Complex Clinical Presentations and their Models

CARLOS FEDER, MD, Internal Medicine; TOMAS FEDER, PhD, Computer Sciences

Most existing diagnostic computer programs are rather internal medicine training or educational programs, incapable to diagnose diseases that afflict actual patients. Some are based on Bayes conditional probability formula; some others are structured with pathophysiologic networks. At best, these programs can deal imperfectly with a single disease able to account for all clinical data (symptoms, physical signs, diagnostic tests, and diagnostic procedures) that a particular patient manifests. These programs fail to process complex clinical presentation, where a combination of diseases or clinical entities afflict simultaneously a single patient, situation frequent in medicine. Our algorithm and computer program are able to process complex clinical presentations, involving concurrent clinical entities and diagnoses; they are also easy to update.

This paper is part of our complete medical diagnostic system, described in detail in our book [2].

We summarize here only basic concepts of previous publications; for better understanding of this paper, the reader is encouraged to consult these publications.

Sensitivity (S) is the cornerstone of our diagnostic system. A practical way to calculate S of a specific clinical datum for a given disease is to determine statistically the fraction of patients afflicted by this disease who manifest the clinical datum:

$$Sensitivity = \frac{Number\ of\ disease\ cases\ manifesting\ the\ clinical\ datum}{Total\ number\ of\ disease\ cases}$$

Positive predictive value (PP value) is the best index to determine the strength with which a specific clinical datum present in a patient supports a specific diagnosis. Our algorithm calculates PP value with the following equation:

$$PP\ value\ i = \frac{Si}{S1 + \cdots + Si + \cdots + Sn} \quad (1)$$

Where PP value i = positive predictive value of the clinical datum for the disease i under consideration

Si = sensitivity of the clinical datum for the disease i under consideration

S1 ... Sn = sensitivities of the same clinical datum for corresponding diseases

Disease model, as defined in our system, is an abstract concept that comprises all clinical data that can be manifested by all patients with a specific disease. A single patient typically never manifests all clinical data that the disease potentially can provoke. Integration of a specific disease model with all of its possible manifestations requires statistical study of a large patient population. *Each clinical form, stage, degree, or complication of a disease has its own disease model*. Because death and iatrogenic diseases are diagnoses that must be established clinically, the corresponding disease models must also be created.

Probability (P) of a diagnosis is calculated with our novel **mini-max procedure**, core of our diagnostic system, considering PP value of clinical data present (favoring corresponding diagnosis) and S of clinical data absent (disfavoring diagnosis). These values are processed by a specific formula:

$$Pi = \frac{PP\ value\ i\ (1 - Si)}{PP\ value\ 1\ (1 - S1) + \ \ldots + \ PP\ value\ i\ (1 - Si) + \ \ldots + \ PP\ value\ n\ (1 - Sn)} \qquad (2)$$

Where P_i = probability of a diagnosis i

$PP\ value_i$ = positive predictive value of the clinical datum present

S_i = sensitivity of the clinical datum absent

$PP\ value_1 \ldots PP\ value_i \ \ldots \ PP\ value_n$ = positive predictive value of the same clinical datum present for each respective diagnosis in the differential diagnosis list

$S_1 \ldots S_i \ldots S_n$ = sensitivity of the clinical datum absent for each respective diagnosis in the differential diagnosis list

We confirmed that mini-max procedure is superior to Bayes formula and other probabilistic or rating methods. Detailed explanations and examples can be found in our previous publication [2].

Cost to obtain a clinical datum involves, in our context, not only expense but also risk and discomfort resulting from the required test or procedure. We assign to each clinical datum one of four overall cost categories: no cost (clinical data typically obtained through medical history and physical examination), small cost (*e.g.*, obtained through routine laboratory analysis, ECG, and other ancillary studies), intermediate cost (*e.g.*, colonoscopy, lymph node excision biopsy), and great cost (*e.g.*, liver biopsy, laparoscopy, laparotomy). **Benefit** of a clinical datum is measured by the magnitude of change it produces in the probability (P) of the respective diagnosis, in turn depending on the magnitude of PP value of clinical data present, which increase P, and the magnitude of S of clinical data absent, which decrease P. The mini-max procedure calculates these P for corresponding diagnoses.

Best cost-benefit clinical data are recommended at each diagnostic stage to be investigated next in a patient, based on mini-max procedure, that predicts, based on probabilistic calculations, which set of such data would end the diagnostic quest, more efficiently and at lowest cost. Recommended best cost-benefit clinical data are typically quite numerous, mandating the need to heuristically reduce its number; this is achieved in part by certain **parameters** described elsewhere [2] that can be set at empirically values by the user. A tradeoff exists in each of these parameters: moving the value in one direction may significantly reduce the number of recommended data, but reducing also slightly the accuracy of diagnostic result, and *vice versa*. The effects that these parameters have on the recommendation of best cost-benefit clinical data are shown in diverse output files mentioned later.

The diagnostic process comprises several levels of complexity. Related clinical data cluster to a syndrome, simple syndromes comprising only a few clinical data coalesce to a complex syndrome or disease, and sometimes to a yet more complex clinical presentation comprising disease causes or complications, where the relation of clinical data with diseases becomes less obvious.

