DATABASE, REGISTRIES AND MONITORING OF CONGENITAL AND GENETIC MALFORMATIONS

متابعة قاعدة معلومات لتسجيل والأمراض الوراثية

Dr. M.Nour*, Prof.Dr. N.Hegazi*, Prof.Dr. S.Temtamy*, Dr. E.H.Aboul-Ezz*, Prof.Dr. N.Abdel-Megid*, and Prof.Dr. A.El-Hazmi**

*** Dept. of Computers & Human Genetics at National Research Center, Cairo,Egypt**
**** College of Medicine and King Khalid University Hospital, Riyadh, Saudi Arabia.**

ملخص:

إنطلاقا من التقدم التكنولوجي والذى أصبح ضرورة هامة للنمو الإقتصادى فقد لوحظ تقدم ملموس فى تطوير وإستخدام الإمكانيات التقنية فى الجانب الطبى. ولقد جهات الدول المتقدمة إلى إستخدام قواعد البيانات لتسجيل ومتابعة مشاكل المواليد . ولقد أصبح من الضرورى الآن تسجيل تلك البيانات على المستوى القومى والتى يكون لها تأثير كبير فى مجال التأهيل والتحكم فى منع إنتشار الأمراض . وفى هذا المجال تم تعاون علمى وبحثى بين مركز الأمير سلمان لأبحاث الإعاقة بالمملكة العربية السعودية والمركز القومى للبحوث ممثلا فى قسم الحاسبات وقسم الوراثة البشرية لتصميم قاعدة بيانات لمتابعة مشاكل المواليد ومعرفة حجم تلك المشاكل وتصنيفها من خلال خطة تهدف إلى دراسة حالة المعاقين وأسباب إعاقتهم . ومن هذا المنطلق تم تصميم قاعدة بيانات بتعاون بين متخصصى الوراثة البشرية ومتخصصى علوم الإجتماع وعلوم الحاسبات بهدف تحديد المشاكل الوراثية فى المنطقة العربية والتى تفيد فى منع إنتشار مشاكل أكثر لمعاقين آخرين .

وهذا البحث يستخدم قاعدة معرفة لطب الأطفال لإقتراح نظام يحتوى على ثلاثة أجزاء أساسية هى نموذج قاعدة المعرفة ونموذج معالجة المعرفة ونموذج الوسيط بين الحاسب والمستخدم .

والتركيز هنا على قاعدة المعرفة للأغراض الوراثية التى تسمح لنموذج معالجة المعرفة بالتشخيص وإعطاء النصائح من خلال التعرف على الصفات الغير عادية والتى تساعد المشخص أيضا على إكتشاف أمراض جديدة لها صفات معينة . وهذا النظام يمكن تصميمه لمشاكل طبية أخرى .

ABSTRACT

With technical and industrial advances which have become a significant part of today's economic growth, remarkable advances have been made in the development and use of technical capabilities that enhance the endeavor of the medical profession. The developed countries have adopted the essential "corner stone " of database and registries development for monitoring birth defects. The use of National and Regional Registries has become "a must " to have a database that is necessary for rehabilitation, control and prevention programs. In Saudi Arabia , towards this objective, a study to establish a database and national registry to monitor birth defects and determine frequency, type, distribution and magnitude of the problem has been initiated in coordination with the Department of Human Genetics and the Computer department at the National Research Centre aimimg at developing an Arab Regional Registry for Congenital and Genetic

Malformation. The Prince Salman Centre for Disability Research (PSCDR) as well as the outpatientt clinic for the human genetics department have developed a comprehensive research plan to study various aspects of handicaps and to assess the present care delivery status and consequently devise the most appropriate and applicable plan of care Specific forms have been designed for collection of relevant information. The completeness and accuracy of reporting is ascertained by close supervision and adequate training of the teams (composed of doctors, nurses and social workers) and a system by which uniformity of diagnosis can be ascertained. It is expected that the data derived from this study will establish a regional registries of the handicapped, incorporating congenital and genetic anomalies in Arabs and will provide the basis for further research in prevention and care programs combating the handicap problem. Towards this goal, this paper uses a pediatric knowledge base acquired from medical geneticists and proposes a system containing mainly three modules: a knowledge base module, a knowledge processing module, and an interface module. This endeavor concentrates mainly on the genetic symptoms knowledge base representation that allows the knowledge processing module to perform diagnosis, advise according to the abnormal features and assist in detecting new syndromes. The system can be extended and generalized for other medical problem domains

