Diagnosis of Glaucoma by Indirect Classifiers

A. Peters[1], B. Lausen[1], G. Michelson[2], O. Gefeller[1]

[1]Department of Medical Informatics, Biometry and Epidemiology, [2]Department of Ophthalmology and Eye Hospital, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

Summary

Objectives: Demonstration of the applicability of a framework called indirect classification to the example of glaucoma classification. Indirect classification combines medical a priori knowledge and statistical classification methods. The method is compared to direct classification approaches with respect to the estimated misclassification error.

Methods: Indirect classification is applied using classification trees and the diagnosis of glaucoma. Misclassification errors are reduced by bootstrap aggregation. As direct classification methods linear discriminant analysis, classification trees and bootstrap aggregated classification trees are utilized in the problem of glaucoma diagnosis. Misclassification rates are estimated via 10-fold cross-validation.

Results: Indirect classification techniques reduce the misclassification error in the context of glaucoma classification compared to direct classification methods.

Conclusions: Embedding a priori knowledge into statistical classification techniques can improve misclassification results. Indirect classification offers a framework to realize this combination.

Keywords

Glaucoma, classification, decision trees

1. Introduction

Medical decision making is often a complex process based on several high dimensional measurements obtained by different examinations. Usually, a priori information about the relationship between the disease and the outcome of some of these examinations is given. Statistical classification methods that mimic the process of medical decision making should pay attention to the distinction between such a priori knowledge and the information from measurements required to predict the parameters defining the diagnosis. A classification rule constructed in such a way can be based on a reduced set of examinations although it makes use of the full information of the data, thereby reducing patients' time and costs for the diagnostic procedure.

A framework called indirect classification has been suggested by Hand et al. (1) to allow the incorporation of medical a priori knowledge into a statistical classification method. It assumes that the outcomes of the examinations are subdivided into three groups of variables: those to be used predicting the diagnosis, those to be used defining the diagnosis and the final diagnostic variable itself. The indirect classification process is executed in two separate steps. In the first step, prediction models for the defining variables embedding all other variables of potential influence are created based on a learning sample. In the second step, these defining variables are classified according to a deterministically known classifying function to yield the final medical diagnosis. The medical a priori knowledge is used twofold in this approach: (i) it is the criterion for the subdivision of variables into the different groups, (ii) it determines the fixed classifying function used in the second step of the procedure.

To illustrate the practical application of indirect classification and to demonstrate the gain achieved by its utilization we apply the method to glaucoma diagnosis. Classification trees (2) are used as a statistical tool to perform the indirect classification approach. Bootstrap-aggregation (bagging) of multiple classification trees stabilizes the method (3, 4). Furthermore, the direct classifiers linear discriminant analysis, classification trees and bagged classification trees are performed. Results of indirect versus direct classification rules are compared with respect to the estimated misclassification error.

The paper is organized as follows: section 2 describes the diagnosis of glaucoma, i.e. the given a priori information about the disease. An overview of the statistical classification methods is given in Section 3. These methods are evaluated in Section 4 and discussed in Section 5.

2. Glaucoma Diagnosis

A priori knowledge incorporated in the framework of indirect classification is given by the diagnosis of glaucoma. Glaucoma is a slowly progressive and irreversible disease, which affects the retinal nerve fiber layer, for details see (5) and (6). It is the second most frequent cause of blindness worldwide and occurs generally in a population of elderly people.

Glaucoma diagnosis is usually based on a multitude of variables describing the visual field defect and the loss of retinal nerve fibers, although a unique definition is controversial (7). Using clinical expert judgment and some a priori analysis, we define glaucoma based on loss of rim area $w_{lor}a$, corrected loss variance w_{clv} and contrast sensitivity w_{cs}. Corrected loss vari-
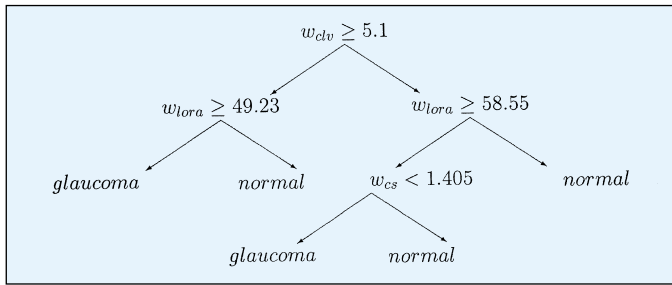
Fig. 1
Diagnosis of glaucoma. Based on the intermediate variables w_{clv} (corrected loss variance), w_{lora} (loss of rim area) and w_{cs} (contrast sensitivity) a patient is diagnosed according to the graph.

