Assessing an AI Knowledge-Base for Asymptomatic Liver Diseases
Ankica Babic1,4, Ulrik Mathiesen2,5, Kristina Hedin3, Göran Bodenmar2, Ove Wigertz4

1 Department of Medical Informatics, Linköping University, Sweden
2 Department of Internal Diseases, Linköping University Hospital, Sweden
3 Department of Infectious Diseases, Linköping University Hospital, Sweden
4 Faculty of Electrical and Computer Engineering, University of Ljubljana, Slovenia
5 Department of Internal Diseases, Oskarshamn County Hospital, Sweden

Discovering not yet seen knowledge from clinical data is of importance in the field of asymptomatic liver diseases. Avoidance of liver biopsy which is used as the ultimate confirmation of diagnosis by making the decision based on relevant laboratory findings only, would be considered an essential support.

The system based on Quinlan's ID3 algorithm was simple and efficient in extracting the sought knowledge. Basic principles of applying the AI systems are therefore described and complemented with medical evaluation. Some of the diagnostic rules were found to be useful as decision algorithms i.e. they could be directly applied in clinical work and made a part of the knowledge-base of the Liver Guide, an automated decision support system.

INTRODUCTION

Building a knowledge-based system to support clinical decisions includes definition and formulation of knowledge which is to be part of the knowledge-base. Among various understandings of the term knowledge we will focus on the knowledge based on expertise. The field of asymptomatic, or hard to determine, liver diseases is an area in which correct clinical decision can lead to a standardization of non-evasive and cost-effective medical examination.

It would be preferable if the liver biopsy could be recommended, or advised against, relying on laboratory findings only. However, clinical practice shows that even if some laboratory findings are slightly raised it is not always possible to determine disease. As a result, ultrasound and liver biopsy are used as additional methods of reaching ultimate confirmation of diagnosis. Due to the risk of discomfort and possible complications as well as the cost of the procedure, biopsy is not performed in all patients. Therefore, availability of knowledge which would enable us to distinguish patients with a trivial cause for the raised values of performed tests from patients who are developing a severe disease, could become an essential clinical support. A system that encompasses such a specific knowledge is likely to perform better than an automated system for established liver diseases. Several decision support systems have been developed to obtain information on the level of accuracy of different procedures necessary to make a correct diagnosis in patients with liver and gallbladder diseases. According to a recent review, however, these charts are very seldom used even in the centres where these systems have actually been developed [1]. One of the reasons might be that the systems are often too elaborate and very often several investigations must be performed to exclude malignant diseases, even in the cases with low probability of such diseases. A decision support system giving guidelines on how to take care of patients with raised liver laboratory values but no obvious clinical symptoms or signs of a liver disease, would probably meet with better response. Since knowledge in this field is scarce, a knowledge-based automated system would be of interest.

In addition to multivariate statistics and other data analyses methods, inductive learning, or machine learning procedures can be used to actively extract knowledge based on a few medical premises only. This enables fast prototyping of medical domain by constructing concept definitions, relation between concepts, or problem formulation [2]. The two best known inductive algorithms presently in commercial use are ID3 and AQ11 [3].

In this paper we shall discuss some aspects of using the system based on the ID3 algorithm [4] for extracting medical knowledge in a form of tree structured rules. Both methodological description and evaluation of the system are given in [5]. Issues related to extracting a domain knowledge, or its conceptualisation, and knowledge appropriateness as seen from the medical experts' point of view are also addressed.

THE INDUCTIVE LEARNING OF DIAGNOSTIC RULES

The inductive learning system is an implementation of a modified version of Quinlan's ID3 algorithm called Assistant Professional (from now on abbreviated AssistPro). It uses a principle of induction to extract rules, i.e. a classifier learning from a training set of examples. Medical data complete with respect to findings, diagnosis, etc. are potential training examples whose attributes correspond to various medical indicators. The result is classification or decision rules that can be used
for the classification of patients into different classes. Knowledge is here represented in the form of binary trees, where a complete decision constitutes a path leading from the root to a leaf of the tree. The decision process starts from the root, follows the subdecisions labelling the tree nodes, and ends up in a leaf, which is labelled with a class. This is called a hard decision tree. An alternative is a soft decision tree which produces class probabilities. A survey of different tree classification methods is found in [6].

