

classifications of unknown structures and the assignment of turns and secondary labels to regions of those structures.

D. List of Relevant Publications

Cohen, F.E., Abarbanel, R.M., Kuntz, I.D. and Fletterick, R.J.: *Secondary structure assignment for α/β proteins by a combinatorial approach*, *Biochemistry*, 22, pp 4894-4909, (October 1983).

At this time, another paper on prediction of "turns" in several classes of proteins has been accepted by *Biochemistry* for publication.

Abarbanel, R.M., Wieneke, P.R., Mansfield, E., Jaffe, D.A., Brutlag, D.L.: *Rapid searches for complex patterns in biological molecules*, *Nucleic Acids Research*, 12, pp 263-280, (January 1984).

Abarbanel, R.M.: *Protein Structural Knowledge Engineering*, *Ph.D. thesis*, University of California San Francisco, (December, 1984).

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations

None.

B. Sharing and Interactions with SUMEX Projects

This project is closely allied with the MOLGEN group, both in computer and scientific interests. Some pattern matching methodology created for the protein data base has been adopted and used in the various DNA knowledge bases. The principal persons in the MOLGEN group have contributed to this project's use and understanding of knowledge base software and resources.

C. Critique of Resource Management

Work continues on the UNIX systems at the University of California, San Francisco and on the Symbolics Lisp Machine there. SUMEX has been used primarily for communications with other researchers.

Resource management remains excellent. The staff are friendly and responsive. Network access, bulletin boards and the mail system have provided a means to collaborate with others doing related work locally as well as in Europe. SUMEX-AIM staff have been most helpful in getting this project started on the Dolphin workstations and in providing an environment where new tools have been made available for use.

III. RESEARCH PLANS

A. Project Goals and Plans

Since the funding for this project has been terminated, remaining work will be supported by Prof. I. Kuntz at UCSF. Development of the KEE based pattern matching and structure inference system continues.

In particular, at this time, an improved general sequence pattern matching facility has been implemented. A hierarchy of pattern types has been developed so that each pattern may inherit methods for evaluation and display, from common ancestor units. Evaluation of patterns and collections of patterns on the 3600 is from 4 to 10 times faster than under Franz lisp on the Vax/750 running UNIX. Display of matches has been made interactive so that the sequence is shown with mouse sensitive regions and pattern symbols allowing a user to determine the reasons for a match. This feedback allows for improved pattern design.

A KEE TellAndAsk operator is being developed that will allow the rule system to interact with the pattern matchers thus allowing inference about patterns and the suggested underlying structure.

Work will continue on this project though at a slow pace due to the other commitments of the principal investigator. As other resources become available, it is hoped that new rule sets may be developed and tested during the next project year.

B. Need for Resources

-- no comment

C. Recommendations

-- no comment

IV.C.3. REFeree Project

REFeree Project

Bruce G. Buchanan, Ph.D.
Computer Science Department
Stanford University

Byron W. Brown, Ph.D.
Dept. of Biostatistics
Stanford University

Daniel E. Feldman, Ph.D., M.D.
Department of Medicine
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. *Project Rationale*

The goal of this project is two-fold: (a) use existing AI methods to implement an expert system that can critique medical journal articles on clinical trials, and (b) in the long term, develop new AI methods that extract new medical knowledge from the clinical trials literature. In order to accomplish (a) we are building the system in three stages.

1. System I will assist in the evaluation of the quality of a single clinical trial. The user will be imagined to be the editor of a journal reviewing a manuscript for publication, but the program will be tested on a variety of readers, including clinicians, medical scientists, medical and graduate students, and clerical help.
2. System II will assist in the evaluation of the effectiveness of the treatment or intervention examined in a single published clinical trial. The user will be imagined to be a clinician interested in judging the efficacy of the treatment being tested in the trial.
3. System III will assist in the evaluation of the effectiveness of a single treatment examined in a number of published clinical trials.

B. *Medical Relevance*

The burden of "keeping up with the literature" is particularly onerous in the practice of medicine and in medical research [30, 31]. Reading the abstracts in a few journals and selecting several key articles for a rapid survey are the best that most clinicians can hope to accomplish each week. The time and effort necessary for a thorough and critical reading of even a few research reports are not available.¹ Sackett reports that to keep up with the 10 leading journals in internal medicine a clinician must read 200 articles and 70 editorials per month [31]. It was also estimated that the biomedical

¹In an informal check on this intuition two of us, with considerable training in analyzing clinical trials (BWB and DEF) timed critical readings of a five page article on a clinical trial in the New England Journal of Medicine [3]. Our times were 30 and 120 minutes.

literature is expanding at a compound rate of 6% to 7% per year, or doubling every 10 - 15 years [31, 28]. Furthermore, even if more time were available the statistical and epidemiological skills necessary for critical reading are not part of most clinicians' repertoires¹; and yet decisions about which therapy to use, what intervention to adopt, or what advice to give patients must be based on a combination of clinical experience and published literature. But the existing literature is often confusing and contradictory [20] and publication in the most prestigious medical journals does not guarantee freedom from serious methodologic flaws and erroneous conclusions [22, 8]. Any assistance to the clinician must deal with both the problem of the vastness of the literature and the quality of the research report. Similar problems are faced by the editors of medical journals, swamped with manuscripts to review and evaluate, and by research scientists and academicians trying to stay abreast of the developments in their fields. How can they cover more and yet evaluate better and more consistently? Clearly any machine assistance would be welcome.

C. Highlights of Progress

This project is just getting started.

Preliminary work has been done on REFEREE [10], a prototype expert system for determining the quality of a clinical trial report, and the efficacy of the intervention evaluated in the trial. REFEREE is written in EMYCIN, a rule-based programming language which allows rapid prototyping of a consultation system that gives advice to a user. It presupposes that a knowledge base about the problem area has been constructed, which usually involves codifying an expert's knowledge.

The basic format of a REFEREE session is fairly simple. The reader is asked a series of questions pertaining to the paper and the study described. The answers given are used to rate the overall quality of the paper and the probable efficacy of the treatment described. (See sample dialogs below).

In the first version of REFEREE, after the program has finished with its chain of questions and deductions, the quality of the paper and the efficacy of the drug are given to the user as a "merit score", an integer between 0 and 10, with 10 indicating the highest quality. Additionally, the user is provided with a series of English language messages indicating the main flaws detected in the paper. The merit score was used because the expert system makes its judgements by using a weighted average of values assigned to each aspect of the paper being critiqued. As the user answers the consultant's questions, the answers are given individual merit scores. For example, if the user's answer indicate that experimental blinding was done correctly, the paper is given a high score in the blinding category. When all merit score assignments have been made, the total merit score is calculated as a weighted average of the categorical merit scores, with those categories that are more crucial to a good paper or clinical trial being given a higher weight.

