Reprinted from *Artificial Intelligence in Medicine,* **3:** 3-20, 1991

Flexible Reasoning about Patient Management using Multiple Models *

William J. Long
M.I.T. Laboratory for Computer Science
545 Technology Square, Room 420A
Cambridge, MA 02139, USA
(617) 253-3508

Keywords: multiple models, causal model, physiologic model, reasoning operator, diagnosis, effect prediction

1 Abstract

This paper reports on the Heart Failure Program, which uses multiple models and multiple reasoning operators to provide patient management information for physicians. The program uses a causal probabilistic knowledge base of pathophysiology for reasoning diagnostically, a quantitative physiologic model for reasoning about the effects of interventions, and a case base for an alternate form of diagnostic reasoning. Using these knowledge bases are reasoning operators to turn patient data into evidence for causal reasoning, using the evidence to assert specific physiologic states, generating a diagnosis or differential from the causal knowledge base or from the case base, using the current diagnostic state to determine what further information would be useful, using the diagnostic state to suggest therapies, predicting the possible effects of therapies, and using the diagnostic hypotheses or the effect predictions to generate graphical explanations for the user. By combining the models and reasoning methods, each potential use of the program benefits because each operator provides conclusions that simplify the task of other operators. The result is a program with uses in many phases of the patient management process.

2 Introduction

Over the past few years we have been developing a program to assist the physician in reasoning about the diagnosis and management of patients with cardiovascular disease characterized by manifestations of heart failure[16]. Since heart failure just means that the disease process makes cardiac output inadequate for the demands of the body, there are many possible causes. This domain is particularly rich in opportunities to reason from causal models because the manifestations are primarily the result of the compensatory mechanisms of the cardiovascular system. When cardiac output becomes inadequate, the system alters the capacitance and volume of fluid compartments to increase the heart input pressure (preload) in an attempt to increase cardiac output. This increased preload, propagated back to the lungs and venous system may lead to the pulmonary congestion or peripheral edema clinically recognized as heart failure. Since a number of disease states can produce the same general picture, the determination of the source of the problem in a particular patient fits very naturally into a paradigm of causal reasoning — in this case linking causes to observed effects to produce a causal explanation. Similarly, reasoning about the effects of an intervention involves causal reasoning because the compensatory mechanisms that produced the manifestations are also affected by the interventions and the overall result depends on how these mechanisms change the physiologic state of the patient.

Reasoning about the causes or effects of the patient state involves reasoning about the mechanisms of the cardiovascular system and the pathophysiology of the diseases. The mechanisms are sufficiently important that we have organized the reasoning of the Heart Failure Program around an integrated physiologic model of the cardiovascular system. The model has probabilistic causal relations and a case base for reasoning diagnostically and quantitative relations for predicting the

*The research reported here was done with my medical collaborators at New England Medical Center, Shapur Naimi, M. G. Criscitiello, Stephen Pauker, and in earlier years of the project, Greg Larsen, Robert Jayes, and Steven Kurzrok. They are responsible for the medical content of the program, although any mistakes in this paper are my own. The work reported here has been supported by National Institutes of Health grants R01 HL 33041 from the National Heart, Lung, and Blood Institute and R01 LM 04493 from the National Library of Medicine.

effects of therapy. This model acts as the organizing structure for the various reasoning operators. The operators use the appropriate part of the model as the knowledge base of medical knowledge to carry out their function and as a template for building the *patient specific model* (PSM) that records the accumulating conclusions about the patient. The organization of the knowledge base and the operators is sketched in figure 1. The operators are identified by boxes and the data they use and produce are the ovals. The three parts of the model are the ovals with heavy borders. This diagram identifies the main interactions but is not intended to be complete. For example, the diagnostic causal model is actually used to some extent by all of the operators.

Figure 1 goes here.

When the program is used to reason about a patient, the diagnostic causal model is the repository for the knowledge to understand the findings. As input data is gathered into the initial PSM by the evidence generator, the model is used to organize and evaluate that data as evidence for the physiologic state of the patient. Some of the input data can be used directly to assert aspects of the physiologic state of the patient. The state determining operator uses logical implication to find such instances and assert them as part of the PSM. However, most aspects of the patient state remain unknown, with multiple possible explanations for the findings.

The next step is reasoning about a differential diagnosis. There are two operators that can do this. The first, the probabilistic diagnostic reasoner, uses the findings and known states in the PSM to identify the likely overall causes and then builds suitable causal explanations for the findings from those causes. These causal hypotheses can be compared by computing their probabilities. The set of ranked causal hypotheses is the differential diagnosis for the case.

The second method of diagnosis is the case-based diagnosis program, CASEY, developed by Phyllis Koton[9, 10]. This makes use of the data base of cases to find a close match to the findings and then uses the physiologic model to resolve the differences, if they are minor. This produces a diagnosis, but depends on there being a close match to the case in the case base.

The differential diagnosis presents alternate explanations for the same findings. Even a single diagnostic hypothesis leaves many physiologic parameter states unknown because there is no need to hypothesize a state to explain the findings. However, it is often important to know the state of these parameters to manage the patient. The diagnosis refinement part of the information suggestion operator uses individual hypotheses and the differential to determine what further input data would be useful for refining the diagnosis.

Once the gaps in the causal disease description have been filled, the next step is to select therapies to manage the patient. This is a two stage process. First, the information suggestion operator looks through the PSM and diagnosis and determines what therapies might be appropriate. Since many of the diseases that cause heart failure are not themselves curable (short of a heart transplant), the appropriate approach is to look for therapies that have the potential to break the causal chains that are producing the undesirable effects.

