19

PUFF: An Expert System for Interpretation of Pulmonary Function Data

Janice S. Aikins, John C. Kunz, Edward H. Shortliffe, and Robert J. Fallat

In this and the next chapter we close this volume with discussions of the two AIM systems that had achieved routine use by the end of the first decade of research in the field. It is important to note that neither requires direct interaction with a physician requesting advice. Thus both systems avoid the significant problems of human engineering and user acceptance that define many of the serious research problems that remain unsolved at present [see Teach and Shortliffe (1981) and a further discussion of these points in Chapter 21]. However, each does provide a glimpse of what lies ahead, and their success at difficult tasks is an encouraging indication of the practical impact that we can eventually expect from this kind of work.

Because MYCIN was designed to keep its knowledge base of rules separate from the program that used them to generate advice (Chapter 5), it was recognized that the program itself could be isolated and used in other domains for which additional rule sets were developed. The resulting EMYCIN system (van Melle, 1980) was used to build several other programs during the late 1970s, in both medical and nonmedical domains [e.g., SACON, a program to provide guidance regarding the use of a computer system to aid in aircraft design (Bennett and Englemore, 1979)]. An early system developed using EMYCIN was PUFF, a collaborative effort between computer scientists from Stanford University, researchers from the Institute of Medical Sciences in San Francisco, and physicians from Pacific Medical Center (PMC).

From *Computers in Biomedical Research*, 16: 199–208 (1983).

For several years pulmonary physiologists at PMC had been toying with ideas for the development of a program to interpret pulmonary function test (PFT) results. They had found it difficult to develop a straightforward algorithm for defining the criteria for test interpretation, however, and as a result were continuing to interpret PFT results by hand when the collaboration with Stanford developed. Working in the EMYCIN environment, they were delighted to find that they could more easily distill their criteria for test interpretation by using the production rule formalism. Within a few months a small experimental system was developed and was shown to perform extremely well for analyzing a subset of PFT abnormalities. Thereafter the rule set was expanded, and, when it had stabilized, the clinicians were eager to implement the system for use at PMC. It had been developed at Stanford on the SUMEX-AIM computer, however, and this was an unrealistic vehicle for providing service computing at a hospital in San Francisco. As is described in this chapter, the PUFF rule set was therefore rewritten into a program using the BASIC language and implemented to run on a minicomputer at PMC. It accordingly became a working tool in the pulmonary physiology lab of this large institution. Its performance and the results of a formal evaluation experiment are described here. In addition, Janice Aikins and her coauthors examine some of the elements of the problem that paved the way for its success and also consider the significant limitations of the solution that warrant further study.

19.1 Introduction

Researchers in the field of artificial intelligence are just beginning to produce systems that capture the specialized knowledge of experts and that use this knowledge to perform difficult tasks. Although the technology is still rather new, a small set of programs now exist as "tools" useful for building these so-called expert systems. This paper describes an expert system, called PUFF, that was built using EMYCIN, a generalization of an earlier medical system named MYCIN. The task chosen for PUFF is described briefly, and the rationale for the appropriateness of this choice is presented. PUFF was initially developed on the SUMEX computer, a large research machine at Stanford University, and was later rewritten in a production version to run on the hospital's own minicomputer. We describe here the history of the PUFF project and its current status, including observations about its limitations and successes. We also take a brief look at the knowledge representation and control structure used for the SUMEX version of the system. Finally, the results of a formal evaluation of the production version of PUFF are presented.

19.2 Task

PUFF interprets measurements from respiratory tests administered to patients in the pulmonary (lung) function laboratory at Pacific Medical Center in San Francisco. The laboratory includes equipment designed to measure the volume of the lungs, the ability of the patient to move air into and out of the lungs, and the ability of the lungs to get oxygen into the blood and carbon dioxide out.[1] The pulmonary physiologist interprets these measurements in order to determine the presence and severity of lung disease in the patient. An example of such measurements and an interpretation statement are shown in Figure 19-1. The test measurements listed in the top half of the figure are collected by the laboratory equipment. The pulmonary physiologist then dictates the interpretation statements to be included in a typewritten report. All of the measurements are given as a percentage of the predicted values for a normal patient of the same sex, height, and weight. The interpretation and final diagnosis are a summary of the reasoning about the combinations of measurements obtained in the lung tests.

