II.A.2.4 PUFF-VM Project

PUFF-VM: Biomedical Knowledge Engineering in Clinical Medicine

John J. Osborn, M.D.
The Institutes of Medical Sciences (San Francisco)
Pacific Medical Center

and

Edward H. Shortliffe, M.D., Ph.D.
Department of General Internal Medicine
Stanford University Medical Center
Stanford University

The immediate goal of this project is the development of knowledge-based programs to interpret physiological measurements made in clinical medicine. The interpretations are intended to be used to aid in diagnostic decision making and in therapeutic actions. The programs will operate within medical domains which have well developed measurement technologies and reasonably well understood procedures for interpretation of measured results. The programs are:

(1) PUFF: the interpretation of standard pulmonary function laboratory data which include measured flows, lung volumes, pulmonary diffusion capacity and pulmonary mechanics, and

(2) VM: management of respiratory insufficiency in the intensive care unit.

The second, but equally important, goal of this project is the dissemination of Artificial Intelligence techniques and methodologies to medical communities that are involved in computer aided medical diagnosis and interpretation of patient data.

I. SUMMARY OF RESEARCH PROGRAM

PUFF:

A. Technical Goals

The task of PUFF program is to interpret standard measures of pulmonary function. It is intended that PUFF produce a report for the patient record, explaining the clinical significance of measured test results. PUFF also must provide a diagnosis of the presence and severity of pulmonary disease in terms of measured data, referral diagnosis, and patient characteristics. The program must operate effectively over a wide range of pathological conditions with a broad clinical perspective about the possible complexity of the pathology.

B. Medical Relevance and Collaboration

Interpretation of standard pulmonary function tests involves attempting to identify the presence of obstructive airways disease (OAD: indicated by reduced flow rates during forced exhalation), restrictive lung disease (RLD: indicated by reduced lung volumes), and alveolar-capillary diffusion defect (DD: indicated by reduced diffusivity of inhaled CO into the blood). Obstruction and restriction may exist concurrently, and the presence of one mediates the severity of the other. Obstruction of several types can exist. In the laboratory at the Pacific Medical Center (PMC), about 50 parameters are calculated from measurement of lung volumes, flow rates, and diffusion capacity. In addition to these measurements, the physician may also consider patient history and referral diagnosis in interpreting the test results and diagnosing the presence and severity of pulmonary disease.

Currently PUFF contains a set of about 250 physiologically based interpretation "rules". Each rule is of the form "IF <condition> THEN <conclusion>". Each rule relates physiological measurements or states to a conclusion about the physiological significance of the measurement or state.

The interpretation system operates in a batch mode, accepting input data and printing a report for each patient. The report includes: (1) Interpretation of the physiological meaning of the test results, the limitation on the interpretation because of bad or missing data; the response to bronchodilators if used; and the consistency of the findings and referral diagnosis. (2) clinical findings, including the applicability of the use of bronchodilators, the consistency of multiple indications for airway obstruction, the relation between test results, patient characteristics and referral diagnosis. (3) Interpretation Summary, which consists of the diagnosis of presence and severity of abnormality of pulmonary function.

C. Progress Summary

<u>Knowledge base</u>:

PUFF is implemented on the PDP-10 in a EMYCIN system which is designed to accept rules from new task domains. A typical rule is:

```
If (FVC>=80) and  (FEV1/FVC<predicted-5) then PEAK FLOW RATES ARE
REDUCED, SUGGESTING AIRWAY OBSTRUCTION OF DEGREE
        if (predicted-15<= FEV1/FVC <predicted-5) MILD
        if (predicted-25<=FEV1/FVC <predicted-15) MODERATE
        if (predicted-35<=FEV1/FVC <predicted-25) MODERATE TO SEVERE
        if (FEV1/FVC <predicted-35)               SEVERE
```

This rule compares the ratio of FEV1, the amount of air that can be forced out in the first second of exhalation with the total "forced vital capacity" (FVC) or total amount of lung volume that can be exhaled. The inability to force out a large percentage of air in the critical first second implies the presence of an obstruction in the airway.

Results:

The results of the PUFF system are reviewed in more detail in the 1978 SUMEX annual report and [Kunz 78]. A version of the PUFF system is now in routine daily use at Pacific Medical Center. Reports are reviewed by a physician pulmonary physiologist. Over 85 % of the reports are accepted by the physician without change; they are signed and entered into the patient record. Most of the remaining reports are edited on-line to modify a small point in the test interpretation.

Table 1 reviews a study of the agreement in severity of diagnoses made by two MD's and by PUFF rules. This study was made with a less complete rule base than what is currently available in the pulmonary lab. In 94% of 144 cases analyzed in a prospective study, the degree of severity (0=none; 1=mild; 2=moderate; 3=moderately-severe; 4=severe) of OAD diagnosed by the first MD was within a single degree of severity of OAD diagnosed by the second MD. In 96% of the 79 cases for which the first MD diagnosed OAD, the second MD diagnosed the severity of OAD within one level of the severity diagnosed by the first MD. Agreement within one degree of severity of the diagnoses by the first and second MD's was substantially lower in RLD and DD cases. These discrepancies occurred because the second MD consistently called RLD more severe than did the first MD, and he consistently did not diagnose diffusion defects when the first MD diagnosed DD of moderate or greater degree.

	Percent Agreement with 1st MD			
	All 144 cases		1st MD made Dx	
Diagnosis	Second M.D.	PUFF Rules	Second M.D.	PUFF Rules
Normal				
OAD	0.94	0.99	0.96	0.97
RLD	0.92	0.97	0.77	1.00
DD	0.87	0.87	0.60	0.80
Total	0.91	0.94	0.86	0.94

Table 1. Percent agreement within one degree of severity of diagnoses Approximately 1500 patients have been interpreted by the system, by two MD's and by the first MD and rules.

In addition to the use of PUFF as a working clinical tool, it has been very useful for evaluation of knowledge representation methods. The original PUFF knowledge base (around 60 rules) represents realistic medical knowledge but is small enough to use for experiments. The PUFF knowledge has been used in the AGE system, the CENTAUR system using a combination of rules and prototypes, and the WHEEZE system, a UNIT-based approach to knowledge-representation.

