Computer Implementation of the New Diabetes Classification Scheme

(From the Faculty of Medicine, Technion, the Israel Institute of Technology and the Diabetes Service and Research Unit, Rambam Medical Center Hatifa, Israel)

F. Wiener, Y. Kanter, S. Kol, D. Barzilai

The classification of diabetes and other categories of glucose intolerance developed by the international workgroup sponsored by the National Diabetes Data Group of the NIH has been implemented on the computer, using a system for the simulation of medical reasoning. The system allows the physician to formulate his medical knowledge as a series of logical inferences and to specify those combinations of clinical findings which confirm or reject each inference. Alphabetical lists of the texts of medical statements appearing in the knowledge base are used to specify patient data. A patient status report is produced by comparing the patient data with the medical logic. The report lists patient data, the inferences confirmed and, if necessary, requests additional data to complete the classification. The system has enabled us to computerize the new classification scheme in a relatively short time and with a minimum of effort. The system has correctly classified nearly 100 patients on our ward and in the endocrinology clinics using data routinely recorded. The computer evaluation is now an integral part of the patient record. It assures that the new classification scheme is applied to all patients in an unbiased and consistent manner.

Key-Words: Diabetes Classification, Computer Simulation of Medical Reasoning, Medical Logic Modules, Diagnosis by Logical Inference, Symptom Combinations

COMPUTERIMPLEMKIERTIUNG DES NEUEN KLAFLIKATIONSSCHEMAS FÜR DIABETES


Schlüssel-Wörter: Klassifikation des Diabetes, Computersimulation medizinischen Denkens, medizinische Logikmodule, Diagnose durch logische Schlußfolgerung, Symptomenkombinationen

Introduction

The rapid increase of knowledge regarding the etiology and pathogenesis of diabetes has led to several recent attempts by both individual authors [3–8] and professional associations [1, 2] to organize these data into a consistent and universally accepted set of criteria for the diagnosis and classification of diabetic disorders.

Diverse criteria, reflecting different clinical approaches, are used to diagnose diabetes. These include glucose levels in blood and urine, abnormalities in glucose tolerance tests, changes in serum lipids, presence or absence of antibodies to the pancreatic islets of Langerhans, hemoglobin A1c levels, and basal membrane thickness measurements. Inconsistencies among the various approaches to evaluating the diabetes patient as well as differences in the formulation of commonly held diagnostic criteria have made it urgent to set up committees to define a new set of criteria for classifying diabetes. The recommendations of the international workgroup sponsored by the National Diabetes Data Group of the National Institutes of Health have recently been published [1]. These recommendations were accepted by the American Diabetes Association as well as by other international bodies including the Expert Committee on Diabetes of the WHO. The European Association for the Study of Diabetes (EASD) has made similar recommendations [2] except for small differences in the glucose levels associated with the various diagnostic categories.

In this paper we report a computer implementation of the new classification scheme proposed by the National Diabetes Data Group. Although patients can readily be classified under this scheme without the aid of a computer, we felt that automatic classification of patients from the data routinely recorded on the ward and in the clinics requires no additional effort on the part of the medical staff and can thus assure more uniform and objective results. Inclusion of the diabetes classification in the computerized case summary also helps to familiarize the new scheme to all those who have access to the patient's chart.
Methods

The new classification scheme for categorizing diabetic patients has been formulated for the computer with the aid of a system for medical reasoning [9]. The system consists of a series of FORTRAN programs which allow formulation of the medical logic (EDIT), verification of the medical statements used in the logic (LIST), construction of data acquisition protocols (PTCL), and calculation of patient data (TEST).

