Uses of Diagnostic Expert Systems in Clinical Care

Peter J. Haug MD., University of Utah Department of Medical Informatics

The development and testing of computerized systems to assist in the diagnostic process is a time honored research activity in medical information science. The focus of the majority of the applications produced is on providing accurate diagnostic suggestions when appropriate clinical information is entered. We believe that diagnostic knowledge has a much wider range of uses than that of simply assigning diagnostic labels. Below we describe three applications which illustrate alternate uses for diagnostic systems. Applications that assist in data collection, assess the quality of medical reports, and extract relevant clinical data from natural language x-ray reports are discussed. We believe that more effort should be directed toward studying the use of diagnostic knowledge bases in processes that help plan diagnostic strategies, in quality assurance applications, and in processes that facilitate all aspects of medical communication.

INTRODUCTION

In 1959, Ledley and Lusted described the application of tools from the realm of symbolic logic and statistical pattern recognition to problems in medicine [1]. They proposed that these tools be used to assist in the diagnostic process and in other problems involving medical decision making. Computer systems were the enabling technology that would bring these tools to the bedside.

Since their key paper, a variety of investigators have produced systems employing different models to diagnose, prognose, and to assist in the medical decision-making process. However, with some notable exceptions, the low capacity and high expense of computing equipment have relegated these efforts to research settings. With the development of economical computer hardware and a growing understanding of the environment in which these applications must exist, an opportune time has arrived to bring diagnostic technologies into the realm of practical medicine.

The premise of this paper is that applications involving diagnostic computing systems must not be limited to diagnosing. A significant portion of the value of these tools lies in the power of well designed models of the diagnostic process to participate in a wide variety of clinical tasks. Physicians clearly exercise their diagnostic knowledge not only when they assign a diagnostic label to a patient, but also during processes as diverse as reading medical reports and critiquing the clinical behavior of their peers. Below we explore a small subset of the range of applications in which diagnostic knowledge can productively participate. In each case, we give an example of a computer program incorporating a diagnostic knowledge base that has already demonstrated useful or interesting behavior.

USES OF DIAGNOSTIC KNOWLEDGE

A variety of applications have been produced that illustrate the use of diagnostic knowledge for purposes outside of the merely diagnostic realm. Unfortunately few of these alternate uses have been well studied. Below we present examples of three applications designed to use diagnostic knowledge in various medical realms. The programs presented were designed to 1) assist in data collection, 2) assess the quality of medical reports, and 3) extract relevant clinical data from natural language reports. These examples come from the HELP Hospital Information System. Our long term goal is to implement large scale diagnostic knowledge bases using a single knowledge representation. We are experimenting with diagnostic models that will support not only the applications described here but also a variety of others whose unifying characteristic is their ability to use diagnostic logic for both diagnostic and non-diagnostic tasks.

1) Assisting Data Collection

Processes that assist in identifying the information necessary to pursue a diagnosis are well described in the medical literature. They are commonly associated with diagnostic programs and typically attempt to direct the process of capturing diagnostic information. Well known examples of programs that can direct queries to a physician user are the INTERNIST [2], DXPLAIN [3], and ILIAD [4] diagnostic consulting programs. Each of these applications incorporates a model of the clinical data gathering process. Unfortunately, the effectiveness of such programs in suggesting which clinical findings to collect has seldom studied. Below we review a comparative study designed to explore whether an interactive history-taking process based on a branching questionnaire would be more effective if driven by a diagnostic knowledge base.

Efforts to direct data collection in the HELP system have concentrated on the patient history. The goal has been to identify tools that could effectively collect a medical history appropriate for use in diagnostic decision support applications. While earlier efforts focused on history appropriate to a wide variety of diseases [5], more recent efforts have focused on acquiring data bearing on pulmonary diseases [6,7]. The results of two studies of these processes are combined in figure 1 to illustrate the effectiveness of these methods.

Three techniques for selecting questions were explored. The first is a simple branching questionnaire. This approach takes full advantage of the hierarchical relationship between more and less specific questions. For instance, if the question "Have you had chest pain with this illness?" is answered "Yes" then more specific questions such as "Is your chest pain brought on by exertion?" will be asked. Alternately, if the answer to the first question is "No", the more specific questions will not be asked.

