Computing and data processing

Construction and application of hierarchical decision tree for classification of ultrasonographic prostate images

R. J. B. Giesen A. L. Huynen R. G. Aarnink
J. J. M. C. H. de la Rosette F. M. J. Debruyne H. Wijkstra

Department of Urology, Biomedical Engineering, University Hospital Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands

Abstract—A non-parametric algorithm is described for the construction of a binary decision tree classifier. This tree is used to correlate textural features, computed from ultrasonographic prostate images, with the histopathology of the imaged tissue. The algorithm consists of two parts; growing and pruning. In the growing phase an optimal tree is grown, based on the concept of mutual information. After growing, the tree is pruned by an alternating interaction of two data sets. Moreover, the structure and performance of the constructed tree are compared to the results using a slightly modified corresponding growing and pruning algorithm. The modified algorithm provides better retrospective and prospective classification results than the original algorithm. The use of the tree for automated cancer detection in ultrasonographic prostate images results in retrospective and prospective accuracy of 77.9% and 72.3%, respectively. Using this tissue characterisation, a supporting tool is provided for the interpretation of transrectal ultrasonographic images.

Keywords—Growing and pruning, Hierarchical decision tree, Prostate, Tissue discrimination, Ultrasonography


1 Introduction

In the diagnosis of prostatic complaints, TransRectal Ultrasoundography (TRUS) plays an important role on the detection of carcinomas in the prostate gland. For an experienced urologist, ultrasonography provides a reasonably accurate tool of carcinomas in the prostate gland. For an experienced Sonography (TRUS) plays an important role on the detection in structures on the experience and expertise of the urologist and their

However, the interpretation of the images is strongly dependent for detecting suspicious lesions in the prostate (Wolf, et al., 1992; Kaye and Lightner, 1993; Gerber et al., 1992). However, the interpretation of the images is strongly dependent on the experience and expertise of the urologist and their capability to distinguish artefacts, benign and malignant structures (Brawer, 1993; Loch et al., 1990; Scardino et al., 1989; Shinohara et al., 1989). When interpreting TRUS images, a urologist uses additional information like Digital Rectal Examination (DRE) and Prostate Specific Antigen (PSA) (Gerber et al., 1992; Wolf et al., 1992; Brawer, 1993; Lee et al., 1989a; Scardino et al., 1989). The combination of TRUS, DRE and PSA, however, does not always provide an objective and accurate indication for the detection and staging of prostate cancer; not all carcinomas are palpable and/or can be visualised. In particular, cancer at an early stage, which is curable and therefore important to detect (Lee et al., 1989a), is difficult to perceive with ultrasound, mainly because of its iso-echogenic character (Shinohara et al., 1989).

We have developed a system to provide additional information to the urologist in the interpretation of ultrasonographic prostate images, the Automated Urologie Diagnostic EXPert system (AUDEX) (Giesen and Huynen, 1991; Huynen et al., 1994, Giesen et al., 1994). This system colour-codes TRUS images according to the probability of malignancy of the imaged tissue. First, textural information in the images is quantified, which is often difficult or impossible for a human observer to perceive. Secondly, this textural information is related to the histopathology of prostate tissue, and the probability for malignancy is calculated.

The textural characteristics are quantified by five parameters (uniformity, contrast, inverse difference moment, entropy and correlation) computed from the co-occurrence matrix (Haralick et al., 1973) in combination with the signal-to-noise ratio (mean/standard deviation). The above five parameters are statistical measures for the two-dimensional spatial dependencies of grey levels in the image. A detailed description of the parameters and technical conditions used for this textural quantification has been presented previously (Huynen et al., 1994).

The correlation between the texture descriptions and histopathology is computed using a binary decision tree
use of a decision tree for prostate tissue discrimination and
determination of the tissue. In this paper, we describe the construction and
use of a decision tree for prostate tissue discrimination and
construction of colour-coded TRUS images.

