Specialized Computer Support Systems for Medical Diagnosis. Relationship with the Bayes' Theorem and with Logical Diagnostic Thinking

Pedro José Negreiros de Andrade

Fortaleza - CE -Brazil

"No one should fool themselves into believing that they can compete with the memory of a computer, and much more than that, with the utilization speed of all memorized data. It is a new world which rapidly arises and which will alter the structure of medical work, and principally, medical ethics" [1].

Specialist computer systems are the outcome, result, or both of the application of what is konwn "Knowledge engineering", one of the subspecialties of artificial intelligence [2]. Such systems use simple techniques of artificial intelligence to simulate the action of human experts. One of the characteristics that an artificial intelligence system has is the capacity to acquire knowledge, i.e., modify itself with the application.

This, as a matter of course, does not happen with the so-called specialist systems, which although perform sometimes comparably to human specialists in the resolution of specific problems, are not capable, usually, of learning, i.e., of exhibiting really intelligent behavior. The greatest interest in these systems dates back to the early '70s, when the artificial intelligence paradigm underwent an important change, very well described by Goldstein [3].

One might deduce that the efficiency of a specialist system depends more on the quantity of knowledge input than the capacity of acquiring knowledge. The fact is that no perfect, complete method has been found yet to create an intelligent environment without sufficient knowledge to think about it. A recent and important advancement in knowledge acquisition is the system based on neuronal network, inspired by biological models, i.e., the anatomical functional basis of the brain [4,5]. But, until this change in the paradigm produces greater results, the specialist system will continue to depend much more on the quality of its basic knowledge than the complexity of its algorithms.

Therefore, within this logic of basic knowledge application, specialist diagnostic systems in medicine were divided, initially, into two types [6]: a) systems based on rules; b) systems based on recognition of patterns.

Most of the initial efforts to apply artificial intelligence techniques to solve real problems in medicine, were founded on rules-based systems. Such programs are fairly easy to create, because their knowledge is catalogued in the shape of rules of the type "*If...then*" used in a chain form with the aim of reaching a conclusion. This type of program easily manages to appear intelligent, allowing also the easy use of the so-called decision tree, very frequently applied in modern medicine. In areas of limited domains, such programs have shown themselves to be fairly effective. One well known example is the Mycin System, developed at Stanford University for antibiotics selection in patients with infection [6]. This program has already been largely tested showing a performance similar to that of specialists in the sphere of infectious diseases. The problem with these rules-based systems is that in more complex areas, such as diagnosis in Internal Medicine, the amount of knowledge is so vast that it is extremely difficult to absorb all of it. Therefore, systems of this type, designed primarily for the medical area, have found a greater application in commercial tasks or techniques, such as "telephone line evaluations". In spite of the obvious utility of these systems in medical programs based on decision trees, the main limitation of their use comes from the difficulty in accepting computer procedures that may place human lives at risk [6].

In view of the difficulties in applying rules-based systems in areas of extensive knowledge, such as Internal Medicine diagnosis in, the solution would seem to be in pattern-matching systems. At the elementary level, such systems function as follows: 1) for each possible illness (hypothetical diagnosis) a response that the patient presents is identified (symptom, sign or findings of complementary examination), and it is evaluated whether it is part of the disease; 2) scores are established for each illness, according to the number of symptoms that are the same as those exhibited by the patient; 3) the illnesses are classified according to these scores; 4) inquiries are made about whether after findings of the illnesses with the highest score are also found in the patients; 5) repeat steps 1 and 2; 6) repeat the procedure for following illnesses [3,4].

Hospital Universitário Walter Cantídio – UFC - Fortaleza
Mailing address: Pedro José Negreiros de Andrade – Rua Tibúrcio Cavalcante, 1445/802 – 60125-100 – Fortaleza, CE - Brazil

If a pattern-matching system were to be developed in this simple format, it would be similar to the manner in which many beginners approach the diagnostic process. As a consequence, it would fail for not addressing the following issues: a) how strongly is a finding (symptom, sign or com–plementary exam results) a reflection of a certain disease [2]; b) how strongly does the absence of a finding eliminate the possibility of a particular illness?; c) which is the prevailing rate of each of the hypotheses in the population under study [2]; d) how strongly does a finding in the patient but not present in the hypothesis invalidate the diagnosis [2]; e) the patient presents with only one or more than one illness [2]; f) if there is more than one illness, would the illness be related [2].

It is known that the experienced physician with a strong acquired knowledge base uses this information with competence during the diagnostic process, this being one of the characteristics that sets him apart from the beginner [8-11].

For this reason, the best known computer programs in the area of differential diagnoses, such as the Internist [i.e], its variation Qmr [13], Dx Plain, Iliad, in addition to the Consultor program, developed by us, incorporate a pattern-matching system, aspects of probabilistic thinking and (in the case of Internist) of a casual relationship. So that specialist systems of this type are better understood, it is necessary that we discuss in more detail the logical basis of the diagnostic process.