The second technique is called decision-driven data acquisition (DDA). In this technique, a frame-based expert system analyses all data available at any point in the patient interview. The individual disease frames determine which additional information they need to evaluate the likelihood of the disease which they represent. Each frame proposes one or more questions. From this list a supervisory process selects a group of five questions which are then presented to the patient. The system passes through this cycle multiple times until criteria are met indicating that no more data is needed.

This system is also aware of the hierarchical relationship between questions. The selection process mediated by the expert system can be seen as an additional filtering process that allows only questions with demonstrable diagnostic relevance to be asked.

A third approach has also been tested. It is similar to the DDA method accept that it has been adapted for use in a setting where the patient cannot come to a terminal. This approach begins when a paper questionnaire containing 20 screening questions is presented to a patient. The answers are entered into the computer and the diagnostic frames are run. They submit their requests for additional data to a filtering process which scores these questions and selects from 0 to 40 of them to print on a second questionnaire. Once the patient has answered this second questionnaire, the process is complete.

The branching questionnaire mode of data collection and the DDA mode were tested in inpatients at the LDS Hospital. Fifty patients took a DDA manage history and 23 received a history managed by the branching questionnaire program. Figure 1 illustrates the results. On average, the DDA mode took a significantly ($p < 0.05$) shorter time to run (8.2 minutes) and asked significantly fewer questions (48.8 questions) than did the branching questionnaire (19.2 minutes and 137 questions, respectively). The two-stage paper questionnaire was tested separately on patients coming to the x-ray department for chest x-rays. It appeared to perform similarly to the interactive DDA mode. It should be noted that there was no significantly difference among the techniques in terms of diagnostic accuracy. Using history alone, all three succeeded in placing the patient's correct disease in a five member differential diagnostic list from 70% to 88% of the time.

2) Assessing The Quality of Medical Reports

A second example of an alternative use of diagnostic knowledge comes from a study of result reporting in the radiology department. The central goal of this project was to develop a technique for measuring the quality of x-ray reporting without requiring the review of radiographs by multiple radiologists. This is in contradistinction to typical approaches for evaluating the accuracy of radiologists. Typically, audit procedures in the radiology department require multiple readings of a select set of x-rays [8,9,10,11,12]. The results of the repeated readings are used to define a gold standard for the films. Then the individual radiologists are compared to this standard.

The technique developed as a part of this project is based on a simple premise. Each examination is a test of the radiologist's accuracy. Instead of comparing the abnormalities reported to a standard formulated through multiple readings, the description in the report is evaluated in comparison to the patient's overall diagnostic outcome. In the case of chest x-rays the standard is the list of final diagnoses attached to the patient's record at the time of discharge. The report generated by the radiologist is successful to the extent that it supports the process that led to these discharge diagnoses.

While a variety of algorithms can be used to link the findings represented in the x-ray report to the final diagnosis, we have demonstrated the success of a variation on Shannon Information Content in discriminating among physicians reading chest x-rays. Our evidence comes from two studies. In the first we used expert systems technologies to demonstrated discrimination in a controlled experiment [13]. In this experiment five x-ray readers read an identical set of 100 films. The assessment driven by diagnostic logic gave results consistent with the differing expertise of the readers and similar to the results of a more standard audit procedure.