Since the desired effects of therapies may not be the only effects and even the desired effects may trigger multiple compensatory mechanisms, the second stage of therapy selection is to predict the overall effect of the therapies. This is done using quantitative constraint equations that relate the values of the parameters and using signal flow analysis to predict the effects of changes in the parameter values. Thus, the therapy prediction operator is able to predict the likely changes in the major parameters in the model given a change in one or more of the possible therapies.

The final operator produces the explanations. This provides graphical explanations for both disease hypotheses and therapy effects. The method for explaining a disease hypothesis is simply to graph the causal relations among the physiologic state nodes and the abnormal input findings in the hypothesis. This displays all of the causal mechanisms involved in producing the input findings and is an effective way of reviewing the hypothesis for logical consistency. Explaining the therapy predictions is done by identifying the pathways through the model that have the greatest influence on the predicted change of a parameter of interest. In the process, the explanation identifies parameters that are important in determining the change.

Thus, these operators work together with the various aspects of the knowledge base to provide mechanisms to assist the physician in the diagnosis of the patient, the refinement of that diagnosis, the selection of therapy, predicting what the therapy might do, and understanding the justifications for the reasoning.

The Heart Failure Program is still in an active state of development, with some parts more thoroughly developed than others. The diagnosis algorithm has been developed over a couple of years and has been applied to a hundred or more cases with considerable success. The therapy prediction algorithm has also undergone considerable development, although the testing has been less extensive, primarily because it is difficult to collect suitable cases to really test it. The case based reasoner has been tested with a case base of forty cases and collection efforts have been completed to test it on a case set of 240 cases. The information and therapy selection operators are simpler and have not been tested in any formal way. The program is implemented on a Symbolics 3650 in Lisp.

This paper will discuss the structure of the knowledge base and then discuss the different operators, how they make use of the information in the knowledge base, and how they interface with one another.

3 Knowledge Base

There are three different models that make up the knowledge base of the Heart Failure Program. Two of these are integrated and form a static physiologic knowledge base and the third is built dynamically from cases. The first is the diagnostic causal model. This model consists of pathophysiologic states connected by probabilities representing the likelihood of causation. The second is the quantitative physiologic parameter model used for therapy prediction. It consists of parameters related by equations representing the constraints between the parameters. The final model is the case base of completed diagnoses built by the case based reasoner and organized for matching against new cases. It is also used for diagnosis and offers an alternate way of producing a diagnosis.

The following sections give a description of each of these models and highlight the relationships that bind them into a unified whole.

3.1 Diagnostic Causal Model

The diagnostic causal model is a clinical level physiologic model of the cardiovascular system. The intent of the model is to represent the causal relations needed to generate explanations of the findings corresponding to the physician's understanding of the physiologic mechanisms. The model can be divided in several ways: parameters versus states, states versus measures, and knowledge base versus PSM.

The model consists of *parameters* representing diseases such as myocardial infarction (heart attack), therapies such as hydralazine (which reduces blood pressure), as well as the physiologic parameters such as heart rate, cardiac output, pressures in the left and right heart, renal function, and so forth. The disease parameters are either primary causes or the important diagnostic entities in the model. The primary causes do not require further causal explanation, although some do have causal explanations within the model (*e.g.* anemia, below). Therapies are included in the model because they often have side effects that contribute to the patient condition, may be needed to explain the absence of expected findings, and because they provide the knowledge base for the therapy selection process. Most of the parameters in the model are the intermediate physiologic parameters that detail the causal mechanisms from the diseases to the findings. In this model there are often lengthy causal chains sometimes exceeding ten parameters in length. The primary reason for the long causal chains is the need to reason about the effects of therapy as well as do diagnostic reasoning when multiple diseases are involved. Thus, the parameters are the interface between the diagnostic model and the model for therapy prediction.

The basic structures in the diagnostic model represent qualitative *states* of the parameters. The qualitative states are typically *present* and *absent* for diseases and other conditions and *low*, *normal*, and *high* for physiologic parameters. These parameter states are linked by probability relations. The probabilistic representation was chosen because of the nature of the information available to make a diagnosis. The quantitative and logical relationships among the parameters do not provide enough constraint to allow diagnosis or even very many useful conclusions, but the experiential knowledge of cardiologists about the frequency of diseases and their effects provides the knowledge necessary to draw useful conclusions in a probabilistic framework. Besides the parameter state nodes, representing the physiology, there are *measures* representing the observables: the history items, the symptoms, the physical exam findings, and laboratory results. These measure structures provide the interface for the interpretation of patient input data.

One of the parameter states is the presence of *anemia*. Part of the definition of *anemia* in the file that produces the knowledge base is:

```
(defstate anemia
        causes (primary (0.05 (prange age 0.04 70 0.08))
                P+ (renal-insufficiency-chronic prob 0.3))
        measure ((CBC (prob anemic 1.0))
                  (dyspnea (prob on-exertion 0.5)) ...))
```

This states that the probability of anemia without other causes is 0.04 if age is less than 70 and 0.08 if age is greater. If the age is unknown, 0.05 is used. If chronic renal insufficiency (kidney disease) is present, the probability that it causes anemia is 0.3. The $P+$ indicates that it is a possible cause for the anemia. For some states there are also worsening factors, that can increase the likelihood of a state, but not cause it themselves, and correcting factors (usually therapies), that decrease the probability of a state. The *measures* are the observable facts about the patient. Values of the measures, such as dyspnea on exertion (shortness of breath), are linked to the parameter states by probabilities. The definition specifies that anemia causes dyspnea on exertion 50% of the time and would always be apparent if a CBC (complete blood count) were done. The *measures* also have definitions which contain such things as the probability that abnormal values might exist without there being a cause within the model, *e.g.*, dyspnea on exertion can exist for reasons outside of the domain of the model. The measures themselves are used in the input menu to gather the patient data.

