applied sub-symbolic methods are nearest neighbor, Bayesian classifier, and (non-symbolic) clustering.

– *Data abstraction methods* are intended to support specific knowledge-based problem solving activities (data interpretation, diagnosis, prognosis, monitoring, etc.) by gleaning out the useful abstractions from the raw, mostly numeric data. *Temporal data abstraction methods* represent an important subgroup where the processed data are temporal. The derivation of abstractions is often done in a context sensitive and/or distributed manner and it applies to discrete and continuous supplies of data. Useful types of temporal abstractions are trends, periodic happenings, and other forms of temporal patterns. Temporal abstractions can also be discovered by visualization. The abstraction can be performed over a single case (e.g., a single patient) or over a collection of cases.

The scope of this paper regarding data mining is *machine learning methods*, with the emphasis on symbolic concept learning and Bayesian classification. In data abstraction, the scope of the paper is limited to *temporal abstraction methods*.

2.4. IDA for medical diagnosis

A typical diagnostic process is the following. In an interview the patient's anamnestic data is obtained and after the preliminary examination of the patient the physician records the status data. Depending on the anamnestic and the status data, the patient takes additional laboratory examinations. The diagnosis is then determined by the physician who takes into account the whole available description of the patient's state of health. Depending on the diagnosis the treatment is prescribed and after the treatment the whole process may be repeated. In each iteration the diagnosis may be confirmed, refined, or rejected. The definition of the final diagnosis depends on the medical problem. In some problems the first diagnosis is also the final, in some others the final diagnosis is determined after the results of the treatment are available, and in some problems there is no way to obtain a completely reliable final diagnosis. For example, in the problem of the localization of the primary tumor the final diagnosis can always be obtained with an operation where the location of the primary tumor is verified, although this "examination" is avoided and replaced with other laboratory tests unless it is really necessary to obtain the verified diagnosis. And in urology, in the problem of diagnosing the type of incontinence, in practice the final diagnosis is never obtained as there is no practical way to verify it.

Medical diagnosis is known to be subjective and depends not only on the available data but also on the experience of the physician, his intuition and biases, and even on the psycho-physiological condition of the physician. Several studies have shown that the diagnosis of a patient can differ significantly if the patient is examined by different physicians or even by the same physician at different times (different day of the week or different hour of the day).

2.4.1. Machine learning

Machine learning methods can be used to automatically derive diagnostic rules from the descriptions of the patients treated in the past for which the final diagnoses were verified. Automatically derived diagnostic knowledge may assist physicians to make the diagnostic process more objective and more reliable.

Typically, automatically generated diagnostic rules slightly outperform the diagnostic accuracy of physician specialists when physicians have available exactly the same information as the machine. Table 1 provides a comparison of the performance of two machine learning algorithms, the naive Bayesian classifier and a decision tree induction algorithm Assistant [8], with the average performance of four physician specialists in three different medical diagnostic problems: the localization of the primary tumor (PRIM), the diagnosis of thyroid diseases (THYR), and rheumatology (RHEU).

Table 1
The comparison of performance of different classifiers in three medical domains.

Classifier	PRIM		THYR		RHEU	
naive Bayes	49%	1.60bit	70%	0.79bit	67%	0.52bit
Assistant	44%	1.38bit	73%	0.87bit	61%	0.46bit
physicians	42%	1.22bit	64%	0.59bit	56%	0.26bit

The following are the brief descriptions of the diagnostic problems (see also [45]).

– **Localization of primary tumor:** The medical treatment of patients with metastases is much more successful if the location of the primary tumor in the body of the patient is known. The diagnostic task is to determine one of 22 possible locations of the primary tumor on the basis of age, sex,