The final result of this calculation is a number between 1 and 10 which serves as a quality measure for the paper or the treatment. A 1 indicates low quality; a 10 indicates the highest quality. An integer as a final result, however, can be very cryptic. It is usually quite difficult, given just an integer, to understand or believe the findings of the consultant. It was discovered quite early that users, when presented with just the bare merit score of the paper, would want to know why the paper was rated in the way it was. For this reason, English language statements are given to the user, indicating the nature of the main flaws of the paper. In each category, if the calculated merit score is

¹A recent survey of the statistical methods used by authors in the New England Journal of Medicine indicated that 42 per cent of the articles surveyed relied on statistical analysis beyond descriptive statistics [6].

found to be less than an arbitrary minimum, this is noted in a sentence or two, and given to the user at the end of the consultation. In this way, the user not only gets an overall picture of the quality of the paper, but also an indication of the general areas in which the paper was found to be lacking.

Several problems were found in the original version of REFEREE. It was discovered that the use of a weighted average precluded the use of EMYCIN's certainty factors. Because of this, the user would often be forced to choose from a fairly limited set of possible answers to the consultant's questions. The lack of versatility implied by this constraint dictated that a new approach which could make full use of EMYCIN's certainty factors should be used.

In order to do this, the old rule base was scrapped, and a new one was written. Instead of deciding on a rating between one and ten to indicate quality, the new version simply decides whether or not the paper in question is of "high academic and scholarly quality", with an EMYCIN certainty factor modifying the conclusion. For example, in the case of a mediocre paper, the program would conclude that the paper was of "high quality", but only with a certainty of say, .5, on a scale between -1 and 1. Though the words "certainty factor" are used for historical reasons, our final number is the equivalent of a merit score.

While at first glance the two approaches seem similar, the second approach was found to be much more flexible and satisfying from the user's standpoint. Since the conclusion is in terms of the program's certainty that the paper's quality is good, the user may incorporate his or her own uncertainty into the dialogue with the program. This was accomplished by asking mainly yes/no questions, and at all times allowing the user to indicate his or her certainty in the answers given. Thus, if the program asks the user if the quality of the paper's literature review was high, he or she can answer simply "yes" or "no", indicating complete confidence in the answers, or modify a yes/no answer with a certainty factor, indicating that he or she is not completely certain. The user's answers, along with the uncertainty indicated by him or her, will be combined by EMYCIN to give a final conclusion on the paper's quality.

As an example, one of the old-style rules might have been something like this: If the user indicates that the literature review is of "poor quality", conclude that the merit of the paper is 3 with a (built-in) weight of 2. After all the merit values had been calculated, a weighted average, (using built-in weights) would be taken to come to the final merit score. In contrast, one of the new rules would be of the form: If the user gives a "yes" answer to the question "Is the literature review thorough and balanced?", conclude that the paper is of good quality with a certainty of .3. While in the first case the user was limited to a set of possible answers (e.g. excellent, good, poor), the second rule gives the user the opportunity to answer either yes or no, and qualify that answer with any degree of certainty desired. If, in the second rule, the user gives a certainty of less than 1 that the literature review was of good quality, the inferred conclusion about the quality of the paper will be automatically downgraded as well. In other words, if the user expresses uncertainty, the conclusion about the quality of the paper will be less certain.

The new approach, in addition to supplying the user with the ability to express varying degrees of uncertainty, also allows for a hierarchical question structure. At any point, if the user is unclear of the appropriate response, the program can prompt with further, more detailed questions, until a conclusion about the original question can be provided. Conversely, whenever a user is willing to give an answer, the program will refrain from dwelling on the issue and omit its long series of sub-questions. In this manner the amount of detail provided can be individualized.

This current version of REFEREE has two hundred rules and has been tested by the present research team on several papers. It is this program that will be expanded as described in Section III-A. Part of a sample consultation is shown below.

-----MEDICINE-1-----

The first paper of MEDICINE-1 will be referred to as:

-----PAPER-1-----

-----STATISTICS-1-----

- 1) What is the size of the control sample?
** 25
- 2) How many of the subjects in the control sample responded to treatment?
** 14
- 3) What is the size of the test sample?
** 23
- 4) How many of the subjects in the test sample responded to treatment?
** 23

...

-----PLANNING-1-----

- 9) Was there an explicit stopping rule defined before the experiment was run?
** N

-----RANDOMIZATION-1-----

- 10) Was there any mention of the use of randomization in patient assignment?
** Y
- 11) Was the assignment of subjects in the experiment performed blindly?
** UNK

...

-----BLINDING-1-----

- 16) Was the experiment double blinded, or was any mention made of blinding in the experiment?
** Y
- 17) Was there any mention of an effort to make the placebo and medication as similar as possible?
** N

...

The strength of the evidence indicating the efficacy of PAPER-1 is as follows:

There is some evidence for efficacy, but further study is needed.

The general quality of the paper is as follows:

The current paper is of poor quality.

The flaws of the current paper are as follows:

A stopping rule was not defined or was not adhered to in the experiment.
The measures taken to evaluate subject compliance were inadequate or non-existent.
Subjects were not randomly assigned treatment groups, seriously weakening the validity of the conclusions.
Though an effort was made to blind the experiment, the techniques used were not effective.

The final calculated efficacy of the drug as indicated by the given clinical trial (between 0 and 10, with a score of 10 being the highest) is as follows:

5.

The final merit of the current paper is as follows:

3.

- 23) Are there any other papers on MEDICINE-1?

** N

- 24) Do you want the results of this consultation output to a file?

** N

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations

Dr. D. Feldman is a physician and epidemiologist at the Stanford Center for Disease Prevention. Prof. B. Brown is currently teaching a Medical School class on reading medical journal articles.

B. Interactions with other SUMEX-AIM projects

Our interactions have all been through the Knowledge Systems Laboratory where we have discussed design and implementation issues.

C. Critique of Resource Management

The SUMEX staff has been most cooperative in helping get this project started. We have tried to place few demands on the SUMEX staff, but have received prompt answers to all questions.

III. RESEARCH PLANS

A. Goals & Plans

It is proposed to construct three computer-based expert systems to assist a variety of different readers in the evaluation of an extensive but well defined area of the medical literature, clinical trials. It is further proposed to test the hypothesis that such programs will enable a variety of users to read the literature on clinical trials more more critically and more rapidly.

The expert systems will be developed using the EMYCIN programming environment and the production rule approach followed successfully in previous expert systems [11, 17, 21, 24, 4].