19.3 Rationale

PUFF's task is to interpret such a set of pulmonary function (PF) test results, and to produce a set of interpretation statements and a diagnosis for the patient. The problem of developing an automated pulmonary function interpretation system was chosen for several reasons:

1. The interpretation of pulmonary function tests is a problem that occurs daily in most hospitals, so a computer program that captures the expertise involved in interpreting these tests, and that can assist in providing interpretations, fills a practical need.
2. The biomedical researchers at Pacific Medical Center (PMC) were interested in the problem and were eager to work with us on developing a solution. It was possible that such a system could enhance the effectiveness of patient care and the pulmonary physician's efficiency. In addition, solution of this relatively simple interpretation problem could identify possibilities for further research into more difficult interpretation tasks.

[1] Measurements include spirometry and, optionally, body plethosmography, single breath CO diffusion capacity, and arterial blood gases. Measurements can be made at rest, following inhalation of a bronchodilator, and during exercise.

PRESBYTERIAN HOSPITAL OF PMC
CLAY AND BUCHANAN, BOX 7999
SAN FRANCISCO, CA. 94120
PULMONARY FUNCTION LAB

WT 40.8 KG, HT 161 CM, AGE 65 SEX F
REFERRAL DX-
*** TEST DATE 5-13-76

		PREDICTED (+/-SD)	OBSER(%PRED)	POST DILATION OBSER(%PRED)
INSPIR VITAL CAP	(IVC) L	2.7(0.6)	2.3 (83)	
RESIDUAL VOL	(RV) L	2.0(0.1)	3.8 (193)	3.1 (154)
FUNC RESID CAP	(FRC) L	2.9(0.3)	4.6 (158)	3.9 (136)
TOTAL LUNG CAP	(TLC) L	4.7(0.7)	6.1 (129)	5.5 (116)
RV/TLC	%	42.	62.	55.
FORCED EXPIR VOL	(FEV1) L	2.3(0.5)	1.5 (66)	1.6 (71)
FORCED VITAL CAP	(FVC) L	2.7(0.6)	2.3 (85)	2.4 (88)
FEV1/FVC	%	82.	64.	66.
FORCE EXP FLOW	200-1200L/S	3.6(0.8)	1.8	1.9
FORCED EXP FLOW	25-75% L/S	2.6(0.5)	0.7	0.7
FORCED INS FLOW	200-1200L/S	2.5(0.5)	2.5	3.4
AIRWAY RESISTANCE (RAW)	(TLC = 6.1)	2.5	1.5	2.2
DF CAP-HGB = 14.5 (DSBCO)	(TLC = 4.8)	23.	17.4 (72)	

****

INTERPRETATION: THE VITAL CAPACITY IS LOW, THE RESIDUAL VOLUME IS HIGH AS IS THE TOTAL LUNG CAPACITY, INDICATING AIR TRAPPING AND OVERINFLATION. THIS IS CONSISTENT WITH A MODERATELY SEVERE DEGREE OF AIRWAY OBSTRUCTION AS INDICATED BY THE LOW FEV1, LOW PEAK FLOW RATES AND CURVATURE TO THE FLOW VOLUME LOOP. FOLLOWING ISOPROTERANOL AEROSOL THERE IS VIRTUALLY NO CHANGE.

THE DIFFUSING CAPACITY IS LOW INDICATING LOSS OF ALVEOLAR CAPILLARY SURFACE.

CONCLUSIONS: OVERINFLATION, FIXED AIRWAY OBSTRUCTION AND LOW DIFFUSING CAPACITY WOULD ALL INDICATE MODERATELY SEVERE OBSTRUCTION AIRWAY DISEASE OF THE EMPHYSEMATOUS TYPE. ALTHOUGH THERE IS NO RESPONSE TO BRONCHODILATORS ON THIS ONE OCCASION, MORE PROLONGED USE MAY PROVE TO BE MORE HELPFUL.

PULMONARY FUNCTION DIAGNOSIS: OBSTRUCTIVE AIRWAY DISEASE, MODERATELY SEVERE, EMPHYSEMATOUS TYPE.

FIGURE 19-1 Verbatim copy of pulmonary function report. The data were obtained from equipment and the interpretation dictated by an expert physician.

3. PF data interpretation was a problem that the artificial intelligence researchers were particularly interested in solving in order to demonstrate the generality and power of expert system techniques. Putting a system into clinical use would contribute to the credibility of those techniques, and also would show their promise and limitations in clinical practice. Earlier AI programs had demonstrated competence, but their use had required large amounts of professional time simply for data input. PUFF, however, produced PF data interpretations automatically without the necessity for user interaction. Thus we hoped that PUFF would be used by the clinical staff.