Clinical entity: a generic term for any element of a *complex clinical presentation*, such as a cause, lesion, syndrome, complication, disease, clinical form, stage, or degree.

The algorithm thus far presented uses *probabilistic* calculations with mini-max procedure and best cost-benefit clinical data to determine the probability of each differential diagnosis. It will work well

with simple clinical entities, such as uncomplicated diseases or syndromes where clinical data typically are interrelated and linked to a single cause or lesion. Some examples of such simple diseases or clinical entities include bronchitis, asthma, gastroenteritis, hyperthyroidism, obstructive jaundice, and renal failure. At this diagnostic stage, a single final diagnosis accounts for all manifested clinical data.

In an actual patient, the clinical picture might be more complicated; as a fact, severely ill patients in intensive care unit often have multi-organ involvement, present multiple and proteiform clinical data, and may mandate consultations with several specialists. For example, coronary artery disease, acute myocardial infarction, congestive heart failure, shock, *and* thromboembolism in a single patient. A specific disease can manifest diverse clinical forms, stages, degree of severity, and clinical presentations complicating the diagnostic process. This situation makes impossible to determine the sensitivity (S) of each clinical datum for the entire *complex clinical presentation* because this would involve multiple clinical forms, concurrent diseases, and multiple pathogenic and pathophysiologic mechanisms. It would require analyzing a statistically significant number of cases with identical combinations of clinical entities; it also would take us into an unmanageable computer complexity. Accordingly, probabilistic methods are unsuitable for processing complex clinical presentations; indeed, to my knowledge, no commercial diagnosis programs that can accomplish this exist. A *categorical* method for processing complex clinical presentations is simple and feasible.

For this reason our algorithm, with its heuristic principles and moderate use of probability, diagnoses first only relatively simple syndromes, clinical entities, or diseases. Let the diagnostic algorithm produce as many final diagnoses of simple concurrent diseases, syndromes, complications, etc. as the clinical data dictate. These concurrent diagnoses might be unrelated or, more frequently, related by specific pathogenic or pathophysiologic mechanisms, many of them currently elucidated, but sometimes only suggested by statistical correlations. Regardless of whether these mechanisms are known or suggested, those diagnoses that present any kind of relation or association, are listed together in input files that we call *complex clinical presentation models*. The integration of such models is purely *categorical*, as it does not require any probabilistic calculation; they are stored in the database. Such complex clinical presentation models, although numerous, are not excessive, and are described in any authoritative medical textbook; they include diagnoses of diseases, syndromes, complications, clinical forms, evolutive stages, and others.

A *complex clinical presentation model* comprises related *clinical entities* and *diseases*; clinical data are excluded from this definition because they are elements of a *disease model*.

Many other algorithms employ tree and network structures that extend from cause of disease to clinical data and *vice versa*, placing probabilities on nodes, branches, and leaves. Most such structures are complex and required years to assemble. We suspect that such structures are difficult to update and would need to be redesigned every few years. In contrast, our algorithm can relatively easily be updated at any time, by simply updating in disease models the S values of clinical data, adding or deleting clinical data when necessary, or adding or deleting disease models.

In summary, the entire diagnostic process is achieved in 2 steps:

Step 1. Probabilistic processing of clinical data matches patient clinical data with clinical data in disease models yielding a differential diagnosis list. Mini-max procedure, best cost-benefit clinical data next to investigate, and discrimination between competing diagnoses and concurrent diagnoses achieve as many concurrent final diagnoses of clinical entities as are required to account for all manifested clinical data. Then, the algorithm proceeds to Step 2.

Step 2. Categorical processing of clinical entities matches confirmed final diagnosis with complex clinical presentation models in the database. If a match is found, all the related diagnoses in this model are included in the differential diagnosis list to be processed in the usual way by min-max procedure and recommended best cost-benefit clinical data, being confirmed or ruled out. The same complex clinical presentation models enable establishing whether concurrent diagnoses are related or unrelated, when respectively a linking model exist or not.

Complex clinical presentations managed by our new program

Complex clinical presentation models are categorical combinations of clinical entities linked by pathophysiologic or statistically significant correlations; they can be created, displayed, and modified with input file *Complex Presentation Models*, which are part of the database. These models have at least four functions: (1) process associated diagnoses to preclude overlooking some of them; (2) manage interactions (masking) among diseases and drugs; (3) distinguish related from unrelated concurrent diagnoses; and (4) health assessment and early detection of occult diseases.

1. *Associated diagnoses*

The purpose of this function is to preclude overlooking diagnoses, which might be of crucial importance for the global treatment of a patient, when no clinical data present supporting them have been entered so far in the computer. Such diagnoses, which current P = 0, are suggested by association with confirmed diagnoses in the complex clinical presentation model, and will be processed even if not included yet in the differential diagnosis list. This is achieved through the following steps:

1. Create complex clinical presentation models—*Complex Presentation Models*—listing in each, all diagnoses that present a possible pathophysiologic or statistically significant link (diseases, complications, evolutive stages, etc.). Our program, assigns to each complex clinical presentation model a letter M, a number (Mxxxx), and an appropriate title (*e.g.*, CARDIOVASCULAR). Each diagnosis in the model has its letter D, corresponding number (Dxxxx) and name (*e.g.*, AORTIC DISSECTION, MYOCARDIAL INFARCTION, PULMONARY EMBOLISM,…).