ance and contrast sensitivity are criteria to assess the visual field defect, whereas loss of rim area indicates the loss of retinal nerve fibers. A graphical representation of the glaucoma diagnosis is given in Fig. 1.

More formally, a function of ordinal variables specifies the exact definition of glaucoma. This definition is translated into a function $g(\mathbf{w}_n)$. Using the classifying function $g(\mathbf{w}_n)$ a patient is classified as normal if $g(\mathbf{w}_n) \leq 0$ and as glaucomatous if $g(\mathbf{w}_n) > 0$, where $\mathbf{w}_n \in \mathbf{R}^3$ and
$$g(\mathbf{w}_n) := w_{n:clv}w_{n:lora} + (1-w_{n:cs})(w_{n:lora}-1)(1-w_{n:clv}) \text{ with} \quad (1)$$
$$\mathbf{w}_n = (w_{n:lora}, w_{n:cs}, w_{n:clv}) \text{ and}$$

$$w_{n:clv} := \begin{cases} 0, & w_{clv} < 5.1 \\ 1, & w_{clv} \geq 5.1 \end{cases},$$

$$w_{n:cs} := \begin{cases} 0, & w_{cs} < 1.405 \\ 1, & w_{cs} \geq 1.405 \end{cases},$$

$$w_{n:lora} := \begin{cases} 0, & w_{lora} < 49.23 \\ 1, & 49.23 \leq w_{lora} < 58.55 \text{ . (2)} \\ 2, & w_{lora} \geq 58.55 \end{cases}$$

Several examination tools assess the visual field (perimetry) and the morphol-

ogy of the optic nerve head (ONH). An examination for measuring the visual field defect is for example the flicker test (8) which provides the variables w_{clv} and w_{cs}.

The ONH can be assessed by a two dimensional fundus photography (papillometry). Loss of rim area w_{lora} is a variable obtained by this examination. It describes the amount of affected nerve fibers. A three dimensional topographical analysis of the ONH can be performed using the Heidelberg Retina Tomograph (HRT) (9, 10). Three series of 32 confocal images (256 × 256 pixels) are acquired using a 670 nm diode laser and a scan angle of 10°. The resulting topographies allow the calculation of a mean topography. All calculations of disc parameters are processed with HRT-Software Version 2.01 (11). A standard reference plane placed 50 mm posterior to the mean height of the contour line is set. The disc margin is defined in a temporal segment between 350° and 356°. On the topographic images, the optic disc was outlined along the inner margin of the scleral ring of Elschnig, see (Fig. 2).

The HRT examination provides further information about the optic nerve head morphology compared with papillometric results.
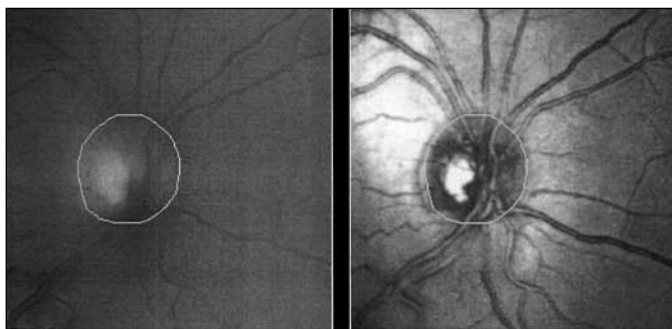

Fig. 2
Illustration of the black/white-coded image of the topography and the reflectivity of the surface.

3. Classification Methods and Assessment of the Misclassification Error

In this section we discuss direct and indirect classification methods more formally. Direct and indirect classifiers are indicated with superscript d and ind respectively. We give the notation of bagged direct classification methods following (3) and introduce bagging for indirect classifiers.