4. PF data interpretation was a problem that was large enough to be interesting (the biomedical researchers did not know how to solve it, and the AI researchers did not know whether their techniques would be appropriate) and small enough that a pilot project of several months' duration could concretely demonstrate the feasibility of a longer development effort. Furthermore, the amount of domain-specific knowledge involved in pulmonary function testing is limited enough to make it feasible to acquire, understand, and represent that knowledge.
5. The domain of pulmonary physiology is a circumscribed field: the data needed to interpret patient status are available from the patient's history and from measurements taken in a single laboratory. Other large bodies of knowledge are not required in order to produce accurate diagnoses of pulmonary disease in the patient.[2]
6. All the data used in the laboratory at PMC were already available in a computer; the computer data were known to be accurate, reliable, and relevant to the interpretation task. The clinical staff in the PF lab were already receptive to the use of computers within their clinical routines.
7. Pulmonary physiologists who interpret test measurements tend to phrase their interpretations similarly from one case to the next. One goal of PUFF was to generate reports from a set of prototypical interpretation statements, thus saving the staff a great deal of tedious work. The staff themselves would not be displaced by this tool because their expertise still would be necessary to verify PUFF's output, to handle unexpectedly complex cases, and to correct interpretations that they felt were inaccurate.

19.4 Project History and Status

This research developed from work done on the MYCIN system (Chapter 5). That program used a knowledge base of production rules (Davis and King, 1977) to perform infectious disease consultations. PUFF was initially built using a generalization of the MYCIN system called EMYCIN (van Melle, 1979). EMYCIN, or "Essential MYCIN," consists of the domain-independent features of MYCIN, principally the rule interpreter, explanation, and knowledge-acquisition modules (Shortliffe et al., 1975). It provides a mechanism for representing domain-specific knowledge in the form of production rules, and for performing consultations in that domain. Just as MYCIN consists of EMYCIN plus a set of facts and rules about diagnosis and therapy of infectious diseases, PUFF consists of the EMYCIN programs plus a pulmonary disease knowledge base.

[2] This was a problem in MYCIN, a related system for determining the diagnosis and therapy for infectious disease cases. The results produced by the system often suffered because it lacked knowledge about related diseases that were also present in the patient.

EMYCIN (and hence the EMYCIN version of PUFF) is written in Interlisp (Teitelman, 1978) and runs on a DEC KI-10 at the Stanford SUMEX-AIM computer facility. In order to run PUFF on a PDP-11 at Pacific Medical Center, a second version of the program was created after the EMYCIN version had been refined. This was done by translating the production rules into procedures and writing them in the BASIC language. Conversion to BASIC was an advantage because the PDP-11 was located on the same site as the laboratory, and its schedule could be easily controlled to support production operation by the system users. However, as a result of the conversion, the production and development versions of PUFF became incompatible, and modifications made to one system were sometimes difficult to make in the other.

The PDP-11 version is now routinely used in the pulmonary function laboratory and provides lung test interpretations for about ten patients daily. Since the system became operational in 1979, it has interpreted the results of over 4000 cases. The BASIC code is currently being converted again so that it will run on a personal computer.

The form of the interpretations generated by PUFF is shown in Figure 19-2. This report is for the same patient as in Figure 19-1, seen several years later. As in the typed report, the pulmonary function test data are set forth, followed by the interpretation statements and a pulmonary function diagnosis. The pulmonary physiologist checks the PUFF report, and, if necessary, the interpretation is edited on-line prior to printing the final report for physician signature and entry into the patient record. Approximately 85% of the reports generated are accepted without modifications. The change made to most others simply adds a statement suggesting that the patient's physician compare the interpretation with tests taken during previous visits. For example, statements such as "These test results are consistent with those of previous visits" or "These test results show considerable improvement over those in the previous visit" might be made. PUFF was not designed to represent knowledge about multiple visits, so this kind of statement must be added by the pulmonary physician.

19.5 Observations

PUFF is a practical assistant to the pulmonary physiologist, and thus is a satisfactory and exciting result of the research done with production rule consultation systems. PUFF's performance is good enough that it is used daily in clinical service, and it has the support of both the hospital staff and its administration. However, improvements could be made in the following areas:[3]

[3]Many of these problems are also present in other rule-based systems; they motivated the development of the experimental CENTAUR system (Aikins, 1980; 1983).