2. Create in the database—*Disease Models*—disease models for the diagnoses mentioned in the previous paragraph, with their corresponding clinical data and sensitivities (S), if these models are not already included.

3. Enter patient's clinical data present and absent in respective *Present Data* and *Absent Data* files and run the *Diagnostic Program*.

4. After the necessary program iterations, for each *confirmed* final diagnosis, the algorithm searches all *complex clinical presentation models* for a match of this confirmed diagnosis with at least one similar diagnosis in the mentioned models. If such a match is established, all the linked diagnoses of the model are included in the differential diagnosis list, if not already included, to be processed for presence or absence. Best cost-benefit clinical data for these diagnoses will be recommended and once selected and investigated, enter them in respective *Present Data* or *Absent Data* files and run *Diagnostic Program* again.

5. The diagnostic program—*Diagnostic Program*—will calculate P of each diagnosis listed in the complex clinical presentation model. Diagnoses inside the model are sorted by decreasing P, the greatest of these P is assigned as P of the entire model, and models are also sorted by decreasing P.

Each linked diagnosis is processed with the usual mini-max procedure, to become a confirmed concurrent diagnosis or to be ruled out. The result of the process is displayed in the output files *Complex Short* that shows linked diagnoses with their P, and *Complex Comprehensive* that shows linked diagnoses with their P *and* best cost-benefit clinical data recommended for further processing. Linked diagnoses that at the current diagnostic iteration have no supporting clinical data present, will not show P, because this value must be calculated based on such clinical data, but will show the recommended best cost-benefit clinical data in *Complex Comprehensive*.

The diagnostic quest requires processing only diagnoses related to confirmed diagnoses (final diagnoses); calculations of P for diagnoses related to non-confirmed diagnoses would be too numerous, cumbersome, and irrelevant. Nevertheless, our program offers both options: (1) *Complex Comprehensive* output file recommends best cost-benefit clinical data for diagnoses in complex clinical presentation models related to all the diagnoses in the differential diagnosis list, confirmed or not, leaving to the user decide to which level he wants to process such diagnoses. (2) *Global Overview, Abridged Global Overview,* and other parameter sensitive output files recommend best cost-benefit clinical data for diagnoses in complex clinical presentation models related only to diagnoses that reached confirmation threshold.

Those diagnoses in a matched complex clinical presentation model that were not included yet in the differential diagnosis list by previously collected clinical data must be included in the differential diagnosis list because of their links to confirmed diagnoses. Their probabilities are calculated in the usual way with mini-max procedure, but not being supported by any initially collected clinical datum, this calculation must rely exclusively on best cost-benefit clinical data. Consequently, information of greatest PP value and S of nonexistent previous supporting clinical data, otherwise displayed between the diagnosis title and the best cost-benefit clinical data in *Complex Comprehensive* output file, is missing for these diagnoses. This information will be displayed only after at least one of the recommended best cost-benefit clinical data is entered in *Present Data* input file, and the *Diagnostic Program* is run again. When some of these diagnoses reach a confirmatory P, they become concurrent diagnoses.

We confirmed that diagnoses, which P is calculated *probabilistically* with mini-max procedure, must be kept simple and pure (not contaminated with causes, complications, etc.); when two or more diagnoses are confirmed, they must be combined *categorically*. An example of how violation of this rule affects results follows: acute aortic dissection sometimes produces the complication myocardial infarction; because of this fact, we erroneously included increased troponins as a clinical datum for acute aortic dissection, when it actually is an exclusive clinical datum for myocardial infarction. As a result, increased troponins incorrectly confirmed acute aortic dissection. Troponins should have been listed only as a clinical datum for myocardial infarction; acute aortic dissection and myocardial infarction should have been diagnosed as concurrent clinical entities, and then be linked categorically as cause and complication, by matching the corresponding complex clinical presentation model they share.

A match of a confirmed diagnosis (*e.g.*, emphysema) with only one similar diagnosis in a complex clinical presentation model suffices to select this model and include all its related diagnoses in the differential diagnosis list, even if only one of them (*e.g.*, pneumothorax) may be confirmed as a concurrent complication.