Formulation of the medical knowledge for the computer begins by defining decision pathways for each issue to be covered by the system. These pathways may be independent, interconnected by common items of patient data, or mutually exclusive. For the diabetes classification scheme the following pathways are defined:

1. Diabetes, clinical
   - Diabetes, laboratory
2. Diabetes, secondary
   - Diabetes due to pancreatic problems
   - Diabetes due to hormonal disorders
   - Diabetes due to drugs and chemical agents
   - Diabetes due to insulin receptor abnormalities
   - Diabetes due to genetic syndromes
   - Diabetes due to nongenetic causes
3. IDDM (Insulin-dependent diabetes mellitus)
4. NIDDM (Non-insulin-dependent diabetes mellitus)
   - NIDDM, nonobese
   - NIDDM, obese
5. Gestational diabetes
6. Glucose tolerance, normal
   - Nondiagnostic elevated FPG (Fasting plasma glucose)
   - Probably abnormal FPG
   - Impaired glucose tolerance
   - Impaired glucose tolerance, obese
   - Impaired glucose tolerance, nonobese
7. Previous abnormality of glucose tolerance
8. Potentially abnormal GT
   - Potentially abnormal GT, obese

Each pathway consists of one or more inferences to be drawn from patient data. The inferences may be sequential as in NIDDM, or parallel as in secondary diabetes.

Diabetes secondary to infectious diseases is not included in the decision pathway for secondary diabetes, since it was not mentioned in the National Diabetes Data Group's classification [1]. Although infectious diseases may lead to the early appearance of diabetes due to the physiological stress caused by the infection, most people with infectious disease do not develop diabetes. In the special case of pancreatic problems, pancreatitis is included as a criterion since it may lead to pancreatic beta cell involvement which can cause diabetes.

The conditions for confirming or rejecting an inference are defined in a medical logic module (Figures 1 and 2). These conditions are logical combinations of patient data and, if necessary, of previously confirmed inferences. The combinations sufficient for confirmation, rejection, or implication (which generates a request for additional data) are expressed by assigning weights to each factor in the module such that each sufficient combination will have a total of 50 or more points. The points contributed by each condition towards each of these decision functions is given in the right, left, and middle columns, respectively.

The medical logic modules for defining diabetes are shown in Figure 1. Clinical diabetes is defined as an unequivocal elevation of plasma glucose together with any one of the four clinical findings listed. Laboratory diabetes is defined by a positive answer to any one of the three conditions stated, viz. 2 or more measurements of FPG greater than 140 mg/dl, or 2 or more measurements of FG in whole blood greater than 120 mg/dl, or 2 or more abnormal oral GTT results. Either laboratory or clinical diabetes confirm the inference DIABETES. This inference is also confirmed by an impaired glucose tolerance in the presence of renal or retinal microvascular damage. The next module defines sec-

![Figure 1: The medical logic for confirming the presence of diabetes. Each medical logic module states the inference under consideration, followed by statements of patient data or previous inferences required to confirm or reject the inference. The module numbers preceding each module are assigned by the system according to the scheme described in the text. Each medical statement is categorized by its clinical (SM, SG, HX, LB) or logical (ID, DS) type. The statements are in free text and may be continued on as many lines as needed](image-url)
Figure 2: The logical pathway for differentiating between IDDM and NIDDM and for separating NIDDM into subclasses. The three columns of numbers express the conditions for rejecting, implicating, and confirming each inference. A conclusion is reached when a total of 50 or more points have been accumulated by the presence (Y) or absence (N) of the diagnostic criteria. The implicate function generates a request for additional data when the inference has not been confirmed or rejected.

<table>
<thead>
<tr>
<th>201</th>
<th>ID</th>
<th>ICDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>ICDM</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>IDY</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>10Y</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>DIADESES</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>DIADESES, SECONDARY</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>32Y</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>17Y</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>17Y</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>17Y</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>17Y</td>
</tr>
</tbody>
</table>

Secondary diabetes as diabetes in the presence of endocrinopathies or other hyperglycemic conditions. For the diabetes to be secondary, however, it must be directly attributable to the endocrinopathy or the hyperglycemic condition.

If the diabetes is judged to be primary (see Fig. 2), IDDM is diagnosed if insulinaemia is present together with the finding that the patient is ketosis-prone under basal conditions. If the patient is not ketosis-prone under basal conditions, the patient is classified as NIDDM, either obese or nonobese depending on the patient’s body weight.

