MEDAS, a Bayesian diagnostic system, includes a Feature Dictionary that contains descriptions of signs, symptoms, and test results, along with alternate medical terminology used in various fields and parts of the country. The Feature Dictionary has now been expanded to include feature interactions to support test selection and treatment protocols. The feature interactions required to provide this flexibility are drug side-effects and allergic reactions, contraindications to tests and treatment, and drug interactions. Although the primary use of the dictionary is to store the feature information to be used by the MEDAS diagnostic system, adding further functionality allows other applications, both inside and outside of MEDAS to be supported. Generic and brand names of drugs appear as synonyms in the dictionary. This widens our abilities to translate data from one system to another.

INTRODUCTION

MEDAS (Medical Emergency Decision Assistance System) uses multi-membership Bayesian inference to estimate the probabilities of relevant disorders. MEDAS was originally developed at the University of Southern California. During that time, the project was supported by NASA. For the past five years, research has continued at the University of Health Sciences/The Chicago Medical School and the system has grown from approximately 70 disorders and 700 features (i.e., signs, symptoms, test results, etc.) for one domain to 200 disorders and 7000 features for six domains [17]. This growth has led to the development of the TOOLBOX (a knowledge engineering tool that facilitates ease of entering data for the medical expert) and the Feature Dictionary.

As the MEDAS system has evolved over the years, the complexities of the medical terminology involved have become increasingly evident. Currently, physicians of different specialties enter sets of features into separate knowledge bases using terminology pertinent to their own areas of emphasis. Often the features in each of the knowledge bases are the same but the terminology used to describe them differs. To alleviate this problem, the features from all of these individual knowledge bases have been combined to create a Feature Dictionary [1,2]. This dictionary is a relational database designed using the Entity Relationship (or E-R) database modeling technique. It provides a means for transcending the medical language barrier by recognizing any feature entered into a knowledge base no matter which of the terms a particular expert has used. In addition to providing a means to merge the current knowledge bases using consistent common terminology, the Feature Dictionary provides a means to evaluate and compare knowledge bases, and translate between MEDAS knowledge bases and other medical knowledge bases. It also provides a framework for the collection and analysis of data for epidemiological research independent of the MEDAS system.

Our experience with the Feature Dictionary has suggested additional uses for this knowledge bank. By extending the current model of the Feature Dictionary, we can support complex feature interactions and maintain multiple feature synonyms. In addition, we can classify sets or groupings of features for queries and identify common characteristics between features.

In fact, the Feature Dictionary can provide many new functions. Among those functions that have been identified so far are the following:

- Support test selection and treatment protocols by adding the following three kinds of feature interactions:
  -- Drug side-effects and allergic reactions
-- Contraindications to tests and treatment
-- Drug interactions with other drugs

- Support applications of the Feature Dictionary for information retrieval such as:
  -- Medical thesaurus development
  -- Feature sub-classifications for queries
  -- Medical document classification and cataloguing

THE FEATURE DICTIONARY

Originally MEDAS was designed to work only in an emergency room setting and we had only one knowledge base -- a set of emergency room disorder patterns. Now, there are a growing number of MEDAS knowledge bases including Obstetrics / Gynecology, Internal Medicine, Psychiatry, and Oncology. When experts create knowledge for a MEDAS knowledge base, it is essential that the system recognize a consistent set of features not only within a single domain (or area of specialty), but also when features appear in a variety of domains. Dombal's research [3] has shown that the performance of a Bayesian system is crucially dependent on the consistent labeling of features. The terms that experts use to refer to particular features vary, depending on an expert's specialty, preference, or regional vernacular. These differences created the need for a Feature Dictionary which can translate medical terms into those meaningful to the system.

The Feature Dictionary was designed with the most up-to-date database design criteria. An E-R (Entity Relationship) diagram has been developed (Figure 1) for the MEDAS system which shows the logical relationship between the major entities [1,2]. Entities are represented by rectangles, the associated attributes by ovals, and the relationships by diamonds. Entities are the central concepts in the database. For MEDAS these are features and disorders. A feature is any item of information about a patient that can be used in making a diagnosis, beginning with age, sex, and other items that appear in the physical examination and medical history, the chief complaint and other current symptoms, and test results. A disorder is represented in a MEDAS knowledge base as a pattern of features.