Let $\mathbf{L} = \{(y_i, \mathbf{w}_i, \mathbf{x}_i), i = 1, \dots N\}$ denote a learning sample of N independent observations. The observations of the learning sample consist of three classes of variables: those to be used to predict the diagnosis (*explanatory variables*), denoted by $\mathbf{x}_i = (x_{i1}, \dots, x_{ip}), \in \mathbf{R}^p$, those to be used defining the diagnosis (*intermediate variables*), denoted by $\mathbf{w}_I = (w_{i1}, \dots, w_{iq}), \in \mathbf{R}^q$, and the final diagnosis itself (*response variable*), denoted by y_i, where $y_i \in \{1,2\}$ has the values "glaucoma" or "normal".

The intermediate variables are related to the explanatory variables by a function $f: \mathbf{R}^p \rightarrow \mathbf{R}^q$ that is $\mathbf{w}_i = f(\mathbf{x}_i) + \epsilon$, where ϵ is a random vector. Note that the functional relationship between explanatory and intermediate variables is not known. In contrast, a deterministically known function $g: \mathbf{R}^q \rightarrow \{1,2\}$ defines the final diagnosis, which classify an observation based on the intermediate variables only: $y_i = g(\mathbf{w}_i)$. In our framework of glaucoma classification this function is given in (1) and displayed in Fig. 1.

When performing the indirect classification method the relationship between explanatory and intermediate variables $\mathbf{w} = f(\mathbf{x})$ is unknown, therefore it has to be estimated. Models are created to predict the values of the intermediate variables based on the explanatory variables. In a second step the classification is performed using the deterministically known classifying structure $g(\mathbf{w})$ between intermediate and response variables. The indirect classification approach is displayed in Fig. 3.

We use classification trees (CTREE^{ind}) for modeling the functional relationship between the set of explanatory and intermediate variables. For each intermediate

variable a classification tree is constructed by a recursive evaluation of all binary splits of every explanatory variable. We use the Gini-index as measure for impurity, see (2). The tree growing is stopped if a node contains less than 20 observations or if none of the evaluated splits decreases the overall misclassification error by a factor of at least 0.01. In the framework of direct classifiers, described below, bootstrap aggregation (bagging) or boosting leads to a substantial reduction of misclassification error in many applications (3, 4, 12).

Bagging for indirect classifiers is performed as follows: a random sample of size N is drawn with replacement from the learning sample **L**. Classification trees are computed based on this sample. Predictions are calculated and classified as normal or glaucoma based on the fixed classifying function $g(\mathbf{w})$. This procedure is iterated k times. Hence, there are k predicted classifications. A patient is classified as normal or glaucoma if the majority of predicted classifications based on the bootstrap samples are normal or glaucoma, respectively. We denote the bagged indirect classifier by bagged-CTREE$^{\text{ind}}$.

In contrast to indirect classification, direct classification rules are trained based on the set of explanatory and response variables only. Hence, the class membership variable y is treated as a nominal variable, the known underlying classification structure consisting of the set of intermediate variables is ignored. The function $h(\mathbf{x}) = y$ is estimated instead of estimating $\mathbf{w} = f(\mathbf{x})$ and applying the known classification rule $g(\mathbf{w})$. Figure 4 displays the direct classification approach.

Linear discriminant analysis (LDA$^{\text{d}}$), classification trees (CTREE$^{\text{d}}$) and bagged classification trees (bagged-CTREE$^{\text{d}}$) are applied as direct classification methods. Classification trees are constructed based on the explanatory variables, the dependent variable is the class membership variable y. The stopping rules for the constructed classification trees in the direct approach are the same as described above. To reduce the misclassification error, classification trees are combined with bagged classification, for more details see (3).
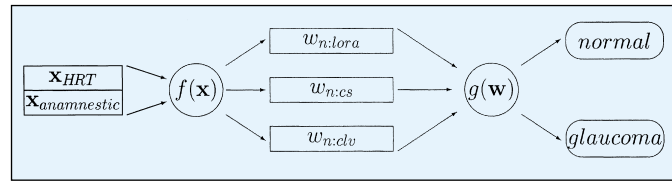

Fig. 3 Indirect classification in the example of glaucoma classification: Models are constructed based on the explanatory variables $\mathbf{x} = (x_{HRT}, x_{anamnestic})$ to predict the intermediates $\mathbf{w} = (w_{n:lora}, w_{n:clv}, w_{n:cs})$. Suspects are classified according to $g(\mathbf{w})$ as glaucoma or normal.