The measure for complete blood count has the following definition:

```
(defmeasure CBC
          type multi-value
          format (:case :val (normal :val :meas)(t :val))
          specificity (high-WBC .9 high-HCT .8)
          constraints (xor normal (* (xor anemic high-HCT) high-WBC)))
```

This is a test that can provide information about both the white cell count (WBC) and the hematocrit (HCT), so it can have more than one value and the legal combinations of values it can have are determined by the *constraints* clause. Both high WBC and high HCT can be caused by conditions that are outside of the domain of the model, so the *specificity* clause specifies the fraction of cases that need to be explained by the model. The program uses this number to estimate the prevalence of this condition as a primary entity. The *format* clause provides the information needed to print the values of the CBC measure as part of any textual representation.

In addition to the parameter, state, and measure structures, which constitute the fixed diagnostic knowledge base, there are structures that are computed from these when the model is loaded to generate the enhanced model. The enhanced model adds structures to represent the links between the states and structures to represent the causal paths through the states. The links are used to record the conditions of causality and simplify reasoning about causes and effects. The paths are generated to speed up the process of diagnosis and reasoning about the probabilities of states. Because of the high degree of connectivity in the model, there are many more path structures than any other type of structure but these allow for operations such as rapid intersections of causal paths and hypotheses that make diagnosis feasible. When the model is loaded, there are many additional slots in the structures that are computed from the rest of the knowledge base, such as the prevalence of findings (above) that make the job of the reasoning operators feasible.

The PSM is generated from the enhanced model to represent the patient when data is entered. The parts of the PSM used in diagnosis consist of structures representing the *values* of the measures entered on input, copies of the parameter states called *nodes*, and links connecting the nodes to each other and to the measure values representing the specific relations including the probabilities between nodes and the measure values as constrained by the input. Thus, the PSM consists of a personalized set of node and value structures connected by link structures, ready to accept the results of the reasoning operators.

It would be possible to include tables of probabilities for each combination of the possible values of the incoming links to each node but in practice a default rule for combining the probabilities on individual links is adequate. In fact, actual data on the probabilities of combinations of causes is very sparse. The combining rule used is the "noisy-or"[19] except for worsening factors, which require another cause, and correcting factors, which decrease the probability. Thus, if causes are P, worsening factors W, correcting factors C, and primary probability is p_0, the probability of a node is:

$$\exists i | i \in P \wedge (p_i = 1.0) \Rightarrow 1.0, \exists i | i \in P \Rightarrow (1 - \prod_{i \in P,W} (1 - p_i)) \prod_{i \in C} (1 - p_i), else \Rightarrow p_0$$

Similarly, each measure value has a probability of being produced by a subset of the nodes in the model, which is computed in the same way.

The diagnostic part of the model covers all of the common causes of heart failure. It also has non-cardiac diseases that cause the same symptoms or complicate the hemodynamic situation such as pulmonary, renal, liver, or thyroid diseases, anemia and infection. The model has been designed for clinical relevance. We have included the parameters and states that make sense to the clinician and provide the significant distinctions for diagnosis and therapy. As a result, diagnostic hypotheses explain findings in terms that make sense to the physician and the therapy predictions are easily related to clinical concerns.

3.2 Prediction Constraint Model

The therapy prediction model uses quantitative constraint equations to specify the relations among the physiologic parameters and consists of a subset of the parameters in the diagnostic model. Since the effects of therapy are determined by the parameters that govern the short-term hemodynamic state of the patient, these are the only ones that are needed in the model. These relations capture the changes that take place in minutes, hours, and to some extent days. Including mechanisms that function over longer time periods would be difficult because of the external factors that influence longer terms changes and not as important for the acute management decisions that are the primary domain of the program.

Like the diagnostic model, the prediction model has a fixed part specified by the file that defines the model, an enhanced part that is computed when the model is loaded, and additions to the PSM that are computed for each patient case. In the fixed part of the model, equations are specified for each parameter that relate the parameter to other parameters in the model. For example, the primary slots of the definition of blood pressure are as follows:

```
(defparam blood-press
          equation (+ (* cardiac-output systemic-vascular-resistance)
                      ra-press)
          measure ((arterial-pressure (value mean-arterial-pressure)))
          ...)
```

This definition provides the equation for the relationship between blood pressure, cardiac output, systemic vascular resistance, and the right atrial pressure. This equation includes right atrial pressure (which is often ignored) because it has significant effects in many heart failure situations when the blood volume is high. Parameters are included in the equations if they are important for predicting the clinical hemodynamic effects of the important therapies. The equations in the model conform to the usual notion of causality in the cardiovascular system (basically related to the direction of blood flow), even though the equations are actually directionless. This makes the relations easier to understand for the physician. The measure clause indicates that the current value of blood pressure is determined by the computed mean arterial pressure from the input data on arterial pressure (which provides the systolic and diastolic pressures). This is the source of the quantitative value used in the equations.

Notice that these equations must be consistent with the causal probability relations of the diagnostic model, but neither is a substitute for the other. The causal probabilities represent a summary of experience with what happens in actual cases that is not apparent from the equations and the equations provide constraints that make it possible to combine the effects of many mechanisms in the feedback system.

The equations are used by the therapy prediction operators to determine the gains on the links between the parameters. The gains are computed as the partial derivative of the equation with respect to the parameter on the link and the link structures become part of the enhanced model. The enhanced model also includes all of the paths through the parameters and the feedback loops that are in the model. These are analogous to the causal paths in the diagnostic model and allow for the rapid computation of the prediction of changes given therapies. The PSM also has parts that correspond to the prediction model, but these are implemented as slot values for the parameter values and the computed changes and are not separate structures.