MEDAS, like any Bayesian system, uses binary features. That is, a feature is marked as either present (+) or absent (-). The original Feature Dictionary recognized only one kind of feature interaction called "binary feature Y interaction." When a feature was entered for a given patient, other related features were automatically set to positive or negative depending on the nature of the disorder. For example, whenever the feature "female" is
marked as present (or '+'), the related feature "male" is marked as absent (or '-'). The presence of binary feature interactions in the Feature Dictionary thus simplified the feature entry process for the user. Binary feature interactions are also used by the MEDAS inference engine to avoid using dependent features in making estimates of disorder probabilities.

Each disorder is represented by a set of binary features called the disorder pattern. The expert physician who designs the disorder pattern begins by selecting a set of relevant features Fj. The next step is to provide two probability estimates $P_{ij}$ and $\overline{P}_{ij}$ for each feature. $P_{ij}$ is the additional probability that $F_j$ will be observed given that disorder $D_j$ is present in the patient. $\overline{P}_{ij}$ on the other hand, is the conditional probability that $F_j$ will be observed given that disorder $D_j$ is not present in the patient [181.

Since MEDAS uses a pattern recognition strategy, it is important to store the conditional probabilities that given features will be present in a patient with a specific disorder and to have a consistent set of features that can be used by various medical domains. The decision to store this information in a database made it much easier to maintain and update. We have now developed on top of this database a complete set of knowledge engineering tools called the TOOLBOX.

The TOOLBOX is a set of knowledge engineering tools designed to facilitate ease of entry of the features, feature relationships, and disorders into the Feature Dictionary. The TOOLBOX provides a full-screen interface to the dictionary so that the physician can load information into the database without having any formal knowledge of the underlying database constructs and without any reference as to how the data is actually stored. In fact, the TOOLBOX is a screen-driven application which allows the medical expert to concentrate on entering and updating feature and disorder information easily. The Feature Dictionary supports the TOOLBOX user by providing synonyms for standard feature names, definitions of terms, and normal ranges for test results.

Additional benefits derived from the Feature Dictionary include [1]:

- Providing definitions for terminology foreign to the user
- Evaluating the performance of competing knowledge bases
- Comparing knowledge bases for the same domain
- Allowing senior medical students to create their own knowledge bases using the same set of terms as the domain experts
- Allowing patient information to be translated into formats demanded by other medical expert systems developed using M-1 or EMYCIN
- Using existing clinical databases to translate patient data into the standard format of MEDAS knowledge
- Collecting data for epidemiological research

The TOOLBOX and the Feature Dictionary make it easy for the expert physician to maintain the knowledge base. They also make it possible for the system to transfer information from one knowledge base to another. Figure 2 shows an E-R diagram of the Feature Dictionary supporting multi-domain knowledge.
FEATURE INTERACTIONS TO SUPPORT TREATMENT

Shortly after MEDAS was developed, a set of treatment protocols were also implemented. These protocols could function as part of the MEDAS system or as a stand-alone subsystem [4,5]. The protocols were simply a list of actions to be performed by the physician. All of the information was stored in text files that could be updated with a text editor. A protocol file often contained questions to be answered before a particular line of treatment was followed. Questions served to determine any defects in the patient that might contraindicate a specific treatment before appropriate measures were recommended for that patient. Recently a relational database was designed to store protocols [6,7]; questions were identified with corresponding MEDAS features, so that if the feature information were already recorded it could be provided to the physician immediately.