In order to make the decision more 'manageable' and clearer to the reader and reduce the influence of noise and inconsistencies in the data, the process of tree pruning is of vital importance. This has an impact on the reliability and complexity of the tree, as well as on the classification accuracy.

To assess complexity, we will be using the number of leaf nodes of the classifier, the reason being that the total number of nodes is of the same order as twice the number of leaves. The accuracy can be defined as a classifier's ability to correctly classify not-yet seen test examples (test set validation).

An optimal classifier is defined as the one which minimises the complexity of the tree while simultaneously maximising the accuracy of classification. Since minimum complexity and maximum accuracy are most often contradictory requirements, we usually have to settle with a compromise between the two.

Building the Tree
The basic idea behind the tree growing procedure is to perform a successive binarization, i.e. split into two halves the given training data set. For this purpose we have defined the concept of impurity, by which we mean the following:

Definition 1. The impurity of a training data set is a quantity with the following properties:

1. \( i \geq 0 \) always
2. \( i = 0 \) only when a single class label appears among the examples (max purity)
3. \( i = \) its maximum value, only when all possible class labels appear in equal proportion among the examples (min purity).

The third item in the definition is true under the assumption that all the classes are of equal importance. Given that we start with a training data set with purity \( i \), a 'clever' split would be the one that maximizes the reduction in impurity due to the splitting of data into two halves. Since a split might lead to one half dominating over the other in size, the larger half should be considered as more 'important' than the smaller one. Therefore the following measure is used in practice to guide the search for a clever split:

\[
\Delta i = i - p_L i_L - p_R i_R,
\]

where \( p_i \) is the proportion of examples falling into the half we hereby call \( L \) and \( i_L \) is the impurity of the same half. Thus \( p_R \) and \( i_R \) are defined correspondingly for the remaining half which we will call \( R \).

In each step of the tree growing algorithm, the learning data at the current node is binarized, i.e. split into two separate halves \( L \) and \( R \). This is done by selecting a rule, or condition, based on a single attribute of the learning examples. In practice, the condition is simply of the kind:

\[
\text{If the attribute value } \leq \tau, \text{ then the example belongs to } L \text{ else it belongs to } R
\]

where \( \tau \) is some constant real values threshold. Thus, the original problem of finding a split is turned into finding a threshold value.

All real data and especially medical data are naturally corrupted by non-systematic errors like noise in either the attribute values or the class assignments. Noise significantly affects the building of trees in the way that trees tend to become overly large; this is sometimes referred to as overtraining. This influences both the readability and the accuracy of the tree. The most obvious way to prevent the tree from becoming overly large, is to find some clever way to stop the growth of the tree while it is being built. This is sometimes referred to as prepruning or forward pruning. The intention of pruning is to find the 'right' compromise between complexity and accuracy of the tree.

INDUCTIVE LEARNING IN THE AREA OF ASYMPTOMATIC LIVER DISEASES

In our previous research based on multivariate statistics [7], the AssistPro system was used to evaluate the efficiency of the statistically-based decision support. In order to avoid doubtful classifications, the accuracy threshold was settled. Only patients classified with probability of 0.7 were denoted as suitable for the study. For this reason, a soft decision classifier was constructed and found to be sufficiently successful. At the same time it has not demanded an experienced knowledge engineer to give methodological guidance. A standing concern, however, was medical relevance of the knowledge so developed. Therefore, we are conducting a case study to exemplify some of the advantages and limitations of using inductive learning. Following some of the guidelines given in [8], we have made the assessment of medical appropriateness and relevance.