2. *Disease and drug interactions (masking)*

Drugs often interact, one enhancing or reducing the effects of another. Drugs also may adversely alter clinical data of a disease. In a somewhat similar manner, concurrent diseases may interact, one reducing (*masking*) or less frequently enhancing a clinical datum of another. Let's consider some examples:

- Chest pain of acute myocardial infarction may be masked by concurrent diabetes, strong analgesics, or advanced age.
- A positive tuberculin reaction may be rendered negative by a concurrent acquired immune deficiency syndrome (AIDS) or a drug (*e.g.*, a corticosteroid).
- A systolic hypertension may be reduced by concurrent acute myocardial infarction or shock.
- Inflammatory symptoms of rheumatic diseases or appendicitis may be suppressed by corticosteroids or antibiotics.
- Diseases that affect liver function are able to produce a false negative cholecystogram, even with a normal gallbladder, because of the incapacity of the liver to concentrate the contrast media. This case should be considered a masking situation, where a liver disease masks or cancels a clinical datum for a normal gallbladder, and should be processed accordingly.
- Typical hypophosphatemia of primary hyperparathyroidism is masked by a concurrent renal failure, produced by this disease, raising phosphate to a false normal level.

Disease and drug interactions are dangerous, because they can mask important clinical data and result in misdiagnosis. This is especially important for diagnosis of life threatening diseases.

The affected clinical datum typically is diminished in intensity or completely masked, as in some of the above examples; we are dealing with a clinical datum absent that would otherwise be present in the disease. In our diagnostic algorithm, the absence of an expected clinical datum tends to rule out the disease in direct proportion to the S of this datum. In the first example, chest pain for acute myocardial infarction has a great S (occurs frequently). With the mini-max procedure, absence of chest pain, a consequence of concurrent diabetic neuropathy, would greatly reduce the P of myocardial infarction and the missed diagnosis could have dismal consequences. Accordingly, if a concurrent disease cancels a clinical datum of the primary disease, the S of this clinical datum must be proportionally reduced, to diminish its rule-out power. A practical solution is to consider chest pain S = 0 whenever myocardial infarction is suspected in a diabetic patient; this is equivalent to eliminate chest pain from diagnostic consideration. In this case the diagnosis of myocardial infarction must be achieved with other clinical data present such as an ECG and cardiac enzymes.

Masking occurs infrequently; therefore only those diagnoses and clinical data known to be susceptible to masking are processed for interaction. Clinical data *present*, either initially collected or recommended as best cost-benefit clinical data, obviously are not masked. Masking refers to a clinical datum *absent*, posing a dilemma whether it is genuinely absent or masked by a concurrent disease or drug. Each clinical datum susceptible to be masked has associated a list of drugs and diseases able to mask it. Only clinical data absent with great S are relevant because only they significantly reduce P of a diagnosis. Summarizing, a clinical datum potentially masked must be detected, have a great S for the corresponding diagnosis, and found absent. When a clinical datum *absent* of *great* S is processed, the algorithm checks whether it is susceptible to be masked. If so, potentially interacting diagnoses are added to the differential diagnosis list to be confirmed or ruled out, and the user is asked whether the patient is receiving specific drugs capable of interaction. If any of these drugs or diagnoses is confirmed, S of the clinical datum susceptible to be masked is reduced to zero, which is equivalent to delete it from the corresponding diagnosis, and total P of the diagnosis is calculated with other clinical data present. This is valid only for the diagnosis which clinical datum is masked, whereas the same masking diseases may be not masking for the same clinical datum in other diseases. When a clinical datum assumed absent is found present, it is disregarded.

Because only certain specific clinical data of specific diseases are susceptible to be masked by specific concurrent diseases or drugs, our previous algorithm described in our former book [1] flags such clinical data in the corresponding disease models and lists the potential masking diseases and drugs.

Our new algorithm [2] handles the problem in a way that is similar to the complex clinical presentations described above, because some related concurrent diseases or drugs with masking property must be processed for presence or absence. With our new diagnostic program, these diagnoses and drugs are included in specific complex clinical presentation models together with the diagnosis that comprises the clinical datum susceptible to be masked, processing masking through the following steps:

1. Create a complex clinical presentation model—in *Complex Presentation Models* input file—for each diagnosis comprising a clinical datum susceptible to be masked (*e.g.*, chest pain of myocardial infarction, which can be masked by concurrent diabetes, potent analgesics, or advanced age). The model, numbered Ixxxx, is given an appropriate title (*e.g.*, MYOCARDIAL INFARCTION WITH MASKED CHEST PAIN) and includes the following items processed like potentially concurrent diagnoses, numbered Dxxxx: (1) Diagnosis with a potentially masked clinical datum (*e.g.*, MYOCARDIAL INFARCTION WITH MASKED CHEST PAIN), (2) The potentially concurrent diagnoses (DIABETES, MASKING DRUGS, ADVANCED AGE) that, if confirmed, could mask the clinical datum (chest pain). Creating these complex clinical presentation models including the diagnoses with clinical data susceptible to be masked, is equivalent to flagging these diagnoses and clinical data.

2. Include, if not already included, in the database—*Disease Models*—disease models for the diagnoses mentioned in the previous paragraph, with their corresponding clinical data: MYOCARDIAL INFARCTION WITH MASKED CHEST PAIN, DIABETES, ADVANCED AGE, and MASKING DRUGS, the latter considered a diagnosis; a single masking drug is sufficient to confirm the diagnosis MASKING DRUGS. Now, the database includes the original myocardial infarction *without* masked chest pain and the added myocardial infarction *with* masked chest pain, the latter disease model omitting this clinical datum (chest pain).