D. Relevant Publications

[1] "A Physiological Rule-Based System for Interpreting Pulmonary Function Test Results", J.C. Kunz, R.J. Fallat, D.H. McClung, B.A. Votteri, J.S. Aikins, H.P. Nii, L.M. Fagan, E.A. Feigenbaum, HPP 78-154, Stanford Heuristic Programming Project, 1978.

[2] "Prototypes: An Approach to Knowledge Representation for Hypothesis Formation", Aikins, J.S., HPP-79-10 (working paper), 1979. Also Int. Joint Conf. on Artif. Intell., Tokyo, Japan, August, 1979.

[3] "A Physiological Rule-Based System for Interpreting Pulmonary Function Test Results", J.C. Kunz, R.J. Fallat, D.H. McClung, B.A. Votteri, J.S. Aikins, H.P. Nii, L.M. Fagan, E.A. Feigenbaum, Proceedings of Computers in Critical Care and Pulmonary Medicine, IEEE Press, 1979.

[4] "The Art of Artificial Intelligence: Themes and Case Studies of Knowledge Engineering", E.A. Feigenbaum, Proceedings of the IJCAI, (1977). (Also Stanford Computer Science Department Memo STAN-CS-77-612).

E. Funding Support

PUFF-VM is supported by NIH grant GM24669 for $164,000 from 1 September 1978 - 30 August 1981. Some indirect costs are included in this total. A renewal application is pending.

VM:

A. Technical Goals

The Ventilator Manager program (VM) interprets the clinical significance of time varying quantitative physiological data from patients in the ICU. This data is used to manage patients receiving ventilatory assistance. An extension of a physiological monitoring system, VM (1) provides a summary of the patient's physiological status appropriate for the clinician; (2) recognizes untoward events in the patient/machine system and provides suggestions for corrective action; (3) suggests adjustments to ventilatory therapy based on a long-term assessment of the patient status and therapeutic goals; (4) detects possible measurement errors; and, (5) maintains a set of patient-specific expectations and goals for future evaluation. The program produces interpretations of the physiological measurements over time, using a model of the therapeutic procedures in the ICU and clinical knowledge about the diagnostic implications of the data. These therapeutic guidelines are represented by a knowledge base of rules created by clinicians with extensive ICU experience. Physiological measurements are collected at the bedside every 2-10 minutes and processed by the PMC computer system. The PMC and SUMEX computers are currently linked by TYMNET and special data communication programs written by SUMEX staff member, William Yeager. the program's interpretation is sent back to terminals in the ICU for evaluation by research clinicians.

B. Medical Relevance and Collaboration

To assist in the interpretation process, VM must be able to recognize unusual or unexpected clinical events (including machine malfunction) in a manner specifically tailored to the patient in question. The interpretation task is viewed as an ongoing process in the ICU, so that the physiological measurements must be continually reevaluated producing a current clinical picture.

This picture can then be compared with previous summary of patient status to recognize changes in patient condition upon which therapy selection and modifications can be made. The program must also determine when the measurements are most likely to be sensitive to error or when external measurements would be of diagnostic significance.

VM offers a new approach toward more accurate recognition of alarm conditions by utilizing the history and situation of the patient in the analysis. This is in contrast to the use of static limits applied to measurements generated to fit the "typical patient" under normal conditions. Our program uses a model of interpretation process, including the types and levels of conclusions drawn manually from the measurements to provide a summary of patient condition and trends. The program generated conclusions are stated at levels more abstract than the raw data; for example, the presence of hemodynamic stability/instability rather than in terms of heart rate and mean arterial pressure. When the data is not reliable enough to make these conclusions, additional tests may be suggested. The recognition of important conclusion for which external verification is sought, will also elicit the suggestion for confirming tests from the program.

C. Progress Summary

VM is provided with the values of 30 physiological measurements on a 2- or 10-minute bases by an automatic monitoring system. The output is in the form of suggestions to clinicians and periodic summaries (see example case below).

<u>Example</u> <u>Case</u>:

The following case demonstrates typical output of the system. The data used in this example were obtained from a post-cardiac surgery patient from the ICU at Pacific Medical Center. The terms VOLUME, ASSIST, CONTROLLED MANDATORY VENTILATION (CMV), and T-PIECE refer to specific types of ventilatory assistance. The output format is: (a) ..time of day.., (b) generated comments for clinicians, starting with "**", and (c) commentary in {}.

```
 ..1350.. ..1351..
 ** SYSTEM ASSUMES PATIENT STARTING VOLUME VENTILATION.
                                        {monitoring started}
 ** HYPERVENTILATION                    {diagnostic conclusions
 ** TACHYCARDIA                          based on monitored data}
 ** PATIENT HYPERVENTILATING.           {suggested therapy based on
 ** SUGGEST REDUCING MINUTE VOLUME       diagnosis}
 ..1400..
 . . .
 ..1450..
** HYPERVENTILATION
** TACHYCARDIA
** PATIENT HYPERVENTILATING.
** SUGGEST REDUCING MINUTE VOLUME
 ..1500..
** HYPERVENTILATION
** PATIENT HYPERVENTILATING.
** SUGGEST REDUCING MINUTE VOLUME


Current conclusions:               {summary information}
HYPOTENSION PRESENT for 41 MINUTES
HYPERVENTILATION PRESENT for 33 MINUTES
SYSTOLIC B.P. LOW for 46 MINUTES
{etc.}

Conclusions:       {time of day}    |.....|.....|.....|.
                                   13    14    15    16
HEMODYNAMICS -- STABLE                        ====
HYPERVENTILATION -- PRESENT            =     == == ====
HYPOTENSION -- PRESENT                   ====  ========
TACHYCARDIA -- PRESENT                 ==== ==

patient is on ASSIST                         =====   ==
patient is on CMV                        ====     ===
patient is on VOLUME                   ==
patient is on NOT-MONITORED        ====

Goal is CMV                            ================
Goal is VOLUME                     ====
                                   |.....|.....|.....|.
                                   13    14    15    16
```

The availability of new measurements requires updated interpretations based on the changing values and trends. As the patient setting changes--e.g., as a patient starts to breathe on his own during removal (weaning) from the ventilator--the same measurement values lead to different interpretations. In order to properly interpret data collected during changing therapeutic contexts, the knowledge base includes a model of the stages that a patient follows from admission to the unit through the end of the critical monitoring phase. Recognition of the appropriate patient context is an essential step in determining the meaning of most physiological measurements.

