The MICROSTROKE Expert System for Stroke Type Diagnosis

Klaus Spitzer, PhD, MD, Andreas Thie, MD,
Louis R. Caplan, MD, and Klaus Kunze, MD

MICROSTROKE is a prototype expert system designed to categorize and diagnose stroke types based on clinical information. The knowledge base of MICROSTROKE includes information from large stroke registries. The system first queries the physician-user for details of the patient's history, information about the onset of stroke, accompanying symptoms, and pertinent neurologic findings and then sums the individual data items, factors in the a priori odds, and arrives at the probabilities of different stroke types for a given patient. Specific diagnosis of stroke type includes thrombosis, embolus, lacune, intracerebral hemorrhage, and subarachnoid hemorrhage. Stroke type diagnoses by MICROSTROKE were correct in 72.8% of 250 cases in the Hamburg Stroke Data Bank. MICROSTROKE runs on any MS-DOS microcomputer and is intended as a practical aid for physicians not fully familiar with the diagnosis of stroke types. (*Stroke* 1989;20:1353–1356)

For the bedside assessment of stroke type, knowledge of the frequency distributions of signs, symptoms, and ecological data associated with the different stroke types can be of prime importance. Neurologists collect historical data, neurologic signs, and symptoms to arrive at a "best guess" as to stroke type, which then forms the basis for performing further diagnostic procedures such as computed tomography (CT scan) or cardiologic or cerebrovascular tests.[1]

During the last decade, much research effort has been devoted to the development of expert systems to cope with complex medical decision-making. An expert system is "an embodiment within a computer of a knowledge-based component, from an expert skill, in such a form that the system can offer intelligent advice or make an intelligent decision about a processing function."[2] Such a system uses expert knowledge to attain high levels of performance in a narrow problem area.[3]

We present MICROSTROKE, the prototype of an expert system for computer-supported stroke type diagnosis based only on clinical and historical patient data available at the bedside. MICROSTROKE serves as an aid in the diagnostic work-up of stroke as both a stroke patient data bank and as an educational tool in clinical teaching. Another expert system, TOPOSCOUT, has been independently developed for stroke localization.[4]

From the Neurologische Universitätsklinik Hamburg-Eppendorf, Hamburg, Federal Republic of Germany (K.S., A.T., K.K.) and the Department of Neurology, Tufts–New England Medical Center, Boston, Massachusetts (L.R.C.).

Presented in part at the 13th International Joint Conference on Stroke and Cerebral Circulation, February 18–20, 1988, San Diego, California.

Address for correspondence: Dr. Spitzer, Neurologische Universitätsklinik Hamburg-Eppendorf, Martinistraße 52, D-2000 Hamburg, Federal Republic of Germany.

Received August 4, 1988; accepted May 16, 1989.

Materials and Methods

MICROSTROKE includes three knowledge data bases. The first contains frequency distributions of clinical and ecological parameters for the stroke types thrombosis, embolus, lacune, intracerebral hemorrhage (ICH), and subarachnoid hemorrhage (SAH). Data are derived from the Stroke Data Bank,[5] the Michael Reese Stroke Registry,[6] and the Harvard Cooperative Stroke Registry[7] (Figure 1). The frequency distribution tables contain relative frequencies of single items (symptoms or historical data) or of a set of alternative symptoms for all stroke patients and for each stroke type, respectively.

A second knowledge data base is implemented using rule-based information coding. Rule-based systems depend on the hypotheses that expert knowledge consists of many independent, situation-specific rules and that computers can simulate expert reasoning by stringing these rules together in chains of deduction.[8] The "if" part of a rule (the premise) contains the pattern or attributes that must be matched for the rule to be used. The "then" part (the conclusion) contains an assertion to be made when the premise is satisfied. A typical rule is "If there is hypertension at onset and early course of deficit is gradual smooth progression of symptoms,

	all	thr	lac	emb	ICH	SAH
Onset on arising.	31	40	50	17	13	16
Physical Stress at onset.	4	1	1	5	10	15
Ordinary daily activity at onset.	59	54	47	68	64	64
Other/unknown.	7	5	2	10	13	6