Logic of the diagnostic process

"The advancement in technology makes the diagnostic problem a main preoccupation. The final diagnosis has left the sphere of the clinic and has become dependent on very expensive medical technology and is potentially iatrogenic. This reality requires us to act with greater care in using scientific methods in clinical discussion, from the starting point that iatrogeny and medical expenses diminish at the rate of correctly ordered complementary examinations" [14].

Basic principles

The word "diagnosis" comes from the Greek meaning to discern or distinguish. In the medical point of view it has been defined by Mason [14] as a series of intellectual and operational procedures through which one obtains an answer to a clinical problem [14]. In this sense there are several words that complement the word "diagnosis" modifying its meaning: clinical diagnosis, anatomical and pathological diagnoses, radiological diagnosis, electrocardiographical diagnosis, etc. The expression "differential diagnosis" was defined by Harvey [11] as the art of distinguishing one illness from another establishing one or more well defined reasons to explain the alterations presented by the patient. The classic teaching instrument of this form of diagnosis has been the clinical pathological session in which particularly difficult cases are submitted to the analysis of general practitioners or specialists teaching purposes. In clinical practice, as well as in the referred sessions, the sequence by which the differential diagnosis is carried can be summarized in the following stages [11]. 1) Acquisition of data – 1a) clinical history; 1b) physical examination; 1c) laboratory examination. 2) Data analysis; 2a) critical evaluation of obtained data; 2b) listing of findings in order of importance; 2c) selection of one or preferably two to three central findings; 2d) listing of illnesses in which these central findings can be found; 2e) searching for final diagnosis by selection of an illness that better explains the findings of the patient; 2f) revision of all positive data, in order not to leave any finding considered important without an explanation.

The sequence mentioned is not necessarily so rigid. The case analysis usually begins during the collection of data, when the physician uses his experience to detail the findings. Upon a complaint of shortness of breath, a Cardiologist will try to define it "during exercise or rest" in a trial endeavor to make a diagnosis of heart failure. From there on during the physical examination the Cardiologist will look for signs of physical congestion (jugular turgescence, rales), or cardiac enlargement, (apex deviation), third heart sound) to confirm the supposed hypothesis. In a patient with prolonged fever, who presents with a mitral murmur, an experienced physician will search for signs of embolism, paleness, splenomegaly or drumming of the fingers, to confirm a suspicion of infectious endocarditis (IE). The two examples show how data analysis can start early, perfecting the result of the history and of the physical examination.

When data are collected, the physician classifies it by order of importance and he process of differential diagnosis begins, in the sequence proposed previously. Some physicians try many times to characterize combinations of results as *syndromes* with the intent of simplifying the differential diagnosis process. Expressions such as consumptive syndrome, diarrhea syndrome, pulmonary hypertension syndrome, congestive syndrome, infectious syndrome, syndrome of pleural effusion and others are frequently used. They represent, in reality, forms of expressing in a few words a combination of symptoms, signs or laboratory findings thought to be of major importance.

In spite of the extreme didacticism of this approach, some problems may occur during its use. The first one is the wrong choice of the central findings. It requires some experience in definition of findings, that, due to its specificity, deserve to be in the core of the diagnostic process. The erroneous choice of a central finding may make the list of possibilities excessively long, making the determination of a diagnosis more difficult. Besides this, the exclusion of a finding pertaining to the list, is a common mistake, of serious consequences for the diagnostic process [15]. Another common mistake, even among the more experienced, is the premature conclusion of a definitive diagnosis or syndrome, without the existence of unquestionable data to establish it [15]

To better understand how a differential diagnosis is carried out, how the above errors can be avoided and how we can simulate medical thinking in a computer program, it is interesting that we make a critical revision of the tree types of diagnostical thinking customarily used.

Types of diagnostical thinking

According to Sox [16] there are types of basic diagnostical thinking: 1) Physiopathological thinking; 2) pattern-matching thinking; 3) probabilities thinking.

Physiopathological thinking is the most difficult to simulate in a computer program. When a patient describes a history of a fever lasting for six weeks, followed three weeks later by pain of increasing strength in the abdominal right superior quadrant, the physician begins to imagine a growing mass that puts pressure on structures sensitive to pain. The physician starts then to think about a liver abscess or a malign nodule with a necrotic center, causing inflammation.