A binary tree is a multi-stage classifier, which consists of a
root node, a set of non-terminal nodes and a number of
terminal nodes. The complex global classification problem in
the root is solved on a local base; it is split into simpler, local
decision in non-terminal nodes of the tree. Each non-terminal
node represents a decision process, based on a subset of the
available training samples, which determines the descendants
of that node. If a node is a terminal node, a class label is
associated with it. In this way, a multi-dimensional feature
space is recurrently partitioned by defining hyperplanes. Each
resulting hypercube represents a certain class in the classification
problem.

When no information is provided about the conditional
probability density functions of the discriminating features in the
classification problem, the use of a tree-structured classifier is a very suitable technique (BREIMAN et al., 1984; FRIEDMAN, 1977). It is successfully used in different classification problems such as character recognition (SETHI and SARVARAYUDU, 1982; WANG and SUEN, 1987), wave form recognition (BREIMAN et al., 1984; GELFAND et al., 1991), remote sensing (HENRICHON and FU, 1989), pattern recognition (PARK and SKLANSKY, 1990; ROUNDS, 1980), and medical classification and diagnosis (BREIMAN et al., 1984; LANDWEERD et al., 1983; LIN and FU, 1983; MUI and FU, 1993).

The tree construction algorithm described here is based on a
combination of the mutual information function (SETHI and
SARVARAYUDU, 1982), a growing algorithm, and an alternating
pruning algorithm. The iterative Tree Growing and Pruning
(ITGP) algorithm is modified and improved for our application
(GELFAND et al., 1991). For the discrimination between
malignant and benign images structures, both algorithms, ITGP and the modified algorithm, were used and evaluated in the classification of textural features computed from ultrasoundographic prostate images.

2 Materials and methods

For training purposes, a series of TRUS images of puncture
biopsies were recorded by digitising the video signal of the ultrasound scanner.* These images were recorded in a standardised way, such that the exact puncture location in the image was known. The image texture at each biopsy location was computed at three different places; the biopsy was divided in a proximal, central and distal part (HUYSEN et al., 1994). These samples, the texture descriptions in combination with the histopathology of the removed tissue, were used as the input for construction of the decision tree. As the aim was to discriminate between benign and malignant prostate tissue, the construction of the tree classifier dealt with a two category classification problem. However, the methods described for building such a decision tree are also applicable to multi-class problems.

For the construction and evaluation of a classifier, a
measurement has to be provided to express the classification
performance. The tree classification error or resubstitution
estimate \( R(T) \) can be defined as (BREIMAN et al., 1984)

\[
R(T) = \frac{1}{N} \sum_{i=1}^{N} \delta(i) \quad \text{where} \quad \delta(i) = \begin{cases} 1 & \text{if } \text{class } t \text{ predicted} \text{ correctly for sample } i \\ 0 & \text{otherwise} \end{cases}
\]

where \( p(t) \) is the probability of a sample reaching node \( t \) and \( p(j|i) \) is the conditional probability of class \( j \) given node \( i \). When assigning class \( j \) to a node if \( j = m_i \) and \( m_i \) minimises \( R(i) \), the error in a node is always higher than or equal to the sum of the errors in its children. The resubstitution estimate \( R(T) \) is calculated by summing the errors in the terminal nodes. Each split results in a possible reduction in the errors, and therefore leads to a potential decrease in the resubstitution estimate of the tree. However, the complexity (number of decisions) of the tree needs to be controlled to deal with the problem of over-classification. To restrict this number of splits, both ITGP and the modified algorithm do not use a stop criterion, but instead consist of two parts: a growing phase and a pruning phase. In the growing phase a complete optimal tree is grown. This means that all nodes are split until no further splitting is possible or necessary; each terminal node has pure class membership or the number of samples is below a predefined threshold. In the pruning phase, the constructed tree is pruned to reach a compromise between error rate and complexity.