Domain Description
The data used for the knowledge acquisition was collected for 165 consecutive patients chosen because of persistently raised levels of liver transaminases, AST (aspartate aminotransferase) and ALT (alanineaminotransferase), during a period of at least 6 months. At the time of referral, the causes for
The abnormal findings of liver laboratory values were unknown. None of these patients had any clinical symptoms or signs of a chronic liver disease. A standardised questionnaire was used to determine the amount of alcohol consumed, concurrent medication, exposure to potentially toxic substances at work, earlier blood transfusions and any abuse of narcotics. Several laboratory tests were then performed and assessed together with a history of illness to exclude established liver diseases such as clear-cut autoimmune chronic active hepatitis, primary biliary cirrhosis, primary sclerosing cholangitis, alpha-2-antitrypsin deficiency, primary haemochromatosis, thyroid diseases and chronic HBV-hepatitis.

The echogenicity of the liver parenchyma was investigated by ultrasound examination. Only patients with undilated bile ducts and without focal lesions of the liver were included.

At the start of the trial, serological tests for hepatitis C virus infection were not available. Frozen serum samples from all patients were later tested for antibodies against this virus. 25 patients, out of 165, were positive with the second generation antiHCV ELISA test.

A liver biopsy was performed in all 165 patients. Coded liver biopsy specimens were then read by one pathologist. Using an established but slightly modified scoring system [9], two main histopathological findings were scored: inflammation and fibrosis (surplus of connective tissue) and additionally the degree of fatty change (steatosis). No changes were scored as 0, minor histopathological findings as 1, whilst moderate and severe changes were scored greater than 1. Six patients were found to have cirrhosis, which was taken as severe fibrosis.

**A Case Study**

The AssistPro classification tree was reached after a learning phase, which included forward pruning, so that only 7 most informative findings appeared in the tree classifier. In order to make diagnosis easy to perform, the original tree was decomposed into 14 diagnostic rules. Each path from the root to a tree leaf was 'read' as one rule. Therefore, the leaves are assigned a rule label. All the rules can be found in Tables 1. and 2.

<table>
<thead>
<tr>
<th>Rule</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. IF (AST &gt; 0.7) and (PREALB ≤ 0.18) THEN BIOPSY 1</td>
<td></td>
</tr>
<tr>
<td>2. IF (AST &gt; 0.7) and (PREALB &gt; 0.18) and (AGE x PREALB ≤ 0.9000) THEN BIOPSY 0</td>
<td></td>
</tr>
<tr>
<td>3. IF (AST ≤ 0.7) and (AGE x QUOT &gt; 40) and (ALT &gt; 0.7) and (FER ≤ 1.50) THEN BIOPSY 0</td>
<td></td>
</tr>
<tr>
<td>4. IF (AST ≤ 0.7) and (AGE x QUOT &gt; 40) and (AGE x FER ≤ 1.1500) and (ANTIHCV = NEGATIVE) THEN BIOPSY 0</td>
<td></td>
</tr>
<tr>
<td>5. IF (AST ≤ 0.7) and (AGE x QUOT ≤ 40) and (AGE x FER ≤ 1.1500) and (ANTIHCV = POSITIVE) THEN BIOPSY 1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Rules corresponding to medical knowledge

**Medical Discussion on Extracted Rules**

All decision rules are judged in relation to relevant medical knowledge. In the present study we define medical knowledge as associations between laboratory values (for liver tests), Body Mass Index (BMI), age and histopathological findings of inflammation and/or fibrosis in liver biopsies. A rule classification is suggested (A: rules corresponding to medical knowledge, B: not readily understandable) followed by an expert explanation.

**A. Rules that correspond to medical knowledge.**

**Rule 1.** If AST is above normal limit and prealbumin (a low value is a marker in serum of an inflammatory process) is below normal limit, then it is almost certain that more severe degrees of inflammation and/or fibrosis or both is present in the liver (Biopsy 1).

**Rule 2** follows the same path as rule 1 except that the serum value of prealbumin is within normal limits, i.e. no severe changes in liver is likely. Biopsy 0 is supported if the multiplication of age and prealbumin is below 9 000. A low age and a low value of this quotient has in the present study been shown as factors associated with only mild or no changes in the liver (Biopsy 0).