Source: Harvard Stroke Registry. (percentage)

FIGURE 1. *Representation of MICROSTROKE's knowledge base of frequency distributions. all, all stroke types; thr, thrombosis; lac, lacune; emb, embolus; ICH, intracerebral hemorrhage; SAH, subarachnoid hemorrhage.*

Let *type* be an element of $\{thrombus, embolus, lacune, ICH, SAH\}$, $a(type)[n]$ the *account* of *type* after the n-th question, and $p(type)[n]$ the correspondent *probability* of *type* after the n-th question.
The five *accounts* $a(thrombus)[0]=34$, $a(embolus)[0]=31$, $a(lacune)[0]=19$, $a(ICH)[0]=10$, $a(SAH)[0]=6$ represent the a priori odds before any patient information is available. The formula

$$a(type)[n+1] = a(type)[n] \times pr^{wei}$$

with

$pr = 1,$	if answer unknown;
$pr = p(symptom \mid type),$	if answer *yes*;
$pr = p(notsymptom \mid type),$	if answer *no*;
$wei = 1,$	if $weight(question, answer) = 10;$
$wei = 0.5,$	if $weight(question, answer) = 5;$
$wei = 0.25,$	if $weight(question, answer) = 2;$

is used for iterative recalculation after each question is answered. The correspondent probability is given by

$$p(type)[n] = \frac{a(type)[n] \times 100}{\sum_{j=1}^{5} a(type_j)[n]}.$$

The dependent probabilities $p(symptom \mid type)$ and $p(notsymptom \mid type)$ are extracted from the frequency distribution tables of the knowledge base.
The $weight(question, answer)$ represents the weight of a question dependent from the answer given by the user. Weights are stored in the knowledge base.

FIGURE 2. *MICROSTROKE's rules for calculating stroke type probabilities. ICH, intracerebral hemorrhage; SAH, subarachnoid hemorrhage.*

then display warning for ICH." Certain rules include combinations of symptoms that are associated with a high probability of intracranial hemorrhage. At present, MICROSTROKE matches the premises of 11 rules with data from a current patient to prompt a warning for ICH, SAH, or both if they apply.

The third knowledge data base consists of exclusively text information, used by the tutorial module of MICROSTROKE and stored as an American Standard Code for Information Interchange (ASCII) file. This knowledge data base serves only educational purposes and has no influence on calculations of stroke type diagnostic probabilities.

MICROSTROKE acquires knowledge by interactively asking the physician user for details of the patient's history, information about the onset of stroke, and accompanying symptoms in a questionnaire comprising 38 items for which frequency distribution tables in different types of stroke are available. The answers accepted are yes, no, unknown, or an option if a multiple-choice question is presented.

The inference engine of any expert system is the computer program that provides its general problem-solving capabilities. The inference engine is separated from the collection of domain knowledge, the knowledge data base. MICROSTROKE's inference engine calculates probabilities of different stroke types using modified Bayesian inference techniques. Each stroke type is attributed an account. Starting from initial accounts representing the a priori odds for the five stroke types, accounts are recalculated after each question depending on the physician-user's answer. The order of the questions presented depends on the MICROSTROKE mode selected. If there is no laboratory data available, intracranial hemorrhage cannot be excluded, and MICROSTROKE's first goal is to detect signs of ICH or SAH. If intracranial hemorrhage has already been excluded, for example, by CT scan and lumbar puncture, data are acquired to differentiate ischemic stroke types. The accounts are displayed as probabilities of stroke types by multiplying each with a constant, yielding an account sum of 100.

Figure 2 shows the calculation process of MICROSTROKE's inference engine in some detail. It must be emphasized that the weight attributed to an item depends on the physician-user's answer regarding this item. Thus, a positive answer may influence calculation more than a negative one; for example, "coma at onset" raises the account of ICH and SAH more than "normal level of consciousness" lowers it. This strategy and the mode-dependent modification of the questionnaire if intracranial hemorrhage has been excluded distinguishes MICROSTROKE from classical Bayesian-based systems.