The majority of the knowledge of the VM program is concerned with the relations between the various concepts known by the program. These concepts include: measurement values, typical therapeutic decisions, diagnostic labels, and physiological states. The connections between concepts are represented by a form of production rules using the structure "IF premise THEN action."

The rules in VM are of the form:

IF facts about measurements or previous conclusions are true

THEN

1) Make a conclusion based on these facts;
2) Print out suggestions for the clinician;
3) Establish expectations about the future values of measurements.

A sample VM rule is shown below.

```
STATUS RULE: STABLE-HEMODYNAMICS
DEFINITION: Defines stable hemodynamics for most settings
APPLIES to patients on VOLUME, CMV, ASSIST, T-PIECE
COMMENT: Look at mean arterial pressure for changes in
    blood pressure and systolic blood pressure for maximum
    pressures.
IF
  HEART RATE is ACCEPTABLE
  PULSE RATE does NOT CHANGE by 20 beats/minute in 15 minutes
  MEAN ARTERIAL PRESSURE is ACCEPTABLE
  MEAN ARTERIAL PRESSURE does NOT CHANGE by 15 torr in 15 minutes
  SYSTOLIC BLOOD PRESSURE is ACCEPTABLE
THEN
  The HEMODYNAMICS are STABLE
```

Figure 1. Sample VM Interpretation Rule. The meaning of 'ACCEPTABLE' varies with the clinical context--i.e., whether the patient is receiving VOLUME or CMV ventilation, etc. This rule makes a conclusion for internal system use. Similar rules also make suggestions to the user.

During this last year, we started clinical validation of the first medically oriented artificial intelligence program to be used in a continuous (real-time) setting. Our experiences with this project for data interpretation in the intensive care unit, has highlighted some basic research questions for time-oriented systems. We have concentrated on explanation, reasoning, and therapy planning in a changing clinical environment.

Our work in explanation has concentrated on showing the change in reasoning between one instant in time and the next. Expressed symbolically, a description of what has changed may often be more useful than a description of the current chain of reasoning. We are also working on methods to reason with data from different points in time, e.g., revising old hypotheses about the clinical setting when a delayed laboratory result becomes available.

Our third area of time-oriented research is the development of techniques for creating and modifying long-term therapeutic plans. These plans are created by combining management strategies provided by expert clinicians. Based on an analysis of previous response to therapy (e.g., following an oncology protocol), specialized treatment will be suggested for patients unable to follow the standard treatment plan. About twenty strategies, and fifteen appropriate patients have been identified for this task.

D. Relevant Publications

Fagan, L.M., Kunz, J.C., Feigenbaum, E.A. and Osborn, J.J.: A symbolic processing approach to measurement interpretation in the intensive care unit. Proc. Third Annual Symposium Computer Applications in Medical Care, Silver Spring, Maryland, October, 1979, pp. 30-33.

Fagan, L.M., Shortliffe, E.H. and Buchanan, B.G.: Computer-based medical decision making: From MYCIN to VM. Automedica 3(2), 1980.

Fagan, L.M.: VM: Representing Time-Dependent Relations in a Medical Setting, Ph.D. dissertation, Stanford University, 1980.

Osborn, J.J., Fagan, L.M., Fallat, R.J., et al: Managing the data from respiratory measurements. Med. Instrumentation, November-December, 1979. (Winner of the 'Best Article of the Year' Award for AAMI - 1979.)

E. Funding Support

PUFF-VM is supported by NIH grant GM24669 for $164,000 from 1 September 1978 - 30 August 1981. Some indirect costs are included in this total. A renewal application is pending.

II. RESEARCH PLANS

A. Long Range Goals and Plans

The main emphasis of this project has switched from the development of the PUFF system to the extension and evaluation of the VM system. This change is consistent with the goals of the NIH proposal, the current use of PUFF in a clinical setting and the research questions that remain in the VM portion of the project. Some long term interests, such as consensus building between experts, will be examined using both application areas.

The long range goal of the VM project is to develop and evaluate an interpretation system that will improve patient care in the ICU. Toward this goal, we plan to extend the rule set, provide better models of physiology and therapy, and start a formal evaluation of the program's therapeutic advice.

The rule set in VM will be extended to handle a greater number of patients. The current emphasis of the program has been on the management of post-surgical patients with normal pre-operative status. We will continue to concentrate on post-surgical patients, but the knowledge base will be augmented to handle patients with additional problems noted before surgery or those who have an unusual response to therapy after surgery. The majority of this knowledge will be used to create a more detailed classification of the patient population and the corresponding generation of expectations.

These rule set extensions will ultimately be limited by representation of the underlying cardiopulmonary physiology and the therapeutic plans used in the ICU. Still other improvements will come from a better model of the mechanical ventilator and other instrumentation. Each of these models will provide a structure upon which to build the rule base, and are motivated by the special problems of evaluating the patient's status in a dynamic clinical setting. These problems include the evaluation of the relationship between actual and anticipated response to therapy and the the recognition of a particular therapy step in the context of a larger therapeutic plan (e.g., the process of removing a patient from the ventilator when the patient has an underlying lung disease).

We will develop integrated causal models of the physiological response of the patient to changes in patient physiological state. In the clinical case, we assume that the model will consist of abstract characterizations of the active parts of the patient and life-support system. The relation between these active parts will characterize the physiology of the patient organ system or the operation of the mechanical life-support system. The general approach to development of models will be to consider a simple, well-understood but potentially useful case: a static model of ventilation; second, we will consider oxygenation as a more complex physiological process involving a larger number of physiological parameters. The oxygenation model will also include some representation of time dependency using both heuristic notions of time change and mathematical representation of time constant of change.

In order to determine the appropriate areas for these model building activities and to insure acceptance by physicians, a careful prospective validation will be carried out to identify the accuracy of the advice of the program.

III. <u>INTERACTIONS WITH THE SUMEX-AIM RESOURCE</u>

A. Collaborations and Medical Use of Programs via SUMEX

The PUFF-VM project requires very close collaboration between investigators at two institutions separated by fifty miles. This kind of collaboration, in which program development and testing proceeds concurrently on the same application system, requires a computer network facility for sharing of code, data and ideas. SUMEX has been used at PMC for running programs developed concurrently by Stanford and PMC staff, and data has been taken from the PMC computer system and transferred to SUMEX on magnetic tape for program development and testing. The SUMEX staff has developed a cooperating set of computer programs to allow the PMC computer and the SUMEX/2020 systems to actively exchange files and program data and output.