The three programs to be developed are separate, but closely related:

1. System I will assist in the evaluation of the quality of a single clinical trial. The user will be imagined to be the editor of a journal reviewing a manuscript for publication, but the program will be tested on a variety of readers, including clinicians, medical scientists, medical and graduate students, and clerical help.
2. System II will assist in the evaluation of the effectiveness of the treatment or intervention examined in a single published clinical trial. The user will

be imagined to be a clinician interested in judging the efficacy of the treatment being tested in the trial.

3. System III will assist in the evaluation of the effectiveness of a single treatment examined in a number of published clinical trials.

Within the duration of this research it is also proposed to test the first two systems against unassisted evaluations by the various categories of readers. The testing will include a formal testing of the programs by comparing the speed and number of flaws found in using the program with similar measurements on unassisted reading. In addition there will be a more informal evaluation by questionnaire of the subjective impressions of users of the program, ascertaining the likelihood of routine use and the value of such a program to the user.

This proposal with its concentration on clinical trials is regarded as the initial step in a more general research goal - building computer systems to help the clinician and medical scientist read the medical literature more critically.

B. Justification for continued SUMEX use

We will continue to use SUMEX for developing the AI methods. We need EMYCIN at the moment because it provides a good environment for building a rule-based system that may grow to many hundreds of rules. EMYCIN is not available on other machines without substantial cost.

C. Need for other computing resources

In the short term we will not need additional resources. Should we decide to implement a new system in a framework other than EMYCIN, we might seek funding to buy a LISP workstation.

D. Recommendations

Although our use has been small, we find the load average on SUMEX often precludes running test cases during the day. We have no specific recommendation, but would like to have access to small amounts of high quality computer time.

IV.C.4. Ultrasonic Imaging Project

Ultrasonic Imaging Project

James F. Brinkley, M.D.
W.D. McCallum, M.D.
Depts. Computer Science, Obstetrics and Gynecology
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

This report is a summary of the overall accomplishments of the ultrasonic imaging project since it is currently being discontinued. The long range goal of this project was the development of an ultrasonic imaging and display system for three-dimensional modelling of body organs. The models would be used for non-invasive study of anatomic structure and shape as well as for calculation of accurate organ volumes for use in clinical diagnosis. Initially, the system was used to determine fetal volume as an indicator of fetal weight; later it could be adapted to measure left ventricular volume, or liver and kidney volume.

The general method we used was the reconstruction of an organ from a series of ultrasonic cross-sections taken in an arbitrary fashion. A real-time ultrasonic scanner is coupled to a three-dimensional acoustic position locating system so that the three-dimensional orientation of the scan plane is known at all times. During the patient exam a dedicated microcomputer based data acquisition system is used to record a series of scans over the organ being modelled. The scans are recorded on a video tape recorder before being transferred to a video disk. 3D position information is stored on a floppy disk file. In a later system the microprocessor will then be connected to SUMEX where it will become a slave to an AI program running on SUMEX. The SUMEX program will use a model appropriate for the organ which will form the basis of an initial hypothesis about the shape of the organ. This hypothesis will be refined at first by asking the user relevant clinical questions such as (for the fetus) the gestational age, the lie of the fetus in the abdomen and complicating medical factors. This kind of information is the same as that used by the clinician before he even places the scan head on the patient. The model will then be used to request those scans from the video disk which have the best chance of giving useful information. Heuristics based on the protocols used by clinicians during an exam will be incorporated since clinicians tend to collect scans in a manner which gives the most information about the organ. For each requested scan a two-dimensional tolerance region (or plan) derived from the model will be sent to the microcomputer. The requested scan will be retrieved from the video disk, digitized into a frame buffer, and the plan used to direct a border recognition process that will determine the organ outline on the scan. The resulting outline will be sent to SUMEX where it will be used to update the model. The scan requesting process will be continued until it is judged that enough information has been collected. The final model will then be used to determine volume and other quantitative parameters, and will be displayed in three dimensions.

We believe that this hypothesize verify method is similar to that used by clinicians when they perform an ultrasound exam. An initial model, based on clinical evidence and past experience, is present in the clinician's mind even before he begins the exam. During the exam this model is updated by collecting scans in a very specific manner which is known to provide the maximum amount of information. By building an

ultrasound imaging system which closely resembles the way a physician thinks we hope to not only provide a useful diagnostic tool but also to explore very fundamental questions about the way people see.

We developed this system in phases, starting with an earlier version developed at the University of Washington. During the first phase the previous system was adapted and extended to run in the SUMEX environment. Clinical studies were done to determine its effectiveness in predicting fetal weight. In the second phase computer vision techniques were used to solve some of the problems observed in the clinical trials on the first phase.

B. Medical Relevance and Collaboration

This project was developed in collaboration with the Ultrasound Division of the Department of Obstetrics at Stanford, of which W.D. McCallum is the director.

Fetal weight is known to be a strong indicator of fetal well-being: small babies generally do more poorly than larger ones. In addition, the rate of growth is an important indicator: fetuses which are "small-for-dates" tend to have higher morbidity and mortality. It is thought that these small-for-dates fetuses may be suffering from placental insufficiency, so that if the diagnosis could be made soon enough early delivery might prevent some of the complications. In addition such growth curves would aid in understanding the normal physiology of the fetus. Several attempts have been made to use ultrasound for predicting fetal weight since ultrasound is painless, noninvasive, and apparently risk-free. These techniques generally use one or two measurements such as abdominal circumference or biparietal diameter in a multiple regression against weight. We previously studied several of these methods and concluded that the most accurate were about +/-200 gms/kg, which is not accurate enough for adequate growth curves (the fetus grows about 200 gms/week). The method we developed is based on the fact that fetal weight is directly related to volume since the density of fetal tissue is nearly constant. As part of this research we showed that by utilizing three dimensional information more accurate volumes and hence weights can be obtained.

In addition to fetal weight, the first implementation of this system was evaluated for its ability to determine other organ volumes in vitro. In collaboration with Dr. Richard Popp of the Stanford Division of Cardiology we evaluated the system on in vitro kidneys and latex molds of the human left ventricle. Left ventricular volumes are routinely obtained by means of cardiac catheterization in order to help characterize left ventricular function. Attempts to determine ventricular volume using one or two dimensional information from ultrasound has not demonstrated the accuracy of angiography. Therefore, three-dimensional information should provide a more accurate means of non-invasively assessing the state of the left ventricle.

C. Highlights of Research Progress

This section will summarize the major accomplishments of this project during its tenure on SUMEX. These accomplishments are described in detail in the Ph.D. dissertation of J. Brinkley, which is listed in the section on recent publications. The completion of the Ph.D. is the reason this project is now being discontinued.

The initial accomplishment was development of a microprocessor-based data acquisition system for acquiring a series of ultrasound images from a patient. The data acquisition system was designed to allow data to be acquired rapidly because the patient and organ must remain motionless while data is acquired. For this reason the exam was divided into 3 passes: patient exam, data entry and data analysis.