MICROSTROKE's knowledge bases are strictly separate from its inference engine. Thus, frequency tables, weights and selection of questions and their order, combinations of symptoms that invoke the ICH/SAH warning module, and the tutorial text information can be easily modified using any word processor.

MICROSTROKE is written in TURBO-PASCAL. It runs on any IBM-compatible microcomputer.

Results

MICROSTROKE consists of five modules that can be switched by entering simple keystrokes in screen menus. In the first, the physician-user must select an item from the knowledge base menu that includes information on the history, neurologic signs, and accompanying findings, and MICROSTROKE then displays relative frequencies of the item selected for all patients and each stroke type from one of the three stroke registries. Frequency distribution tables (as in Figure 1) are presented.

The second, tutorial, module is used for computer-aided instruction. We have created text screen images that are linked by a logical relation, providing a stepwise explanation of stroke problems (e.g., pathophysiology, diagnosis, and management). MICROSTROKE's text data base has the ability to store complete publications on the subject of stroke (e.g., this paper) to be read on the computer screen during a session. Furthermore, the tutorial module has the capacity for explanation, displaying path-

```
probability of thrombosis:  18  #########
probability of lacune:       2  #
probability of embolus:     21  ##########
probability of ICH:          5  ###
probability of SAH:         54  ###########################
warning:                        ICH or SAH
--------------------------------------------------------------------------
REC 19: Gender male. Younger than 56. No atherosclerosis. No hyperten-
sion at onset. Headache at onset.
--------------------------------------------------------------------------
How about the early course of deficit ? ->1
1 Maximal deficit at onset.
2 Stuttering or stepwise course.
3 Gradual smooth course.
4 Fluctuant course.
5 Early course unknown.
(1-5)
```

FIGURE 3. *Excerpt from typical MICROSTROKE session. ICH, intracerebral hemorrhage; SAH, subarachnoid hemorrhage.*

ways of reasoning, physician-user's instructions, and help screens.

In the third, expert system, module, the physician-user is first asked whether intracranial bleeding has already been excluded. If so, MICROSTROKE omits questions especially designed to detect ICH and SAH. MICROSTROKE starts out with the a priori odds of different stroke types and modifies them according to the physician-user's consecutive answers. Figure 3 shows a typical MICROSTROKE screen. The upper part displays the calculated odds of different stroke types for the patient whose entered characteristics are listed in the middle. MICROSTROKE's questions and the answers that will be accepted by the inference engine appear in the lower part. In this particular case, a warning message appears on the screen, drawing attention to the possibility of intracranial bleeding in this young patient presenting with headache. The patient's data profile is reviewed continuously during the session for patterns of symptoms commensurate with ICH or SAH. At the end of the session, MICROSTROKE has summed up all entered information to present its final odds. The physician-user is asked for a final diagnosis based on laboratory studies, if available.

When all data of authentic cases are stored, including the physician-user's final diagnosis confirmed by laboratory studies, MICROSTROKE updates its stroke registry (fourth module). Data profiles can be replayed for later reviews of particular cases, and information can also be transferred to other programs for statistical analyses or scientific reports. MICROSTROKE's registry has the ability to store large quantities of data, providing easy access for patient care and research.

MICROSTROKE's fifth, control unit, module has access to data files of large stroke registries that are not included in its knowledge base. Cases stored in these stroke registries may be presented to MICROSTROKE in an automatic fashion so that the data of a particular stroke registry, rather than the physician-user, will answer the system's questions. The final odds calculated by MICROSTROKE may then be compared with the diagnosed stroke types in the registry, thereby achieving quality control of the system. Using the control unit, MICROSTROKE's guesses have been prospectively tested for conformity with the final diagnoses of 250 cases in the Hamburg Stroke Data Bank.[9] MICROSTROKE was correct in 72.8% of all, in 11 of 12 SAH (91.7%), in 17 of 27 ICH (63.0%), in 96 of 128 embolic (75.0%), and in 58 of 82 thrombotic strokes (70.7%). Nine ICH, one SAH, and 12 embolic strokes were incorrectly diagnosed as thrombosis, and one lacune was taken for an embolus. The control unit also checks the quality of MICROSTROKE's diagnostic assessment of authentic cases from previous sessions using the built-in stroke registry.