The model includes equations from several sources. Some equations were readily available from the physiology literature, such as the relationship between cardiac output, vascular resistance, and pressures. Others were determined from data in the literature, such as the relation between heart rate and systolic time. Basic hemodynamic relations were also borrowed from existing models such as Coleman's Human Program[2], including the relation between blood pressure and vagal stimulation. Included are equations for the hemodynamic effects of congestive cardiomyopathy (primary muscle disease) and four of the valvular diseases (aortic stenosis, aortic regurgitation, mitral stenosis, and more recently mitral regurgitation). The valvular diseases are modeled by equations relating the pressure drops or regurgitant volumes to the effective valve area from the original hydraulics equations developed by Gorlin[6, 7]. The therapies in the model include representative drugs from most of the major classes of cardiovascular therapies. Each therapy is represented by its direct effects on parameters, with proportions determined by comparing the model predictions to published results. The parameters included in the model are those likely to be measured in the patient or reported in the literature, plus the parameters needed to account for the actions of the usual cardiovascular agents.

3.3 Diagnostic Case Base

The diagnostic case base is a data base of patient cases with completed diagnoses organized as a discrimination net by the findings, to use in case-based diagnostic reasoning. This case base uses a self-organizing memory system implemented by Koton[10] and patterned after the memory system developed by Kolodner[8]. It consists of two kinds of data structures *gens*, which are generalizations representing two or more cases, and *cases*, which contain the identification, description, input findings, and accepted diagnosis of an individual case. These structures are organized as a network with the case structures as terminal nodes. Each *gen* has a *feature list* with the features that occur in at least two-thirds of the cases represented by the gen, a *diffs list* with features used to differentiate among the cases, a *causal list* with the parameter states common to all of the cases, and a node count.

Whenever the correct diagnosis is determined for a case, it is added to the discrimination net. A case structure is generated to represent it and an appropriate location is found for it by following the diffs list paths through the network. At each point in the network, if a gen is found the appropriate features on the diffs list of that gen are followed, if a case is found a new gen is produced from the two cases, if nothing is found the case becomes the new terminal node. Thus, the network is built reflecting the features found in the cases diagnosed by the program. The purpose of this structure is to allow the case-based reasoner to rapidly find the cases that are similar to a new case as a starting point for a diagnosis.

4 Reasoning Operators

The reasoning operators are the active elements of the program. They take the input, turn it into a patient specific model, and generate the diagnostic information and the management information that assists the physician in managing a patient. Each one uses the appropriate part of the knowledge base to transform parts of the PSM or the input into new parts of the PSM or output to the user. The eight operators are discussed in the following sections with emphasis on how they support each other.

4.1 Converting Data to Evidence

The first operator takes the input from the user and starts the PSM. The input is gathered by a dynamically expanding menu allowing textual summarization. It includes the patient history and presenting symptoms, vital signs, physical exam and the laboratory results. There is limited reasoning within the menus that enforces constraints among categorical values and appropriate precision for numeric values. The intention is to capture the information pertinent to the cardiovascular disease without requiring the system to do extensive reasoning beyond the program's intended medical domain and to provide physicians with a display of only the relevant patient information in an effective manner. Thus, it is assumed that the data has been interpreted and filtered by the user. For example, a single blood pressure is entered and assumed to be representative. Also, the program asks for interpreted findings on the electrocardiogram rather than the raw properties of the signal, as is more traditional. The physicians using this interface have found it to be an effective method of capturing and reviewing the data.

The evidence generator turns the data gathered from the menus into the initial PSM by using the data values to create *value* structures. These value structures cause a *node* structure to be created for each parameter state that could cause the value and those nodes in turn cause nodes to be generated for each possible cause of the node. Thus, by creating the value structures, the evidence generator creates the basic PSM structure used for most of the diagnostic reasoning.

4.2 Physiologic State Determination

The next operator does the local processing on the PSM to customize it to the patient by adjusting the probabilities, asserting any nodes that have definite values, and checking for inconsistencies. The first step is to go through the links in the PSM and do all of the partial evaluation possible on the probabilities to take into account input information such as patient age and sex. Then the program can compute the probabilities along paths to prepare the PSM for the probabilistic reasoning of diagnosis.

The local determination of the state of nodes in the patient specific model can be done at several levels. The primary source of information for this determination is the value structures created by the evidence generator. If there is a value with only one cause (and it requires a cause), that cause node is asserted true. If there is a node that always produces a value and that value is asserted false or otherwise inconsistent with the cause, the node is false. Nodes asserted in this way are certain and the automatic state determination stops at this point. However, it is sometimes useful for the user to assume that nodes with high probability are also true. Because of the large number of loops in the causal model, it is not computationally feasible to compute the exact probability of the nodes. It is possible to compute a conservative estimate of the probability of a node by locally comparing the probability of the node being true from the immediate evidence, to the probability that the evidence is caused by other nodes. If the probabilities of the causes are conservatively estimated and the threshold for asserting the node is high, this method identifies nodes that have a high probability from which the user can select ones to be asserted.

These determinations are added to the PSM and constrain it. There is also a truth maintenance system (TMS)[18] that assures that the assertions added to the PSM are consistent and adds any further implications of the nodes. Technically a TMS is not necessary because the probabilities of such nodes would be 1 or 0. In fact, the TMS in effect precomputes when the probabilities would go to 1 or 0 and provides an efficient mechanism for asserting such facts. It also provides a mechanism to track the source of any inconsistencies in the input. As a result, all hypotheses considered later in diagnosis include the nodes known to be true and exclude those known to be false. If a considerable amount of information is known about the patient, this simplifies differential diagnosis. When this operator is finished, the PSM has all of the information from the input and all of its definite implications asserted.

This operator is also used after the differential diagnosis to assert the consensus nodes. That is, the hypotheses with any significant probability are intersected to determine what nodes are common to all of them. These are then added to the PSM with the appropriate justifications to simplify the task of information gathering for diagnosis refinement and therapy selection.