We also use the SUMEX system for purposes other than program development. A joint PMC-Stanford report of VM was prepared entirely through the the word communications and processing capabilities of SUMEX. Investigators from the two institutions have collaborated in writing reports together; the separate contributions are prepared on SUMEX, edited and merged with an exchange of messages but without ever requiring actual meetings. We have also used the system for trading bibliographic information with other AIM users. We have also experimentally run the Internist program using SUMEX.

B. Sharing and Interactions with Other SUMEX-AIM Projects

We have participated in the AIM workshop and had very fruitful interaction with a number of other SUMEX users, directly influencing our perception of important problems and potentially appropriate solutions. Personal contacts at other conferences, at Stanford AI weekly meetings, and at PMC with visiting members of the AIM community, have also been very helpful in keeping abreast of the current thinking of other members of the AI community and with members of the medical community interested in computer based physiological analysis and diagnosis. We believe that the use of a common machine and the existence of the AIM conference encourages increased recognition and better communication with other AIM workers. Within AIM we most closely collaborate with the MYCIN, MOLGEN and DENDRAL projects, who share common space, common techniques, and common attitudes.

C. Critique of Resource Management

The SUMEX community continues to be an extremely supportive environment in which to do research on uses of artificial intelligence in clinical medicine. The community has two equally vital resources -- the people with knowledge and interest in AI and the facility on which AI system development can proceed. They are equally excellent as resources, helping hands when faced with problems, and friendly support for continued productive research. The availability of INTERLISP; of a facility on which routine data processing functions (eg. manipulating magnetic tapes and making long listings) can take place; and of message-sending among remote users are all vital functions for our project. SUMEX provides them in an environment which is friendly and reliable. Management of the SUMEX facility is consistent and excellent.

D. Needs and Plans for Other Computational Resources

The future goals of the project (as described above) will require considerable computational requirements in the near future. These requirements will come in the form of active development of a large INTERLISP program, and extensive testing of the program in a clinical environment. We hope to perform as much of the evaluation work as possible on the 2020. System development of the program will probably continue on

SUMEX during off-hours or be off-loaded to the spare time on the 2020. Most subsidiary text processing tasks have been off-loaded from SUMEX to avoid the high load average situation during the day. The storage of usable versions of the program and the test files used in the evaluation of the program will require about 1000 additional pages on the SUMEX computer.

The on-line evaluation of the program is currently using the 2020 on a stand alone basis one afternoon a week. Batch runs after modifications of the rule set also use computer time during off hours. This level of computer usage should remain during 1981-82.

E. Recommendations for Future Community and Resource Development

We perceive the evolution of our.AI capability as moving from a highly speculative development state, for which the interactive development capabilities of SUMEX are vital, to a more stable but still changing validation-and-evaluation state. Ultimately we foresee rather stable specification of a program for routine clinical use. Thus, we see the need to transfer our AI techniques from the SUMEX PDP-10 to a local host. For this transfer, a principal long-range need is for software systems that will allow us to run AI systems on a mini-computer after they have been developed on the more powerful SUMEX facility. If the validation of PUFF-VM in the PMC clinical setting shows the programs to be effective in health care, then we hope and expect to be able to provide the capability on a routine basis.

We would also like to encourage SUMEX's role as a facilitator of information transfer between AIM users. This can happen by scheduling on-line demonstrations that any other user can "connect to," or by providing a common depository for AI and medicine information. This might take the form of on-line bibliographies, collecting common user packages, or connecting common research interests together. This communication service would complement the technical service facilities currently provided by the SUMEX staff.

II.A.2.5 Rutgers Computers in Biomedicine Project [Rutgers-AIM]

Rutgers Computers in Biomedicine

Rutgers Research Resource--Computers in Biomedicine

Principal Investigator: Saul Amarel
Rutgers University
New Brunswick, New Jersey

I. SUMMARY OF RESEARCH PROGRAM

A. Goals and Approach

The fundamental objective of the Rutgers Resource is to develop a computer based framework for significant research in the biomedical sciences and for the application of research results to the solution of important problems in health care. The focal concept is to introduce advanced methods of computer science - particularly in artificial intelligence - into specific areas of biomedical inquiry. The computer is used as an integral part of the inquiry process, both for the development and organization of knowledge in a domain and for its utilization in problem solving and in processes of experimentation and theory formation.

At present, the total number of investigators who participate in scientific activities of the Resource is 81; of these, 34 have Rutgers appointments, 26 are outside investigators who participate in collaborative research projects that are mainly located at Rutgers, and 21 are investigators from collaborative national AIM projects that are located in different parts of the country. In addition, the Resource has 11 other members in Administrative, Computer Systems/Operations and general programming and secretarial functions. Thus, the Rutgers Resource community numbers at present a total of 92 participants.

Resource activities include research projects (collaborative research and core research) training/dissemination projects, and computing services in support of user projects. The research projects are organized in three main AREAS OF STUDY. These areas of study and the senior investigators in each of these are:

1) Medical Modeling and Decision Making (C. Kulikowski, S. Weiss)

2) Modeling Belief Systems and Commonsense Reasoning (C. Schmidt and N.S. Sridharan)

3) Artificial Intelligence: Representations, Reasoning, and System Development (S. Amarel)

The training/dissemination activities of the Rutgers Resource include sponsorship of the Annual AIM Workshop - whose main objective is to strengthen interactions between AIM investigators, to disseminate research methodologies and results, and to stimulate collaborations and imaginative resource sharing within the framework of AIM. Starting in 1979, the Workshop is being organized and hosted on a rotational basis by the members of the AIM community, in coordination with the Rutgers Research Resource. The Sixth AIM Workshop was organized by the SUMEX-AIM Resource and it was held at Stanford University in August, 1980.

B. Medical Relevance; Collaborations

During 1980-81 we continued the development of a versatile system for building consultation models, called EXPERT. We reached a new milestone in technology transfer by implementing a scheme for the automatic translation of EXPERT models (with some restrictions) into microprocessor compatible algorithmic form. EXPERT is being used extensively in the development of several medical consultation models in collaboration with clinical investigators in rheumatology, ophthalmology, clinical pathology and with researchers in biomedical modeling.