As we began to work on facilities to handle tc selection and treatment recommendations it became essential to add drugs and other treatments as features in the Feature Dictionary. To make it possible to recognize different drug terminology, generic and brand names for drugs were listed as synonyms in the dictionary. With changes to the Feature Dictionary, it became apparent that we would need to establish new feature interactions for drugs to better support test recommendations and treatment protocols since drugs play a major role in treatment. Having standardized feature names facilitates this effort. The feature interactions now added describe treatment contraindications, drug side-effects, allergic reactions, and interactions [14,15,16].

Drug Side Effects and Allergic Reactions

It is becoming increasingly important for MEDAS to be able to assist in the diagnosis of patients suffering from drug side-effects as more and more patients arrive in the clinic or emergency room suffering from drug reactions to both illegal drugs and drugs prescribed by physicians. In general, drug induced disorders are caused by substance abuse, multiple drug use with adverse interactions, or drugs prescribed for another disorder [12,13].
For example, barbiturate, cocaine, and hallucinogen overdoses usually result from drug abuse. Lithium overdoses most often occur in patients who are being treated for a neurological disorder when due to their physical condition, the body is unable to sustain the prescribed dosages. Aplastic anemia is a disorder that can be caused by a variety of drugs prescribed for some other problem the patient is suffering from. Penicillin, streptomycin, busulfan, and chlorpropamide are only a few of these drugs.

Initially, we planned to combine all kinds of drug reactions into one relationship. But after careful analysis, we decided to separate this relationship into two new kinds of feature interactions: the side-effect relationship (which deals with expected and frequent drug reactions) and the adverse affects or allergic reaction relationship (which deals with unexpected and infrequent reactions.) For example, consider the drugs sulfasalazine, ibuprofen, and clonodine. The side-effects of sulfasalazine include drowsiness, constipation, brownish coloration of the urine, and sometimes yeast infections. Its allergic reactions (which are potentially much more serious) range from skin rashes, hives, and itching to headaches, dizziness, diarrhea, abdominal pain, loss of appetite, swollen glands, depression, pancreatitis and sometimes kidney damage. For ibuprofen, common side-effects are weight retention and rust coloration of the urine, while mild allergic reactions include rashes, hives, itching, headaches, and dizziness and the severe adverse effects are liver damage, kidney damage and painful urination. Lastly, the side effects for clonodine are drowsiness and constipation but the allergic reactions involve skin rashes, hives, localized swelling, headaches, dizziness, nausea, and vomiting. More serious side-effects include heart-related disorders, nightmares, depression, and hallucinations.

Having side-effects and allergic reactions in the dictionary allows MEDAS to have more information available when making a diagnosis. This information will have to be used wisely since the mere presence of a drug does not necessarily imply the reason for the existence of a symptom. Therefore this information would not be used to reduce or increase the probability of a feature being present during diagnosis. Rather, it would be used simply to inform the physician that a drug the patient is taking has a side-effect (or allergic reaction) that matches a symptom of the patient's disorder.

In order to support the side-effect and allergic reaction relationships, the Feature Dictionary was enhanced as indicated in Figure 3a. With these new relationships included, the data from the above examples would be stored in the dictionary as illustrated below in Figure 3b.
Contraindications to Tests and Treatment

The next new relationship between features that we are adding to the Feature Dictionary is called Contraindications. The inclusion of the contraindication relationship allows MEDAS to ask appropriate questions before recommending a test or treatment. This is particularly important when the physician is considering certain high risk tests such as angiography. Analysis of the data in the National Stroke Data Bank has shown that angiograms have significantly greater risks for people with heart problems.

As we increase the coverage of drugs in the database and as we begin to look at the features required to supplement treatments, the contraindications relationship, has become increasingly important. This relationship helps MEDAS to guide the physician in determining a patient's suitability for different drug treatment protocols. For instance, there are certain medications (antibiotics, anti-epilepsy drugs, etc.) that are excreted or metabolized by the liver and/or kidney. If a patient has poor kidney function then a medication that is excreted by the kidney would have to be monitored closely if not contraindicated. Precautions would have to be taken since toxic effects would show up if the body does not properly rid itself of such drugs. In instances such as these, the MEDAS system would suggest that the physician monitor the situation closely by looking at the blood level of the drug or even the organ that is directly affected.