**Rule 3**. This branch to the right of the tree represents a no-biopsy decision (Biopsy 0). A normal value of AST is indicative of minor histo-pathological changes. A high multiplication value of age and quotient between AST and ALT as well as ALT-values above normal limits are both associated with more severe changes. But since AST-value is normal it is reasonable that a serum ferritin value below 150 is strong enough marker to predict Biopsy 0, i.e. only minor changes.

**Rule 4**. All four parts of this rule are associated with Biopsy 0, i.e. a normal AST value, a low
multiplication value of age and quotient between AST and ALT, a low multiplication value of age and serum ferritin and a negative test for chronic C virus hepatitis (anti HCV).

**Rule 5.** The first three parts are the same as described for Rule 4 above associated with Biopsy 0 and only minor changes in the liver. Nevertheless, if the anti-HCV test is positive, the likelihood of a drastic development of inflammation and/or fibrosis is drastically increased in spite of the low values found in all other parts of the rule suggesting Biopsy 0. The reason for this is that almost every patient with positive anti HCV tests has more severe degrees of inflammation and/or fibrosis (Biopsy 1).

**B. Rules not readily understandable on the basis of medical knowledge.**

From the methodological point of view these rules are not wrong. If attempts were made to find out in detail which patients are responsible for the different branches of the tree, it is likely that all these rules could be explained. This could be exemplified with Rule 10. The last part of this rule shows that a low serum-albumin value is predictive of Biopsy 0 while a value within normal limits is predictive of Biopsy 1. The reason for this is probably that the second condition of the last part of this rule shows that the multiplication of age and the quotient between albumin and prealbumin is high (above 9000). This quotient is higher with higher albumin values. This is the mathematical basis for the last part of the rule with low or normal albumin value. Nevertheless, from the clinical point of view this last part makes no sense and is likely to be wrong for the whole patient material. Low serum albumin values are clear markers of ongoing inflammation.

Our attempt to clarify differences between statistical associations and clinical implications of rules yielded the following medical comments for Rules 6-14.

**Rules 6+7.** AST within normal limit is predictive of Biopsy 0, age multiplied with quotient AST/ALT above 40 is predictive of Biopsy 1 and ALT within normal limit is an indication of only minor liver defect. The number of patients with values satisfying the conditions of the said three parts of this rule is very small since almost every patient had ALT-value above the normal limit. That an albumin value below (Rule 6) or above the upper normal limit (Rule 7) is of predictive value for liver biopsy findings is hard to explain.

**Rules 8+9.** An AST value within normal limits is associated with only minor defects, the multiplication value of age and quotient of AST and ALT above 40 is indicative of more severe changes, and both ALT above normal limits and ferritin above 150 are also likely to represent more severe degree of changes. However, why GGT value below 1.05 (Rule 8) is prediction for Biopsy 1 and GGT value above 1.4 (Rule 9) is prediction for Biopsy 0 is not understandable. We know from other analysis of the present material that the value of GGT has a very low impact on what can be expected from an histopathological examination of liver biopsy.

**Rules 10+11.** AST value above normal limit is associated with Biopsy 1, prealbumin above 0.18 with Biopsy 0 and age multiplied with the quotient between albumin and prealbumin above 9000 with Biopsy 1. As stated above in the first paragraph of this section about rules of questionable value, the last part of the rule 10+11 is contradictory to medical knowledge. A low serum albumin value should be strong prediction of Biopsy 1, the opposite to the last parts of Rules 10 and 11.

**Rules 12+13+14.** An AST value below 0.7 is associated with Biopsy 0, age multiplied with quotient between AST and ALT below 40 with Biopsy 0 and age multiplied with serum ferritin above 11500 with Biopsy 1. The next step for all three rules again concerns GGT-values. As stated above, these values are poorly connected with the severity of changes in the liver. Moreover, with these rules a GGT value above 1.4 is prediction for Biopsy 1, in rule 9 this limit of GGT was prediction for Biopsy 0. The last parts of rule 12 and 13 are contrary to medical knowledge. A high body mass index (BMI) is clearly associated with only non severe changes in the liver i.e. steatosis. Thus BMI >25 should be prediction for Biopsy 0 and BMI<25 for Biopsy 1.