Pattern-matching thinking is more frequently used as much by medical students as by specialists. It is also easier to simulate in computer programs. Certain findings happen together, and their combination leads the physician to formulate hypotheses. From there onwards, the physician compares the patient data with that of the illness (hypothetical) to check to what degree they might come together. One of the flaws of this type of thinking is the inability to identify an illness when it is present in a form different from that of its classic manifestation [7,14]. The others flaws are similar to the ones commented upon in relation to the computer programs that use rules of simple pattern-matching: a) not taking into consideration how strongly the presence of a finding evokes an illness; b) not taking into consideration how strongly the absence of an expected finding removes a determined hypothesis; c) not taking into consideration the prevailing illness; d) not taking into consideration the non-pertinence of a finding, i.e., how strongly the fact of it not being explained by an illness eliminates it as hypothesis.

Probabilistic thinking is based on the fact that physicians deal with uncertainty to a degree that is hardly comparable to other professionals. Probability, in this case, would only be a way to measure this uncertainty when two physicians say, for example, that a patient probably has pulmonary embolia; one could be thinking of a 30% probability, and the other of 90%. In handling this uncertainty, the physician dealing with it many times counts on laboratory examinations, which alter it, but do not eliminate it. In the probabilistic interpretation of these tests, the experienced physician uses; intuitively, the so-called Bayes Theorem, [16,17] which relates sensitivity, specificity , prevalence and post-test probability. Obviously, the application of such concepts should ideally refer not only to laboratory examinations, but to findings of patient history or physical examination.

In addition to these three classic forms of thinking, two other techniques for making a diagnosis deserve to be mentioned - intuitive thinking and diagnosis by aphorism. Intuitive thinking is probably an unconscious mixture of the three types of thinking already mentioned incorporating other subjective elements that are difficult to define. Diagnosis by aphorism is more specifically a mixture of thinking by pattern- matching with probabilities thinking. Such aphorisms constitute themselves in a series of rules based on experiences by well-respected physicians, for example, the Courvoisier sign ("jaundice due to the obstruction of the common bile duct associated with palpable, painless gallbladder suggests a patient with cancer of the head of the pancreas") Another example is "jaundice associated with paleness and enlarged spleen makes one think of hemolytic anemia". Such aphorisms, however memorable, and encompassing an infinite combination of clinical findings, constitute, in fact, a simplification of complex problems. Besides their memorization, aphorisms contribute little to teaching the logic of the diagnostic process. It is interesting to note that great emphasis has been placed in recent years on the so-called rules of clinical prediction. They have as their objective reducing the degree of doubt in the diagnosis by determining how to use clinical findings to make predictions. A more careful analysis shows that these rules of clinical prediction are the modern, re-edition of old aphorisms [18]. The difference is that such regressions established through the analysis of hundreds of patients, use calculations of probabilities and sophisticated mathematical techniques. Therefore, they have been the focus of great interest in recent years, particularly for their potential use in support systems of medical decision making [18,19].

Due to the importance of probabilistic thinking in the support systems for medicine, we shall discuss it in further detail on the following pages:

Probabilistic thinking in medicine

"When you can measure a phenomenon about which you are talking and express it in numbers, you know something about it. But, when you can not express it in numbers, your knowledge is vague and unsatisfactory. It may be the beginning of knowledge, but you progressed very little toward the state of science". Lord Kelvin (1824-1907).

"The appearances for the mind are 4 types: things are what they seem to be, or are not, and not seem to be, or are, and not seem to be, or are not, and even so seem to be. Identifying correctly all these cases is the task of the wise man." Epictetus (2nd Century AD) [20].

Basic concepts

It is well known that medical knowledge is based more on probabilities than on certainties. Probabilities thinking is so important in the diagnostic process, in the definition of clinical prediction [18], and even in treating complex cases, it is astonishing that practically nothing is taught about it in the medical curriculum. To better understand the quantitative aspects of diagnostic logic, or more specifically probabilistic thinking and the Bayes theorem, a short introduction is necessary to the concepts of sensitivity and specificity, starting from the definition of conditional probability. Within this concept for two events whichever A and B, P.(A) > 0, we define conditional probability of B, A having been established as:

$$P(B/A) = \frac{P(A\_B)}{P(A)}$$

Lets now consider the following situation, synthesized in a 2 x 2 table, where number n_1 people known to have an illness and n_2 people known to be free of the illness undergo a particular clinical test.

Test / Illness	Positive	Negative	Total
(I) Present	a	b	n_1
(i) Absent	c	d	n_2
Total	a + c	b + d	N

We have in this situation the following possible results: positive test and illness present: TP (truly positive) = a cases; negative test and illness absent: TN (truly negative) = d cases; negative test and illness present: FN (false negative) = b cases; positive test and illness absent: FP (false positive) = c cases.

When we ask: if the illness is present, what is the probability of a positive test, or better still, P (T/I)? The answer is given in the following manner:

$$P(T/i) = \frac{P(T\_i)}{P(i)} = \frac{a/N}{n1/N} = \frac{a}{n1} = \frac{a}{a+b} = \frac{PV}{PV+FN}$$

This is what is called estimated sensitivity of the test. When we ask: if there is no illness, what is the probability of a negative test, or better still, P (T/I)? The answer is given in the following manner:

$$P(T/i) = \frac{P(T\_i)}{P(i)} = \frac{d/N}{n2/N} = \frac{d}{n2} = \frac{TN}{TN+PF}$$

This is called the estimated specificity of the test.