A patient's physiological makeup is the key to whether or not a physician prescribes a drug. Consider diuretics, for example. Most of these drugs need good kidney function. If the kidney functions well, the physician would give one drug whereas if the renal function was decreased, a different drug would be prescribed.

When the contraindications relationship is added to the Feature Dictionary, the treatment aspects of the MEDAS system are greatly enhanced. During diagnosis, as a physician selects a treatment, he/she would be immediately informed of the possible contraindications based upon the patient's current physical condition. That way the physician can have more information available prior to making his/her final treatment decisions. This capability to
remind the physician of critical facts would definitely result in better medical care for the patient. This new contraindications relationship is illustrated in Figure 4.

![Figure 4: ER-Diagram after adding Contraindications Relation](image)

**Drug Interaction**

A more specialized, although very important relationship between features is represented by the drug interaction relationship [11]. This is such a complex issue we have considered the use of an expert system or separate system. While we are still debating the advisability of adding drug interactions to the dictionary, we have included this relationship in the dictionary model. See Figure 5 for an illustration.

![Figure 5: ER-Diagram after adding Drug Interactions Relation.](image)

(The Keys for Drug Interactions are three fields - two drugs and one symptom)
Drug interactions are a very complicated issue. Some drugs can potentiate the effect of other drugs whereas some drugs can nullify the affect of other drugs. For example, there are cardioactive drugs, some of which are synergistic (they act to potentiate each other), while others are antagonistic. Whereas one drug may enhance cardiac function the other may be a cardiac depressant. The physician has to observe the effects of the drugs on the patient to see if the proper results are being achieved. For example, one drug may increase the heart rate by 10% and another drug may also increase it by 10% if these drugs are taken alone. But when taken together, the net effect is that the patient's heart rate is increased by 50%.

Some drugs counteract each other. For instance parasympathetic and sympathetic drugs' effects on the heart are undesirable since they can nullify each other and affect other body systems as well. For example, one drug can increase the heart rate by 10% and the other could decrease it by the same amount so that the net result is that the patient's disorder is not being treated and other disorders result from side-effects of the drugs.

Applications of the Feature Dictionary in Information Retrieval

It would be a simple matter to adapt the Feature Dictionary for use as a relational thesaurus in information retrieval. The purpose of a relational thesaurus in a bibliographic information retrieval system is to augment the user's original query with semantically related index terms [8]. In an interactive information retrieval system like that envisioned in the Unified Medical Language System (UMLS) planned by the National Library of Medicine, a relational thesaurus can play an even bigger role. The system can display to the user a number of index terms related to the query terms in different ways and guide the user in formulating a more appropriate query using these terms.

The Feature Dictionary already contains synonyms for signs, symptoms, tests, and drugs as well as definitions for each entry. Each feature is included in a feature category which defines the appropriate ISA or taxonomy relation in which it participates. Other lexical relations could be added by hand or otherwise derived automatically from the descriptions using the methods developed by Evens et al. [9] in their analysis of definitions in Webster's Seventh Collegiate Dictionary.

The Feature Dictionary can also be used in the automatic classification and cataloguing of new medical documents, papers, and reports [10]. The dictionary already contains not only features but disorders. (All of the ICD9 disorders are included as "SETTING OF" features). Thus the Feature Dictionary can be used to classify incoming documents, assign index terms (or keyword phrases), and build an index for a medical document collection.
SUMMARY

The Feature Dictionary extensions have added additional functionality to MEDAS and provided support for other applications. The goal of producing recommendations for test selection and treatment protocols required that we add drugs and other types of treatment as features. Then treatment contraindications and drug reactions and interactions can be represented as feature interactions. Generic and brand names for drugs become synonyms in the Feature Dictionary. In database terms this implies the addition of several new types of relationships between features. The Feature Dictionary now contains enough feature relationships and descriptions so that it could be used as a thesaurus for medical information retrieval system or as a basis for the classification of medical documents.

REFERENCES