3. Run the diagnostic program—*Diagnostic Program*. When supporting clinical data present select a diagnosis with a clinical datum susceptible to be masked, the differential diagnosis list—*Comprehensive Differential Diagnosis List*—will display two similar competing diagnoses: <DIAGNOSIS WITHOUT MASKING> and <DIAGNOSIS WITH MASKING>. However, a problem results at this point: the confirmatory clinical datum (*e.g.*, increased troponins with S = 1.00) of acute myocardial infarction without masking competes with the same confirmatory clinical datum (increased troponins also with S = 1.00) of myocardial infarction with masked chest pain. This yields a total P of only 0.50 for each diagnosis, instead of 1.00 for one of them, because these two S values (each = 1.00) add in the denominator of equation 1, reducing to half the PP value of the mentioned datum and consequently the P of the final diagnosis. To preclude this from occurring, we resort to an artifice, adding in the original disease model in the database (MYOCARDIAL INFARCTION WITHOUT MASKED CHEST PAIN) the confirmatory clinical datum without masking (*increased troponins without masked chest pain*) to the already existing confirmatory clinical datum (*increased troponins*). Similarly, we add in the disease model with masking (MYOCARDIAL INFARCTION WITH MASKED CHEST PAIN) the confirmatory clinical datum with masking (*increased troponins with masked chest pain*) to the already existing confirmatory clinical datum (*increased troponins*). All these clinical data have the same S = 1? For clinical data with small S it is not necessary to repeat these clinical data with and without masking, because they have no relevance in confirming or ruling out diagnoses. Conversely, every clinical datum with great S,

having greater ruling out power (P reaching deletion threshold), must be duplicated the way mentioned above.

Once both competing diagnoses—DIAGNOSIS WITHOUT MASKING> and <DIAGNOSIS WITH MASKING—are in the differential diagnosis list, output files *Global Overview*, *Abridged Global Overview*, *Data Cost Procedure Quantity*, and *Abridged Data Cost Procedure Quantity* will recommend both best cost-benefit clinical data: increased confirming clinical datum with masked clinical datum (*increased troponins with masked chest pain*) and increased confirming clinical datum without masked clinical datum (*increased troponins without masked chest pain*). Which of these two recommended data must be selected and entered in *Present Data* depends on whether at least one of the masking diagnoses was confirmed final or not.

4. Now, it is necessary to establish whether a masking diagnosis can be confirmed. If a diagnosis able to mask a clinical datum of another diagnosis—as establish by *Complex Presentation Models*—is already included in the differential diagnosis list, supported by at least one clinical datum present, it will automatically be processed. However, if no supporting clinical data for this diagnosis were collected so far, all its clinical data (listed in the corresponding disease model) will be displayed as recommended best cost-benefit clinical data in *Complex Comprehensive* and *Comprehensive Differential Diagnosis List*, sorted by cost category and are expected to be quite numerous. A reduced list of these recommended best cost-benefit clinical data can be seen in parameter affected abridged output files. After investigation, these clinical data are entered in *Present Data* or *Absent Data* respectively and saved; the program—*Diagnostic Program*—is run again. Masking diagnosis and its P will be displayed now in the differential diagnosis list.

5. If any masking diagnosis is confirmed (DIABETES, MASKING DRUGS, or ADVANCED AGE), we enter the confirmatory datum with masking (*increased troponins with masked chest pain*) in the list of clinical data present—*Present Data*. If no masking diagnosis is confirmed, we enter the confirmatory datum without masking (*increased troponins without masked chest pain*) in the list of clinical data present—*Present Data*.

6. The diagnostic program—*Diagnostic Program*—is run again and the result displayed in *Comprehensive Differential Diagnosis List*, *Complex Comprehensive*, and several other output files, will confirm one (MYOCARDIAL INFARCTION WITHOUT MASKED CHEST PAIN) or the other (MYOCARDIAL INFARCTION WITH MASKED CHEST PAIN) of the two competing diagnoses. P of the confirmed diagnosis will now equal 1, because although increased troponins alone in both diagnoses continue to compete with each other, troponins with masked chest pain and troponins without masked chest pain are mutually exclusive and are distinct clinical data. The mini-max procedure will yield a P = 1, superseding the lower value of troponins were they processed alone.

Without flagging, how do we know which clinical data are susceptible to be masked and which diagnoses include them? There are three clues:

i. Two similar diagnoses competing in the differential diagnosis list.

ii. Their denominations (with and without masking).

iii. Complex clinical presentation models list the masked and masking diagnoses in input file *Complex Presentation Models*, and output files *Complex Comprehensive* and *Complex Short*.

For *associated diagnoses* we process all diagnoses in the complex clinical presentation model only if related to confirmed diagnoses (those diagnoses with P equal to or greater than the *Confirmation Threshold*). Instead, for *disease and drug interactions* (masking) we process all masking diagnoses in the complex clinical presentation model only if the potentially masked diagnoses reaches a P equal to or greater than the *Cutoff Present* parameter [2] (with reasonable chance to become confirmed by other supporting clinical data). The confirmation of a masking diagnosis and the removal of masked clinical datum from the masked diagnosis will increase P of the latter, precluding it from being ruled out, which otherwise could have occurred were its P unduly reduced by S of such clinical datum if not removed.