1. Introduction

With technical and industrial advances which have become a significant part of today's economic growth, remarkable advances have been made in the development and use of technical capabilities that enhance the endeavor of the medical profession. Among the advances that help in the medical profession are powerful tools that help in diagnosis such as expert systems, artificial intelligence, pattern recognition, data base systems and others. The field of genetics could easily benefit from the advances acquired by computers both in diagnosis, advice and help to the disabled patients themselves.

Our paper is concerned with all the three fields in an attempt to help mankind especially those who have been deprived by genetic problems.

1.1 Computers and Genetic Problem Determination

The genetic problems occur due to various factors; actually the number of these factors is very large in addition to some unknown factors. As a rule of thumb the first important item is to observe and know the factors. As every region has some factors akin to it more than others. It is very wise to study all these factors by collecting them in a registry the input to which is an elaborate questionnaire to all patients or canditates that have had a genetic problem or are liable to one. The number of main questions in this questionnaire depasses 500 questions that can be grouped in some subsets of selected origin. This large number of questions result in sets that can be grouped with some interrelations. Having a medical registry can help in deducting information that classifies some diseases and some syndromes according to some factors among which for example could be the environment, the region, the habits, etc.

This Arab Regional Registry for Genetic Problem has been designed and will run in both the National Research Center and Prince Salman Center for Disability Research center to enable the establishments of a nucleus of a middle east database for genetic problems and an expert system for users in this field.

1.2 Why Computers

- The number of items of interest which were considered is large from the personal history, there are up to 110 sets for each patient on the higher level. Each set represents the status of the set such as eye set, nose set, vision set, ear set, and others. Each one of these sets contains some attributes. It is very difficult to register all the different possibilities of the patient's features manually whatever these features are on the lower level or on the higher level. The manual registering consumes a large amount of time and space. Also, some inconsistency may exist and this may affect the accuracy of the diagnosis process. So, it is better to use the computer instead of doing these registers manually.
- The computer takes into consideration the different levels and different possibilities for any syndrome defects.
- The computer can easily calculate the frequency, type, distribution, and magnitude of the problem under consideration.
- The computer can easily detect new syndromes.
- The computer can retrieve the basic symptoms very fast. Also, it assists the geneticist to conclude the diagnosis of the genetic diseases.

1.3 Artificial Intelligence and Genetic Problems

First, what is Artificial Intelligence(AI) AI is the making of computers to do things that seem to be intelligent. The hope is that more intelligent computers can be more helpful to us, and better able to respond to our needs. Another thing it often means is nonnumeric ways of solving problems. But intelligence is a vague word. One thing it often means is advanced engineering, sophisticated software techniques for hard problems that cannot be solved in an easy way.

Artificial intelligence includes:-

- getting computers to communicate with us in a human language like English,either by printing or a computer terminal (natural language processing.)
- getting computers to remember complicated interrelated facts and draw conclusions from them (inference).
- getting computers to offer us advice based on complicated rules for various situations (expert system).
- getting computers to look through cameras or microscopes and see what is there(vision).
- getting computers to offer us movements around objects in the real world(robotics).

The above functions can be performed. Nowadays, because expert systems are feasible software that, we can work with in complicated field such as that of genetics.

1.4 The Arab Regional Registry of Genetic Problems

The Arab Regional Registry for Genetic Problem can be established in three forms: Filing systems, Data base systems, or Knowledge base systems (Artificial Intelligence systems). Data base and Knowledge base systems (or AI systems) technologies[1] represent the extremes from the point of view of the solutions to information intensive problems .