We use 10-fold cross-validation (10CV) for an unbiased estimate of the misclassification error (13) . This method works as follows: the given learning sample is randomly divided into 10 rather equally sized and mutually exclusive subsets. Predictive models are estimated based on 9 subsets and are used to predict the classification of the 10th subset with direct and indirect classification approaches, respectively. The proportion of misclassification is estimated in each subset and averaged over all 10 possible segmentations.

In the following analysis, this procedure is replicated 50 times for each classification method under consideration.

All calculations of classification trees are performed using the package rpart (14), in the statistical environment R (15).

4. Analysis and Results

Data from a cross-sectional study including 85 glaucomatous and 85 normal eyes from the Erlangen Glaucoma Registry are used (9). Only the measurements of the first examination of one eye of each patient are analyzed. The variables are obtained by HRT, papillometric and visual field examinations and include anamnestic data. Normal and glaucoma subjects are frequency matched by age and sex to adjust for possible confounders. We have 37 men and 48 women per group, the age structure is summarized in Table 1.

For future examinations we assume that only HRT and anamnestic data are available. Hence, only these sets of variables are used as explanatory variables and denoted by $\mathbf{x}_{HRT}$ and $\mathbf{x}_{anamnestic}$ respectively. There are 64 HRT and 7 anamnestic explanatory variables. The HRT parameters include several measurements of the volume and areas of the papilla in certain parts. The anamnestic variables are gender, age, maximally observed intra-ocular pressure, intra-ocular pressure at the examination day, genetic disposition, blood pressure and body mass index of the patients. The intermediate variables are based on the papillometric and visual field variables w_{lora}, w_{clv} and w_{cs} described earlier. Since the exact definition of glaucoma is formulated out of the ordinal information whether cut-points are exceeded or not, the intermediate variables are $w_{n:lora}$, $w_{n:clv}$ and $w_{n:cs}$ as described in Section 2. Table 1 summarizes the values of w_{lora}, w_{clv}, w_{cs} and some anamnestic variables.

For the indirect approach, the classifying function $g(\mathbf{w}_n)$ is known and fixed, see Fig. 1. In order to classify a new observation based on the set of explanatory variables, a predictive model for the intermediate variables has to be fitted. Classification trees are constructed as predictive models for

Fig. 4
Direct classification in the example of glaucoma classification: Glaucoma classification rules are constructed based on the explanatory variables
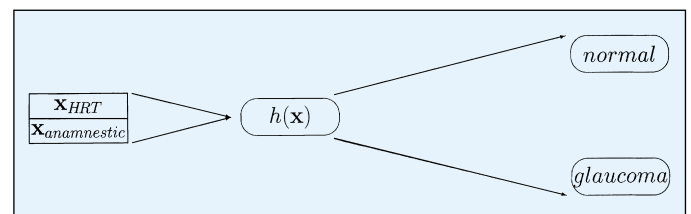$\mathbf{x} = (x_{HRT}, x_{anamnestic})$.

Table 1 Descriptive statistics for diagnostic and anamnestic variables (mean +/- standard deviation).

	normal	glaucoma
corrected loss variance (w_{clv})	2.753 +/- 5.977	42.242 +/- 35.667
contrast sensitivity (w_{cs})	1.434 +/- 0.215	1.137 +/- 0.300
loss of rim area (w_{lora})	45.245 +/- 15.033	69.865 +/- 10.674
age	55.177 +/- 9.604	55.271 +/- 9.534
intra ocular pressure	16.707 +/-3.361	16.393 +/- 3.271
body mass index	25.278 +/- 3.953	25.086 +/- 3.759

this purpose. Note that each intermediate variable is predicted by a univariate predictor, i.e. each intermediate variable is modeled separately. The bagged indirect classifier introduced in Section 3 is calculated using classification trees based on 50 bootstrap samples.

For comparison we applied the direct classifiers LDA[d], CTREE[d] and bagged-CTREE[d] to the same problem. The direct classification methods are performed on the data set of explanatory variables only, i.e. HRT measurements and anamnestic data. Bagging is performed using 50 bootstrap replications.