Problems in <u>rheumatology</u> are particularly important in health care, given the high prevalence and chronic nature of arthritis and related disorders. They also represent an active area of biomedical and clinical research, in which a group of our medical collaborators at the University of Missouri under Dr. Gordon Sharp has been noted for its contributions. The application of A.I. approaches to problems of medical decision making in this domain was facilitated by our collaboration with Dr. Donald Lindberg, Director of the Health Care Technology Center at the University of Missouri.

Our experience with the design of the rheumatology model has shown us that the knowledge engineering tools and know-how that we developed so far in the Resource make it possible to move incrementally and rapidly in the construction of a new medical knowledge base in collaboration with expert clinical researchers. Moreover, this experience is leading us to the development of a methodology for guiding the interaction of medical and computer science researchers in model building. The sequence of developments of consultation models follows a natural progression, aided at every step by an interplay between the clarification of medical concepts and the application of logical methods of model design. Our work in this area is contributing to a better understanding of a central problem in the application of Artificial Intelligence to the design of expert computer-based systems; namely, what are the representations, the processes and the interface facilities that are needed to acquire, augment, and refine knowledge bases of different types by interacting with specialists in a domain.

During the current research period the initial prototype model in connective tissue disease diagnosis developed by Dr. Sharp and his colleagues was subjected to careful review and evaluation by an outside panel of experts. As a result, the model was significantly broadened in scope (of both findings and diagnoses), and a new criterion-based scheme for consistently organizing knowledge in the specialty was devised. The Resource investigators at Rutgers formalized this approach within the framework of EXPERT, and developed a method for updating criterion-based models using the cases stored in their data base. Another important development motivated by the rheumatology project has been the introduction of a special logical feature for reporting the diagnostic interpretations in a sub-language that more closely follows clinical terminology.

In ophthalmology work continues with the development of the neuro-ophthalmological model by Dr. William Hart at Washington University, the glaucoma model by Dr. Y. Kitazawa at Tokyo University and the infectious eye disease model by Dr. Chandler Dawson at the University of California at San Francisco. The elaboration of a specialized treatment strategy and explanation model has received strong support through the infectious eye disease project.

In clinical pathology, we have collaborated with Dr. Robert Galen (Columbia University and Overlook Hospital) in the development of a model for the interpretation of serum protein electrophoretic patterns. The prototype model was tested by specialists in the domain and improved to the point that it could be used to produce an automatic interpretation for the output of a scanning densitometer. In order to facilitate the transfer of the EXPERT advice scheme to the instrument environment, we developed an automatic translating program which produces an algorithmic form of the EXPERT model (which is restricted slightly in its scope of primitives for this purpose). The algorithmic version was in turn automatically translated into a microprocessor assembler code. Thus we have completed a full cycle of transfer of technology from the large scale developmental system of the Resource to a clinically useable instrument.

In biomedical modeling applications, the EXPERT scheme has been used to build preliminary prototype models that will interact with simulation programs, giving advice on the modeling process itself. In collaboration with Dr. William Yamamoto of George Washington University, we are working on a model of respiration, while enzyme kinetics modeling continues to be investigated with Dr. David Garfinkel of the University of Pennsylvania.

C. Highlights of Research Progress

1. Medical Modeling and Decision-Making

Research during the past year has concentrated on problems of representation, knowledge acquisition and treatment planning strategies, resulting in significant extensions of the EXPERT formalism, and its application in specific medical domains (see I.B above). From the developmental perspective, we have achieved a great step forward by producing automatic translation programs for EXPERT models into microprocessors compatible form.

The major achievements of our work can be summarized as follows:

1.1 Technology Transfer

A program for automatically translating an EXPERT model into algorithmic form was first developed. This required a restriction on the logical primitives used in the EXPERT model, which did not prove to be a significant limitation.

A program for converting the algorithmic version of the EXPERT model into a microprocessor assembly language (for the Motorola 6000) was then developed.

The translation programs were tested with prototypes and an actual expert-derived model for serum protein electrophoresis interpretation. The utility and dissemination of expert models will be greatly enhanced once they can be used with clinically useful and routinely applied instruments.

1.2 EXPERT System Changes

a) Procedure-Invocation Capability/Distributed Control

We have added a command within EXPERT that permits suspension of execution of the reasoning strategies while (user defined) procedure is invoked. This can accept and pass back parameters from the model, so that specialized procedural knowledge (in the form of mathematical simulation models, search procedures, pathway tracing procedures, etc.) in a domain can be explicitly coded and its results used by the consultation model.

This capability also allows one EXPERT model to call another, thereby simulating the passage of information and control from one specialist to another when dealing with a complex medical case. This provides an initial focus for studies in <u>distributed knowledge base</u> design and reasoning in consultation.

b) Criterion Based Representation

A particular representation formalism has been developed to encode clinical criteria in a natural manner within the EXPERT scheme. This involves the specification of an intermediate hypothesis structure for possible, probable and definite truth-value modifiers of a disease, in the form of major, minor and exclusionary findings. It is implemented through a special knowledge acquisition front-end which then translates the criteria from tabular form to EXPERT rules.

c) Output Language Capabilities

These have been implemented as special logical functions that can be used to override the conventional EXPERT output with special clinical language templates. The templates are filled with the specific interpretations and supporting findings for each individual case, thus making for more specific (or personalized) reporting of conclusions and advice.

1.3 Knowledge Acquisition and Model Building Strategies

We have developed several heuristic methods for giving advice on the updating of a knowledge base specified in the Criterion-based form, as well as a front end for helping build such a model. The learning capabilities are interactive so that a meta-EXPERT system suggest rules that could be generalized or restricted depending on the performance of the current model on a data base of cases. The user can select a particular modification of the model based on these suggestions (or he can try any other of his choice), and the meta system will test and comment on the performance of the modified model. If the results are not as satisfactory as expected, further changes will be suggested and the process repeated until performance is improved.

1.4 Treatment Planning: Representation and Strategies

An investigation of treatment planning representations has resulted in a new system that emphasizes facilities for encoding sequences of actions (plans of management) and producing explanations of reasoning based on them. Strategies of treatment selection involve a generalized scheme for positive (therapeutic) effects and negative (contraindication, allergy, complication) effects, that are used as the basis of sorting procedures. A system has been implemented and is being tested in the area of infectious eye disease management.