PRESBYTERIAN HOSPITAL OF PMC
CLAY AND BUCHANAN, BOX 7999
SAN FRANCISCO, CA. 94120
PULMONARY FUNCTION LAB

WT 40.8 KG, HT 161 CM, AGE 69 SEX F
REFERRAL DX

**TEST DATE 5-13-80

		PREDICTED (+/-SD)	OBSER(%PRED)	POST DILATION OBSER(%PRED)
INSPIR VITAL CAP	(IVC) L	2.7	2.3 (86)	2.4 (90)
RESIDUAL VOL	(RV) L	2.0	3.8 (188)	3.0 (148)
TOTAL LUNG CAP	(TLC) L	4.7	6.1 (130)	5.4 (115)
RV/TLC	%	43.	62.	56.
FORCED EXPIR VOL	(FEV1) L	2.2	1.5 (68)	1.6 (73)
FORCED VITAL CAP	(FVC) L	2.7	2.3 (86)	2.4 (90)
FEV1/FVC	%	73.	65.	67.
PEAK EXPIR FLOW	(PEF) L/S	7.1	1.8 (25)	1.9 (26)
FORCED EXP FLOW	25-75%L/S	1.8	0.7 (39)	0.7 (39)
AIRWAY RESIST(RAW)	(TLC= 6.1)	0.0(0.0)	1.5	2.2
DF CAP-HGB=14.5	(TLC= 4.8)	24.	17.4 (72)	(74%IF TLC = 4.7)

****

INTERPRETATION: ELEVATED LUNG VOLUMES INDICATE OVERINFLATION. IN ADDITION, THE RV/TLC RATIO IS INCREASED, SUGGESTING A MODERATELY SEVERE DEGREE OF AIR TRAPPING. THE FORCED VITAL CAPACITY IS NORMAL. THE FEV1/FVC RATIO AND MID-EXPIRATORY FLOW ARE REDUCED AND THE AIRWAY RESISTANCE IS INCREASED, SUGGESTING MODERATELY SEVERE AIRWAY OBSTRUCTION. FOLLOWING BRONCHODILATION, THE EXPIRED FLOWS SHOW MODERATE IMPROVEMENT. HOWEVER, THE RESISTANCE DID NOT IMPROVE. THE LOW DIFFUSING CAPACITY INDICATES A LOSS OF ALVEOLAR CAPILLARY SURFACE, WHICH IS MILD.

CONCLUSIONS: THE LOW DIFFUSING CAPACITY, IN COMBINATION WITH OBSTRUCTION AND A HIGH TOTAL LUNG CAPACITY IS CONSISTENT WITH A DIAGNOSIS OF EMPHYSEMA. ALTHOUGH BRONCHODILATORS WERE ONLY SLIGHTLY USEFUL IN THIS ONE CASE, PROLONGED USE MAY PROVE TO BE BENEFICIAL TO THE PATIENT.

PULMONARY FUNCTION DIAGNOSIS:
1. MODERATELY SEVERE OBSTRUCTIVE AIRWAYS DISEASE.
   EMPHYSEMATOUS TYPE.

FIGURE 19-2 Pulmonary function report generated by PDP-11 version of PUFF.

- representation of prototypical patterns,
- addition or modification of rules to represent knowledge not previously encoded,
- alteration of the order in which information is requested during the consultation, and
- explanation of system performance.

The first point refers to the fact that many cases can be viewed as relatively simple variations of typical patterns. PUFF does not recognize that a case fits a typical pattern, nor can it recognize that a case differs in some important way from typical patterns. As a result, PUFF's explanations of its diagnoses lack some of the richness of explanation that physicians can use when a case meets, or fails to meet, the expectations of a prototypical case. The medical knowledge in PUFF is encoded as *rules*. Rules encode relatively small and independent bodies of domain knowledge. The rule formalism makes modification of the program's knowledge much easier than when that knowledge is embedded in computer code. However, additions or modifications to the rules as referred to in the second point have caused difficulties because changes to one rule sometimes affect the behavior of other rules in unanticipated ways. The last two points apply only to the EMYCIN version of PUFF, which runs interactively in a consultation-style, question-and-answer mode with the user. In that system, questions are sometimes asked in an unusual order, and explanations of both the questions being asked of the user and the final interpretation need to be improved.