Data base applications are usually well understood and can be realized through algorithmic methods. They require the maintenance of large collection of facts which may change overtime, but have a somewhat regular structure.
Knowledge base systems, on the other hand are usually applied to problems which are not fully understood and require the use of heuristic inference techniques[2,3]. The individual pieces of information are often much fewer than in database domains. The important considerations for AI formalisms are representational richness and ease of manipulation by the inference mechanism. The knowledge base of the genetic diseases including malformation combines features from data base and AI technologies. It needs to maintain large databases and inference capabilities. i.e. the genetic system can be considered as Knowledge Information Processing System (KIPS). The genetic system can perform some inference operations such as deduction to conclude the patient's diagnosis. Also, it performs the inference operations not only during the retrieval, but during the updating process. This paper describes an approach to designing a knowledge representation to perform diagnosis, monitoring birth defects, and detecting new syndromes in the genetic domain. The genetic knowledge representation system is developed by using the Logic Programming (PROLOG). The knowledge of the genetic system is represented in a production system form into which fast inference capabilities are built. That knowledge base contains a set of facts and rules. The facts were acquired from clinical geneticist's pedigree the observations and patients and symptoms. Some of the patients, genetic history items that were observed are: name, age, sex, origin, pedigree, pregnancy history, delivery history, neonatel history and body features such as skull shape , scalp hair, forehead, face, eye, and others. All the previous items of the patients will be stored in a data base file as a declarative knowledge. The system is supported by a set of rules which can be acquired from the geneticist as a procedural knowledge. The rules will be implemented and operated on the declarative knowledge to perform the diagnosis process. In the following we will describe the general diagnostic methodology and the knowledge representation for genetic problems. Also the implementation of the genetic rule base, the system performance and concluding remarks will be also discussed.

2- The Genetic Diagnosis and Advice Methodology

The main objective is how to help the community in the field of genetic problems. This can be done by giving advice to the medical workers about their patients for cases related to genetic problems to diagnose, advice, or prevent future problems. In this trend, two main items are necessary :the identification of the abnormal features of the patients and the necessary advices and recommendations for patients on the diagnosis results (genetic counseling).

2.1 Identification of the Abnormal Features of the Patients

The genetic diseases contain three main areas. These areas are:Single gene disorders [Malformation Syndromes present in Mendelion Inheritance in Man (MIM)& Inborn Errors of Metabolism (IBM), Chromosomal abnormalities, and Multifactorial disorders.This paper is interested only in the syndromes MI. This phase is interested in studying the main morphological features, and the patterns of inheritance such as: recessive, and dominant inheritance.

The diagnosis process takes more than one step. The first step is how to acquire the symptoms from the patients. This can be performed by making the following issues:

- Suspect discrimination and test generation.
- Critical observations of the patients and accurate recording of features.
- A detailed questionnares for the patient's status such as name, ancestral origin, parentral consanguinity.

The diagnosis starts when some of the findings or features do not match with the normal. Then, the medical geneticist (as a part of confirmation) will advice the patient to do another test(s) such as blood analysis, X-ray, and so on. The abnormal features and observations obtained from these sources will be taken into account besides the signs and symptoms observed before. The amalgamation of the previous tests will ease the identification of the reasons that cause these abnormal features. This identification can be done also by using some medical guidance such as McKusick's MIM. This reference is continuously and is available on line (OMIM). If the cause of the abnormal features are identified, then the system turns to the advice or the guidance phase. Otherwise, it looks for other knowledge to further discriminate between the suspects. The genetic knowledge base in this trend contains two items: Facts, and Rules. The facts represent the genetic findings of the patient which were observed before. The rules represent the necessary conditions for these abnormal features to occur. These rules can be acquired from the expert geneticist or from the medical references/journals which can be considered as one of the most important guidance in syndromes, advice, and diagnosis. If the abnormal features of the patient do not match with any of the syndrome's features (a rare case), then these features will be registered as a new group or a new syndrome.

2.2 Advice or Problem Prevention Step Genetic Counseling

Once the cause of the abnormal features are identified, the advice step begins. This step concentrates on giving the patients advices for treatment purposes or for preventing further genetic problems in their families. The advice by preventing (or not) the marriage of the patients with identical disorders or prenatal diagnosis in subsequent pregnancies are examples. The 'Production System' genetic knowledge representation technique will be presented as in the following section.