The estimated misclassification errors are summarized in Table 2 for all classification methods.

The direct classification results have higher estimated error rates than the indirect ones. Using classification trees the direct approach achieves an estimated classification error of 27.3%, whereas the indirect estimated misclassification rate is 22.8%. The best misclassification result with 19.3% of the direct approaches is achieved by using a bagged classification tree. In contrast, the smallest estimate of 17.9% is achieved by combining indirect classification and bagging. Comparing bagged direct and indirect classifiers, the reduction in estimated misclassification error from 19.3% to 17.9% demonstrates that a priori knowledge can improve the classifier.

5. Discussion

The indirect classification approach is a framework that combines medical and statistical knowledge. A set of measurements is structured and a classification rule is constructed with respect to medical a priori information. The gain of this procedure is a classification rule based on a reduced set of necessary diagnostic tests but incorporating the medical a priori information about the full set of measurements.

The framework of indirect classification has been applied here to the problem of glaucoma classification. The set of variables from different examination tools has been structured into explanatory, intermediate and response variables. The division has been performed considering the important aspect of glaucoma that patients do not usually detect its onset. However, early detection is of main importance, since adequate therapy can slow down the progression of the disease. As it is known that damages in the optic nerve head precede visual field defects of the patients (16), a good classification rule should be based on those measurements, which are able to detect early damages in the retinal nerve fiber layer. The HRT is an appropriate tool to detect early damages (17, 18). Hence, the ideal explanatory variables are HRT and anamnestic variables. Moreover, the definition of the disease is based on the optic nerve head morphology and the visual field defects of the patient. The three intermediate variables employed in the procedure also belong to these two areas.

The application of the indirect approach demonstrates the fruitful synergy of medical knowledge acquisition and statistical classification methods. A similar linkage of medical diagnosis and statistical methods has been performed by (19) to improve classification of vertical facial deformations in the diagnosis of short and long face syndrome.

The advantages of the indirect approach are of interest for various areas of medical decision making. It is an overall difficulty that a patient is investigated with several tools. A natural first aim is to reveal the relationship between different examinations in order to decide whether some medical examinations can be disregarded. The division of variables used in the indirect classification is an approach similar to the framework of graphical modeling (20). Extracted a priori information from this procedure of structuring offers the opportunity to build an indirect classification rule based on a reduced set of examinations.

Indirect Classification

	Estimated Misclassification Error
CTREE[ind]	22.8%
Bagged-CTREE[ind]	17.9%

Direct Classification

	Estimated Misclassification Error
LDA[d]	29.0%
CTREE[d]	27.3%
bagged-CTREE[d]	19.3%

Table 2
Estimated misclassification rates. Mean of 50 iterations of 10-fold cross-validations for each classification method.

However, a statistical difficulty is the choice of an analysis technique, which is able to model the given data adequately. The incorporation of classification trees in a medical context has been discussed by (21-23) and applied in the example of use of glaucoma diagnosis. Combining classifiers with bagging reduces misclassification errors in the direct and indirect approaches. The linkage of indirect classification with recent developments in the classification task, for example boosting, may lead to further improvements of indirect classifiers (cf. (24, 25)). All in all, in the demonstrated example estimated misclassification errors are reduced using indirect classification compared to similar direct approaches. This indicates the gain achieved by incorporating medical knowledge in statistical classification methods. Furthermore, it is demonstrated that the indirect classification approach offers the opportunity to structure the data. The use of the state of the art methodology bootstrap aggregating (bagging) improves indirect and direct classification. Hence, combining the indirect approach with bagging minimizes the misclassification result.