These sensitivity and specificity concepts are quite well known, however, frequently on an intuitive or superficial level. They were used more intensely after World War II from the analysis of the efficiency of radiologic findings in the diagnosis of tuberculosis [21].

A highly sensitive test very rarely will not be positive in people who really have this illness. As an example of a highly sensitive test would be fever in patients with endocarditis, or the anti-nuclear factors in patients with systemic lupus erythematosus (SLE).

A highly sensitive test very rarely will not be negative in people who do not have the illness. As an example of the high specificity test, we would have the presence of a large vegetation on the echocardiogram for the diagnostic effects of endocarditis. To summarize, high sensitivity tests are useful for not confirming an illness when negative, and highly specific tests are useful for confirming illness when positive. Always, when we alter the positivity criteria of a test to make it more specific, we will be, on the other hand, diminishing its sensitivity. The opposite is just as true.

The concept of prevalence refers to the frequency of the illness in the population under study. The concept of predictive value is more complex, refers to the probability of an individual having the illness after the altered results of a test. We can only know the predictive value of a test if we consider altogether information regarding prevalence, sensitivity and specificity of this test mathematically expressed by the Bayes formula.

Bayes' theorem

Although the concepts of sensitivity and specificity are well understood, the same does not apply to the concepts of pre- and post-test probability and, particularly, with the Bayes' theorem, which in spite of being included in the introductory chapters of classic textbooks [8] is rarely read by or taught to medical students.

Recently, the capacity of medical students to interpret quantitatively laboratory examination results was tested through a question on the medical internship test II", a test that has a 70% sensitivity and a 90% specificity. Applied to a patient belonging to a population in which the prevalence of an illness is 1%, the examination results are positive. What is the post-test probability of this patient having this illness?

It was no surprise at all, to see that the majority of the students placed this probability between 70-90%, instead of choosing the correct answer, which was 6,6%. The only students who replied correctly were those who, during rounds in the Cardiology service, had received some information about the Bayes' theorem through its application to ergometry [22].

Thomas Bayes (1702-1761) was an English philosopher, mathematician and theologian, who is considered one of the founders of probability calculation Bayesian analysis would therefore be a theory of a statistical decision factor for the calculation of probability of a proposition based on the original probability and new relevant factors.

Expressed in clinical terms, it is a concept by which the predictive value of a test or clinical findings depends not only on its sensitivity and specificity, but also on the previous probability, (i.e., the prevalence of the illness in the population under study). It is of importance to emphasize that the pioneering study about the application of Bayes' theorem in medicine was published in 1959 by a specialist in computer science, Ledley, and a physician, radiologist, Lusted, the latter already taking part of the works that gave birth to the medical concept of sensitivity and specificity [20,23].

The original Bayes formula, used by statisticians, is as follows:

$$P(Ci/A) = \frac{P(Ci).P(A/Ci)}{\sum_{j=1}^{N} P(Cj).P(A/Cj)} \quad i = 1,\ldots,n$$

A simplified version, applicable to medicine [29] is the following:

$$VP+ = \frac{Pv.S}{Pv.S + (1-Pv).Pv(1-E)}$$

And analogically:

$$PV- = \frac{Pv.(1-S)}{Pv(1-S)+(1-Pv).E}$$

Where: PV + = predictive value (post-test probability) of positive test;

PV - = predictive value (post-test probability) of negative test; Pv = prevalence of illness (or pre-test probability); S = sensitivity of test or clinical finding; E = specificity of test or clinical finding.

An inherent aspect of the Bayesian formula is the possibility that it can be applied in sequence. In other words, after applying the result of a test to the Bayesian formula, the predictive value obtained becomes the new probability of occurrence of the illness. From there on new probabilities should be calculated because of other tests or findings. The only presupposition for this sequential application is that the tests be independent of each other, i.e., that the result of one should not interfere with that of the other. An example of nondependent tests is the depression of ST segment in the ergometry and the presence of the same depression of ST in the Holter. Another example is paleness during physical examination and finding diminished hemoglobin in the blood analysis.

As far as we could find the first practical trial for applying the Bayes theorem to medicine was in Cardiology through the works of Warner et al. [24,25] who used it as an auxiliary to diagnosing congenital diseases. But, the work that created a greater repercussion in our specialty was published by Rifkin and Hood [26] who managed to explain through it the curious differences in credibility of the stress electrocardiogram when applied to men or women, as well as to patients with symptoms or without symptoms. Other works followed, demonstrating from a practical point of view, the strength of employing Bayesian logic in the interpretation of ergometry [17,27] consequently being greatly understood among cardiologist, and soon extended over all areas of Internal Medicine.