**DISCUSSION AND CONCLUSION**

The Artificial intelligence system based on inductive learning principles has demonstrated in this study some desirable properties of a good knowledge elicitation tool. It is intended for fast prototyping of domain knowledge, even when taking into account only a few medical premises. Resulting rules can be a good basis for knowledge representation in a future automated system. The quality of derived knowledge was discussed from the medical assessment point of view. Out of 14 rules formulated, 5 were found relevant and 9 could not be readily understandable on the basis of medical knowledge. This is not to be considered as a poor result of knowledge extraction. In the field of unspecified liver diseases even experts do not have clear decision rules ('know-how') that are based on solid medical knowledge. The main result of knowledge extraction is a possibility to predict, relying on ordinary laboratory values, more severe degrees of histopathological findings with a certainty of at least 0.7. The tree shows that this is possible especially in patients with raised values of both AST or ALT (the right part of the tree). In these patients only complementary laboratory values of prealbumin and albumin are needed for prediction of the findings of light microscopy examination. The quotient between albumin and prealbumin is itself of great importance for this prediction. In patients with raised levels of ALT, but with normal AST at the time of
liver biopsy, the ability to make prediction based on laboratory findings is more difficult to interpret. Most of this latter group of patients had raised values of both AST and ALT at least some times during the six-month observation period preceding liver biopsy. For this group of patients one can probably conclude that the decision tree can be used mainly to avoid unnecessary liver biopsies. The decision tree is based on the results obtained for 165 patients. The results shown in the tree are probably representative for groups of patients living in the socio-economic milieu similar to that found in Sweden. Since patients with chronic hepatitis C virus infection were included in the present material, the advice available in the decision tree is probably representative also for regions with high prevalence of this disease such as Southern Europe.

Pruning of decision tree enables us to take only a few medical findings to recommend (or advise against) biopsy. From the medical point of view, however, it is of interest to check additional values especially immunoglobulins, auto antibodies and serum iron in order to exclude established liver diseases. Liver biopsy is often needed in asymptomatic liver diseases not only as the ultimate confirmation of diagnosis, but also to assess how active the disease is. Moreover, ultrasound examination was performed in all patients primarily in order to exclude focal lesions of the liver and/or dilated bile ducts. The result of changes of eccogenicity of the liver parenchyma did not come up as a significant factor for liver biopsy prediction in the decision tree.

In our earlier publications, we presented and discussed other decision support algorithms based on statistics, importance of outliers of laboratory values. Previous work in knowledge extraction from the same data using multivariate statistics and rough sets [10,11] has proved that these methods were more precise than the inductive learning system used. At the same time, they were more time consuming and methodologically demanding. Finally, all methodologies used are aimed at helping clinicians to identify patients who have severe changes in the liver.

This study looks into some of the important issues concerning objective medical decisions with the aid of knowledge-based systems [12]. The primary objective was to support clinicians' requirements to predict biopsy necessity by means of informational technology based in artificial intelligence.

The study has shown that the inductive learning system can be recommended for the conceptualisation of the medical domain knowledge. Its tree-structured representation has two main advantages: the decomposition of knowledge into maintainable units (rules), and migration of knowledge among experts and support systems (population specific and generic ones). Methodological and medical evaluation performed illustrated that diagnostic rules once obtained can become decision algorithms after assessing their appropriateness. Rules based on a few significant laboratory findings only were enough for a biopsy recommendation. That is a valuable clinical help, since the medical expertise itself does not provide many answers. Therefore, any new knowledge is to be considered and applied, as well as added to the knowledge base of the Liver Guide [13], an automated decision support system. This can directly contribute to the avoiding of invasive medical examination and to the better cost-effectiveness of clinical diagnostics.

References