In the first pass, video ultrasound images are acquired from a commercial ultrasound scanner and stored on a videotape recorder, while position information from the locator is stored on floppy disk. In the data entry pass these scans are recalled from the tape recorder and outlined with the light pen. In the third pass the positions and outlines are sent to SUMEX, where the data analysis occurs. Software at SUMEX generates the 3D position of all outline points and allows them to be displayed graphically.

Before it was possible to use the data it was necessary to ascertain the accuracy of the 3D points. The accuracy of 3D point determination was found to be .6 cm. Individual sources of this error were analyzed and found to come about equally from the scanner resolution and the locator. These results were reported in Brinkley, Muramatsu et al., 1982.

The 3D points form the input to the modelling system. A regular mathematical model must be fitted to the arbitrary data in order to allow accurate volumes to be calculated. Two types of modelling system were developed: a "data-driven" system, which uses simple numerical techniques to interpolate a model to the data, and a "knowledge-driven" system which uses artificial intelligence techniques to overcome many of the deficiencies in the data-driven approach.

A detailed description and engineering evaluation of the data-driven approach can be found in Brinkley, Muramatsu et al., 1982. In the data driven system a series of regularly spaced scans are fitted to whatever data is present. The computer has no knowledge of what it is looking at. Engineering evaluations of this system were done on balloons, kidneys and molds of the human left ventricle, imaged in a water bath. For all three types of objects calculated volumes were generally within 5 percent of measured volume. These results provided justification for continued development, and showed more promise than standard clinical techniques which only use one or two measurements and an assumed shape.

The data-driven system was next evaluated for its ability to predict fetal weight, first *in vitro*, then *in utero*. The *in vitro* results are described in Brinkley, McCallum et al, 1982. In this study the relationship between measured weight and measured volume for a series of 26 dead neonates was shown to be highly linear, thus justifying the use of volume as a measure of fetal weight. The ability of volumes found by head and trunk reconstructions to predict fetal weight was then determined, and found to be quite good ($R=.985$).

The system was then used to predict fetal weight *in utero* as described in Brinkley, McCallum et al, 1983. Forty-one pregnant women were imaged within 48 hours of delivery. A total of 19 ultrasonic measurements were made, including head and trunk volume by reconstruction, as well as many simpler measurements utilized in the literature. These measurements were compared with weight measured at birth. The best combination of measurements was found to be a product of three head diameters, a product of three trunk diameters and trunk volume by reconstruction, giving a standard error of 69 g/kg (against natural log of birthweight). The most popular method in the literature gave a standard error of 106 g/kg suggesting that 3D information could improve weight prediction by about 30 percent.

However, further analysis showed that if the trunk volume by reconstruction was not included the standard error was still 73 g/kg, showing that the volumes by reconstruction were not that useful. This observation led to an evaluation of some of the problems in the data driven system, which in turn led to the need for an artificial intelligence approach.

The basic problems with the data-driven system were noise, missing data and awkwardness. Missing data was especially a problem in the term fetus since it was often impossible to visualize the fetal head and neck. If these data were not present the

resulting volume would be too small because the computer had no way of knowing that it should interpolate an approximate neck or rump volume. The awkwardness came from the fact that it was necessary to outline all the scans with a light pen - usually about 90 minutes for a head and trunk reconstruction.

These problems were all related to the fact that the computer had no knowledge of what it was looking at. The goal of the knowledge-driven program was to give the computer the kind of anatomic knowledge that a radiologist utilizes in order to overcome deficiencies in the data.

The knowledge-driven system is described in Brinkley 1983 and Brinkley 1985. The system was implemented and tested on two shape classes of balloons (round and long-thin). For each balloon class a training set of similarly-shaped balloons was used to give the computer knowledge of the given shape. This training set consisted of ultrasonic reconstructions obtained by the previous system. The knowledge was then used to analyze ultrasound data from a similarly-shaped balloon which was not part of the training set. The initial input to the system consisted of the three-dimensional positions and orientations of a series of ultrasound slices. These slices were previously acquired manually and stored on a video tape recorder. The system was also given the two endpoints of the balloons, which allowed a reference coordinate system to be established. The balloon endpoints interacted with the shape knowledge to define an initial tolerance region, within which the system expected the actual balloon surface to be found. The system's best guess as to the location of the actual balloon surface was the middle of the tolerance region.

Once the initial tolerance region was established an hypothesize-verify paradigm was employed to alternately request a particular ultrasound slice, to provide a tolerance region for an edge detector on that slice, to manually acquire the border of the balloon on that slice, and to update the model by combining the new data with the shape knowledge. This process continued until it was judged that additional slices could contribute no new information.

For an example round balloon (measured volume 267 cc) the initial best guess volume after specifying the endpoints was 242 cc. After one slice best guess volume was 279 cc. After nine slices (out of a possible 30) the system judged that no more slices would be useful; best guess volume was 265 cc. For a different training set of long-thin balloons the final best guess volume for a new reconstruction, after 9 out of a possible 22 slices, was 459 cc, measured volume 461 cc. These results show that learned shape knowledge allowed the system to form a reasonable guess as to the location of the balloon surface even after only two endpoints had been specified.

The overall conclusions of this research are (1) three-dimensional ultrasound data provides accurate volumes at least *in vitro*, (2) 3D data may improve fetal weight prediction by approximately 30 percent (3) use of artificial intelligence techniques, when further developed, hold promise for greatly improving the performance of a three-dimensional organ modelling system.

D. Recent Publications

1. Brinkley, J.F., Muramatsu, S.K., McCallum, W.D. and Popp, R.L.: *In vitro evaluation of an ultrasonic three-dimensional imaging and volume system.* Ultrasonic Imaging, 4:126-139, 1982.
2. Brinkley, J.F., McCallum, W.D., Muramatsu, S.K. and Liu, D.Y.: *Fetal weight estimation from ultrasonic three-dimensional head and trunk reconstructions: Evaluation in vitro.* Amer. J. Obstet. Gynecol. 144(6):715-721, 1982.

3. Brinkley, J.F., McCallum, W.D., Muramatsu, S.K., and Liu, D.Y.: *Fetal weight estimation from lengths and volumes found by ultrasonic three-dimensional measurements*. J. Ultrasound Med. 3:163-168, 1983.
4. Brinkley, J.F.: *Learned shape knowledge in ultrasonic three-dimensional organ modelling*. Second place, student paper competition, Symposium on Computer Applications in Medical Care, Baltimore, October 23-26, 1983.
5. Brinkley, J.F.: *Ultrasonic three-dimensional organ modelling*. Ph.D. Dissertation, Stanford University, Stanford Computer Science Technical report STAN-CS-84-1001, 1984.
6. Brinkley, J.F.: *Knowledge-driven ultrasonic three-dimensional organ modelling*. To be published in *IEEE Trans. Pattern Analysis and Machine Intelligence*, Summer 1985.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Collaborations

We collaborated more with medical people than anyone else. The project was located in the Obstetrics Department at Stanford where W.D. McCallum manages the ultrasound patients. We also collaborated with Dr. Richard Popp in the Division of Cardiology at Stanford.