What these works show, if we take a test with 60% (S=0.6) sensitivity and a specificity of 90% (E = 0.9) and apply it in a population with 1% (PV = 0.01), prevalence of illness, that the probability of a patient with a positive test presents with the illness would be (according to the Bayes formula) the following:

$$PV+ = \frac{0.01 \times 0,}{(0.01 \times 0.6)+(0.99 \times 0.1)} = 0.056 \ (or \ 5.6\%)$$

This situation reflects the probability of a 43-year-old woman without symptoms or a young woman of 33 years with noncardiac thoracic pain, having coronary illness through a positive stress test [17,26-28].

The basic lesson from employing the formula in this case is that it is a waste of money, an unnecessary expenditure, to request an electrocardiographic stress test in this situation, as it would hardly alter the action to be taken [29].

However, if the pre-test probability were 50% (case of a 48-year-old man with atypical angina) the positive stress test would alter the probability of coronary illness to 85% if positive and to 28% if negative.

Another form of calculating the post-test probability is through the use of diagrams. Let's take for example a population with probability of 5% of presenting with an illness, taking a test that has a 70% sensitivity and 90% specificity. The case carries a certain analogy, with the use of an excretory urography in a patient with a probability a little greater than the usual of having renal artery stenosis. In case of, therefore, an abnormal urography, what would be the probability of renal artery stenosis in a patient with a previous probability of 5%? To solve this problem we can construct the following diagram:

```
                    1000 patients
                          I
    950 healthy      (PV=5%)=>   50 with stenosis
    95    855<=(E=90%)  (S=70%)=>  35    15
    FP      VN      (VP)    (FN)
                            VP
```

$$\text{Probability post-test} + = \frac{VP}{VP+FP}$$

$$\text{Probability post-test} + = \frac{35}{35+95} \quad \frac{35}{130} = 26.9\%$$

What we have learned from the use of the Bayesian formula in this case is that the practice of requesting urography for the evaluation of possible renal artery stenosis, a rule in the past, should be suspended, considering that the cost and morbidity of the test would be greater than the gain obtained by it [30]. The consequence is that other tests such as the renogram with captopril began to be used instead and even in patients with significant suspicion of renovascular hypertension.

Of the examples mentioned, the great importance of previous probability is clear, i.e., the prevalence of illness in the interpretation of a test or clinical finding. It is also clear that tests with limited sensitivity and specificity alter significantly diagnostic probability only when applied in individuals with intermediate probability of having an illness [16].

In the situations in which the previous probability is very high or very low, limited tests of sensitivity or specificity do not alter the diagnosis. The exception would be tests with high specificity for individuals with low previous probability or tests of high sensitivity for individuals with high previous probability.

Other forms of utilization of the Bayes' theorem have been proposed, for example the use of nomograms [31], and the 2 x 2 table of statisticians and epidemiologists [15] considered by some to be easier applications [32].

In spite of the fairly recent revelation of the Bayes theorem among internists and Cardiologists, especially for use in the interpretation of ergometric tests [22,32], it is important to emphasize that experienced physicians have always employed it, although intuitively, in all stages of the diagnostic process.

When the older general practitioners affirmed that the clinical findings were more important than laboratory examination results when making judgments they were thinking in the Bayesian way without being aware of it. When they affirmed it to be better to think about common illnesses, even though manifestations or symptoms were the unusual instead of typical manifestations of rarities, they were, intuitively, applying the Bayes' theorem. Using the presence of findings of high specificity to confirm diagnoses and findings of high specificity to remove them, is to apply, although not formally, the Bayesian logic. At the time of technological proliferation, when laboratory examinations are unnecessarily requested and erroneously interpreted, to give due importance to the patient's history and to epidemiological data, is to think in the Bayesian way [22].

Diagnosis based on scores, its relation to the Bayes' theorem and to specialist systems.

The Bayes theorem due to its logical perfection, mathematical exactness

and its possibility of sequential application, therefore seems the ideal instrument for building a bridge between medical thinking and computer science. It has, however, some limitations, especially with regard to its use in specialist systems of differential diagnosis: a) difficulty in defining the previous probability in different situations. Programs designed for the USA would have to be modified for regions like the rural part of the state of Ceará. Besides, the prevalence of certain illnesses varies also according to the institution, being sometimes very low in a general outpatient clinic, not so low in the infirmary of a tertiary hospital and high in anatomical-clinical sessions; a) difficulty in applying the Bayes formula sequentially in multiple tests and, especially in circumstances when the tests are not independent [16,33]; b) difficulty of medical specialists in charge of elaborating the database in understanding the formula.