Inferences are designated as ID (intermediate) or DS (final). The former are not printed in the patient status report (Figures 3 and 4) if they are part of a reasoning pathway in which a subsequent inference has been confirmed. Thus, for example, the inference DIABETES (Figure 1) is designated ID. It will not be printed in the patient status report if any of the subsequent modules DIABETES, SECONDARY, IDDM, or NIDDM, NONOBSE or OBSE are confirmed.

Conditions for requesting additional data in following up the patient may also be specified in the medical logic. If DIABETES has been confirmed, both IDDM and NIDDM are put in the implicated state (as specified by the 50 points in the middle column, Fig. 2). The patient status report (Figure 4) will then request that the conditions in the module whose state is unknown be investigated at the follow-up visit. If the module is rejected, as would be the case if DIABETES, SECONDARY has been confirmed (as specified by the 50 points in the left column, Fig. 2), the request for additional data is not made.

The foregoing shows that the system provides a method for programming medical knowledge using medical statements in natural language, organizing this knowledge into a series of inferences and using threshold logic to define the sufficient combinations of factors for confirming or rejecting each inference.

The medical logic modules are entered into the computer as input to program EDIT. The program assigns an item number to each medical statement in the module according to the following scheme: ID 1—200; DS 201—350; symptoms SM 601—800; signs SG 801—1000; history and progress notes HX 1001-1200; and diagnostic test results LB 1201—1400. The statement texts are collected into a text file and an index is created to access this file. The medical logic is converted into a decision matrix in the computer’s memory. The portion of the decision matrix corresponding to the modules shown in Figure 2 follows:

<table>
<thead>
<tr>
<th>201</th>
<th>4</th>
<th>0</th>
<th>50</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>1209</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>201</td>
<td>1004</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>201</td>
<td>5</td>
<td>50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>0</td>
<td>50</td>
<td>33</td>
</tr>
<tr>
<td>8</td>
<td>1004</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>50</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>202</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>202</td>
<td>803</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>203</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>33</td>
</tr>
<tr>
<td>203</td>
<td>802</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
</tbody>
</table>

Each row in the decision matrix represents a criterion statement in the module. The first column is the item number of the statement of the inference to be considered (e.g. 201—IDDM, 8—NIDDM), and the second column is the item number of the medical statement used as a criterion. The next 3 columns record the weight contributed by the criterion to each of the decision functions: rejection, implication, and confirmation, respectively. If the weight is positive, the points are contributed when the criterion is present (Y). A negative weight indicates that the points are contributed if the criterion is absent (N), e.g., the absence of item 1004—KETOSIS-PRONE UNDER BASAL CONDITIONS will contribute 17 point to confirm inference 8—NIDDM.

Program LIST produces an alphabetized list of the medical statement texts for each type (ID, DS, SM, SG, HX and LB) of statement. Each entry in the list consists of the statement identification number, its text, and the identification number of each module in which the statement appears.

The alphabetized lists, or, if more convenient, a questionnaire derived from these lists and produced by program PTCL are then used to specify case descriptions. Each case consists of a series of finding numbers and the finding state, Y if present, N if absent.

The cases are fed into the computer as input to program TEST. Patient data is evaluated by comparing it with the decision matrix (i.e., the medical logic modules) and accumulating weights towards each inference in sequence. If an inference is confirmed or rejected, its identification number and state (Y or N, respectively) are added to the
patient record and are thus available for analyzing subsequent modules. Program TEST then prints the patient status report which lists the clinical findings, the latest confirmed inference in each decision pathway, and each implicated inference with its corresponding request for additional data.

After printing the report, the system stores the patient data on the disk. When new data becomes available, it is appended to the patient file and a new status report is generated. Since only the latest entry for a given finding is considered in assessing the patient's status, a finding which was positive (Y) initially can be entered as negative (N) in the subsequent data (Figure 4).

Since the programs are written in FORTRAN, the system can be run on a computer of any size which supports FORTRAN and which has on-line storage on disk or diskettes. Versions of the system have been implemented on a number of IBM computers, on the Xerox Sigma 3, and on the PDP/11 series. A minimum of 24K memory is required. Cases can be entered either in batch mode or interactively at a computer terminal. Processing time for each case will depend on the size of the computer, but even with mini- or microcomputers no more than 2 to 3 seconds are required. The operating costs, therefore, are mainly for personnel to enter the data and to attach the output to the patient's chart.