4.3 Differential Diagnosis

A differential diagnosis is a set of hypotheses that could account for what is known about the patient. The process of differential diagnosis starts with the input findings and the nodes with known values as determined by the previous reasoning operators. Each hypothesis in the differential is an explanation of these findings in terms of causal pathways through the model. The differential consists of the hypotheses found that have the highest probability.

The appropriate mechanism for finding the differential depends on the nature of the diagnostic causal model. The model is similar to those investigated by Pearl[19] as Bayesian probability networks. The difference is that this model has forward loops (excluded by Pearl) and nodes with multiple paths between them (handled only in exponential time by Pearl's methods). We investigated modifications to Pearl's algorithm and to our model. However, even eliminating the forward loops in an earlier model version there were still about 40 links that would have to be cut to analyze multiple paths between nodes. Thus, Pearl's algorithm would require weighted summing of about 2^{40} solutions, which is completely infeasible. Lauritzen and Spiegelhalter's algorithm[11] might improve on this as well as more recent efforts to provide fast implementations[1], but the computation would remain infeasible. Indeed, Cooper has shown that the problem is NP-hard[3]. Thus heuristic methods are necessary to handle the complex networks required to represent this domain.

We have developed a mechanism to generate the differential diagnosis based on the causal paths from primary nodes to the findings. Since only about 50 of the 150 nodes in the model are primary (having a non-zero probability of existing without some other cause), all of the paths from these nodes to all others are generated and stored when the model is loaded. The probabilities along those paths are computed when the patient data is entered. Given these pathways, the task is to find sets of pathways that together account for all of the findings that need a causal explanation. In comparing hypotheses we discovered that the natural notion of *different* hypotheses requires that they differ in some significant node, nodes which we have labeled *diagnostic*. The algorithm is as follows: 1) collect the abnormal values from the input and the abnormal nodes asserted to be true in the PSM; 2) find all of the diagnostic or primary nodes that could account for each finding; 3) rank the diagnostic and primary nodes by the number of findings they account for; 4) use the nodes that account for the largest or nearly largest number of findings as seeds for small covering sets of primary nodes; 5) for each covering set, order the findings by the difference between the first and second highest probability path to it; 6) for each finding, find the best path from the partial hypothesis and add it to the hypothesis; 7) and then prune the hypothesis of unneeded primary nodes and extra paths that decrease the probability. Finally, the probabilities of the hypotheses are computed by multiplying the probabilities of the nodes given the other nodes in the hypothesis and they are rank ordered and presented to the user. These probabilities could be normalized by the probability of the findings but that is an intractable problem and unnecessary as long as we are only rank ordering hypotheses. The algorithm is discussed in detail in a paper[12].

This approach to diagnosis differs considerably from others that have appeared in the literature. Reggia's minimal set covering approach[20] ignores the fact that the best hypothesis may not be minimal and would not find the hypothesis in figure 2. Other approaches to diagnosis based on digital circuit analysis[4, 21] assume that every node is primary and every node can be measured. If every node were treated that way, a network of this size would be computationally intractable.

Figure 2 goes here.

Our mechanism is effective for producing a meaningful set of hypotheses for the findings in the cases we have tested and it usually takes a couple of minutes on a Symbolics 3650 workstation. The user can compare the hypotheses, see explanations, and consider the differences. Figure 2 is the display of the first of five hypotheses in the differential generated for an actual patient with findings that included rales (fluid sounds in the lungs), pedal edema (ankle swelling), high BUN (poor renal function), nausea, S3 (abnormal heart sound), and runs of VT (arrhythmia). The display graphically presents the complete explanation for the findings and provides a textual summary of the case at the bottom of the screen. In the display the findings are in lower case, intermediate nodes in upper case, primary nodes in bold face, primary probabilities in parentheses, causal probabilities on links and $W+$ indicating worsening factors that increase the probability and $P-$ indicating correcting factors that decrease it. This hypothesis accounts for the findings with congestive cardiomyopathy and renal insufficiency while the second hypothesis accounts for the findings with congestive cardiomyopathy alone. Those hypotheses nicely capture the actual physician's initial dilemma: whether the high BUN was the result of renal disease or inadequate blood flow to the kidney. Other hypotheses included valve disease, which is an important consideration. This hypothesis illustrates several features of the algorithm: 1) it handles multiple causes; 2) it handles multiple pathways between nodes; 3) findings can be left unexplained (the murmur); and 4) findings caused by therapies (digitalis toxicity here) are handled. (This example is discussed in more detail in[15].)

This method of generating hypotheses is heuristic and indeed it is possible to construct networks where it does not find the best answer. However, only the search is heuristic, not the use of probabilities in ranking hypotheses. Thus, if a better diagnosis is found, it is easy to test that it is better. As a result, it is possible to add hill-climbing techniques or even alternative diagnosis generators and be able to compare the probabilities of the diagnoses. We have tested over 100 actual cases thus far as well as many created cases and have found the algorithm to be effective. On one set of 42 cases, collected while developing the algorithm, the performance was tabulated. In 31 of these the program produced reasonable hypotheses. In five, the hypotheses were almost right but parts of the mechanisms were inappropriate. In the other six cases the best hypothesis was missed. There were two main reasons for these problems: 1) the program did not reason appropriately with the temporal relationships between cause and effect, and 2) it did not handle severity relations appropriately. These problems are part of our research agenda.