These difficulties would be the probable explanation for the fact that programs based on a rigid application of the Bayes' theorem function better in more limited areas of medical knowledge. One way to minimize them would be the use of the Bayes' theorem converted into a system of scores, as proposed by Rembold and Watson [34].

According to these authors, converting the probability in chances and making the equation linear by use of the corresponding natural logarithm, we would have the conversion of fractional values into whole numbers, which would be called weights or scores, which are to be understood and utilized. The formula proposed by Rembold and Watson would be the following:

$$1n \text{ odds post-test } (+) = 1n \text{ odds pre-test } + 1n \frac{S}{1-E}$$

$$1n \text{ odds post-test } (-) = 1n \text{ odds pre-test } + 1n \frac{1-S}{E}$$

(1n = natural logarithm, S = sensitivity; E = especificity).

Even before entering into the mathematical merit of this proposition, it becomes attractive for suggesting that by attributing positive scores for tests or clinical findings in function, principally, of specificity (if present or positive) and negative scores in function of sensitivity (if absent or negative) one would be, in a way, applying the Bayes' theorem.

Another reason to look at this type of proposal favorably is that the idea of making diagnoses using scores is a long familiar subject to clinicians, with an enormous amount of literature about it. In Cardiology, perhaps Jones [35] was the first to propose, and have accepted, a form of diagnosis analogous to scoring.

The so-called Jones criteria (two major signs or one major and two minor as a condition for a diagnosis of rheumatic fever) are easily transformable into scores. In this case it is sufficient to attribute the value of 5 for the major signs and 2 for the minor signs and consider 9 as the sum of points needed to indicate a probable diagnosis of rheumatic fever. The major signs are obviously of greater specificity, the minor of lesser specificity. Still within Cardiology, in addition to rheumatic fever, such diversified illnesses as coronary heart disease [29], heart failure [36], mitral valve prolapse [37], endocarditis [38] and pericardial effusion have already received proposed criteria for diagnosis through the adding up of scores or something similar.

In rheumatology the criteria of the American College of Rheumatology for diagnosis of rheumatic arthritis [39] and SLE [40] are well known. In endocrinology it is interesting to remember the Wayne criteria of points for the diagnosis of hyperthyroidism [41]. The potential list is enormous, suggesting that diagnosis by scores has a strong logical base (of a Bayesian nature, however without the mathematical rigor of using the formula) becoming an instrument of extremely simple utilization in differential diagnosis software.

For all these reasons, programs, such as the Internist - I and its variant QMR, try to make diagnoses through the use of scores that are attributed to the relation between the clinical findings and the illnesses. One of these scores refers to the evoking strength, which according to the authors would present grossly the concept of positive predictive value, strongly influenced by the specificity and the prevalence. The second one refers to the frequency with which the illness is found, which relates to the sensitivity of the finding.

The Internist program has been tested and reported upon in a classic article [12], showing lower performance compared with those responsible for discussion of cases in the anatomical-clinical sessions of The New England Journal of Medicine, but comparable to the one responsible for patients (clinical diagnosis). In a follow-up evaluation, its variant QMR [42] had a satisfactory performance in terms of being auxiliary to the diagnosis in real cases in university hospitals. Besides these, Dx - Plain [43] of the Harvard University and Iliad [44], are based on similar principles, the latter being rigorously Bayesian.

The Consultor program, initially developed in the area of Cardiology [45] and then extended to all Internal Medicine [46] makes important analogies with the referred systems. Consul-

tor has been evaluated initially through its submission to medical residency tests [46] and, more recently, through the anatomical-clinical sessions of the Walter Cantidio University Hospital and of The New England Journal of Medicine (in this instance only Cardiology cases). The system had, based on these evaluations, a very acceptable performance in Internal Medicine and, particularly, in Cardiology [1, 47] . The four North American systems comparatively evaluated also had an acceptable, comparatively similar performance [48,49]. Some of Consultor's limitations, however, caused specialists in the discipline to conclude that such systems should be used only by physicians capable of identifying and using the pertinent information and ignoring the nonpertinent information supplied by the system.

A series of inquiries can be made in relation to the specialist support systems of medical diagnosis. Could such systems curtail the development of clinical thinking [2]. Should their use by layman or paramedics be permitted or not? Could systems like these replace internists in the future? In relation to the first question, our experience is that the utilization of specialist diagnostic support systems by medical students helps students understand more intensely concepts like sensitivity, specificity and prevalence and their relation to the Bayes' theorem, sharpening, rather than dulling clinical thinking. Obviously, a more definitive evaluation would require that the diagnostic performance of students trained in such systems and that of students not exposed to the programs discussed be compared. The second question is more difficult to answer. Although no

knowledge should, in principle, be privileged to a corporation, the current actual tendency of people in charge of specialist systems is not to give access to the laymen. This is fundamentally due to two reasons: the most important is that one can not yet fully trust the predictive diagnoses of these systems, which could harm patients if used by ~~not~~ inadequately trained individuals; the other is that the specialist systems' performance requires accurate gathering and interpretation of clinical findings, which are not found, at the moment, outside the medical profession.

