Original Article

Automated Screening for Three Inborn Metabolic Disorders: A Pilot Study

Authors
Kavitha S, MSc
Bioinformatics Division, Department of Botany, Bharathiar University, Coimbatore 641046
Sarbadhikari SN, MBBS, PhD
TIFAC-CORE in Biomedical Technology, Amrita Vishwa Vidyapeetham, Amritapuri 690525
Ananth N Rao, PhD
Metabolic Disorders Laboratory, Amrita Institute of Medical Sciences and Research Centre, Elamakkara, Kochi - 682026

Address For Correspondence
Dr. S. N. Sarbadhikari, MBBS, PhD
TIFAC-CORE in Biomedical Technology,
Amrita Vishwa Vidyapeetham,
Amritapuri 690525, India.
E-mail: supten@gmail.com

OJHAS Vol 5 Issue 3(1): Automated Screening for Three Inborn Metabolic Disorders: A Pilot Study www.ojhas.org/issue19/2006-3-1.htm

1

Abstract:

Background: Inborn metabolic disorders (IMDs) form a large group of rare, but often serious, metabolic disorders. **Aims**: Our objective was to construct a decision tree, based on classification algorithm for the data on three metabolic disorders, enabling us to take decisions on the screening and clinical diagnosis of a patient. **Settings and Design**: A non-incremental concept learning classification algorithm was applied to a set of patient data and the procedure followed to obtain a decision on a patient's disorder. **Materials and Methods**: Initially a training set containing 13 cases was investigated for three inborn errors of metabolism. **Results**: A total of thirty test cases were investigated for the three inborn errors of metabolism. The program identified 10 cases with galactosemia, another 10 cases with fructosemia and the remaining 10 with propionic acidemia. The program successfully identified all the 30 cases. **Conclusions**: This kind of decision support systems can help the healthcare delivery personnel immensely for early screening of IMDs. **Key Words:** Decision support techniques, Metabolic diseases, Computer-assisted diagnosis, Expert system.

Introduction:

Inborn Metabolic Disorders (IMDs) (1) are inherited disorders that are caused by a defect in a single gene. Some IMDs produce relatively unimportant physical features or skeletal abnormalities. Others produce serious disease and even death. Most inborn errors of metabolism are monitored by routine blood or urine tests.

Galactosemia (1) is the inability of the body to use the simple sugar galactose (causing the accumulation of galactose 1-phosphate), which then reaches high levels in the body, causing damage to the liver, central nervous system and various other body systems.

Hereditary fructose intolerance (1) is a metabolic disease caused by the absence of an enzyme 1-phosphofructaldolase. Propionic acidemia (1) is an inherited disorder in which the body is unable to process certain proteins and lipids properly. The gene defect for Propionic Acidemia is an autosomal recessive genetic trait and is unknowingly passed down from generation to generation.

Many of the inborn errors of metabolism can be treated (1) effectively. Treatment depends on the enzyme defect itself and how readily the compounds involved can be eliminated or replaced.

Decision trees (2) are excellent tools for making decisions where a lot of complex information needs to be taken into account. They also help us to form an accurate, balanced picture of the risks and rewards that can result from a particular choice.

A decision tree is a representation of a decision procedure for determining the class of a given instance. In many problems chance (or probability) plays an important role. Decision analysis is the general name that is given to techniques for analyzing problems containing risk/uncertainty/probabilities. Decision trees are one specific decision analysis technique. A decision tree takes as input an object or situation described by a set of properties, and outputs a yes/no decision.(2)

Research on information retrieval and database management systems has been advancing very quickly over the past few decades, especially in healthcare applications. One such information system was the Internet platform oriented program developed by Hofestadt et al (3) that provided complex and specific knowledge about inborn errors of metabolism. Another

OJHAS Vol 5 Issue 3(1): Automated Screening for Three Inborn Metabolic Disorders: A Pilot Study www.ojhas.org/issue19/2006-3-1.htm

2

program developed by Kauert et al (4) provided diagnosis and therapy for IEMs. This information system could be helpful for scientists and physicians working in the field of inborn metabolic diseases. PROLOG, another computer aided expert system developed by Pince et al (5) was used to obtain uniform and fast phenotyping and reporting of dyslipoproteinemia.

Metanet (6) is a neural network based computer program for diagnosis of inherited metabolic diseases in children, using plasma or urinary amino acid results as input wherein the expert system asks a question to the user regarding the presence or absence of common clinical and/or biochemical abnormalities. Using both the amino acid data and the answers to the questions, the MetaNet program provides a provisional diagnosis. The diagnostic output indicates the degree of confidence of the program in the diagnosis.