4.4 Case Based Diagnosis

CASEY, developed as a doctoral thesis by Phyllis Koton[10], adds a case-based diagnostic reasoning operator to the Heart Failure Program. This is an alternative diagnosis generator, allowing a comparison of different methodologies. It also allows the user to find similar cases for comparison. The CASEY operator uses the values from the input, the causal diagnostic model, and the case base. The first step in case-based diagnosis is using the input findings to find similar cases in the case base. CASEY has no predetermined list of important findings, so all of the findings are used at this stage to search for matches. From the causal diagnostic model, CASEY has a mechanism for assessing the similarity between findings. For example, rales on the physical exam and vascular redistribution on a chest X-ray are both evidence for the same physiologic state, cardiac pulmonary congestion. Thus, CASEY can find and justify the matches, not because the findings exactly match, but because the findings can provide evidence for the same physiologic states. Once a partial match has been found, the next step is to try to adapt the stored case to the new findings. There may be parts of the old case for which there is no support in the findings or findings that require additions to the old causal model. CASEY has a set of operators that take care of these situations. The simplest situation is new findings which add support for states already in the causal hypothesis. If the finding can not be explained by an existing state, CASEY makes use of the mechanism in the diagnostic reasoner to look for a good pathway from some part of the hypothesis to the new finding. If this will require new diagnostic states or primary causes, the search is abandoned and the match is rejected. If when all of the new findings have been added, there are parts of the hypothesis that are unsupported, CASEY has a pruning procedure that eliminates states until the whole structure is supported. The result of this process is a hypothesis that completely explains the findings and can be treated in the same way as a hypothesis from the causal diagnostic reasoner.

A case-based reasoner also needs to learn from the cases that are handled. Thus, CASEY needs a source of correct answers. After diagnosis has taken place, using CASEY, the causal reasoner, or a user guided diagnostic procedure, the user can assert that a diagnosis is correct and allow CASEY to add it to the case base. The structure and entry process for the case base is described in section 3.3.

CASEY is quite conservative in its approach to revision, but even with 50 cases in the initial database (including about ten different primary diseases) it was able to find sufficiently close matches to satisfactorily diagnose cases 80% of the time. Usually, the diagnoses generated with the probabilistic reasoner are better, but there are a few instances where CASEY has found a better hypothesis. The interaction between the case-based reasoner and the causal model reasoner gives CASEY the leverage needed to handle cases when the superficial similarity of the findings may be quite small and add the potential efficiency of the associational reasoning to the system. Indeed, CASEY is the first effective combination of case-based reasoning with an extensive causal model.

4.5 Additional Information Suggestion

When the program has produced a hypothesis that the user has accepted as the diagnosis, there still may be unknown aspects of the patient state. Perhaps the user accepted only part of the hypothesis because of uncertainties about data or because additional complicating factors have not yet been ruled out. Resolving these issues may uncover easily treatable aspects of the disease or affect the selection of appropriate therapy.

The mechanism we use for examining these possibilities and looking for additional patient data that would be useful in refining the diagnosis is quite simple. For a single diagnosis it traverses the diagnostic causal chains of the hypothesis looking for unknown states with links into or out of the chain that are likely. For a differential it starts with the nodes that are different among the hypotheses. From these it searches for measurements (observations or tests) that might clarify the state of the node (positive or negative). If there are no measurements, it examines the possible causes of the node for measurements to clarify the situation at that level. This mechanism identifies the measures that could have a bearing on the situation and lists them for the user. The justification for these suggestions are the nodes that they will clarify. Thus, this operator uses the PSM with the diagnosis to produce a list of suggested measurements with justifications.

4.6 Therapy Suggestion

When the user has decided on the diagnosis, the next step is to determine appropriate therapies. The program looks for candidate therapies by searching along the causal chains in the diagnosis that lead to undesirable outcomes. Therapies are included in the model as having corrective effects on the nodes (as well as possible detrimental effects). The therapies with the potential to break some of the causal paths are collected as a list of potential therapies. This approach allows the program to find therapies that are appropriate even though the findings that usually trigger their consideration are absent. For example, hydralazine decreases the systemic vascular resistance, so it is commonly used to decrease blood pressure. However, in a patient with primary cardiac muscle dysfunction, the systemic vascular resistance can be too high even though the blood pressure is normal when the cardiac output is low. In such a patient, the use of hydralazine will be suggested because of the high systemic vascular resistance and is quite appropriate. Since the therapies typically have multiple effects or the primary effect has multiple potential consequences, it is necessary to determine what the effects of the therapy actually will be. That is done by the therapy prediction algorithm. There are also other considerations when selecting therapies for the patient, such as side-effects outside the domain of this model, requirements for monitoring drug effects, and known patient sensitivities, that are not covered by the program.

4.7 Therapy Effect Prediction

Therapy effect prediction uses the physiologic state determined by the diagnosis and the values of parameters to estimate the effect of adding or removing one or more therapies. The mechanism for prediction of the effects of therapy is based on signal flow analysis[17] and computes the changes in the parameters from steady state to steady state. The choice of a constraint representation for therapy prediction rather than a probabilistic representation is appropriate because there is usually enough information available to solve the constraints and because the complex interactions among the parameters make the generation of a probabilistic model extremely difficult.

The advantage of the signal flow mechanism over other means of solving the equations is that it produces a record of the paths of influence on the parameters and their relative contribution. This record provides the basis for an explanation of the change and may in the future allow a way to deal with some of the uncertainty in the relationships and measurements. Initially, we applied this mechanism to a model with qualitative relations on the links between parameters. This worked well in our early tests in which the actions of drugs in the normal patient were compared to the model predictions[14], but we had considerable difficulty extending the model to account for the behavior of mitral stenosis, the first valvular disease considered. Since most of the parameters have known *quantitative* relationships to other parameters and the computational mechanism supports quantitative reasoning, we converted the model to use quantitative relations. To do so required adding the handling of integrated parameters and compensating for non-linearities as well as developing a new model.

The prediction constraint model conforms to the usual notion of causality in the cardiovascular system. Since one normally thinks of the hemodynamic relations on the right and left sides of the heart separately, there is a problem in representing the equality of left and right outputs in steady state. Physiologically blood volume shifts between circulations equalizing the outputs. To capture this we have levels (integrated variables) representing the volume in each circulation.