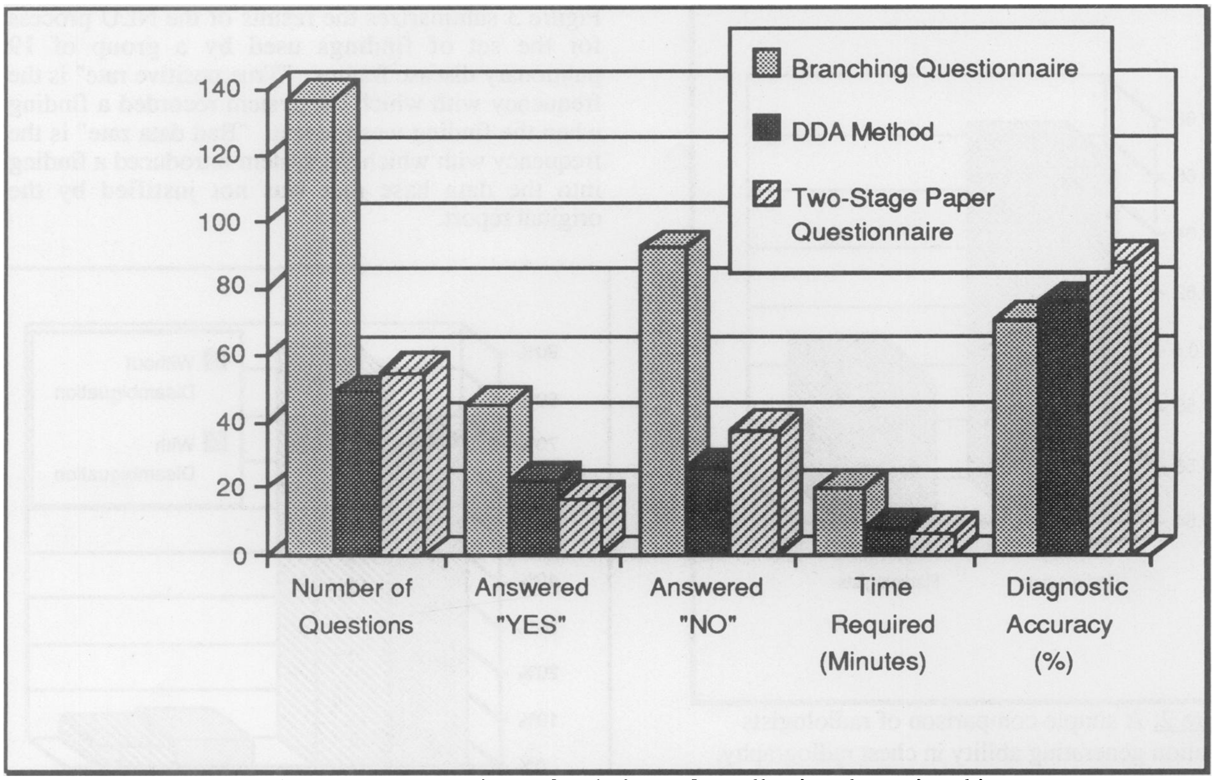

Figure 1: A comparison of techniques for collecting the patient history.

In a second study of this audit technique, we extended the test environment into the realm where we hope to use it clinically. We tested a group of radiologists following their standard procedure for interpreting radiographs. Each chest x-rays was reported only once as a part of the radiologist's daily work. The goal of this study was to test the ability of a knowledge based approach to measure the quality of x-ray reporting without requiring repeated reporting of the radiographs.

This technique uses a measure called the Outcome-Directed Information Content (ODIC) to assess the amount of information any x-ray report contributes to the process of recognizing the diseases that a patient has. The ODIC is described in detail elsewhere [13]. It is a modified version of the Shannon information content measure and is designed to assess both the positive information contributed by x-ray findings relevant to a patient's disease and the negative information contributed by findings which do not apply to any of the patient's illnesses. X-ray readers can be compared based on the bits of information produced.

Figure 2 illustrates a simple experiment where 651 chest x-ray reports generated by a group of radiologists are compared to the patients' discharge diagnoses using the ODIC. The radiologists were grouped according to whether they had additional (post residency) training in chest radiology. The "Trained" radiologists produced 11% more information than the "Untrained" radiologists (0.664 bits as opposed to 0.589 bits, significant at $p < 0.005$).

The average ODIC successfully discriminates these groups. However, it is an overall measure. Examination of the interaction between the groups of radiologists and disease subgroups indicates that the ODIC can also discriminate at the level of different diseases ($p > 0.05$). This suggests that the ODIC can not only discriminate overall quality of x-ray interpretation, but it may also be of use to pinpoint the specific diseases for which an individual radiologist is failing to generate effective information.

3) Extracting Coded Data From Medical Reports

A final example of alternate uses of diagnostic decision logic comes from the realm of natural language understanding (NLU). Here, we have explored the use of the knowledge embedded in a frame-based, diagnostic system to encode and place in a clinical data base significant facts represented in the free-text of x-ray reports [14].