B. Sharing and Interactions with SUMEX projects

Mostly personal contacts with the Heuristic Programming Project and Medical Information Science Program at Stanford. The message facilities of SUMEX have been especially useful for maintaining these contacts.

C. Critique of Resource Management

In general SUMEX has been a very usable system, and the staff has been very helpful.

III. RESEARCH PLANS

A. Project Goals and Plans

The major conclusion from the research leading to the Ph.D. is that the current hardware we use for three-dimensional location is not accurate enough to permit further work on organ modelling. For this reason I have proposed several alternative methods of utilizing 3D medical image data, including 3D CT, NMR or ultrasound. All these modalities produce 3D arrays of data which would be much easier to use than arbitrary slices.

Given this type of data, fairly straightforward extensions of the model representation developed for balloons could be used for the heart or kidney. The basic idea would be to have the human operator indicate three organ landmarks within the 3D data, then let the computer utilize learned shape knowledge to selectively "biopsy" portions of the 3D data in order to define the actual organ instance. Since the data would be available as a 3D array, the edge detection process could take place along a one-dimensional tolerance region rather than on a two-dimensional slice. Since all forms of medical images are becoming available as 3D arrays this seems like a better approach than the selection of individual slices.

Depending on the interest of engineers in providing 3D data much of the AI modelling could still be done on SUMEX. Many of the AI techniques could also be developed for 2D images for knowledge-driven border detection. However, there are no plans to continue this research at present.

B. Justification and requirements for continued SUMEX use

The goals of this project seem to be compatible with the general goals of SUMEX, i.e., to develop the uses of artificial intelligence in medicine. The problem of three-dimensional modelling is a very general one which is probably at the heart of our ability to see. By developing a medical imaging system that models the way clinicians approach a patient we should not only develop a useful clinical tool but also explore some very fundamental problems in AI.

The availability of a large well supported facility like SUMEX was very useful for developing this system.

C. Needs and plans for other computing resources beyond SUMEX-AIM

Judging from our present experience it appears that SUMEX could not handle the amount of data required for image processing on digitized ultrasound scans. The recent advent of relatively powerful microprocessors and personal LISP machines makes these machines very attractive for further development. SUMEX could still act as a communications crossroads, however.

D. Recommendations

Since any further research on this project would require dedicated image processors we would hope to see these kind of systems being developed by the SUMEX resource. Projects that would be of direct interest are networks (such as ETHERNET), personal computer stations, graphics displays, etc.

IV.D. Pilot AIM Projects

Following is a description of the informal pilot project currently using the AIM portion of the SUMEX-AIM resource, pending funding, full review, and authorization.

In addition to the progress report presented here, an abstract is submitted on a separate Scientific Subproject Form.

IV.D.1. PATHFINDER Project

PATHFINDER Project

Bharat Nathwani, M.D.
Department of Pathology
University of Southern California

Lawrence M. Fagan, M.D., Ph.D.
Department of Medicine
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

Our project addresses difficulties in the diagnosis of lymph node pathology. Five studies from cooperative oncology groups have documented that, while experts show agreement with one another, the diagnosis made by practicing pathologists may have to be changed by expert hematopathologists in as many as 50% of the cases. Precise diagnoses are crucial for the determination of optimal treatment. To make the knowledge and diagnostic reasoning capabilities of experts available to the practicing pathologist, we have developed a pilot computer-based diagnostic program called PATHFINDER. The project is a collaborative effort of the University of Southern California and the Stanford University Medical Computer Science Group. A pilot version of the program provides diagnostic advice on 80 common benign and malignant diseases of the lymph node based on 150 histologic features. Our research plans are to develop a full-scale version of the computer program by substantially increasing the quantity and quality of knowledge and to develop techniques for knowledge representation and manipulation appropriate to this application area. The design of the program has been strongly influenced by the INTERNIST/CADUCEUS program developed on the SUMEX resource.

A group of expert pathologists from several centers in the U.S., have showed interest in the program and helped to provide the structure of the knowledge base for the PATHFINDER system.

B. Medical Relevance and Collaboration

One of the most difficult areas in surgical pathology is the microscopic interpretation of lymph node biopsies. Most pathologists have difficulty in accurately classifying lymphomas. Several cooperative oncology group studies have documented that while experts show agreement with one another, the diagnosis rendered by a "local" pathologist may have to be changed by expert lymph node pathologists (expert hematopathologists) in as many as 50% of the cases.

The National Cancer Institute recognized this problem in 1968 and created the Lymphoma Task Force which is now identified as the Repository Center and the Pathology Panel for Lymphoma Clinical Studies. The main function of this expert panel of pathologists is to confirm the diagnosis of the "local" pathologists and to ensure that the pathologic diagnosis is made uniform from one center to another so that the comparative results of clinical therapeutic trials on lymphoma patients are valid. An expert panel approach is only a partial answer to this problem. The panel is

useful in only a small percentage (3%) of cases; the Pathology Panel annually reviews only 1,000 cases whereas more than 30,000 new cases of lymphomas are reported each year. A Panel approach to diagnosis is not practical and lymph node pathology cannot be routinely practiced in this manner.

We believe that practicing pathologists do not see enough case material to maintain a high-level of diagnostic accuracy. The disparity between the experience of expert hematopathology teams and those in community hospitals is striking. An experienced hematopathology team may review thousands of cases per year. In contrast, in a community hospital, an average of only 10 new cases of malignant lymphomas are diagnosed each year. Even in a university hospital, only approximately 100 new patients are diagnosed every year.

Because of the limited numbers of cases seen, pathologists may not be conversant with the differential diagnoses consistent with each of the histologic features of the lymph node; they may lack familiarity with the complete spectrum of the histologic findings associated with a wide range of diseases. In addition, pathologists may be unable to fully comprehend the conflicting concepts and terminology of the different classifications of non-Hodgkin's lymphomas, and may not be cognizant of the significance of the immunologic, cell kinetic, cytogenetic, and immunogenetic data associated with each of the subtypes of the non-Hodgkin's lymphomas.

In order to promote the accuracy of the knowledge base development we will have participants for multiple institutions collaborating on the project. Dr. Nathwani will be joined by experts from Stanford (Dr. Dorfman), St. Jude's Children's Research Center -- Memphis (Dr. Berard) and City of Hope (Dr. Burke).