1.5 Other Research and Development

a) Data Base Compatibility of EXPERT

A method for transferring case data from an existing data base into EXPERT form was developed and implemented for our thyroid data base.

b) Model-Based Knowledge Acquisition

A scheme for prompting the acquisition of EXPERT-type rules by a model of domain function has been investigated.

2. Modeling of Belief Systems and Commonsense Reasoning

2.1 Recognition of Commonsense Plans: Plan Generation and Revision

The human information processing system possesses a richly structured long-term memory which contains an enormous store of general knowledge, concepts and facts about the domains of commonsense discourse. During this period our research has focused on ways in which this type of information can be represented and used in plan generation and revision.

In plan generation we have developed a rule-based representation for normative and normalcy knowledge. A plan generator has been implemented which uses this knowledge as a basis for default reasoning so that planning may proceed despite the unavailability of knowledge about the specific planning context.

In both planning and plan revision we have explored the ways in which default assumptions and the hierarchy of commonsense concepts may be used. By using the hierarchy of concepts, we have been able to develop strategies for dealing with uncertain and incomplete knowledge which at the same time provides the basis for guiding future refinement and revision of a plan hypothesis.

Further, we have reimplemented our plan generator so that (a) it is organized as a pair of cooperative processes that propose plan alternatives and criticize the proposals, (b) both processes can be interactively guided, and (c) user may call for replanning by functioning as a critic.

2.2 Reasoning with Commonsense Concepts

In the area of goal-directed inference about individuals we have studied ways of representing both the intension and extension associated with hierarchically structured commonsense concepts. Using this representation we have investigated strategies of plausible inference and the attendant problem of evaluating the consistency of the resulting inferences and proposed strategies for dealing with incomplete and/or inconsistent inferences.

In a recent report, we have explored a formal construct called PERMISSIVENESS of a default rule as an aid to resolving inconsistencies among default rules. Permissiveness requires no prior probabilities or additional domain knowledge.

3. Artificial Intelligence: Representations and System Development

3.1 AIMDS: System for Studying Knowledge Representations and Commonsense Reasoning

The main objective of this research during the current period were: to continue experimental work on the present major <u>applications</u> of AIMDS; to enhance the <u>description method</u>, so that it becomes possible to construct complex descriptions, including the ability to describe system behavior to itself; to develop new <u>processing frameworks</u> for extended power and usability of the system; and to improve the <u>AIMDS implementation</u> for efficiency, reliability and quality of user interface. Substantial progress was made in each of these areas. The two main applications of AIMDS remain in the area of <u>common-sense reasoning</u> studied in the Resource, and in studies of <u>legal reasoning</u> that are funded by the NSF. The AIMDS system provides a solid foundation for our collaborative research on commonsense reasoning; it continued to be used this period in the representation of psychological concepts of plan, expectation, interpretation, as well as of common sense concepts of objects, places, persons, goals and of world knowledge governing the use of such concepts. In the past, the two AIMDS functions MAKE and FIND accepted partial descriptions of individuals in a domain. There were no functions that constructed descriptions. One of the significant new capabilities that was introduced this period is functions for building and manipulating descriptions within AIMDS. Also, improvements in AIMDS implementation which were developed this period include powerful methods for optimizing memory use. The development of interesting and convenient user interfaces has occupied a subordinate role in our project. But the recent introduction of a general purpose display front-end for AIMDS has begun to fill a gap. The display, called RELDISP, displays a fixed 7x7 matrix on the top half of the screen; using the bottom half as the traditional scroll mode user interaction panel. The rows and columns of the matrix display the seven most recently referenced instances in the associative memory, and the entries of the matrix display the name of the relation asserted or denied between any pair of instances. The display is quite effective in focusing the viewers' attention during a demonstration of any of the AIMDS application programs. We have yet to gather experience in its use for the programmer who develops application programs. The RELDISP package was also used to conduct a demo of the plan recognition project during the Sixth Annual AIM Workshop held at Stanford. Presently, a revised user manual is in preparation which has expanded sections that describe important new additions to the AIMDS system. The new manual and the improved user interface facilities are expected to contribute to further dissemination of the AIMDS system in the Resource and in the national AIM community.

3.2 Expertise Acquisition and Theory Formation

There are two main research activities in this subproject. One, which was initiated in the previous period, is focusing on the task of acquiring good control strategies for heuristic search via experimentation -- within the framework of a given problem representation. The second, which we have been pursuing for several years, is concerned with improvements in a problem solving system via shifts in problem representation. Related to the second approach, is the problem of how to acquire and use models at different levels of resolution in a given domain in order to increase problem solving performance (expert performance) in the domain. Problems of theory formation are at the heart of the processes of expertise improvement and acquisition that we have been studying here. During this period, we made substantial progress in these studies.

3.3 Knowledge Acquisition and Man-Machine Interfaces

In this subproject we have two research activities. The first, in which most of our effort was concentrated in the current period, is concerned with Natural Language Interfaces for Consultation Systems. This work has resulted this year in a doctoral dissertation by Vic Ciesielski, a research assistant in the Resource. The second activity is focusing on a problem of language acquisition in the context of a student/teacher dialogue.

3.4 Programming Environment for AI Research

Our work on the development of a supportive programming environment for AI research has continued this period with emphasis on the development of an enhanced new version of the Rutgers/UCI LISP.

D. Up-to-Date List of Publications

This following is an update of publications in the Rutgers Resource for the period 1980 and 1981 (only publications not listed in previous SUMEX annual reports are presented here).

Amarel, S., (1980) "Initial Thoughts on Characterization of Expert Systems" Rutgers Technical Report CBM-TM-88.

Amarel, S., (1981) "Review of Characteristics of Current Expert Systems" Rutgers Technical Report CBM-TM-89.

Amarel, S. (1981) "Problems of Representation in Heuristic Problem Solving; Related Issues in the Development of Expert Systems" Rutgers Technical Report CBM-TR-118.

Banerji, R.B. and Mitchell, T., (1980) "Description Languages and Learning Algorithms: A Paradigm for Comparison" Rutgers Technical Report CBM-TR-107.

Ciesielski, V. (1980) "A Methodology for the Construction of Natural Language Front Ends for Medical Consultation Systems" Rutgers Technical Report CBM-TR-112.

Galen, R., Weiss, S., Kulikowski, C.A., (1981) "The Interpretation of Serum Protein Electrophoresis by Computer" Medical Laboratory Observer, (in press).

