectionist approach, *International Journal of Neural Systems*, Vol. 1(3), 1990.

*Table 1:* The appropriateness of various algorithms for medical diagnosis.

classifier	performance	transparency	explanations	reduction	miss. data handling
Assistant-R	good	very good	good	good	acceptable
Assistant-I	good	very good	good	good	acceptable
LFC	good	good	good	good	acceptable
naive Bayes	very good	good	very good	no	very good
semi-naive Bayes	very good	good	very good	no	very good
backpropagation	very good	poor	poor	no	acceptable
$k$ -NN	very good	poor	acceptable	no	acceptable

Table 2: Semi-naive Bayes: an explanation of a decision in the femoral neck fracture recovery problem.

Decision = No complications (correct)

Attribute value	For decision (bit)	Against decision (bit)
Age = 70 - 80	0.07	
Sex = Female		-0.19
Mobility before injury = Fully mobile	0.04	
State of health before injury = Other	0.52	
Mechanism of injury = Simple fall		-0.08
Additional injuries = None	0.00	
Time between injury and operation > 10 days	0.42	
Fracture classification acc. to Garden = Garden III		-0.30
Fracture classification acc. to Pauwels = Pauwels III		-0.14
Transfusion = Yes	0.07	
Antibiotic profilaxis = Yes		-0.32
Hospital rehabilitation = Yes	0.05	
General complications = None		-0.00
<b>Combination:</b>	0.21	
Time between injury and examination < 6 hours AND Hospitalization time between 4 and 5 weeks		
<b>Combination:</b>	0.63	
Therapy = Artroplastic AND Anticoagulant therapy = Yes		

*Table 3:* Results of various classifiers in the ischaemic heart disease diagnosis (Kukar, 2001). The percentage of reliably diagnosed cases together with the amount of wrongly classified cases is given both for the positive and negative cases.

(a) Stepwise calculation of posttest probabilities.

(b) Using all attributes at once to calculate posttest probabilities.

(c) Using all attributes at once to evaluate the reliability of classification of single new cases.

classifier	positive cases		negative cases	
	reliable (%)	errors (%)	reliable (%)	errors (%)
physicians	73	3	46	8
semi-naive Bayes (a)	79	5	46	3
Assistant-I (a)	79	5	49	8
neural network (a)	78	4	49	8
semi-naive Bayes (b)	90	7	81	11
Assistant-I (b)	87	8	77	6
neural network (b)	86	5	66	9
naive Bayes (c)	<b>89</b>	<b>5</b>	<b>83</b>	<b>1</b>
semi-naive Bayes (c)	<b>91</b>	<b>6</b>	<b>79</b>	<b>2</b>
Assistant-I (c)	77	18	55	18
Assistant-R (c)	<b>81</b>	<b>5</b>	<b>77</b>	<b>2</b>
$k$ -NN (c)	64	12	80	12
neural network (c)	81	11	72	11