3. Knowledge Representation for Genetic Problems

The Knowlegde Representation (KR) can be considered as the case of any AI applications[4]. The KR systems combine two main elements: Data structures and Interpretive procedures for using the knowledge embedded in the data structures. The two elements of the KR system are essential. The use of knowledge is one of the most important knowledge issues. For this purpose, it is better to acquire the knowledge of the genetic diseases from their correct sources. Knowledge acquisition occurs at two levels, the first level is concerned with structuring facts in a data base. This can be obtained from the patient's signs and symptoms and other discriminate tests and analysis. The second level is concerned with relating information to previously stored information. The knowledge base can be used for reasoning at different levels. One level is formal reasoning according to the rules of propositional calculus. Another kind of reasoning comes very naturally to humans, but have been extremely difficult for machines is reasoning by

analogy and generalization and abstraction. Another kind of reasoning is procedural reasoning as represented by production rules which will be considered in this work. Production systems are very important due to the following;

- Production systems represent not only the genetic knowledge, but also their corresponding actions.
- Production systems provide a language in which the representation of the knowledge base systems is natural.
- The study of human behaviour protocols originally led to production concepts formulation.

The production system technique which was selected to represent the genetic knowledge base contains three main components[5,6]. These components are: Global data base, Knowledge base, and Control structures as shown in figure 1.

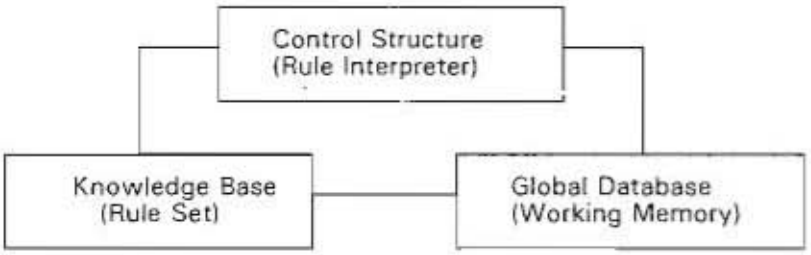

Figure 1: Production System Components

The global database is the main data structure of production systems. The data base is the basic structure on which the production rules operate. It is a dynamic structure, continually changing as a result of the operation of the production rules. The global data base of the genetic problems can be referred as the context of the patient's symptoms and features. The global database contains the general observations of the patient which can be represented in the form of relational database files (or sets). Each set has a name, and some attributes. Each set is concerned with some related symptoms of the patient. e.g. the eye set includes all symptoms regarding the eye's symptoms. Also the nose set includes all symptoms regarding the nose's symptoms. Also, the other genetic features sets include their corresponding symptoms. Each set contains a link with its related sets as shown in figure 2.

The production system also contains a set of production rules. Production rules have a condition part (IF) and an action part (THEN). The condition part contains the necessary symptoms and conditions to occur for the corresponding actions or conclusions. This conclusion may be in the form of syndrome diagnosis or in a form of advice given to the patient. The genetic knowledge base system contains a huge amount of rules related to the different types of syndromes and diseases, the production rules become applicable and subject to being fired by the control system.

The control structure represents an interpreter logic program to control the order in which genetic production rules are fired and resolve conflicts if more than one rule is applicable. The control system repeatedly applies rules (from the geneticist experience) to the data base (patient's features) until a description of the diagnosis is produced. The rule

interpreter can be considered as an inference engine to conclude or detect the diagnosis of the different diseases. It aslo reasons how and why the diagnosis is obtained. It also detects the occurance of a goal state and records the rules which have been applied to reach it for later reference[7] The control structure or the rule interpreter is implemented in PROLOG programming. This logic program is operated on the genetics knowledge base (patient's features and geneticist medical experience) to obtain the conclusion (advice or syndrome diagnosis).