Goldberg, R.N. and Weiss, S.M., (1981) "An Experimental Transformation of a Large Expert Knowledge-Base" Proceedings of the Fourteenth Hawaii International Conference on Systems Sciences, pp. 828-836, (1981). (to appear in Journal of Medical Systems.) Rutgers Technical Report CBM-TR-110.

Goldberg, R. and Weiss, S.M. (1980) "Constructing an Expert Knowledge Base for Thyroid Consultation Using Generalized A.I. Techniques" Proc. Hawaii International Conference on Systems Science, Honolulu, Jan. 1980.

Goodson, J.L. (1980) "A Process for Evaluating Tree-Consistency" Rutgers Technical Report CBM-TR-111.

Kastner, J. and Weiss, S.M., (1981) "A Precedence Scheme for Selection and Explanation of Therapies" Rutgers Technical Report CBM-TM-90.

Kulikowski, C.A., (1980) "Artificial Intelligence Methods and Systems for Medical Consultation" IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-2, No. 5, Sept. 1980, pp. 464-476.

Kulikowski, C.A., Weiss, S.M. and Galen, R.S., (1980) "On the Diagnostic Frontier: Expert Consultation by Computer" Diagnostic Medicine, Nov/Dec. 1980, pp. 89-90.

Kulikowski, C.A. and Ostroff, J.H. (1980) "Constructing an Expert Knowledge Base for Thyroid Consultation Using Generalized A.I. Techniques" Proc. Annual Symposium on Computers in Health Care, Washington, Nov. 1980.

Lindberg, D. and Weiss, S.M. (1980) "On the Diagnostic Frontier: Expert Consultation by Computer" MEDINFO '80.

Lindberg, D., Sharp, G., Kingsland, L., Weiss, S., Hayes, S., Ueno, H., Hazelwood, S. (1980) Proceedings of the Third World Conference on Medical Informatics, pp. 1311-1315 (1980).

Mitchell, T., Utgoff, P. and Banerji, R. (1980) "Learning Problem-Solving Heuristics by Experimentation" Rutgers Technical Report CBM-TR-114.

Mitchell, T.M. (1980) "The Need for Biases in Learning Generalizations" Rutgers Technical Report CBM-TR-117.

Nagel, D., (1980) "An Experiment in Extracting Some Properties of Binary Relations" Rutgers Technical Report CBM-TM-84.

Nagel, D., (1980) "Some Considerations on Extracting Definitional Information about Relations" Rutgers Technical Report CBM-TM-85.

Nudel, B. and Utgoff, P. (1981) "A Bibliography on Machine Learning" Rutgers Technical Report CBM-TR-120.

Politakis, P. and Weiss, S.M., (1981) "A System for Empirical Experimentation with Expert Knowledge" Rutgers Technical Report CBM-TM-91.

Politakis, P. and Weiss, S. (1980) "Designing Consistent Knowledge Bases: An Approach to Expert Knowledge Acquisition" Rutgers Technical Report CBM-TR-113.

Politakis, P., Weiss, S. and Kulikowski, C.A. (1980) "Designing Consistent Knowledge Bases for Expert Consultation Systems," 13th Annual Hawaii International Conference on Systems Science, January, 1980, pp. 675-683.

Politakis, P., Weiss, S., (1980) "Toward a Systematic Methodology for the Design of Expert Consultation Systems", Proceedings ACM Computer Science Conference, p. 22, (1980).

Schmidt, C.F., (1980) "The Role of Object Knowledge in Human Planning" Rutgers Technical Report CBM-TM-87.

Schmidt, C.F. (1980) "Plan Recognition and Revision: Understanding the Observed Actions of Another Actor" Rutgers Technical Report CBM-TR-115.

Smith, D. (1980) "FOCUSER: A Strategic Interaction Paradigm for Language Acquisition" Rutgers Technical Report CBM-TR-116.

Sridharan, N.S., (1980) "Representational Facilities of AIMDS: A Sampling" Rutgers Technical Report CBM-TM-86.

Sridharan, N.S., Schmidt, C.F., and Goodson, J. L. (1980) "The Role of World Knowledge in Planning" Rutgers Technical Report CBM-TR-109.

Sridharan, N.S., Schmidt, C.F., and Goodson, J.L. (1981) "Reasoning by Default" Rutgers Technical Report CBM-TR-119.

Weiss, S.M. and Kulikowski, C.A. (1980) "On the Diagnostic Frontier: Expert Consultation by Computer" MEDINFO '80.

Weiss, S.M., Kern, K., Kulikowski, C.A. and Uschold, M.F. (1981) "A Guide to the Use of the EXPERT Consultation System. revised 1981.

Weiss, S., Kulikowski, C.A., (1980) "Data Base Aids for Designing Consultation Systems", Proceedings of the Third World Conference on Medical Informatics, pp. 794-797 (1980).

Weiss, S., Kulikowski, C.A., (1981) "Expert Consultation Systems: The EXPERT and CASNET Projects, Pergamon INFOTECH State of the Art Report on Machine Intelligence, (in press).

Weiss, S., (1980) "Tools for Knowledge Engineering", Proceedings Workshop on Expert Systems, pp. 59-64 (1980).

Weiss, S. (1981) Article on The EXPERT System, to appear in the Knowledge Engineering Chapter in "Expert Systems," editors: Hayes-Roth, F., Waterman, D., (in press).

Weiss, S., Kulikowski, C.A., and Galen, R. (1981) "Developing Microprocessor Based Expert Models for Instrument Interpretation" Rutgers University Department of Computer Science Tech. Memo, (1981).

E. Funding Support

The Rutgers Resource is funded through an NIH grant entitled "Rutgers Research Resource on Computers in Biomedicine" -- grant number P41RR643. The Principal Investigator is Dr. Saul Amarel, Professor, Chairman Department of Computer Science, and Director of the Laboratory for Computer Science Research, Rutgers, the State University of New Jersey. This grant is in the first year of its fourth 3-year renewal extending from Dec. 1, 1980 through Nov. 30, 1983. The total direct cost awarded for the 3-year period is $1,395,580 with first year funding of $495,079 in direct costs from Dec. 1, 1980 through Nov. 30, 1981.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations and Dissemination; Interactions

The SUMEX-AIM facility provides one of the nodes where some of our medical collaborators can access programs developed at Rutgers.

An important responsibility of the Rutgers Resource within the national AIM community is to sponsor dissemination and training activities.