Results

The patient status report for 2 illustrative cases is shown in Figure 3. The first case is a patient from the internal medicine follow-up clinic. The fasting glucose measurements confirm DIABETES, but the fact that the patient is on steroids and that the diabetes appeared in conjunction with the steroid administration leads to the conclusion that this is a DIABETES DUE TO DRUGS AND CHEMICAL AGENTS. The second case is from the endocrinology clinic. Since the patient's thyrotoxicosis is inactive and the present diabetes is not associated with an endocrinopathy or any other cause of hyperglycemia, the diabetes is primary. The presence of insulinopenia and the fact that the patient is ketosis-prone under basal conditions classifies this case as an IDDM.

The third case, shown in Figure 4, is a patient seen in the diabetes clinic. Although the patient was on oral hypoglycemics initially, this finding is set to N subsequently since insulinopenia developed and the patient is now being treated with insulin. The current data, however, is sufficient only to confirm the presence of diabetes. The patient report requests data on whether the patient is ketosis-prone under basal conditions in order to classify him as either IDDM or NIDDM. In the next report, after it was established that the patient is not ketosis-prone, the patient is classified as NIDDM, NONOBES.

Discussion

The fact that the programs which comprise the system for simulating medical reasoning are independent of medical content makes the system applicable to a wide range of medical problems. Using this system, medical logic has been formulated and successfully applied to patient data in investigating the coronary patient [10], in tumor classification and treatment selection in breast cancer [11] and in the assessment of nephrologic diseases [12]. Since knowledge of computer programming is not necessary for using the system, it can be readily used by physicians and medical students to investigate their particular area of interest.

The basic philosophy underlying our system that medical diagnosis is a logical rather than a statistical process, together with our implementation of WORTHY's observation [13] that the physician is not satisfied to merely eliminate contending hypotheses but also requires positive identification of the patient's disorders, has made the system

\begin{verbatim}
PATIENT WL

CLINICAL FINDINGS
1201 Y FASTING GLUCOSE, W/CAF, HLD > 120 MGL/DL, 2X OR MORE
1203 Y HYPERGLYCEMIA
1200 Y ENDOCRINE PATHELE ASSESS HYPERGLYCEMIC COND/SYNDRONE PRESENT
1303 Y DIABETES ASSOCIATED WITH ENDOCRINE PATHELE ASSESS HYPERGLYCEMIA

BASED ON YOUR FINDINGS, THE FOLLOWING INFERENCES CAN BE MADE
206 DIABETES DUE TO DRUGS AND CHEMICAL AGENTS

PATIENT TF

CLINICAL FINDINGS
1016 Y THYROTOXICOSIS, NON-ACTIVE
1206 Y FASTING GLUCOSE, W/CAF, PLASMA > 140 MGL/DL, 2X OR MORE
1205 Y INSULINOPENIA
1203 Y DIABETES ASSOCIATED WITH ENDOCRINE PATHELE ASSESS HYPERGLYCEMIA
1205 Y KETONEURIA
1004 Y KETOSIS-PRONE UNDER BASAL CONDITIONS
1004 Y INSULIN ADMINISTRATION (40 UNITS PER DAY AND LESS)
803 Y BODY WEIGHT WITHIN NORMAL RANGE
1001 N RENAL OR RETINAL MICROVASCULAR DAMAGE

BASED ON YOUR FINDINGS, THE FOLLOWING INFERENCES CAN BE MADE
201 IDDM
\end{verbatim}

Figure 3: Patient Status Report. The text of each clinical finding is printed together with its identification number (as assigned by the system, cannot) and its state. Y = present; N = absent. These, followed by the conclusions drawn by system
CLINICAL FINDINGS
1200 Y FASTING GLUCOSE, VENOUS PLASMA > 140 MG/DL, 2 X OR MORE
1039 Y PATIENT ON A LOW SUGAR DIET
1040 Y ORAL HYPOGLYCEMICS ADMINISTRATION
1002 N ENDOCRINOLOGY OR OTHER HYPERGLYCEMIC CONDITIONS/SYNDROME PRESENT
603 Y BODY WEIGHT WITHIN NORMAL RANGE
1001 Y RENAL OR RETINAL MICROVASCULAR DAMAGE