Even though PUFF does exhibit certain limitations, the representation of pulmonary knowledge as production rules allows the encoding of interpretive expertise that previously was difficult to define because it is heuristic knowledge of the expert. EMYCIN on the SUMEX computer provided an excellent environment for acquiring, encoding, and debugging this expertise. However, it would have been inefficient and somewhat impractical to use the interactive EMYCIN version of PUFF in a hospital setting. The simplicity of EMYCIN's reasoning process made the translation into BASIC procedures a feasible task, thus allowing the hospital's own computer staff to take over maintenance of the system.

The BASIC version of PUFF runs in batch mode and does not require interaction with a physician. We believe that this system was readily accepted by the pulmonary staff for several reasons. First, the program's interpretations are consistently accurate. Second, explanations of diagnoses are appropriately detailed so that the user has confidence in the accuracy of correct diagnoses and enough information with which to recognize and modify incorrect diagnoses. Third, less physician time is required to produce consistently high-quality reports using the system than is required to analyze and dictate case reports without it. Finally, the program is well integrated into the routine of the laboratory; its use requires very little extra technician effort.

19.6 Overview of EMYCIN-PUFF

19.6.1 Knowledge Representation

The knowledge base of the EMYCIN-PUFF system consists of (a) a set of 64 *production rules* dealing with the interpretation of pulmonary function tests and (b) a set of 59 *clinical parameters*. The production version (BASIC-PUFF) has been extended to include 400 production rules and 75 clinical parameters. The clinical parameters represent pulmonary function test results (e.g., TOTAL LUNG CAPACITY and RESIDUAL VOLUME), pa-

```
RULE011

IF:  1) A: The mmf/mmf-predicted ratio is between 35 and 45, and
        B: The fvc/fvc-predicted ratio is greater than 80, or
     2) A: The mmf/mmf-predicted ratio is between 25 and 35, and
        B: The fvc/fvc-predicted ratio is less than 80
THEN:  1) There is suggestive evidence (.5) that the degree of
          obstructive airways disease as indicated by the MMF
          is moderate, and
       2) It is definite (1.0) that the following is one of the
          findings about the diagnosis of obstructive airways
          disease: Reduced mid-expiratory flow indicates
          moderate airway obstruction.

PREMISE:  [$AND ($OR ($AND (BETWEEN* (VAL1 CNTXT MMF) 35 45)
                           (GREATER* (VAL1 CNTXT FVC) 80))
                     ($AND (BETWEEN* (VAL1 CNTXT MMF) 25 35)
                           (LESSP* (VAL1 CNTXT FVC) 80]
ACTION:  (DO-ALL (CONCLUDE CNTXT DEG-MMF MODERATE TALLY 500)
                 (CONCLUDETEXT CNTXT FINDINGS-OAD
                               (TEXT $MMF/FVC2) TALLY 1000))
```

FIGURE 19-3 A PUFF production rule in English and LISP versions.

tient data (e.g., AGE and REFERRAL DIAGNOSIS), and data that are derived from the rules (e.g., FINDINGS associated with a disease and SUBTYPES associated with the disease). There may be auxiliary information associated with the clinical parameters, such as a list of expected values and an English translation used in communicating with the user.

The production rules operate on associative <attribute object value> triples, where the attributes are the clinical parameters, the object is the patient, and the values are given by the patient data and lung test results. Questions are asked during the consultation in an attempt to fill in values for the parameters.

The production rules consist of one or more premise clauses followed by one or more action clauses. Each premise is a conjunction of predicates operating on associative triples in the knowledge base. A sample PUFF production rule is shown in Figure 19-3.

The rules are coded internally in LISP. The user of the system sees the production rules in their English form, which is shown in the upper part of the figure. The English version is generated automatically from templates, as is described in van Melle (1979).

19.6.2 Control Structure

The EMYCIN-PUFF control structure is primarily a goal-directed backward chaining of production rules. The goal of the system at any time is to determine a value for a given clinical parameter. To conclude a value for a clinical parameter, the program tries a precomputed list of rules whose actions conclude values for the clinical parameter [refer to van Melle (1979) for details].