3. *Related and unrelated concurrent diagnoses*

Complex clinical presentation models also distinguish related from unrelated concurrent diagnoses. When two or more concurrent diagnoses are included in a single complex clinical presentation model, by definition these diagnoses are related. Conversely, if no single model exists that includes (relates) the concurrent diagnoses, they are unrelated.

4. *Health assessment and early detection of disease (occult diseases)*

Health or normality is a diagnosis by exclusion of all possible diseases. A healthy patient has no complaints and no abnormal clinical data. However, patients sometimes ignore or underestimate their symptoms, or even hide them for social or legal reasons. Also, many diseases are occult, at least in their early stages. Health diagnosis mandates a comprehensive history and physical examination identical to that performed for diseased patients, only that we have no initial clues for possible disease. If some clinical datum is unveiled during the history or physical examination, even though it might seem irrelevant or unimportant (*e.g.*, a mild tension headache), we are obliged to enter it in the computer, so that the diagnostic program can evaluate it. If the medical history and physical examination are so far completely normal, the patient still could have an occult or incipient disease, such as diabetes, lipid abnormalities, or cancer, which often is asymptomatic in its early stages. Early detection of these occult diseases offers a better chance for cure. Consequently, despite a normal history and physical examination, health assessment and early detection of disease mandate additional clinical studies to improve diagnostic accuracy.

Health diagnosis poses the same problem as diagnosing overt disease, namely how many and what kind of tests and procedures would provide reasonable confidence that the patient indeed is healthy and all possible occult diseases have been ruled out. No ideal solution exists to this problem, because medicine is an inexact science. Even were a patient subjected to all currently available tests and diagnostic procedures—a practical impossibility—a disease still could be missed. When one should stop considering further diagnostic efforts is unclear. For an apparently normal person, one cannot request biopsies of all his organs, a laparoscopy, or other invasive and costly procedures. The limit of such efforts depends on patient age and gender, risk factors, financial status, willingness to submit to recommended procedures, insurance company approval, involved liability, and many other factors.

Malingering and hypochondriasis are diagnoses that a computer algorithm should be able to detect. It is suspected whenever clinical data do not converge to a single diagnosis or combination of diagnoses. No diagnosis reaches a confirmation threshold. This situation mandates a comprehensive work-up, including a comprehensive medical examination, laboratory tests, or other diagnostic procedures, which are expected to be within normal limits.

The way our algorithm deals with occult disease or absence of disease is as follows. Sometimes the occult disease diagnosis is already included in the differential diagnosis list, supported by at least one clinical datum present shared with some other apparent diseases or is associated with a confirmed diagnosis, in which case it will automatically be processed. However, to preclude missing an occult disease that has no single clinical datum present to support it and no relation to any confirmed diagnoses, a specific complex clinical presentation model is created, comprising all diagnoses with serious prognosis, having the potential to remain occult for some time, and occurring frequently. We named this model *Occult Diseases*. A comprehensive database is expected to include already all the disease models corresponding to such diagnoses. Detecting occult diseases is achieved through the following steps:

1. Create a complex clinical presentation model preceded by letter O—in *Complex Presentation Models* input file—named OCCULT DISEASES, listing all diagnoses with serious prognosis and having the potential to remain occult for some time.

2. Include in OCCULT DISEASES model a diagnosis preceded by letter D, named OCCULT DISEASES ACTIVE in addition to all potentially occult diseases.

3. Include OCCULT DISEASES ACTIVE diagnosis in the database—*Disease Models*—comprising only one clinical datum preceded by letter C, named *activate occult diseases*, in *no cost* category, assigning it an S = 1. Because this clinical datum is not manifested by any other diagnosis, it will yield a PP value of 1 (equation 1) and consequently a P of 1 (equation 2) for the mentioned diagnosis and for the model.

4. If the user wants to check for occult diseases, he must activate the complex clinical presentation model OCCULT DISEASES; this is accomplished by entering the clinical datum *activate occult diseases* in the *Present Data* input file and running the program—*Diagnostic Program*. As mentioned above, this clinical datum present, will confer a P = 1 to the "diagnosis" OCCULT DISEASES ACTIVE that will be displayed as confirmed on top of all other diagnoses in the differential diagnosis list, acting as an alert that occult diseases model is activated.

5. For each diagnosis listed in this *Occult Diseases* model, if no supporting clinical data for this diagnosis were collected so far, all its clinical data (in the corresponding disease model) will be displayed as recommended best cost-benefit clinical data, sorted by cost category. These clinical data are expected to be quite numerous in *Comprehensive Differential Diagnosis List*, but will be limited by parameter affected abridged output files.

6. Then the user decides up to which cost level he wants to select these best cost-benefit clinical data, based on patient and social conditions.

7. After investigating the selected best cost-benefit clinical data, enter them in respective *Present Data* or *Absent Data* and save.