The ITGP algorithm proposes an iterative process of both
growing and pruning with an interaction between two populations. The available samples in the training set are divided into two equal-sized subsets. These populations alternately grow and prune a tree. The first population is used to construct a complete optimal tree. Based on the errors made by classifying the second population with the built tree, the tree is pruned. The decision of whether to prune is based on a comparison of the error in a node with the sum of the errors in its subtree. If a node error is equal to or less than the sum of the errors of the terminal nodes in the subtree, the split was unnecessary or even disadvantageous. The subtree is then pruned in order to eliminate the unnecessary split. After pruning the optimal tree, the two populations are exchanged and the process is repeated; the tree is then grown with the second population and pruned with the first one. Gelfand et al. have shown that these alternating processes of growing and pruning converge, and that in the final tree a good compromise is reached in the correspondence of the classification of the two populations (GELFAND et al., 1991).

In our modified algorithm, the interaction between the two populations is limited to the pruning phase. Particularly in the case of small data sets, the possible different distributions of the samples over two populations result in a wide variation in the structure of the resulting tree. Therefore, we suggest growing a complete optimal tree using all samples in the training set, and after building this complete tree, splitting the population into two equal sized subpopulations. The tree is then alternately pruned by the populations on the same criterion as used in the ITGP algorithm. The pruning process stops when two trees pruned consecutively are equal, which means that the resulting tree is the optimal one for both populations. In this way, the decisions made in the non-terminal nodes of the tree are based on all samples, and the structure of the tree is less affected by the distribution of the samples over the populations.

Moreover, to control the number of non-terminal nodes (splits) of the tree, a complexity cost \( a \) was introduced in both algorithms. This factor is a constant which weighs the number of terminal nodes. It is an additive function on the number of terminal nodes, and therefore does not directly affect the depth of the tree (maximum number of decisions in a path from the root to the terminal nodes). The estimated

* Kretz Combison 330 with a VRW 77 Ak 7.5 MHz multi-phase rectal transducer
complexity error \( R_d(T) \) of a (sub)tree can be calculated by (Breiman et al., 1984; Gelfand et al., 1991)

\[
R_d(T) = \sum_{\text{terminal}} R_d(t) = \sum_{\text{terminal}} R(t) + N_{\text{terminal}} \alpha
\]

\[
R_s(t) = R(t) + \alpha
\]

where \( N_{\text{terminal}} \) stands for the number of terminal nodes.

In the growing phase, both with the ITGP and the modified algorithm, we used the concept of mutual information (Sethi and Sarvarayuda, 1982) for the selection of an optimal split in a node. This function, which is applicable in a multi-feature multi-class environment, provides a measure of the information gained at a certain split criterion in a node. The best split in a node can be found by calculating the mutual information over all possible splits in all feature dimensions and selecting the one with maximum information.

As true values of the various probabilities used for computation of the mutual information and the node errors were absent, an estimation based on the samples in the data set had to be used. Assuming the probabilities in a node to be proportional to the occurrence of the several classes in the node, a very unbalanced distribution can have disadvantageous consequences for the classification; the ratio of the misclassification rates of the various classes is proportional to the ratio of the occurrence of the classes in the data set, resulting in a disparity in the error rates. Therefore, it is important to adjust for unbalanced priors, to ensure an equalisation of the individual class misclassification rates. In both the growing and the pruning phases, we introduced misclassification weights \( W_i \) for a correction of the distribution of the learning samples over the classes. These weight factors are inversely proportional to the \( a \) priori probability of the corresponding class in the training set. In the computation of probabilities in a node, the number of samples of a class is multiplied with the corresponding weight:

\[
P(c_i^T) = \frac{N(c_i^T)W_i}{\sum_j N(c_j^T)W_j}
\]

where \( N(c_i^T) \) is the number of samples of class \( i \) in node \( k \). In this way, all probabilities are normalised to equal priors of all classes in the learning set.

After constructing a decision tree, all terminal nodes were labelled with a probability for malignancy. This probability was computed according to the distribution of the training samples reaching the node. Using these probabilities, new ultrasonographic images were analysed and colour-coded by placing a window (approximately 5 x 5 mm) over the image, computing the textural features in this window and classifying them with the decision tree. The resulting probability for malignancy was translated to colour, using a colour scale ranging from blue, representing 0 probability for malignancy, to red, representing 100% probability for malignancy, and then projected into the original image. Next the window was shifted, and this process was repeated until a specified region of interest had been analysed.