If the rules fail to conclude a value for a parameter, a question is then asked of the user in order to obtain that value. An exception to this process occurs for parameters labeled ASKFIRST. These represent information generally known by the user, such as results of pulmonary function tests. For these parameters it is more efficient simply to ask a consultation question than to attempt to infer the information by means of rules.[4]

19.7 Evaluation of the BASIC-PUFF Performance System

The knowledge base from the original performance version of PUFF was tested on 107 cases chosen from files in the pulmonary function laboratory at Pacific Medical Center. Those 107 cases formed a representative sample of the various pulmonary diseases, their degrees, and their subtypes. Modifications were made to the knowledge base, and the cases were tried again. This iteration continued until our collaborating expert was satisfied that the system's interpretations agreed with his own. At this point the system was "frozen," and a new set of 144 cases was selected and interpreted by the system. All 144 cases also were interpreted separately by two pulmonary physiologists (the expert working with us and a physician from a different medical center).

The results of the comparison of interpretations by each diagnostician are presented in the table in Figure 19-4. The table compares close agreement in diagnosing the severity of the disease, where close is defined as differing by at most 1 degree of severity. Thus, for example, diagnoses of mild (degree = 1) and moderate (degree = 2) are considered close, while mild and severe (degree = 3) are not. Further, a diagnosis of normal is not considered to be close to a diagnosis of a mild degree of any disease.

The table shows that the overall rate of agreement between the two physiologists on the diagnoses of disease was 92%. The agreement between PUFF and the physician who served as the expert to develop the PUFF knowledge base (MD-1 in the table) was 96%. Finally, the agreement between PUFF and the physician who had no part in the development of the PUFF knowledge base (MD-2) was 89%. Figure 19-5 shows the distribution of diagnoses by each diagnostician. The number of diagnoses made by each diagnostician does not total 144 because patients were often diagnosed as having more than one disease.

[4]In the BASIC version of PUFF implemented at PMC, all of the test data are known ahead of time so that "asking a question" merely entails retrieving another datum from a stored file.

DIAGNOSIS	PERCENT AGREEMENT MD-1 MD-2	MD-1 PUFF	MD-2 PUFF
NORMAL	92	95	92
OAD	94	99	94
RLD	92	99	85
DD	90	91	85
TOTAL	92	96	89
(S.D.)	(1.63)	(3.83)	(4.69)

Diseases: Normal = Normal Pulmonary Function
OAD = Obstructive Airways Disease
RLD = Restrictive Lung Disease
DD = Diffusion Defect

FIGURE 19-4 Summary of percent agreement in 144 cases.

DIAGNOSIS	DIAGNOSTICIAN MD-1	MD-2	PUFF
NORMAL	31	26	30
OAD	79	85	89
RLD	52	45	55
DD	53	35	52

FIGURE 19-5 Number of diagnoses by each diagnostician for 144 cases.

19.8 Conclusions

The PUFF research has demonstrated that if the task, domain, and researchers are carefully matched, then the application of existing techniques can result in a system that successfully performs a moderately complicated task of medical diagnosis. Success of the program can be measured not only in terms of the system's technical performance, but equally importantly, by the ease and practicality of the system's day-to-day use in the lab for which it was designed. Rule-based representation allowed easy codification and later modification of expertise. The simplicity of the rule interpreter in the Interlisp version facilitated translation into BASIC and implementation on the hospital's own PDP-11 machine. Using EMYCIN allowed the researchers to move quickly from a point where they found it difficult even to describe the diagnostic process to a point where a simple diagnostic model was implemented. Having a diagnostic model allowed them to focus on individual issues in order to improve that model. Although PUFF does not itself represent new artificial intelligence techniques, its success is a testimonial for EMYCIN. In addition, its simplicity has facilitated careful analysis of EMYCIN's rule representation and control structure and has led to other productive research efforts (Aikins, 1980; 1983; Smith and Clayton, 1980).

ACKNOWLEDGMENTS

The PUFF research team consists of an interdisciplinary group of physicians and computer scientists. In addition to the authors, these have included Larry Fagan, Ed Feigenbaum, Penny Nii, Dr. John Osborn, Dr. B. J. Rubin, and Dianne Sierra. We also thank Dr. B. A. Votteri for his help in evaluating PUFF's performance and Doug Aikins for his editorial help with this paper.

The research was funded in part by NIH grants MB-00134 and GM-24669. Computer facilities were provided by the SUMEX-AIM facility at Stanford University under NIH grant RR-00785. Dr. Shortliffe is supported by research career development award LM-00048 from the National Library of Medicine. Dr. Aikins was supported by the Xerox Corporation under the direction of the Xerox Palo Alto Research Center.