8. Run *Diagnostic Program* again until the potentially occult diagnoses are confirmed or ruled out, or until cost becomes prohibitive. If all recommended best cost-benefit clinical data, up to a reasonable cost level, are absent, ruling out or reducing considerably P of occult diagnoses, one can reasonably assume that the patient is healthy.

COMMENTS

Most of previous existing computerized diagnosis programs are not able to diagnose complex clinical presentations, where diverse diseases and clinical entities afflict simultaneously a single patient. Our diagnostic system solves this problem resourcing to complex clinical presentation models, listing all possible combinations of related diseases and clinical entities. These models enable diverse applications: precluding overlooking diagnoses associated with concurrent confirmed diagnoses, detecting disease and drug interactions (masking), distinguishing related from unrelated concurrent diseases, and health assessment and early detection of disease (occult diseases).

CONCLUSIONS

Our algorithm and program, although somewhat complex, is straightforward, especially when compared to other attempts in this field. It emulates a clinician's diagnostic reasoning. It is logical and mathematically simple. Bayes formula is used with modifications, because it is unable to process properly interdependent clinical data (as are most symptoms) and concurrent diseases. To facilitate implementation and updating of the algorithm, we tend to avoid complicated tools of artificial intelligence, such as causal, hierarchical, and probabilistic trees and networks. The algorithm freely uses heuristic procedures, so as to preclude excessive proliferation of clinical data and diagnoses. It promises to be user friendly because it is expressed in natural language, is rational, and readily understandable. Determination of accurate sensitivity of clinical data and integration of clinical entities into complex clinical presentation models will be labor-intensive. A complete database with all known diseases, clinical data, clinical presentations, and other information can be created; this major task will require a dedicated team of medical specialists.

REFERENCES

[1] FEDER C. Computerized Medical Diagnosis: A Novel Solution to an Old Problem. Infinity Publishing, West Conshohocken, Pennsylvania, 2006

[2] FEDER C, and FEDER T. A Practical Computer Program that Diagnoses Diseases in Actual Patients. Infinity Publishing, West Conshohocken, Pennsylvania, 2008

LEDLEY RS and LUSTED LB. Reasoning Foundations of Medical Diagnosis. Science, 130 (9): 9-21, July 3, 1959

HENRION M, PRADHAN M, DEL FAVERO B, HUANG K, and O'RORKE P: Why is diagnosis using belief networks insensitive to imprecision in probabilities? Twelfth Conference on Uncertainty in Artificial Intelligence, Portland, OR, 446-454. 1996. SMI-96-0637

MYERS JD, POPLE HE, and MILLER RA. INTERNIST: Can Artificial Intelligence Help? In: Connelly, Benson, Burke, Fenderson, eds. Clinical Decisions and Laboratory Use. Minneapolis: University of Minnesota Press, 1982: 251-269

LUSTED LB. Introduction to Medical Decision Making. Springfield, Illinois, Charles C Thomas, 1968

LUSTED LB. Twenty Years of Medical Decision Making Studies. CH1480-3/79/0000-0004$00.75. 1979 IEEE

POPLE HE. Heuristic Methods for Imposing Structure on Ill-structured Problems: The Structuring of Medical Diagnosis. In: Szolovits P, ed. Artificial Intelligence in Medicine, AAAS Symposium Series, Boulder, Colorado: West-view Press, 1982: 119-185

MARTIN J: Computer Data-Base Organization. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1977

CHAKRAVARTY S and SHAHAR Y. A Constraint-Based Specification of Periodic Patterns in Time-Oriented Data. Sixth International Workshop on Temporal Representation and Reasoning (TIME-99), Orlando, FL, 29-40. 1999. SMI-1999-0766

ALTMAN RB. AI in Medicine: The Spectrum of Challenges from Managed Care to Molecular Medicine. AI Magazine 20(3):67-77, 1999. SMI-1999-0770

BLEICH HL. Computer-Based Consultation: Electrolyte and Acid-Base Disorders. The American Journal of Medicine 53: 285-291, November 1972

BLOIS MS, TUTTLE MS, and SHERERTZ DD. RECONSIDER: A Program for Generating Differential Diagnoses. IEEE: 263-268, 1981

CHAKRAVARTY S and SHAHAR Y. Acquisition and Analysis of Periodic Patterns in Time-Oriented Clinical Data. 2000. SMI-2000-0822

DE DOMBAL FT, LEAPER DJ, STANILAND JR, McCANN AP, and HORROCKS JC. Computer-Aided Diagnosis of Acute Abdominal Pain. British Medical Journal 2: 9-13, 1972

ESHELMAN L, EHRET D, McDERMOTT JP, and TAN M (1987.) MOLE: A Tenacious Knowledge Acquisition Tool. International Journal of Man-Machine Studies, 26: 41-54

GENNARI J, MUSEN MA, FERGERSON RW, GROSSO WE, CRUBEZY M, ERIKSSON H, NOY NF, TU SW: The Evolution of Protégé: An Environment for Knowledge-Based Systems Development. 2002. SMI-2002-0943