C. Highlights of Research Progress

C.1 Accomplishments This Past Year

Since the project's inception in September, 1983, we have constructed several versions of PATHFINDER. The first several versions of the program were *rule-based* systems like MYCIN and ONCOCIN which were developed earlier by the Stanford group. We soon discovered, however, that the large number of overlapping features in diseases of the lymph node would make a rule-based system cumbersome to implement. We next considered the construction of a *hybrid system*, consisting of a rule-based algorithm that would pass control to an INTERNIST-like scoring algorithm if it could not confirm the existence of classical sets of features. We finally decided that a modified form of the INTERNIST program would be most appropriate. The original version of PATHFINDER is written in the computer language Maclisp and runs on the SUMEX DEC-20. This was transferred to Portable Standard Lisp (PSL) on the DEC-20, and later transferred to PSL on the HP 9836 workstations. Two graduate students, David Heckerman and Eric Horvitz, designed and implemented the program.

C.1 The PATHFINDER knowledge base

The basic building block of the PATHFINDER knowledge base is the disease profile or *frame*. The disease frame consists of *features* useful for diagnosis of lymph node diseases. Currently these features include histopathologic findings seen in both low- and high-power magnifications. Each feature is associated with a list of exhaustive and mutually exclusive *values*. For example, the feature *pseudofollicularity* can take on any one of the values *absent*, *slight*, *moderate*, or *prominent*. These lists of values give the program access to *severity* information. In addition, these lists eliminate obvious interdependencies among the values for a given feature. For example, if *pseudofollicularity* is *moderate*, it cannot also be *absent*.

Evoking strengths and frequencies are associated with each feature-value pair in a

disease profile. We are experimenting with different scales for scoring each feature-value pair, and several methods for combining the scores to form a differential diagnosis. A disease-independent import is also assigned to each feature-value but only a two-valued scale is used. This is because, in PATHFINDER, imports are only used to make boolean or yes/no decisions (see below). In addition to import, PATHFINDER utilizes the concept of *classic* features for a disease -- within each disease frame, the pathologist marks those feature-value pairs which are considered to be part of the classic pattern of the disease.

The PATHFINDER knowledge base contains information about obvious association between features. This information is of the form: "Don't ask about feature x unless feature y has certain values." For example, it wouldn't make sense to ask about the degree or range of follicularity if there are no follicles in the tissue section. The feature links also serve to identify interdependencies among features. Feature interdependence is a problem because it can lead to inaccuracies in scoring hypotheses.

The prototype knowledge base was constructed by Dr. Nathwani. During the beginning part of 1984, we organized two meetings of the entire team including the pathology experts to define the selection of diseases to be included in the system, and the choice of features to be used in the scoring process.

D. Publications Since January 1984

Horvitz, E.J., Heckerman, D.E., Nathwani, B.N. and Fagan, L.M.: *Diagnostic Strategies in the Hypothesis-directed PATHFINDER System, Node Pathology*. HPP Memo 84-13. Proceedings of the First Conference on Artificial Intelligence Applications, Denver, Colorado, Dec., 1984.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations and Program Dissemination via SUMEX

Because our team of experts are in different parts of the country and the computer scientists are not located at the USC, we envision a tremendous use of SUMEX for communication, demonstration of programs, and remote modification of the knowledge base. The proposal mentioned above was developed using the communication facilities of SUMEX.

B. Sharing and Interaction with Other SUMEX-AIM Projects

Our project depends heavily on the techniques developed by the INTERNIST/CADUCEUS project. We have been in electronic contact and have met with members of the INTERNIST/CADUCEUS project, as well as, been able to utilize

information and experience with the INTERNIST program gathered over the years through the AIM conferences and on-line interaction. Our experience with the extensive development of the pathology knowledge base utilizing multiple experts should provide for intense and helpful discussions between our two projects.

The SUMEX pilot project, RXDX, designed to assist in the diagnosis of psychiatric disorders is currently using a version of the PATHFINDER program on the DEC-20 for the development of early prototypes of future systems.

C. Critique of Resource Management

The SUMEX resource has provided an excellent basis for the development of a pilot project. The availability of a pre-existing facility with appropriate computer languages, communication facilities (especially the TYMNET network), and document preparation facilities allowed us to make good progress in a short period of time. The management has been very useful in assisting with our needs during the start of this project.

III. RESEARCH PLANS

A. Project Goals and Plans

Collection and refinement of knowledge about lymph node pathology

The knowledge base of the program is about to undergo revision by the expert, and then will be extensively tested. A logical next step would be to extend the program to clinical settings, as well as possible extensions of the knowledge base.

Other possible extensions include: developing techniques for simplifying the acquisition and verification of knowledge from experts, creating mapping schemes that will facilitate the understanding of the many classifications of non-Hodgkin's lymphomas. We will also attempt to represent knowledge about special diagnostic entities, such as multiple discordant histologies and atypical proliferations, which do not fit into the classification methods we have utilized.

Representation Research

We hope to enhance the INTERNIST-1 model by structuring features so that overlapping features are not incorrectly weighted in the decision making process, implementing new methods for scoring hypotheses, and creating appropriate explanation capabilities.

B. Requirements for Continued SUMEX Use

We are currently dependent on the SUMEX computer for the use of the program by remote users, and for project coordination. We have transferred the program over to Portable Standard Lisp which is used by several users on the SUMEX system. While the switch to workstations has lessened our requirements for computer time for the development of the algorithms, we will continue to need the SUMEX facility for the interaction with each of the research locations specified in our NIH proposal. The HP equipment is currently unable to allow remote access, and thus the program will have to be maintained on the 2060 for use by all non-Stanford users.

C. Requirements for Additional Computing Resources

Most of our computing resources will be met by the 2060 plus the use of the HP9836 workstation. We will need additional file space on the 2060 as we quadruple the size of our knowledge base. We will continue to require access to the 2060 for communication purposes, access to other programs, and for file storage and archiving.

D. Recommendations for Future Community and Resource Development

We encourage the continued exploration by SUMEX of the interconnection of workstations within the mainframe computer setting. We will need to be able to quickly move a program from workstation to workstation, or from workstation back and forth to the mainframe. Software tools that would help the transfer of programs from one type of workstation to another would also be quite useful. Until the type of workstations that we are using in this research becomes inexpensive (\$5000 or less), we will continue to need a machine like SUMEX to provide others with a chance to experiment with our software.