Discussion

MICROSTROKE is designed to serve as a computer-based diagnostic tool and a knowledge base for stroke type diagnosis. MICROSTROKE is intended as a practical aid for practitioners and neurologists involved in stroke management on wards, in stroke units, and in emergency rooms. Our expert system is hardware-independent, allowing it to be run on a wide variety of small computers, including laptops. These computers are now available in almost all medical institutions. Application of computer-assisted methods for faster and more accurate diagnostic routines could save time, permitting physician-users to give more time to patient care or teaching.[10] MICROSTROKE's ability to detect intracranial hemorrhage with high sensitivity could prompt physician-users to make appropriate diagnostic tests very early during the clinical course.

Simulating a physician's decision-making process, MICROSTROKE incorporates modified Bayesian statistics based on frequency distributions of symptoms combined with information stored in rule-based knowledge data banks.[11–13] In contrast to statistically oriented expert systems, a class of expert systems based on inexact reasoning has been developed by investigators in artificial intelligence, taking into account the judgmental, subjective, and nonprobabilistic characteristics of medical knowledge.[14–16] TOPOSCOUT, an expert system for stroke localization, belongs to this class of expert systems.[4] Though there are major objections to the use of Bayesian techniques in medical diagnosis,[17,18] Bayesian classification systems may be more valuable than is commonly believed for some medical problems.[19,20] The major advantage of computer programs based on Bayesian statistics is their ability to incorporate data from large patient registries. For example, MOS is a module that deduces stroke type by using clinical and historical data from the Michael Reese Stroke Registry.[21] Like MOS, many expert systems developed in the medical field are connected to clinical data banks.[22] Furthermore, MICROSTROKE exhibits self-learning characteristics, making use of its stroke registry module, so that with each new patient, the amount of stored information increases. If these data are made available to the inference engine, MICROSTROKE simulates the growing clinical experience of a human expert.

Since expertise is the cornerstone upon which an entire expert system is built,[23] the true power of MICROSTROKE lies in its knowledge data base. Yet, the process of extracting expertise from an expert physician and transforming it into a formal structure that can be understood and used by an expert system is a long, difficult, and error-prone task.[16,24] For this reason, we decided to implement only published information from large stroke registries. We are aware that important clinical items are missing from MICROSTROKE's questionnaire, but we preferred to start out with a limited but well-documented knowledge base. This restriction of the knowledge base was mainly responsible for the 27.2% incorrect stroke type diagnoses, especially in ICH and thrombotic strokes. For example, due to the lack of discriminating questions, MICROSTROKE revealed difficulties in distinguishing patients with cerebral infarcts who presented with impaired levels of consciousness from patients with ICH. As work progresses, MICROSTROKE's knowledge base will be enlarged by published cases that have been studied in greater detail to improve its performance. Nevertheless, in converting this prototype expert system into a high-performance one we may meet with unexpected difficulties.[25]

MICROSTROKE suggests that computer-assisted stroke type diagnosis is feasible at the bedside.[26] A prospective study currently in progress at Tufts–New England Medical Center and Hamburg University Hospital must be concluded to appraise MICROSTROKE's true value. In the future, efforts to improve MICROSTROKE's performance will include expansion of its knowledge bases, incorporation of its own experience contained in its stroke registry, and further problem-specific modification of the underlying inference techniques.