Integrated relationships can be handled because the derivative of an integrated parameter in steady state is zero. This provides the additional constraint needed to determine the level of the parameter. The steady state assumption is justified because the time constants of concern differ by an order of magnitude or more. Typically, we are either interested in the changes that take place in minutes (immediate hemodynamic changes) or in the fluid balance involving renal function which takes place in days. Other changes require weeks or longer to take place. Thus, physiologic mechanisms with longer time constants than the time of interest can be ignored. With this extension the procedure for determining the change in all parameters requires two steps: determine the levels of the integrated parameters necessary for their derivatives to be zero, then use these values plus the original change to determine the final values of all of the parameters.

The second extension to the reasoning was to handle non-linearities in the relationships between parameters. A non-linearity implies that the gain between two parameters varies over the amount of the change. For example, in the relation determining the blood pressure (section 3.2) the effect of cardiac output depends on the systemic vascular resistance, which will also change. The algorithm uses the initial gain to determine the changes, but that is not always adequate. Our solution is to adjust all of the gains to be the average gain over the range of the change and iterate until the final values conform to the constraint equations. This approach has theoretical limitations, but it usually converges rapidly to a consistent solution. Once a solution is proposed, its adequacy can easily be tested by verifying that all of the equations are satisfied.

The model has been validated by comparing published data to the model predictions. The data came from papers in the literature in which patients with one of the diseases were given a therapy or exercised and the hemodynamic variables are reported before and after the intervention. When the hemodynamic data are fairly complete, the initial values of the model parameters can be computed or estimated and the model can simulate the patient. Our efforts in validating the model were very fruitful[13]. The model proved sufficient to account for the average behavior reported in each of the five papers studied. With only two or three minor exceptions, the predictions were within the errors of the mean reported for all of the modeled parameters, once appropriate distributions of direct effect were determined for the therapies and the exercise the patients experienced.

Therapy prediction starts with the PSM as determined so far. Since this is a quantitative model, the primary source of data is the parameter values that are entered in the input. For example, if the user enters a heart rate of 110 in the input, the diagnostic reasoning treats that as a high heart rate, but the therapy prediction reasoning uses the actual value. Many of the other parameter values needed in the model are computable from those that the user enters, and the program automatically computes these. Estimates of other parameter values are determined from the diagnosis. Most of these are simple to determine. For example, all of the diseases that are not in the diagnosis are given the parameter value zero. However, when important parameters such as the cardiac output have to be estimated, the program picks a number in the appropriate range that is consistent with the other parameters that influence it. Once a complete and consistent initial state is determined, the program can be used to predict the effect of one or more therapies. This is done by indicating a change in the value of the desired therapy (an increase, a decrease, or multiple therapies). The primary effects of examples of most of the important classes of therapies have been determined by our study of the literature and are included in the model. However, other therapies can be considered if the user knows the primary effects by specifying changes in the values of the primary effect parameters. The program then shows the predicted changes in all of the parameters. One problem is determining the appropriate dose for comparing effects. There is a mechanism in the program that will adjust the dose until a desired change is achieved in a particular parameter. That allows the user to answer such questions as, "if enough of the therapy is given to increase the cardiac output to 5.0 L/min, what will happen to the other parameters?"

Figure 3 goes here.

Figure 3 shows the prediction of the exercise response of a patient with mitral stenosis (actually the average data for 10 patients). The display excludes unneeded diseases and associated parameters and links with no or very small gain. The parameters are organized by type and region of circulation with arrows showing the circulatory flow. The data in the paper included the heart rate, cardiac output, left ventricular systolic pressure, left ventricular end diastolic pressure, pulmonary artery pressure, and pulmonary wedge pressure[5]. The rest of the initial parameter values (shown at the bottoms of the parameters) were computed from these to run the model, making assumptions consistent with the state of the patient. For this example, the amount of exercise applied was chosen to produce the cardiac output of 8.0 L/min reported in the paper. The changed values are printed at the tops of the parameters. The predictions for the other reported parameters were all within the errors of the mean.

4.8 Explanation

The final operator is the explanation operator. This takes the PSM and the results of other operators and provides the user with a way of understanding them. There are two kinds of explanation produced by the program, one that provides information about the causal model and the diagnostic hypotheses and one that allows the user to explore the therapy predictions.

An example of the kinds of graphical explanations provided for diagnostic hypotheses was shown in figure 2. The diagnostic hypothesis consists of two lists of nodes, a list of nodes true in the hypothesis and a list of false nodes. These are explained to the user by graphing the causal relations in the list of true nodes. This kind of explanation is a rich source of information. It justifies the hypothesis by proposing the mechanisms by which they might have been produced in physiologic terms understandable to the physician. This helps the physician to see what assumptions are being made and therefore may identify aspects that need verification and whether the hypothesis is really appropriate given what the physician knows.

This method of graphical explanation is also useful for a number of other aspects of understanding the analysis and the model. Because parts of the display can be highlighted or italicized, it can be used for comparing two hypotheses, generated with the same or different input or with variations in the causal model. This makes it readily apparent what is common and what is different about the hypotheses. It is also useful for examining the conclusions of other operators. One can display the nodes with definite values and see what must be explained by the diagnostic process. It can display suggested therapies to see where their expected effects intersect the diagnosis. It can also be used to show where additional information might affect the diagnosis. This graphical method of display is also a good way of exploring the causal model for model development or to gain a better understanding of the model.