It seems therefore evident that the freeing of specialist systems like Consultor, QMR, Iliad and Dx-Plain for use by layman could bring more risks than benefits to patients, as the algorithm of these systems needs to be improved, as a long way still continues to separate them from a behavior really similar to medical thinking [49-52] . Such is the case, particularly for Consultor [28], by the incorporation of a greater degree of physiopathological "thinking" to its logic and by establishment of the relation of cause and effect between illnesses. Possibly, this will only be achieved through the substitution of the present computer language to one more appropriate for artificial intelligence. But, even if the performance of these systems improves, making them really intelligent and independent, it would be difficult to accept the possibility that they might become a substitute for physicians. Due to the humanitarian aspect of medicine, the need for detailed data collection and the capacity for compassion for our fellow man, human beings will always have a secure a position in the care of patients.

References

1. Ribeiro JC. Aula da Saudade. Rev Med Univ Fed Ceará 1976; 15: 67-70.
2. Genaro S. Sistemas Especialistas: O Conhecimento Especializado. Rio de Janeiro - São Paulo: Livros Técnicos e Científicos Editora, 1986.
3. Goldstein J, Papert S.Cogn Sci, 1977; 1: 84. Apub Duda RO, Sotliffe EH. Expert systems research. Science 1983; 220: 261-20.
4. Baxt WB. Use of an artificial neural network for the diagnosis of miocardial infarction. Ann Intern Med 1991; 115: 843-8.
5. Ortiz J, Sabbatini RM, Gheffer CGM, Silva CES. Uso de redes neurais artificiais na avaliação da sobrevida na insuficiência cardíaca. Arq Bras Cardiol 1995; 64: 87-90.
6. Duda RO, Shotliffe EH. Expert system research. Science 1983; 220: 261-8.
7. Slozovitz P, Patil R, Schwartz WB. Artificial inteligence in medical diagnosis. Ann Intern Med 1988; 108: 80-7.
8. Pauker SG. Clinical decision making. In: Wyngarden JB, Smith Jr LH, Bennett JC (Ed). Cecil's Textbook of Internal Medicine. Philadelphia: WB Saunders, 1992: 68-73.
9. Braunwald E. Introduction to clinical medicine. In: Braunwald et al (Ed). Harrison's Principles and practice of Internal Medicine. New York: McGraw-Hill, 1987: 1-5.
10. Johns RJ, Hazzard WR. Clinical information and problem solving. In: Harvey AM, et al (Ed). The Principles and Practice of Medicine. Norwalk-Connecticutt: Appleton Century Crafts, 1984: 5-27.
11. Harvey AM, Bordley J. Diferential Diagnosis - Abridgement of the 2nd Ed. Philadelphia: WB Saunders, 1972: 1-18.
12. Miller RA, Pople Jr, HE, Myers JD. Internist I, an experimental computer based diagnostic consultant in general internal medicine. N Eng J Med 1982; 307: 468-75.
13. Miller RA, McNeil MA, Chalinor SM, Masarie FE, Myers JDP. The internist1/ quick medical reference project- Status report. West J Med 1986; 145: 816-22.
14. Rodrigues PMM. Lógica diagnóstica. Ceará Médico 1981; 3: 71-2.
15. Voytovich AE, Ryppey RM, Sufreddem AS. Premature conclusions in diagbnostic reasoning. J Med Educ 1985; 60: 302-7.
16. Sox Jr HC. Medical decision making. In: Barondness JA, Carpenter G, Harvey AM. Diferential Diagnosis. Philadelphia: Lea & Fediger, 1994: 9-22,
17. Diamond GA, Forrester JS. Analisys of probability as an aid in the clinical diagnosis of coronary artery disease. N Eng J Med 1979; 300: 1350-8.
18. Wasson JH, Sox HC, Neff RK. Clinical prediction rules: aplications and methodological standards. N Eng J Med 1980; 392: 1109-17.
19. Yauker SG, Kassirer JP. The threshold aproach to clinical decision making. N Eng J Med 1980; 302: 1109-17.
20. Fletcher RH, Fletcher SW, Wagner RH. Epidemiologia Clínica: Bases Científicas da Conduta Médica. Porto Alegre: Artes medicas, 1991
21. Lusted LB. Introduction to Medical Decision Making. Springfield-Illinois: CC Thomas, 1968.
22. Andrade PJN. Análise bayesiana: razões para utilizá-la. Rev Cir Cardiovasc 1990; 3: 36-8.
23. Ledley RS, Lusted LD. Reasoning foundations of medical diagnosis. Science 1959; 130: 9-21.
24. Warner HR, Toronto F, Veasy LG. Experience with Bayes's theorem for computer diagnosis of congenital heart disease. Ann NY AcadSci 1964; 115: 558-67.
25. Warner HR, Toronto AK, Vezsy LG, Stephenson RA. Mathematical aproach to medical diagnosis : aplication to congenital heart disease. JAMA 1964; 177: 177-83.
26. Rifkin RD, Hood Jr WB. Bayesian analisys of the eletrocardiogram exercise testing. N Eng J Med 1977; 297: 681-5.
27. Wegner DA, Ryan TJ, McCabe CH. Exercise stress testing: correlations between history of angina, ST response and prevalence of coronary artery disease in the coronary artery surgery study (CASS). N Eng J Med 1979; 30: 230-5.
28. Andrade PJN. Apresentação e avaliação de um programa de computador de auxílio ao diagnóstico médico. Tese de Livre-Docência-UFC, 1995.
29. Gibbons RJ, Balady GJ, Beasley JW, et al. Guidelines for exercise testing. A report of the American College of Cardiology/American Heart Association Task