Finally, we want to emphasize that physicians disagree among themselves as much as computer programs disagree with doctors,[17] especially in the diagnosis of stroke type.[27]

References

1. Caplan LR, Stein RW: *Stroke. A Clinical Approach.* Boston, Butterworths, 1986
2. Forsyth R (ed): *Expert Systems.* London, Chapman & Hall, 1984
3. Waterman DA: *A Guide to Expert Systems.* Reading, Mass, Addison-Wesley, 1986
4. Spitzer K, Thie A, Caplan LR, Kunze K: The TOPOSCOUT expert system for stroke localization. *Stroke* 1989; 20:1195–1201
5. Foulkes MA, Wolf PA, Price TR, Mohr JP, Hier DB: The Stroke Data Bank: Design, methods, and baseline characteristics. *Stroke* 1988;19:547–554
6. Caplan LR, Hier DB, D'Cruz I: Cerebral embolism in the Michael Reese Stroke Registry. *Stroke* 1983;14:530–536
7. Mohr JP, Caplan LR, Melski JW, Goldstein RJ, Duncan GW, Kistler JP, Pessin MS, Bleich HL: The Harvard Cooperative Stroke Registry: A prospective registry. *Neurology* 1978;28:754–762
8. Schwartz B, Patil RS, Szolovits P: Artificial intelligence in medicine. Where do we stand? *N Engl J Med* 1987; 316:685–688
9. Spitzer K, Becker V, Thie A, Kunze K: The Hamburg Stroke Data Bank: Goals, design and preliminary results. *J Neurol* 1989;236:139–144
10. Richard RH: Evaluation of a medical data system. *Comput Biomed Res* 1970;3:415–425
11. Davis R, Buchanan B, Shortliffe E: Production rules as a representation for a knowledge based consultation program. *Artif Intell* 1977;7:15–45
12. Hudson DL, Estrin T: Derivation of rule-based knowledge from established medical outlines. *Comput Biol Med* 1984; 14:3–13
13. Clancey WJ: The epistemology of a rule-based expert system—A framework for explanation. *Artif Intell* 1983; 20:215–251
14. Shortliffe EH: *Computer-Based Medical Consultations:* MYCIN. New York, Elesevier, 1976
15. Szolovits P: *Artificial Intelligence in Medicine.* Boulder, Colo, Westview Press, 1982
16. Torasso P: Knowledge based expert systems for medical diagnosis. *Stat Med* 1985;4:317–325
17. Bouckaert A: Medical diagnosis: Are expert systems needed? *Int J Biomed Comput* 1987;20:123–133
18. Szolovits P, Pauker SG: Categorical and probabilistic reasoning in medical diagnosis. *Artif Intell* 1978;11:115–144
19. Chard T: Human versus machine: A comparison of a computer 'expert system' with human experts in the diagnosis of vaginal discharge. *Int J Biomed Comput* 1987;20:71–78
20. Zagoria RJ, Reggia JA: Transferability of medical decision support systems based on Bayesian classification. *Med Decis Making* 1983;3:503–509
21. Hier DB, Atkinson GD, Perline R, Hill H, Evans M, Desai B, McCormick WC, Caplan LR: Can a patient data base help build a stroke diagnostic expert system? *Med Inf* (Lond) 1986;11:75–81
22. Hier DB, Caplan LR, Hill H, Evens M, Sinha A: A microcomputer-based expert system to assist in localization of anatomic damage after stroke (abstract). *Neurology* 1984; 34(Suppl 1):83
23. Hayes-Roth F, Waterman DA, Lenat DB (eds): *Building Expert Systems.* Reading, Mass, Addison-Wesley, 1983
24. Mars NJ, Miller PL: Knowledge acquisition and verification tools for medical expert systems. *Med Decis Making* 1987; 7:6–11
25. Alvey PL, Myers CD, Greaves MF: High performance for expert systems: I. Escaping from the demonstrator class. *Med Inf* (Lond) 1987;12:85–95
26. Tuhrim S, Reggia JA: Feasibility of physician-developed expert systems. *Med Decis Making* 1986;6:23–26
27. Gross CR, Shinar D, Mohr JP, Hier DB, Caplan LR, Price TR, Wolf PA, Kase CS, Fishman IG, Calingo S: Interobserver agreement in the diagnosis of stroke type. *Arch Neurol* 1986;43:893–898

KEY WORDS • cerebrovascular disorders • diagnosis • expert systems