To compare the ITGP algorithm and the modified version, both retrospective (resubstitution of the training samples) and prospective (classification of ‘new’ samples), the classification error \( R(T) \) was used (eqn. 1). Like the growing and pruning phases, this proportion of wrongly classified samples was computed after normalisation to equal priors (eqn. 3). \( 1 - R(T) \) then provides a measure for the accuracy, which is equal to the average of sensitivity and specificity.

A total of 198 images of biopsies have been analysed, with 139 (70.2%) of benign and 59 (29.7%) of malignant tissue. Each biopsy location was analysed by taking a proximal, central and distal area, resulting in a collection of 594 (3 x 198) six-dimensional data samples used to build a decision tree with both tree construction algorithms. As classification of the training samples after resubstitution in the built tree is often not a reflection of the classifier’s practical performance (in the case of over-classification), the available samples were divided into two populations. One set, consisting of two-thirds of all samples (279 benign and 117 malignant), was used to construct the tree and the other set (38 benign and 60 malignant) for prospective evaluation. Owing to the mutual dependency of the three samples from one biopsy, the triplets are kept together both in the distribution of the samples over the training and prospective set, as well as in the division of the training set into two populations for the tree construction. With both algorithms a series of trees were calculated as a function of the complexity factor \( a \). To check the influence of the distribution of the samples over the different populations, for each a 100 random distributions were used to build trees. A threshold of three samples was used as the stop criterion for splitting a node.

### 3 Results

For both algorithms the results are presented as the average over 100 random distributions of the samples over the populations. The mean retrospective and prospective accuracy \( 1 - R(T) \), corrected for unbalanced prevalence, are plotted in

<table>
<thead>
<tr>
<th>( n = 100 )</th>
<th>retrospective</th>
<th>prospective</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>accuracy, %</td>
<td>accuracy, %</td>
</tr>
<tr>
<td>ITGP</td>
<td>( a = 0.006 )</td>
<td>75.76 [2.98]</td>
</tr>
<tr>
<td>modified</td>
<td>( a = 0.008 )</td>
<td>77.94 [2.62]</td>
</tr>
</tbody>
</table>

Fig. 1 Average (\( n = 100 \)) retro- and prospective accuracy (%) of trees constructed with ITGP and modified algorithm as function of complexity factor \( a \)
Fig. 2. Ultrasonographic prostate image with proven malignancy.

Fig. 3. Colour-coded TRUS image; computer analysis corresponds to histopathological examination; region marked malignant (red) was proven to be cancer.

4 Discussion and conclusion

The algorithms produce equal sized and balanced trees. However, the modified algorithm provides significantly better results than the original algorithm. Whether this improvement is also of clinical significance is questionable. However, compared with the ITGP algorithm the structure of the constructed trees using the modification varies less with different distributions of the samples over the populations. Using the modified algorithm, a reduction in the standard deviation of 38.8% can be seen for the number of terminal nodes and 32.1% for the depth of the tree. Moreover, the modified algorithm is computationally quicker. In the modified algorithm, the growth of one complete optimal tree needs more computation time than when using the ITGP algorithms, because the tree is built with all available training samples. However, the growing phase is applied only once, whereas in the ITGP algorithm the number of growing phases is unknown; the process is repeated until two subtrees pruned consecutively are equal. As the total computation time for pruning is comparable for both methods, the modified algorithm provides a computationally quicker hierarchical tree construction; in our application, it saves up to 45% of computation time.

Nevertheless, there are some disadvantages associated with the modification. In the ITGP node labelling and pruning are based on disjoint data samples. Therefore, an internal error rate can be estimated (based on the disjoint data samples in the growing phase) (GELFAND et al., 1991). By growing with all training samples, however, such internal error rate estimation is not provided. In addition to the resubstitution error, which is often an overestimation for the practical performance, only cross-validation could be used for error estimation. However, the latter introduces a computational burden (BREIMAN et al., 1984). Therefore, we have chosen to split the available data.