SUBSEQUENT FINDINGS
1040 N ORAL HYPOGLYCEMICS ADMINISTRATION
1206 Y INSULIN RESISTANCE
1042 Y INSULIN ADMINISTRATION (21 TO 40 UNITS PER DAY)

BASED ON YOUR FINDINGS, THE FOLLOWING INFERENCES CAN BE MADE

4 DIABETES

FOR PATIENT FOLLOWUP, OBTAIN THE DATA LISTED BELOW EACH ITEM

201 (40) IDDM
1004 (10) KETOSIS-PHORE UNDER BASAL CONDITIONS

8 (23) NIDDM
1004 (17) KETOSIS-PHORE UNDER BASAL CONDITIONS

CLINICAL FINDINGS
1206 Y FASTING GLUCOSE, VENOUS PLASMA > 140 MG/DL, 2 X OR MORE
1039 Y PATIENT ON A LOW SUGAR DIET
1040 Y ORAL HYPOGLYCEMICS ADMINISTRATION
1002 N ENDOCRINOLOGY OR OTHER HYPERGLYCEMIC CONDITIONS/SYNDROME PRESENT
603 Y BODY WEIGHT WITHIN NORMAL RANGE
1001 Y RENAL OR RETINAL MICROVASCULAR DAMAGE

SUBSEQUENT FINDINGS
1040 N ORAL HYPOGLYCEMICS ADMINISTRATION
1206 Y INSULIN RESISTANCE
1042 Y INSULIN ADMINISTRATION (21 TO 40 UNITS PER DAY)

SUBSEQUENT FINDINGS
1004 N KETOSIS-PHORE UNDER BASAL CONDITIONS

BASED ON YOUR FINDINGS, THE FOLLOWING INFERENCES CAN BE MADE

262 NIDDM, NONCEASES
15 NIDDM, NONCEASES, CN DIET
213 NIDDM, NONCEASES, CN INSULIN (21 TO 40/DAY)

Figure 4: Sequence of status reports on a single patient. Since patient data is stored on the system disk and updated after each visit, the system prompts the user on the data to be obtained at the follow up visit. The status of findings observed at a previous visit can be changed at the follow up visit and only the latest entry for a given finding is used to classify the patient. For patient follow up, the text of the inferences under consideration is printed together with the data to be obtained. The prints already accumulated and those to be contributed by the new data are shown in parentheses.

tentiously suitable for computerizing the new scheme for classifying diabetes. The diagnostic criteria were directly incorporated into medical logic modules and the entire scheme was computerized in a relatively short time and with a minimum of effort. The initial formulation was applied to several test cases. The modular structure of the medical logic simplified revision and expansion of the medical knowledge base to give full expression to the new diabetes classification scheme as well as to our concepts which supplement this scheme.

The computer system for diabetes classification was applied to the records of scores of patients. Each patient was correctly classified in accordance with the National Diabetes Data Group criteria. The computer evaluation now forms part of the patient record both on our ward and in our outpatient clinics. The computer system assures that the new classification is applied to all patients in an unbiased and consistent manner.

References


Addresses of the authors: Fred Wiener, Ph. D., Senior Lecturer, Faculty of Medicine, Technion, P. O. B. 9649, Haifa 31096, Israel; Yoram Kanter, M. D., Ph. D., Senior Lecturer, Faculty of Medicine, Technion, and Head, Diabetes Service and Research Unit, Rambam Medical Center, Haifa; Shahar Kol, M. D., Diabetes Service and Research Unit, Rambam Medical Center, Haifa; David Barzilai, M. D., Professor and Dean, Faculty of Medicine, Technion, and Head, Dept. of Internal Medicine »C«, Rambam Medical Center, Haifa.