Effect of Different System Thresholds on the Performance of the Medical Expert System CADIAG-II/COLON

A. Rappelsberger [1], K.-P. Adlassnig [1], C. Lagor [2], W. Scheithauer [3], G.-V. Kornek [3]

[1] Department of Medical Computer Sciences, Section on Medical Expert and Knowledge-Based Systems, University of Vienna Medical School, Spitalgasse 23, A-1090 Vienna, Austria, e-mail: andrea.rappelsberger@akh-wien.ac.at

[2] Department of Medical Informatics, The University of Utah, Salt Lake City, USA, e-mail: Charles.Lagor@m.cc.utah.edu

[3] Department of Internal Medicine I, Division of Oncology, University of Vienna Medical School, Währinger Gürtel 18-20, A-1090 Vienna, Austria

Background

CADIAG-II (Computer-Assisted DIAGnosis, version 2) is a computer-assisted consultation system to support the differential diagnostic process in internal medicine. It is able to propose diagnoses based on patient's given symptoms, signs, and test results and—if possible—to confirm or exclude them. If necessary, suggestions for subsequent medical investigations are offered. Every step of this consultation system is explained in a detailed manner. CADIAG-II/COLON deals with the subarea of colon diseases; it contains 37 diseases considering 436 signs of illness. 2200 relationships between these signs and diseases were incorporated into the knowledge base, from that 1152 for the characterization of the frequency of occurence of a symptom with a disease and 1048 as value for the strength of confirmation for a disease. Two thresholds A and B, respectively, determine whether a documented relationship for the frequency of occurrence and strength of confirmation, respectively, will be considered in the inference process. By doing this weak relationships may be disregarded. A further threshold (H) determines the extent of hypothesis generation itself. The threshold H has to be reached by at least one strength of confirmation value of a patient's symptom to a diagnosis to have this diagnosis be generated as hypothesis.

Objective

The aim of the present study [1] was to evaluate CADIAG-II/COLON's hypothesis generation with different values of A, B, and H and to determine whether an optimal setting may be recommended to the user.

Material and Methods

The study included 103 cases wih a total of 119 clinical diagnoses of colon diseases. CADIAG-II/COLON generated a total of 30 hypotheses lists for each case making use of six different settings for A and B and five for H, respectively. The basic evaluation consisted of comparing the hypotheses lists with the respective diagnoses. Four different decision criteria defining a true positive result were employed for interpretation. In order to obtain information on the general performance of the system receiver operating characteristic (ROC) curves were generated and kappa indices were calculated.

Results

The results confirm trends which were already described in a similar study on CADIAG-II/PANCREAS [2]. The increase of true positive and false positive ratios as a result of the decrease of H could be verified. Once again the tendency towards worsening the sys-

tem's performance by using increasingly strict chosen decision criteria could be observed. Evaluation revealed that the threshold values are interdependent as A and B determine a lower threshold for H (and vice versa). Using the most liberal settings of thresholds (H=0, A=0.01, B=0.01) CADIAG-II/COLON generated 94% true positives, whereby the correct hypothesis ranked first in 53% of cases.

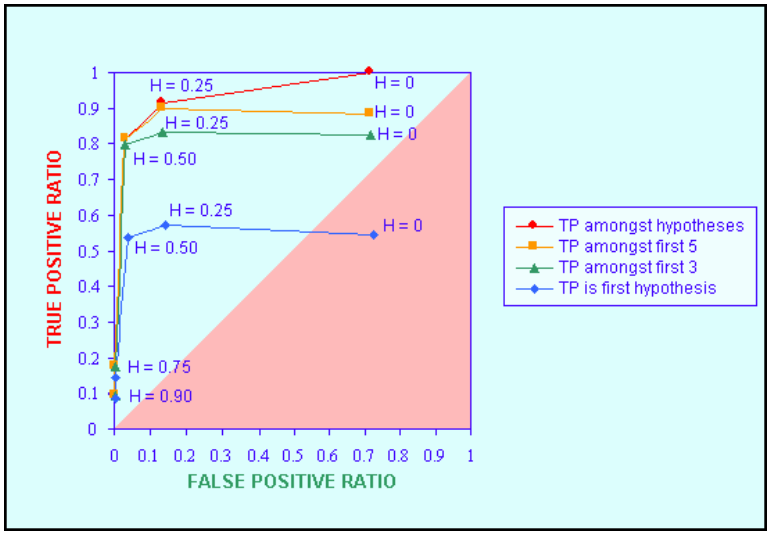

Fig. 1: Effect of different threshold values

Technical Specification

The online consultation system CADIAG-II is programmed in CICS/VSE command level language and PL-I. It is embedded in the time-sharing environment of the medical information system WAMIS [3] and thus available to some selected clinics and institutes of the Vienna General Hospital. It runs on an IBM 2003 under VSE/ESA and uses several VSAM index-sequential files to store not only the knowledge base of CADIAG-II but also the patient data, which are collected via the integrated routine medical documentation and laboratory system.

Conclusion

According to the ROC curves an optimal performance may be achieved with the most liberal definition of a true positive value (hit if clinical diagnosis is among the generated hypotheses) and threshold settings at H=0.25, A=0.01, and B=0.01 (Fig. 1). Kappa statistics showed a fair agreement in this range. Nevertheless it seems to be advisable not only to adapt the choice of threshold values to the respective clinical situation and the resulting need for sensitivity and specifity out of it, but also to consider the individual threshold settings of possibly single suspected diseases.

References

[1] Lagor, C., Adlassnig, K.-P., Kornek, G.-V. & Scheithauer, W. (1998) Optimal Threshold Settings in CADIAG-II/COLON. In Trappl, R. (Ed.) *Cybernetics and Systems'98*, Austrian Society for Cybernetic Studies, Schottengasse 3, A-1010 Vienna, Austria, 163-168.

[2] Adlassnig, K.-P. & Scheithauer, W. (1989) Evaluation of Medical Expert Systems Using ROC Curves. *Computers and Biomedical Research,* 22, 297-313.

[3] Grabner, G. (1985) (Hrsg) *WAMIS – Wiener Allgemeines Medizinisches Informations-System,* Springer-Verlag, Berlin.