Table 2 Average classification results (sensitivity, specificity, positive and negative predictive value) of the optimal series trees constructed with both ITGP and modified algorithms

<table>
<thead>
<tr>
<th>n = 100</th>
<th>sensitivity, %</th>
<th>specificity, %</th>
<th>positive predictive value, %</th>
<th>Negative predictive value, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITGP a = 0.006</td>
<td>retrospective</td>
<td>71.98</td>
<td>79.59</td>
<td>59.60</td>
</tr>
<tr>
<td></td>
<td>prospective</td>
<td>62.86</td>
<td>76.98</td>
<td>54.28</td>
</tr>
<tr>
<td>modified a = 0.008</td>
<td>retrospective</td>
<td>73.51</td>
<td>82.37</td>
<td>63.62</td>
</tr>
<tr>
<td></td>
<td>prospective</td>
<td>64.63</td>
<td>79.90</td>
<td>58.30</td>
</tr>
</tbody>
</table>

Table provided by the modified algorithms, the maxima in the prospective results were used $a_{\text{max}}$. The structure and performance of these trees are presented in Tables 1 and 2. The series of trees ($n = 100$) are comparable for both methods, the modified algorithm provides a computationally quicker hierarchical tree construction; in our application, it saves up to 45% of computation time.

Concerning the classification accuracy, both retrospective (77.94%) and prospective (72.27%), the modified algorithm provided better classification results than the ITGP algorithm ($p < 0.001$).

Fig. 2 shows an example of an ultrasonographic prostate image with a histological proven malignancy. The corresponding colour-coding is presented in Fig. 3. It can be seen that, using the computer analysis, the iso-echoic malignancy (undetectable with normal TRUS) is marked as cancerous.
into a training set and a test set to evaluate the actual performance. By using such a set, real prospective performance is obtained.

A total of 198 images recorded from puncture biopsies were used. Reasons for biopsy were an abnormal DRE, elevated PSA (≥10 ng ml⁻¹) and/or suspected TRUS. Using these three clinical indicators for malignancy, the following classification results were obtained for cancer detection: a sensitivity of 97%, specificity of 22%, and a positive (negative, respectively) predictive value for both of 35%. Comparing these results with the prospective results obtained with automated analysis (sensitivity of 65%, specificity of 80%, positive (negative, respectively) predictive values of 58% and 84%), it can be seen that using the latter provides better performance for cancer detection. Although the sensitivity using the clinical indications is higher than that using computer analysis, the very low specificity introduced lower predictive values. The positive predictive values for routine prostate examination (DRE, PSA, and/or TRUS) is comparable with those reported in the literature; values between 29% and 42% are presented (CHODAK et al., 1986; LEE et al., 1989; SHINOHARA et al., 1989; BRAWER, 1993). Comparing the results of automated tissue characterisation to those found in similar studies about image processing in the prostate is difficult, because either only the sensitivity (98-1%) is reported (BERTERMANN et al., 1989; LOCH, et al., 1990), or the settings and conditions of the study are not clearly described (ZIELHEL ET AL., 1985).

In summary, we have provided a reasonably accurate methodology for constructing a multi-stage tree classifier. This classifier has been used to relate textural features from ultrasonographic prostate images to histopathology. In this way, additional information is provided for the interpretation of transrectal ultrasonographic prostate images, and the diagnostic value of TRUS can be increased.

Acknowledgment—The authors like to thank Professor J. M. Thijssen, PhD, MSc, for his helpful criticism in the preparation of this paper.

References

GIESEN, R. J. B. and HUYSEN, A. L. (1991): 'Ultrasonographic tissue discrimination; automatic detection of prostate carcinoma.' Master Thesis, Department of Computer Science, University of Twente, Enschede, The Netherlands

Author's biography

Robert J. B. Giesen was born in Kerkrade, The Netherlands, in 1967. He received his MSc degree in Computer Science and Biomedical Engineering in 1991 from the University of Twente, Enschede, The Netherlands. Since 1991 he has been working at the BioMedical Engineering Group of the Department of Urology at the University Hospital Nijmegen, The Netherlands. He received his PhD in 1995. His research interests are in medical image analysis, computer science and ultrasonography.