The focus of our efforts in this area continues to be centered around the AIM Workshops and sessions on AIM research at national and international conferences.

In order to increase the dissemination of AIM work within specialty fields of medicine, we have also presented tutorial papers at relevant conferences.

1) Sixth AIM Workshop (1980):

This past year the AIM Workshop was hosted by Stanford University, according to the new scheme of rotation. Dr. E. H. Shortliffe organized the Workshop, coordinating with Dr. C. A. Kulikowski for the Rutgers Resource, which continues as the main sponsor: the budget was made available from the Rutgers Resource to cover most local arrangements at Stanford and travel funds for the participants.

2) AAAI Conference:

Resource investigators presented papers and were panelists at the first AAAI Conference that followed the workshop. This gave an opportunity for wider dissemination of the AIM research in the artificial intelligence community.

3) Expert Systems Workshop:

A strong representation of Resource investigators at this workshop (which was organized by Rand) enabled the contribution of ideas and experience from the biomedical applications of the Resource to be documented in the upcoming book that will cover this new general area of research.

4) MEDINFO:

The medical modelling and decision making group of the Resource presented a paper on the EXPERT system and its updating capabilities at this important international conference. The conference was co-chaired by one of the groups close medical collaborators, Dr. Donald Lindberg, who also presented the joint work on rheumatology model development using the Rutgers EXPERT system.

5) Hawaii International Conference on Systems Sciences:

Investigators from the Resource's medical group received the best paper award in the medical decision session at this Conference for the contribution entitled: "An Experimental Transformation of a Large Expert Knowledge Base".

6) Fourth Annual Symposium on Computers in Health Care:

Research on large data base testing via consultation models applied in the area of thyroid diseases was presented by members of the medical group of the Resource.

B. National AIM Projects at Rutgers

The national AIM projects, approved by the AIM Executive Committee for the 1980 grant year, who use the Resource system are listed here. BRIGHT is continuing under the direction of Dr. W. Gordon Walker of Johns Hopkins University, mostly for research in clinical medicine. The CONGEN and DENDRAL projects, under Dr. Djerassi and Dr. D. Smith, are predominantly at SUMEX with the Resource used when the system is overloaded, as a backup for demonstrations, and to allow for ease in collaborations with Dr. San Felippo of Rutgers Chemistry Department. The INTERNIST project has been using the Resource for backup during demos, etc, all text-editing functions and when SUMEX is down for extended periods. Most of the actual research on the Ohio State project is continuing to be done at Rutgers under Dr. Chandrasekaran and Artificial Intelligence models of clinical reasoning are being developed by Dr. Greenes of Harvard University on the Rutgers system. Dr. Garfinkel, heading the Penn project, has continued his work on modeling of metabolic pathways on the Rutgers system, in collaboration with Resource investigators. During this period, the AIM Executive Committee has authorized a group of about 25 researchers in biomedical modeling and simulation (from various parts of the country) to access the Rutgers/AIM system on a pilot basis -- to help them formulate collaborative research programs.

C. Critique of Sumex-AIM Resource Management

Rutgers is currently using the Sumex facility for three purposes:

- communication with other researchers working at Sumex, and with Sumex staff.

- building collaborative ophthalmology models. The researchers working on this project are from ophthalmology groups at a number of institutions around the country.

- backup computing for demonstrations, conferences and site visits.

In the first two cases Sumex is being used because it provides a convenient center for communication among people spread around the country. This is true because of its comparatively good Tymnet facilities, as well as because of the good support that its staff provide for remote users, particularly those who are not experienced computer users.

Our usage is currently running at less than 200 connect hours per year, with an overall connect/CPU ratio of about 42. This usage has gone down since last year because of the fact that Rutgers is now again on the ARPAnet. Much of our cooperation with users and staff at Sumex can be done using mail and file transfer through the ARPAnet. This does not require Rutgers users to log into Sumex. Such cooperation continues to be important for us, even though it is no longer reflected in our usage statistics.

The quality of the support given by Sumex staff continues to be very high. Our users report that the load at Sumex continues to be very high, despite attempts by staff to control it. However the problem has not been enough to cause serious difficulties for most of our activities.

III. RESEARCH PLANS

A. Project Goals and Plans

We are planning to continue along the main lines of research that we have established in the Resource to date. Our medical collaborations will continue with emphasis on development of consultation systems in rheumatology, ophthalmology and endocrinology. The psychological component of work on belief systems and common sense reasoning will be phased out, while the basic AI issues of plan recognition/generation and default reasoning will receive increased attention. Our core work will continue with emphasis on further development of the EXPERT framework and also on AI studies in representations and problems of knowledge and expertise acquisition. We also plan to continue our participation in AIM dissemination and training activities as well as our contribution -- via the RUTGERS/LCSR computer -- to the shared computing facilities of the national AIM network.

B. Justification and Requirements for Continued SUMEX Use

Continued access to SUMEX is needed for:

1) Backup for demos, etc.

2) Programs developed to serve the National AIM Community should be runnable on both facilities.

3) There should be joint development activities between the staffs at Rutgers and Sumex in order to ensure portability, share the load, and provide a wider variety of inputs for developments.

C. Needs and Plans for Other Computing Resources beyond SUMEX-AIM

Beyond the current SUMEX-AIM facility there is need for access to a more 'personal' type of computing facilities (e.g., Dolphins). In addition SUMEX might provide a high quality output device (e.g., line printer or XGP) for the community.

D. Recommendations for Future Community and Resource Development

We strongly support the efforts of Sumex staff to develop personal computing resources. Our own local planning efforts have been helped considerably by the assistance of Sumex staff. As physical facilities are spread out in more and more places, we believe that centers such as Sumex can play a growing role in keeping track of technology, planning, and technical support of users. It may also prove convenient for Sumex to serve as a center for mail and other communications activities carried out by personal computers at remote sites.

We believe that computing is going to move towards smaller machines spread around more centers. As this happens, there is some danger that the scientific community which has been built up around the Sumex-AIM computing resource will tend to become fragmented. In this context it becomes increasingly important for the AIM project to sponsor meetings and other kinds of visits involving members of the community. It also becomes important to separate Sumex's role in communications, coordination, and technical consultation from their role as suppliers of computing time. The latter role is likely to decrease outside the Stanford community as time goes on. It is therefore important to make sure that the former continues to be supported, and in fact to grow.