S1=Personal Information

Name	Age	Address	Phone	Birth-Order	Origin	Occuption

S2=History of present illness

Name	Onest of Signs	Congenital

S3=Pedigree

Name	No.of Sibs	Status of each

S4=Pregnacy History

Name	Drugs	Infection	X-Ray	Bleeding Time	Prem. Delivery

S5=Delivery History

Name	Trauma	Anoxia	Cephalic Breech	Normal/Delayed

S6= Milestones of development

Name	Motor Devel.	Mental Devel.	Speech	Dentition	Deffness

S7=Cranium

Name	Normal	Hydrocephaly	Macrocephaly	Microcephaly

S8=Scalp Hair

Name	Normal	Sparse	Alopecia	Fair

S9=Forehead

Name	Wide	Narrow Bulging

S10=Eye

Name	Depth Set	Hypertelirism	Hypotelirism	Proninent Eye

Figure 2: Genetic sets entities

The genetic problems take into account all the different organs of the patients. These organs exceed approximately 110 entities (the genetics history sheet designed by National Research Center human geneticist Temtamy) such as facial features eye, forehead, history of present illness, pregnancy history, delivery history, and others. Each one of these entities can be considered as a set. Each set contains a set of features and arguments as shown in figure 2. The abnormal features on the higher level (sets level) may identify syndromes or defects. Also, the abnormalities on the lower level may specifically identify genetic syndromes. If the different permutation of all sets including all attributes or arguments were taken into account, it is noticed that a huge number of syndromes exist. In case of representing these sets and their corresponding attributes and relations, the computer can help the geneticist to diagnose, help the clinical geneticist and the patient, or detect new syndromes, i.e the different types of possibilities from the patient's features can be obtained on the higher and lower level as shown in figure 3.

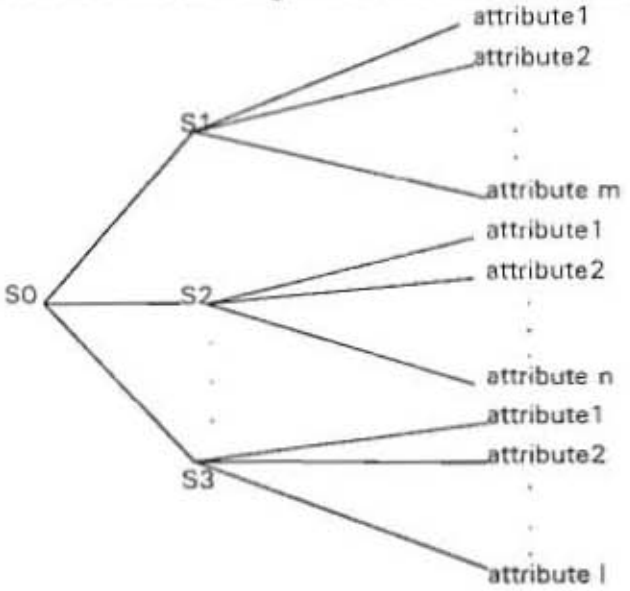

Figure 3: A hierarchy of the main sets and their corresponding arguments

The releationship and linking between the different types of sets and their corresponding arguments may produce a huge amount of syndromes as shown in figure 4.

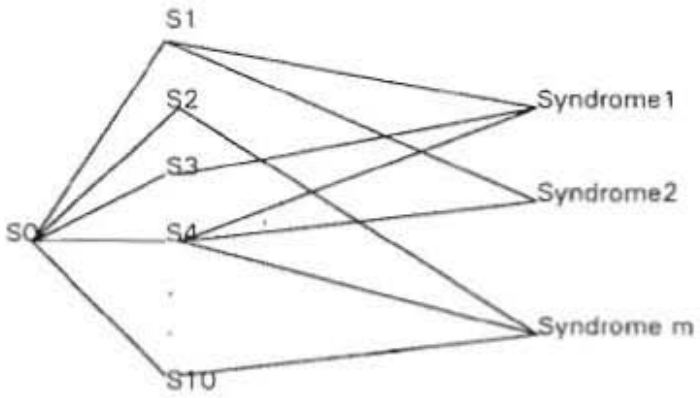

Figure 4: Different syndrome possibilities exist on the higher level (set level) for 10 organs

The types of syndromes permutation exist in the lower level will be very large. In the current state up to 6000 cases related to different patients are edited from the deptartment of human genetics (NRC) to the computer.
The inference engine[8] uses the experience of the geneticist to classify and conclude the different syndromes related to their corresponding features. In the following section the production rules implementation will be discussed

4. Production Rules Implementation

There are different types of rules regarding the genetics problems. These rules can be transferred to the computer from the geneticist, medical references such as MIM, documents, and medical journals[9,10,11] Each rule and their corresponding arguments are assigned certainty factors. By joining the different symptoms and the abnormal features of the patient, the production systems can identify the syndrome type. The following rules can be considered as a sample of the genetics rule base.