IV.D.2. RXDX Project

RXDX Project

Robert Lindsay, Ph.D.
Michael Feinberg, M.D., Ph.D.
Manfred Kochen, Ph.D.
University of Michigan
Ann Arbor, Michigan

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

We are developing a prototype expert system that could act as a consultant in the diagnosis and management of depression. Health professionals will interact with the program as they might with a human consultant, describing the patient, receiving advice, and asking the consultant about the rationale for each recommendation. The program uses a knowledge base constructed by encoding the clinical expertise of a skilled psychiatrist in a set of rules and other knowledge structures. It will use this knowledge base to decide on the most likely diagnosis (endogenous or nonendogenous depression), assess the need for hospitalization, and recommend specific somatic treatments when this is indicated (e.g., tricyclic antidepressants). The treatment recommendation will take into account the patient's diagnosis, age, concurrent illnesses, and concurrent treatments (drug interactions).

B. Medical Relevance and Collaboration

There has been a growing emphasis in American psychiatry on careful diagnosis using clearly defined clinical criteria (Feighner, et al., 1972; Spitzer, et al., 1975, 1980; Feinberg and Carroll, 1982, 1983). These efforts have led to several sets of criteria for the diagnosis of psychiatric disorders. The "St. Louis" criteria (Feighner, et al., 1972) were succeeded by the Research Diagnostic Criteria (RDC), formulated by researchers from St. Louis and New York (Spitzer, et al., 1975). The RDC led directly to the criteria that are now quasi-official in American psychiatry, DSM-III (Spitzer, et al., 1980). All of these criteria lists were based on a combination of clinical opinion and literature review, and use a decision-tree approach to making a diagnosis. These diagnostic systems have been shown to be acceptably reliable, but their validity remains untested. Other groups have used a multivariate statistical approach to diagnosis. Roth and his colleagues (Carney, et al., 1965) published a discriminant index for distinguishing "endogenous" from "neurotic" depressed patients. This work was repeated by Kiloh, et al. (1972) with much the same results, confirming the findings of Carney, et al. (1965).

We have done similar work, deriving two discriminant indices for separating endogenous depressed patients (unipolar or bipolar) from nonendogenous (neurotic) patients. We cross-validated these indices in separate groups of patients, and also validated them against an external standard, the dexamethasone suppression test (Feinberg and Carroll, 1982, 1983). At the same time, we and others have been further developing this and other biological measures that may differentiate between patients with endogenous and nonendogenous depression. These include neuroendocrine tests such as the dexamethasone suppression test (DST) and quantitative studies of sleep using EEG. Carroll, et al. (1981) have shown that the DST is abnormal in about 67%

of patients with endogenous depression (melancholia) and only 5-10% with nonendogenous (neurotic) depression. Kupfer, et al. (1978) and Feinberg, et al. (1982) have similar results with EEG studies of sleep. These biological markers may be useful for routine clinical use, and can certainly be used as external validating criteria to test the performance of different clinical diagnostic methods, including those mentioned above. Furthermore, we have developed biological criteria for "definitely endogenous" depression and "definitely nonendogenous" depression based on DST and sleep EEG. (Carroll, et al., 1980). Our goal is to use these criteria as an external validating criterion for assessing the performance of various new or different diagnostic schemes, in particular an expert system of the sort we are developing.

C. Highlights of Research Progress

We examined two other SUMEX-based psychiatry projects, the BLUEBOX project of Mulsant and Servan-Schreiber (1984), and the HEADMED project of Heiser and Brooks (1978, 1980). Mulsant and Servan-Schreiber visited us at Michigan and discussed the rationale and progress of their project. Heiser also visited with us and agreed to collaborate with our project as a consultant.

At Michigan, we encoded the Hamilton Rating Scale (Hamilton, 1967) into EMYCIN rules. This is the standard scale (in English) for rating the severity of depression, and many of the items in it are relevant to our consultant program. We moved our work to the AGE system, breaking the Hamilton scale into its component subscales and adding other components to determine patient demographic information, personal and family psychiatric history, and other rating scale information. We then introduced other knowledge sources to construct a differential diagnosis list for psychiatric illnesses based on our expert's taxonomy and methods. We are now focussing on rules that discriminate endogenous from non-endogenous depression. Concurrently we are developing a treatment knowledge base on a LISP workstation. Thus far, the treatment knowledge base contains information about drug therapies, including types, dosages, activities, interactions, and side effects.

We have conducted interviews with patients recently admitted to the University of Michigan Adult Psychiatric Hospital. They are interviewed by Feinberg and the interviews are observed by Lindsay plus a group of psychiatric residents, psychiatrists and psychologists. After the interview, Feinberg is debriefed by Lindsay, and then the others discuss the case. These data are the initial source of the expert knowledge base for our consultant.

D. List of Relevant Publications

This project has not yet produced any publications. The following list contains the references cited above, including our previous publications relevant to the RxDx Project.

1. Carney, M. W. P., Roth, M. and Garside, R. F.: *The diagnosis of depressive syndromes and the prediction of ECT response*, Brit. J. Psychiatry, 111, 659-674, 1965.
2. Carroll, B. J., Feinberg, M., Greden, J. F., Haskett, R. F., James, N. McL., Steiner, M., and Tarika, J.: *Diagnosis of endogenous depression: Comparison of clinical, research, and neuroendocrine criteria*, J. Affect Dis., 2, 177-194, 1980.
3. Carroll, B. J., Feinberg, M., Greden, J. F., Tarika, J., Albala, A. A., Haskett, R. F., James, N. McL., Kronfol, Z., Lohr, N., Steiner, M., de Vigne, J-P, and Young, E.: *A specific laboratory test for the diagnosis of melancholia, Standardization, validation, and clinical utility*. Arch. Gen. Psychiatry, 38, 15-22, 1981.