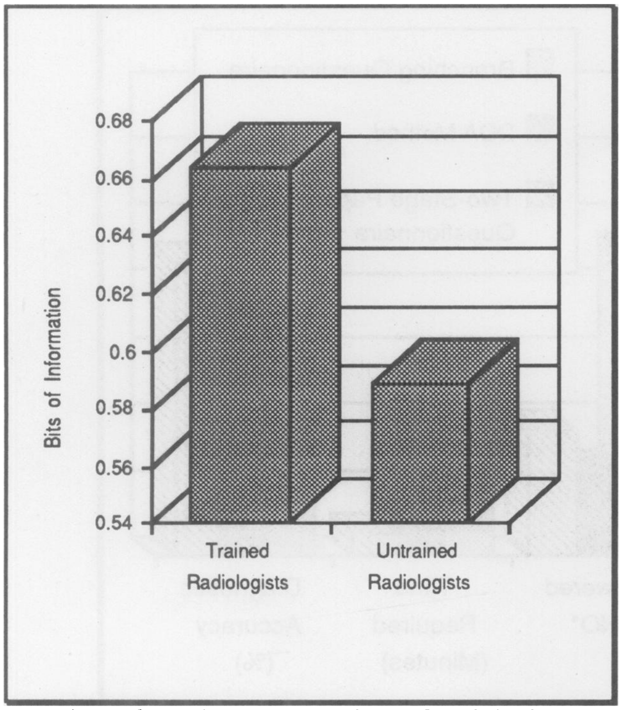

Figure 2: A simple comparison of radiologists information generating ability in chest radiography.

The diagnostic system provides two services to the NLU system. First it defines the data elements that the language processor understands. These typically consist of finding/location pairs. For instance, in the frame for pneumonia, data structures are specified for "localized infiltrate" in the "right upper lobe", "right middle lobe", "right lower lobe", "left upper lobe", "left lower lobe", or "lingula". The language system searches for these concepts (using an appropriate thesaurus to recognize synonyms) in sentences in the x-ray report. When it can match the concepts in an acceptably discreet utterance, it will create a data base record containing appropriate codes from the HELP data dictionary.

While this simple semantic approach is successful a portion of the time, the complex character of natural language can defeat the process. If the system can find only part of the information necessary to create a data record, the element found is marked as ambiguous and a process is invoked that attempts to disambiguate the statement. This process uses the expert system directly to determine whether the proper combination of concepts can be inferred from the information available in the rest of the report. If, for instance, the radiologist has indicated that pneumonia is present, then the system may be able to infer that the infiltrate mentioned is a localized rather than a generalized infiltrate by analyzing the effect of the alternative interpretations on the probability of this disease.

Figure 3 summarizes the results of the NLU process for the set of findings used by a group of 19 pulmonary disease frames. "True positive rate" is the frequency with which the system recorded a finding when the finding was present. "Bad data rate" is the frequency with which the system introduced a finding into the data base that was not justified by the original report.

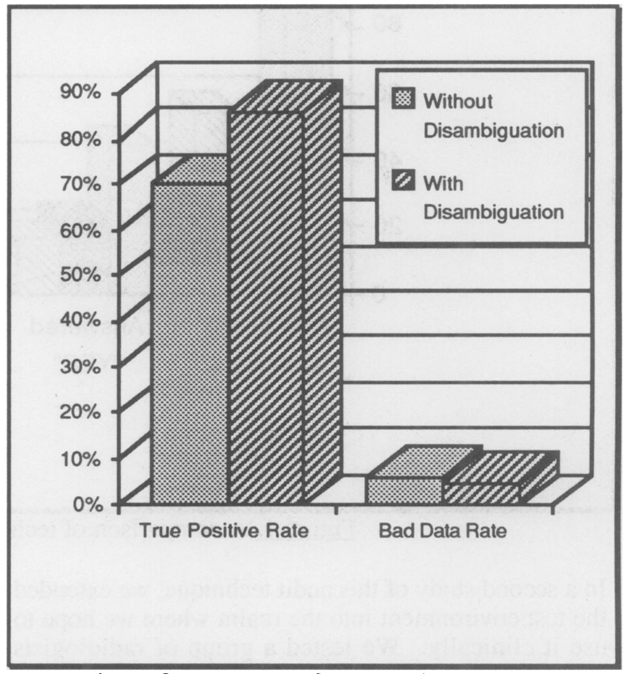

Figure 3: Accuracy of a Natural Language Understanding System with and without disambiguation.