RULE 1

IF drug=value1 and infection time=value2 and x_ray duration time=value3 and bleeding time=value4 and delivery history is delayed

THEN syndrome 1.

RULE 2

IF the delivery status is trauma and the mental development is retarded and development is delayed and the C-T of the brain is abnormal

THEN syndrome 2.

RULE 3

IF the forehead is high and the nose is prominent and

the face is round and the eye is hyperteloirism and delayed motor
developmet with blephraphimosis
THEN syndrome 3

5. System Performance

In fact the inferece engine of the genetics knowledge base and the system performance in general can be characterized by some features as follows,

- The genetics knowledge base system has the ability to perform inference operations. The detection or the recognition of the diagnosis can be considered as one of these inference operations.
- The system can perform some statistical operations such as frequency distribution, grouping, and magnitude of the problem has been initiated. These figures are very significant from the geneticist's point of view.
- The system is a forward chaining, i.e it starts with the abnormal features of the patient and then goes forward to conclude the diagnosis.
- The system has the ability to overcome any ambiguity by acquiring extra symptoms and information of the patients to obtain a correct diagnosis.
- The genetics knowledge base is modular, i.e it is designed in a structured form. This makes the system more flexible to add or update or extend any module at any time.
- The system performance is reasonable and it can be generalized for other medical problems.
- The system can be modified to be used as learner for novice geneticists.

6.Conclusions

A Knowledge based system pertaining to genetics diagnosis was presented. The system uses a huge amount of knowledge acquired from the geneticist. The production system was selected to represent the genetic knowledge base. The diagnosis and advice can be done by using a forward chaining mechanism. The knowledge base accepts the features of the patient then the inference engine (or the control structure interpreter) concludes the corresponding diagnosis. The genetic knowledge base system helps the geneticist to present advices to the patient as the system identifies the main features and characteristics of the Arab environment.

7. References

1- Fumihiro Maruyama, Tamio Mano, Kazuhi Hayashi, Taeko Kakuda, Nobuaki Kawato, and Takao Uehara, 'Prolog-based expert system for logic design', proceedings of the international conference on FGCS, Tokyo, Japan, 1984.

2- Fumio Mizoguchi, Hayato Ohwada, and Yoshinori Katayama 'LOOKS: Knowledge representation system for designing expert system in a logic programming framework', proceedings of the international conference on FGCS, Tokyo, Japan, 1984.

3- Grian Piero Zarri, 'Intelligent information retrieval: An interesting application area for the new generation comp uters systems', proceedings of the international conference on FGCS, Tokyo, Japan, 1984.

4- Hiroshi Motoda, Naoyuki Yamada, and Kenichi Yoshida, 'A Knowledge based system for plant diagnosis', proceedings of the international conference on FGCS, Tokyo, Japan, 1984.
5- Morris W. Firebaugh, 'Artificial Intelligence: A Knowledge-based approach', pws-kent publishing company, 1989.
6- Samuel O.Aletan, 'Artificial intelligence languages and architectures: past, present, and future', Artificial intelligence methods and applications: advanced series on artificial intelligence-vol.1, edited by Nikolaos G. Bourbakis, 1992.
7- Jun Gu, 'On optimizing a search problem', dept. of electrical engineering, university of calgary, Canda, Artificial intelligence methods and applications: advanced series on artificial intelligence vol.1, edited by Nikolaos G. Bourbakis, 1992.
8- A.Y.Al-hawaj, M. Hamed, 'Design of an expert system for teaching programming fundamentals', proceedings of the 10th national computer conference, computer center, King-Abdulaziz university, Jeddah, 1988.
9- M.G.Khayat, H.A.Al-Muhtaseb, 'Knowledge representation in natural language system' , proceedings of the 10th national computer conference, computer center, King Abdualaziz university, Jeddah, 1988.
10- Ahmed Sharaf Eldin, and Sameh M.H.Khalil, 'A Proposed expert system for alarm processing and fault identification in power plants' proceedings of the 10th national computer conference, computer center, King Abdulaziz university, Jeddah, 1988.
11-McKusick, V.A Mendalian Inheritance in M-Catalog of Autosomal Dominant Autosomal dominant Autosomal Recessive and X-linked Disorders. John Hopkins University press 1992.