4. Feighner, J. P., Robins, E., Guze, S. B., Woodruff, R. A., Winokur, G., and Munoz, R.: *Diagnostic criteria for use in psychiatric research*, Arch. Gen. Psychiatry, 26, 57-63, 1972.
5. Feinberg, M. and Lindsay, R. K.: *Expert systems*. Proceedings of the NCDEU Annual Meeting, Key Biscayne, Florida, May 1985.
6. Feinberg, M. and Carroll, B. J.: *Separation of subtypes of depression using discriminant analysis: I. Separation of unipolar endogenous depression from non-endogenous depression*, Brit. J. Psychiatry, 140, 384-391, 1982.
7. Feinberg, M. and Carroll, B. J.: *Separation of subtypes of depression using discriminant analysis. II. Separation of bipolar endogenous depression from nonendogenous ("neurotic") depression*, J. Affective Disorders, 5, 129-139, 1983.
8. Feinberg, M. and Carroll, B.J.: *Biological markers for endogenous depression in series and parallel*, Biological Psychiatry 19:3-11, 1984.
9. Feinberg, M. and Carroll, B.J.: *Biological and nonbiological depression*, Presented at Annual Meeting of the Society of Biological Psychiatry, Los Angeles, May, 1984, Abstract #81.
10. Feinberg, M., Gillin, J. C., Carroll, B. J., Greden, J. F., and Zis, A. P.: *EEG studies of sleep in the diagnosis of depression*, Biological Psychiatry, 17, 305-316, 1982.
11. Heiser, J. F. and Brooks, R. E.: *Design considerations for a clinical psychopharmacology advisor*, Proc. Second Annual Symp. on Computer Applications in Medical Care. New York: IEEE, 1978, 278-285.
12. Heiser, J. F. and Brooks, R. E.: *Some experience with transferring the MYCIN system to a new domain*, IEEE Trans. on Pattern Analysis and Machine Intelligence, PAMI-2, No. 5, 477-478, 1980.
13. Kiloh, L. G., Andrews, G., and Neilson, M.: *The relationship of the syndromes called endogenous and neurotic depression*, Brit. J. Psychiatry, 121, 183-196, 1972.
14. Kupfer, D. J., Foster, F. G., Coble, P., McPartland, R. J., and Ulrich, R. F.: *The application of EEG sleep for the differential diagnosis of affective disorders*, Am. J. Psychiatry, 135, 69-74, 1978.
15. Mulsant, B. and Servan-Schreiber, D.: *Knowledge engineering: A daily activity on a hospital ward*, Computers in Biomedical Research, 1984.
16. Spitzer, R. L., Endicott, J. and Robins, E.: *Research diagnostic criteria*, (2d ed.) New York State Department of Mental Hygiene, New York Psychiatric Institute, Biometrics Research Division, 1975.
17. Spitzer, R. L.: (Ed.) *Diagnostic and statistical manual of mental disorders*, (3d ed.). Washington, D. C.: American Psychiatric Association, 1980.
18. Van Melle, W.: *The EMYCIN Manual*, Computer Science Department, Stanford University, Report HPP-81-16, 1981.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaboration and Program Dissemination via SUMEX

We have established via SUMEX a community of researchers who are interested in AI applications in psychiatry. We also have used the message system to communicate with other AI scientists at SUMEX and elsewhere.

B. Sharing and Collaboration with other SUMEX-AIM Projects

Our use of EMYCIN and AGE has been of major importance. In addition, we have worked with Dr. Larry Fagan to learn about his Pathfinder program. We used that program, on SUMEX, to obtain some information for the RxDx project by applying it to data we previously collected on depression symptom frequencies.

C. Critique of Resource Management

We have been using EMYCIN and AGE in our work, and have found these programs very valuable, saving us many hours of programming in LISP. There are some problems with them, many of which center around discrepancies between the versions described in the manuals and the versions actually running on SUMEX. We would suggest that software be more strongly supported than is now the case, if it and SUMEX are to be even more useful to beginners in AI in Medicine.

SUMEX itself has been invaluable. We don't have ready access to any other machine of equal computing power which also has a strongly supported LISP available. Specifically, the LISP compiler available on the Amdahl 5860 here differs from those used at major AI centers such as Stanford and MIT. We have also made good use of the ARPANET connections that SUMEX offers. Feinberg spent a month of his sabbatical working with Prof. Peter Szolovits at MIT, learning about AI in Medicine. This visit was arranged using computer mail through SUMEX. Lindsay and Feinberg were able to continue their collaborative work while the latter was in Cambridge, using the same medium. The alternative would have been days lost in the mails and many dollars spent on phone calls. We have also been able to get help with problems that arise with EMYCIN and AGE using computer mail.

Most of the limitations of SUMEX, and they are often severe, derive from the necessity to access it via TYMNET. Response time is often impossibly slow, and even at its best the delays are annoying and frustrating, even for editing and debugging. For example, editing is limited to a primitive line editor, since EMACS interacts with the network XON/XOFF handshaking in a disastrous way. The staff has not been helpful in solving these network related problems, probably because they do not have to live with them in their own interactions with the system. In any case, many of the problems are beyond the reach of the Sumex staff. The future of long-haul network collaborations depends critically on increased bandwidth and faster response times.

It would have been helpful to us to obtain the AGE system that runs on a Xerox 1108. However, the \$530 price, though perhaps modest in comparison to its development costs, was beyond the reach of our budget. It would be helpful if distribution costs for software could be held under \$100.

III. RESEARCH PLAN

A. Project Goals and Plans

Our immediate objective is to develop an expert system that can differentiate patients with the various subtypes of depressive disorder, and prescribe appropriate treatment. This system should perform at about the level of a board-certified psychiatrist, i.e. better than an average resident but not as well as a human expert in depression. Eventually, we plan to enlarge the knowledge base so that the expert system can diagnose and prescribe for a wider range of psychiatric patients, particularly those with illnesses that are likely to respond to psychopharmacological agents. We will design the system so that it could be used by non-medical clinicians or by non-psychiatrist MD's as an adjunct to consultation with a human expert. We plan also to focus on problems of the user interface and the integration of this system with other databases.

B. Justification and Requirements for continued SUMEX use

The access to SUMEX resources is essentially our sole means of maintaining contact with the community of researchers working on applications of AI in medicine. Although we plan to move our system to local workstations as soon as we are able, the communications capability of SUMEX will continue to be important.

We anticipate that our requirements for computing time and file space will continue at about the same level for the next year.

C. Needs and Plans for Other Computing Resources

As our project evolves and we run into the limitations of the time-shared SUMEX facility, we anticipate employing different expert systems software. At this time, we are not at a stage to say exactly what that will be, but our project is not sufficiently large that we will be able to mount such a software development project ourselves, so we will depend on development and support elsewhere. Ultimately, when our consultant is made available for field trials and clinical use, it will need to be transported to a personal computer that is large enough to support the system yet inexpensive enough to be widely available. A LISP machine is an obvious candidate. While current prices of the necessary hardware are too high, computer prices are continuing to drop. Our design strategy is to avoid limiting ourselves and our aspirations to that which is affordable today; instead we will attempt to project the growth of our project and the price-performance curve of computing such that they meet at some reasonable point in the future.

D. Recommendations for Future Community and Resource Development

Valuable as the present SUMEX facilities are to us, they are in many ways limited and awkward to use. The major limitation we feel is the difficulty and sometimes the impossibility of making contact with everyone who could be of value to us. We hope that greater emphasis will be put on internetwork gateways. It is important not only to establish more of these, but to develop consistent and convenient standards for electronic mail, electronic file transfers, graphic information transfer, national archives and data bases, and personal filing and retrieval (categorization) systems. The present state of the art feels quite limiting, now that the basic concepts of computer networking have become available and have proved their potential.

We expect that the role of the SUMEX-AIM resource will continue to evolve in the direction of increased importance of communication, including graphical information, electronic dissemination of preprints, and database and program access. The need for computer cycles on a large mainframe will diminish. We hope to have continued access to the system for communication, but do not anticipate continued use of it as a LISP computation server beyond the next year or eighteen months.