The explanation of the therapy prediction operator requires a somewhat different approach. The causal links as determined by the model equations are represented in the initial display, providing an overview of the model. When therapy prediction is done, the display shows the expected changes in the parameters as well. Still, that does not provide the user with any understanding of why those changes should take place. Indeed, with simulation based methodologies for predicting the changes, there is no good way of sorting out the relative importance of different influences on the changes. With the signal analysis based approach, all of the influences are recorded on the pathways through the model. This allows the program to identify the major influences in any expected change. In figure 3, this is shown. The user asks to see the major influences of exercise on cardiac output under these conditions by selecting *highlight* on the *cardiac output* parameter menu and the program highlights the two pathways that have the largest effect on this parameter. This helps the user to see which relations are most influential in determining what will actually happen.

Both of these methods of explanation have the advantage that they provide a lot of information about the conclusions of the program. In essence, they answer many questions without the user having to ask them. This high bandwidth communication has proven to be an excellent method of model development as well as an effective method for explaining analyses to the user.

5 Summary

Thus, the Heart Failure Program consists of a set of knowledge bases and a number of operators that take input about a patient from the user, build a patient specific model, and provide mechanisms for the user to understand the analyses of the operators. This structure of many operators and multiple reasoning strategies, provides a flexibility that does not limit the ways in which the user can apply the program. The user is able to concentrate on questions of diagnosis, if that is what is important in the case, or move on to the potential effects of therapies if that is what is important. In real cases, it is clear that these issues are closely interrelated and the program allows the user to move back and forth between diagnosis and management considerations as appropriate. The other advantage of this structure is the power gained by each operator by building on the conclusions of previous operators. For example, the diagnostic process is simplified because definite conclusions have already been made and these can be handled as if they were additional input constraining the possible conclusions.

References

[1] S. K. Andersen, K. G. Olesen, F. V. Jensen, and F. Jensen: 1988, "HUGIN — A Shell for Building Bayesian Belief Universes for Expert Systems," *Eleventh International Joint Conference on Artificial Intelligence*, pp 1080-1085.

[2] Coleman, T. G. and Randall, J. E.: 1983, "HUMAN: A Comprehensive Physiological Model," *The Physiologist*, **26:** 15-21.

[3] Cooper, G. F.: 1987, "Probabilistic Inference Using Belief Networks is NP-Hard," Tech Report KSL-87-27, Knowledge Systems Laboratory, Stanford University.

[4] deKleer, J. and Williams, B. C.: 1987, "Diagnosing Multiple Faults," *Artificial Intelligence* **32:** 97-130.

[5] Giuffrida, G., *et al*: 1979, "Hemodynamic Response to Exercise After Propranolol in Patients With Mitral Stenosis," *The American Journal of Cardiology*, **44:**1076-1082.

[6] Gorlin, R. and Dexter, L.: 1952, "Hydraulic Formula for Calculation of the Cross-Sectional Area of the Mitral Valve During Regurgitation," *American Heart Journal*, **42:** 188.

[7] Gorlin, R. and Gorlin, G.: 1951, "Hydraulic Formula for Calculation of Area of Stenotic Mitral Valve, Other Cardiac Values and Central Circulatory Shunts," *American Heart Journal*, **41:** 1.

[8] Kolodner, J. L.: 1983, "Maintaining Organization in a Dynamic Long-Term Memory," *Cognitive Science*, **7:** 243-280.

[9] Koton, P. A.: 1988, "A Medical Reasoning Program that Improves with Experience," *Symposium on Computer Applications in Medical Care Conference*, pp. 32-37.

[10] Koton, P. A.: 1988, *Using Experience in Learning and Problem Solving,* MIT PHD Thesis.

[11] Lauritzen, S. L. and Spiegelhalter, D. J.: 1988, "Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems," *Journal of the Royal Statistical Society B* **50:** 157-224.

[12] Long, W.: 1989, "Medical Diagnosis Using a Probabilistic Causal Network," *Applied Artificial Intelligence,***2:**283-299.

[13] Long, W. J., Naimi, S., Criscitiello, M. G., and Jayes, R.: 1986, "Using a Physiological Model for Prediction of Therapy Effects in Heart Disease," *1986 Computers in Cardiology Conference*, pp 15-20.

[14] Long, W. J., Naimi, S., Criscitiello, M. G., and Kurzrok, S.: 1986, "Reasoning About Therapy from a Physiological Model," *Proceedings of the Fifth Congress on Medical Informatics*, pp 756-760.

[15] Long, W. J., Naimi, S., Criscitiello, M. G., and Larsen, G.: 1988, "Differential Diagnosis Generation from a Causal Network with Probabilities," *1988 Computers in Cardiology Conference*, pp. 185-188.

[16] Long, W. J., Naimi, S., Criscitiello, M. G., Pauker, S. G., and Szolovits, P.: 1984, "An Aid to Physiological Reasoning in the Management of Cardiovascular Disease," *1984 Computers in Cardiology Conference*, pp. 3-6.

[17] Mason, S. J.: 1956, "Feedback Theory — Further Properties of Signal Graphs," *Proceedings of the IRE*, **44:** 920-926.

[18] McAllester, D. A.: 1982, "Reasoning Utility Package User's Manual, Version One," MIT/AIM-667, Massachusetts Institute of Technology.

[19] Pearl, J.: 1986, "Fusion, Propagation, and Structuring in Bayesian Networks," *Artificial Intelligence*, **29:** 241-288.

[20] Reggia, J. A., Nau, J. S., and Wang, Y.: 1983, "Diagnostic Expert Systems Based on a Set Covering Model," *International Journal of Man-Machine Studies* **19:** 437-460.

[21] Reiter, R.: 1987, "A Theory of Diagnosis From First Principles," *Artificial Intelligence* **32:** 57-95.

Figure legends

Figure 1: Overview of Heart Failure Operators and Data Structures

Figure 2: Congestive Cardiomyopathy and Renal Insufficiency Hypothesis

Figure 3: Prediction for Mitral Stenosis with Exercise