Force on Practice Guidelines (Committee on Exercise Testing) J Am Coll Cardiol 1997; 30: 260-315.

30. Cheilin MD, Solokox M, McIlroy MB. Clinical Cardiology. Lange Medical Publishers 1994: 280.

31. Fagen TJ. Nomogram for Bayes's theorem. N Eng J Med 1975; 243: 57.

32. Brito AHX. Análise bayesiana do teste ergométrico. Arq Bras Cardiol 1991; 56: 97-103.

33. Sox HL. Probability theory in the use of diagnostic tests. Ann Intern Med 1986; 104: 60-6.

34. Rembold C, Watson DW. Pos-test probability by weights: a simple form of Bayes's theorem. Ann Intern Med 1988; 108: 115-20.

34. Jones TD. The diagnosis of rheumatic fever. JAMA 1944; 126: 481-2.

35. McKee PA, Castelli WB, McNamara PM. The natural history of congestive heart failure: the Framingham study. N Eng J Med 1971; 285: 1441-6.

37. Perloff JK, Child JS, Edward JE. New guidelines for the clinical diagnosis of mitral valve prolapse. Am J Cardiol 1986; 57: 1124-8.

38. Von Reyv CF, Levy DS, Arbeit RD, Friedlan G, Crompacks CS. Infective endocarditis: an analisys based on strict case definition. Ann Intern Med 1981: 94: 505-18.

39. Arnett FC, Edworthy SM, Chlock DA. The American Rheumathism Association 1987 revised criteria for the classification of rheumathoid arthritis. Arthrits Rheum 1987; 31: 315-20.

40. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982; 25: 1271-8.

41 Wayne EJ. Clinical and metabolic studies in thyroid disease. Brit Med J 1960; 1: 78-82.

42. Bankowitz RA, McNeil MA, Chalinor SM, Parker RC, Kappoor WN, Miller RA. A computer-assisted medical diagnostic consultation service. Ann Intern Med 1989; 110: 824-32.

43. Batnett GO, Cimino JJ, Hupp JA, Hoffer EP. Dxplain: an evolving diagnostic decision system. JAMA 1987; 87: 67-74.

44. Warner HR Jr. Iliad: moving medical decision-making into new frontiers. Method Inf Med 1989; 28: 370-2.

45. Andrade PJN, Menezes HVB, Teixeira CAC, Colares FAN, Lima JMC. Projeto Hipócrates - Sistema de auxílio ao diagnóstico diferencial em Cardiologia. Sessões de Temas Livres: VIII Congresso Norte-Nordeste de Cardiologia, 1988.

46. Andrade PJN, Menezes HVB, Teixeira CAC, et al. - Avaliação de um *software* de diagnóstico diferencial em Medicina Interna e Cardiologia. Arq Bras Cardiol 1993; 60: 285-8.

47. Andrade PJN, Menezes HVAB, Rocha ELA, Magalhães HA. Avaliação com ênfase em Cardiologia de um software de diagnóstico diferencial em Medicina Interna. Arq Bras Cardiol 1996; 67(supl 1): 110.

48. Berner ES, Webster GD, Shugerman AA, et al. Performance of four computer-based diagnostic systems. N Eng J Med 1994; 330: 1792-6.

49. Kassirer JP. A report on computer-assisted diagnosis - the grade. CN Eng J Med 1994; 330: 1884-5.

50. Barnett GO. The computer and clinical judgement. N Eng J Med 1982; 307: 493-4.

51. Sabatini RME. O diagnóstico médico por computador. Informédica 1993; 1: 5-10.

52. Schwartz P, Papil RM, Slozovits P. Artificial inteligence in medical diagnosis: where do we stand? N Eng J Med 1987; 316: 685-7.