GINI RA and FEDER C. Informatica Clinica: Presente y Futuro. La Semana Medica (Argentina) 157: 113-128, 1980

GORRY GA, KASSIRER JP, ESSIG A, and SCHWARTZ WB. Decision Analysis as the Basis for Computer-Aided Management of Acute Renal Failure. The American Journal of Medicine 55: 473-484, 1973

GORRY GA, PAUKER SG, and SCHWARTZ WB. The Diagnostic Importance of the Normal Finding. The New England Journal of Medicine 486-489, March 2, 1978

GREENS RA, PELEG M, BOSWALA AA, TU S, PATEL VL, and SHORTLIFFE EH. Sharable Computer-Based Clinical Practice Guidelines: Rationale, Obstacles, Approaches, and Prospects. Medinfo, London, UK, 2001. SMI-2001-0860

HUANG K, HENRION M: Efficient Search-Based Inference for Noisy-OR Belief Networks: Top Epsilon. Proceedings of the Twelfth Conference of Uncertainty in Artificial Intelligence, 325-331. Aug 1996, Portland, OR. SMI-96-0640

ILIAD® 4.5 Diagnostic and Reference Tool for Physicians and Medical Professionals. User Guide. 1998

JAAKKOLA TS, JORDAN MI: Variational Probabilistic Inference and the QMR-DT Network. Sun May 9, 16:22:01 PDT 1999

LUDWIG DW. INFERNET – A Computer-Based System for Modeling Medical Knowledge and Clinical Inference. Proceedings of the Fifth Annual Symposium on Computer Applications in Medical Care: 243-249, November 1981

MIDDLETON B, SHWE M, HECKERMAN D, HENRION M, HORVITZ E, LEHMANN H, & COOPER G: Probabilistic Diagnosis using a reformulation of the INTERNIST-1/QMR Knowledge Base II. Evaluation of Diagnostic Performance. Section on Medical Informatics Technical report SMI-90-0329, Stanford University, 1990

MILLER RA, POPLE HE, and MYERS JD. INTERNIST-I, An Experimental Computer-Based Diagnostic Consultant for General Internal Medicine. The New England Journal of Medicine, 1982: 468-476

MUSSEN MA, GENNARI JH, and WONG WW: A Rational Reconstruction of INTERNIST-I Using PROTÉGÉ-II. Nineteenth Annual Symposium on Computer Applications in Medial Care. New Orleans, LA, 289-293. 1995. SMI-95-0574

MUSEN MA: Modeling for Decision Support. Section on Medical Informatics. Stanford University School of Medicine. Stanford, CA 94305-5479. SMI-98-0739

NOY NF and MUSEN MA. Ontology Versioning as an Element of an Ontology-Management Framework. Stanford Medical Informatics, Stanford University, 251 Campus Drive, Stanford, CA 94305, USA. SMI-2003-0961. March 31, 2003

PATRICK EA: Decision Analysis in Medicine: Methods and Applications. CRC Press, West Palm Beach, FL, 1979

PERLROTH MG, and WEILAND DJ. Fifty Diseases: Fifty Diagnoses. Year Book Medical Publishers, 1981

POPLE HE, MYERS JD, and MILLER RA: Dialog: A Model of Diagnostic Logic for Internal Medicine. Fourth International Joint Conference on Artificial Intelligence. Tbilisi, Georgia, URRS, 3-8 September 1975, Volume Two. 1975: 848-855

PRADHAN M, DAGUM P: Optimal Monte Carlo Estimation of Belief Network Inference. Proceedings of the Twelfth Conference of Uncertainty in Artificial Intelligence, 446-453. Aug 1996, Portland, OR. SMI-96-0638

SHORTLIFE EH. Computer-Based Medical Consultations: MYCIN. American Elsevier Publishing Company, 1976

SHORTLIFFE EH. The Next Generation Internet and Health Care: A Civics Lesson for the Informatics Community. In C.G. Chute, Ed., 1998 AMIA Annual Symposium, Orlando, FL, 8-14. 1998. SMI-98-0730

SHWE M, MIDDLETON B, HECKERMAN D, HENRION M, HORVITZ E, LEHMANN H, & COOPER G. Probabilistic Diagnosis Using a Reformulation of the INTERNIST-1/QMR Knowledge Base I. The Probabilistic Model and Inference Algorithms. Methods of Information in Medicine, 30(4):241-255, 1991. SMI-90-0296

SZOLOVITS P and PAUKER SG. Categorical and Probabilistic Reasoning in Medical Diagnosis. Artificial Intelligence. 11: 115-144, 1978

TU SW et al: Modeling Guidelines for Integration into Clinical Workflow. Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. SMI-2003-0970

VAN BEMMEL JH, MUSSEN MA (eds): Medical Informatics. Springer, 1997

WEINSTEIN MC and FINEBERG HV: Clinical Decision Analysis. W. B. Saunders Company, 1980

WEISS S, KULIKOWSKI CA, and SAFIR A. Glaucoma Consultation by Computer. Comput. Biol. Med. 8: 25-40, 1978