By using the expert system as a source of context, the NLU system was able to improve its success in recognizing and encoding diagnostically important clinical findings. The disambiguation process significantly increased the true positive rate (70% to 86%) while leaving the bad data rate at an acceptable level (6% to 5%).

DISCUSSION

The three examples discussed above are part of a much larger set of applications that can benefit from the information stored in diagnostic knowledge bases. To the extent that human cognition provides a model, a knowledge of the relationship between diseases and clinical findings can support a variety of useful behaviors. The medical expert applies his diagnostic knowledge every day for many tasks besides assigning disease states to his patients. This fund of expertise drives the process of seeking new clinical information, allows critical assessment of the quality of available information, and is the foundation for much of the communication a physician has with his peers.

If researchers can extend their ability to use computerized medical knowledge into these same realms, the value of the diagnostic systems which they develop will increase dramatically. If, however, medical expert systems continue to be designed and studied principally as assistants in the process of assigning a diagnostic label, their impact on medical care will remain modest at best. We believe that the development of true, multi-use, diagnostic knowledge bases is one of the keys to bringing the fruits of the information sciences to medicine. Applications that use them will function in the diagnostic planning, quality assurance, and clinical communication systems of the future.

Supported in part by Grant RO1LM05323 from the National Library of Medicine.

References

[1]. Ledley RS, Lusted LB. Reasoning Foundations of Medical Diagnosis. Science (1959); 130:9-21.

[2]. Miller RA, Pople HE, Myers JD. INTERNIST-I, An experimental computer-based diagnostic consultant for general internal medicine. NEJM (1982); 307:468-476.

[3]. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. DXplain - An evolving diagnostic decision support tool. JAMA (1987); 258:67-74.

[4]. Lincoln, MJ, Turner CW, Haug PJ, Warner HR, Williamson JW, Bouhaddou O, Jessen SG, Sorenson D, Cundick RC, Grant Morgan. Iliad Training Enhances Medical Students' Diagnostic Skills. Journal of Medical Systems (1991); 15:93-110.

[5]. Warner HR, Rutherford BD, and Houtchens B. A Sequential Bayesian Approach to History Taking and Diagnosis. Computers and Biomedical Research. (1972); 5:256-262.

[6]. Haug PJ, Warner HR, Clayton PD, Schmidt CD, Pearl JE, Farney RJ, Crapo RO, Tocino I, Morrison WJ, and Frederick PR. A Decision-Driven System to Collect the Patient History. Computers and Biomedical Research (1987); 20:193-207.

[7]. Haug PJ, Rowe KG, Rich T, Farney R, Pearl J, Schmidt CD, Clayton PD. A Comparison of Computer-Administered Histories. Proceedings AAMSI Congress (1988); 21-25.

[8]. Herman PG, et al. Disagreements in chest roentgen interpretation. Chest (1975); 68:278-282.

[9]. Yerushalmy J. Reliability of chest radiology in the diagnosis of pulmonary lesions. Am J Surg. (1955); 89:231-240.

[10]. Koran LM. The reliability of clinical methods, data, and judgments (second of two parts). NEJM (1975); 293:695-701.

[11]. Rhea JT, Potsaid MS, DeLuca SA. Errors of interpretation as elicited by a quality audit of an emergency radiology facility. Radiology (1979); 132: 277-280.

[12]. Baines CJ, McFarlane DV, Wall C. Audit procedures in the national breast screening study: Mammography interpretation. J Can Assoc Radiol (1986); 37:256-260.

[13]. Haug PJ, Clayton PD, Tocino I, Morrison JW, Elliot CG, Collins DV, and Frederick PR. Chest Radiography: A Tool for the Audit of Report Quality. Radiology (1991); 180:271-276.

[14]. Haug PJ, Ranum DL, and Frederick PR. Computerized Extraction of Coded Findings from Free-Text Radiology Reports. Radiology (1990); 174:543-548.