References

1. Hand DJ, Li HG, Adams NM. Supervised classification with structured class definitions. Computational Statistics & Data Analysis 2001; 36: 209-25.
2. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. California: Wadsworth; 1984.
3. Breiman L. Bagging predictors. Machine learning 1996; 24 (2): 123-40.
4. Breiman L. Arcing classifiers. The Annals of Statistics 1998; 26 (3): 801-49.
5. Coleman AL. Glaucoma. The Lancet 1999; 354: 1803-10.
6. Weih LM, Nanjan M, McCarty CA, Taylor HR. Prevalence and predictors of open-angle glaucoma: Results from the visual impairment project. Ophthalmology 2001; 108 (11): 1966-72.
7. Lee BL, Bathija R, Weinreb RN. The definition of normal-tension glaucoma. J Glaucoma 1998; 7 (6): 366-71.
8. Horn FK, Jonas JB, Korth M, Jünemann A, Grundler A. The full-field flicker test in early diagnosis of chronic open-angle glaucoma. Am J Ophthalmology 1997; 123 (3): 313-9.
9. Mardin CY, Horn FK, Jonas JB, Budde WM. Preperimetric glaucoma diagnosis by confocal scanning laser tomography of the optic disc. Br J Ophthalmology 1999; 83 (3): 299-304.
10. Swindale NV, Stjepanovic G, Chin A, Mikelberg FS. Automated analysis of normal and glaucomatous optic nerve head topography images. Investigative Ophthalmology Visual Science 2000; 41 (7): 1730-42.
11. Heidelberg-Engineering. Heidelberg Retina Tomograph: Bedienungsanleitung Software version 2.01. Heidelberg: Heidelberg Engineering GmbH; 1997 .
12. Freund Y, Schapire R. Experiments with a new boosting algorithm. Machine Learning: Proceedings of the Thirteenth International Conference. 1996: 148-56.
13. Schiavo RA, Hand DJ. Ten more years of error rate research. International Statistical Review 2000; 68 (3): 295-310.
14. Therneau TM, Atkinson EJ. An introduction to recursive partitioning using the rpart routine. Technical Report 61. Rochester: Section of Biostatistics, Mayo Clinic; 1997.
15. Ihaka R, Gentleman R. R: A Language for data analysis and graphics. J Computational Graphical Statistics 1996; 5: 299-314.
16. Eid TM, Spaeth GL, Katz LJ, Azuara-Blanco A, Agusburger J, Nicholl J. Quantitative estimation of retinal nerve fiber layer height in glaucoma and the relationship with optic nerve head topography and visual field. J Glaucoma 1997; 6 (4): 221-30.
17. Iester M, Mikelberg FS, Courtright P, Drance SM. Correlation between the visual field indices and heidelberg retina tomograph parameters. J Glaucoma 1997; 6 (2): 78-82.
18. Kamal DS, Viswanathan AC, Garway-Heath DF, Hitchings RA, Poinoosawmy D, Bunce C. Detection of optic disc change with the Heidelberg retina tomograph before confirmed visual field change in ocular hypertensives converting to early Glaucoma. Br J Ophthalmology 1999; 83 (3): 290-4.
19. Hammond P, Hutton TJ, Nelson-Moon ZL, Hunt NP, Madgwick AJA. Classifying vertical facial deformity using supervised and unsupervised learning. Methods Inf Med 2001; 40: 365-72.
20. Cox DR, Wermuth N. Multivariate dependencies. Models, analysis and interpretation. London, UK: Chapman & Hall; 1996.
21. Dusseldorp E, Meulman JJ. Prediction in medicine by integrating regression trees into regression analysis with optimal scaling. Methods Inf Med 2001; 40: 403-9.
22. Lausen B, Sauerbrei W, Schumacher M. Classification and regression trees (CART) used for the exploration of prognostic factors measured on different scales. In: Dirschedl P, Ostermann R, editors. Computational Statistics. Heidelberg: Physica-Verlag; 1994, 483-96.
23. McQuatt A, Sleeman D, Andrews PJD, Corruble V, Jones PA. Discussing anomalous situations using decision trees: A head injury study. Methods Inf Med 2001; 40: 373-9.
24. Hothorn T, Lausen B. Double-Bagging: Combining classifiers by bootstrap aggregation. Pattern Recognition 2002; 36 (6): in press.
25. Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. J American Statistical Association 2002; 97: 77-87.

Correspondence to:
Dr. Olaf Gefeller
Department of Medical Informatics
Biometry and Epidemiology,
Friedrich-Alexander-University Erlangen-Nuremberg
Waldstrasse 6
D-91054 Erlangen, Germany
E-mail: olaf.gefeller@rzmail.uni-erlangen.de