The objective of the present study was to construct a decision tree from the clinical data on certain inborn errors of metabolism so that the predictions with the tree can be used for early diagnosis and interventions.

To the best of our knowledge, this is the first report from India for screening IMDs by automated classifiers.

Materials and Methods:

Subjects

The inborn errors of metabolism handled in this study were Galactosemia, Fructosemia and Propionic Acidemia. The real data of patients was obtained from the hospital to which the authors are affiliated.

The cases had been referred from the Pediatrics and Neurology departments of the hospital, and were confirmed by chemical analysis from serum/urine.

Moreover, the incidences of such disorders are not high and the clinical features are rarely documented properly.

Decision Tree For The Problem

A decision tree is constructed by looking for regularities in data. A decision tree is an arrangement of tests that prescribes an appropriate test at every step in an analysis. More specifically, decision trees classify instances by sorting them down the tree from the root node to some leaf node, which provides the classification of the instance. Each node in the tree specifies a test of some attribute of the instance, and each branch descending from that node corresponds to one of the possible values for this attribute. An instance is classified by starting at the root node of the decision tree, testing the attribute specified by this node, then moving down the tree branch corresponding to the value of the attribute. This process is then repeated at the node on this branch and so on until a leaf node is reached.

The basic classification algorithm (7) was a non-incremental concept learning method that produced a hypothesis in the form of a decision tree. The algorithm accepted a training set of positive and negative examples of a concept. The decision tree and rules produced from applying the algorithm attempted to give a generalization that could be used to classify objects not in the original training set. The classification algorithm could be summarized as follows.

- **The inputs:** Features in a training set (T)
- **The outputs:** A decision tree
- **Step 1:** If all elements in T are positive then create a 'yes' node and halt.
- **Step 2:** If all elements in T are negative then create a 'no' node and halt.

OJHAS Vol 5 Issue 3(1): Automated Screening for Three Inborn Metabolic Disorders: A Pilot Study www.ojhas.org/issue19/2006-3-1.htm

3

- **Step 3:** Otherwise select an attribute F with values V_1, V_2, V_3, ... , V_n. Partition T into subsets T_1, T_2, T_3, ..., T_n according to their values on F. Create a branch with F as parent and T_1 *etc*. as child nodes.
- **Step 4:** Apply the procedure recursively to each child node.

The input to this decision tree was a set of instances containing information about the patients and their medical conditions. The output enabled us to decide the type of inborn errors of metabolism in the patient. The input data consisted of seven attributes namely seizures, mental Retardation, gastro intestinal infection, respiratory tract infection, liver dysfunction, hypoglycemia, and hepatomegaly with four values each (-, +, ++, +++). The values -, +, ++, +++ indicate absent, mild/low, moderate and severe respectively. Decision trees often used in medical diagnosis structurally consist of two types of nodes: non terminal (intermediate) and terminal (leaf). The former corresponds to questions asked about the characteristic features of the diagnosed case. These may be factorial, *e.g*., "Does the patient have symptom X?", or ordinal, *e.g.*, "Is symptom X absent, mild, moderate or severe?", or continuous, *e.g.*, "What is the patient's median motor latency at the wrist?". (The above questions, though not actually related to IEMs, are just examples to indicate the mode of questioning in any diagnostic process.) Questions about the characteristic features of the cases used in this study were ordinal *i.e.,* questioning whether the symptom is absent, mild, moderate or severe. The best value was selected for the intermediate node. The terminal node, on the other hand, generated decisions/diagnosis.

A stepwise decision-making process, where a single question was asked each time and, depending on the answer, a different branch of the tree containing another set of questions was asked, achieved diagnosis. The root of the tree contained the first diagnostic question asked by the classifier. Depending on the answer for each particular case, a different branch of the tree is traversed, arriving at a decision node. This method of building decision trees was called top-down induction of decision trees and was based on recursive partitioning of data. In brief, at each recursive step, one question was asked which selects a best value. This question partitioned the design set, and the process was repeated for each subsets, until they reached a final decision. This was the method of applying the classification algorithm for the patient's real data to build the decision tree classifier in order to understand the decision process.

Results:
Initially a training set containing 13 cases was investigated for the three inborn errors of metabolism. Using a small set of training data for classification is an essential step in any decision tree construction process. This is because the initial classification of data using a training set can give an idea about the categorization of the future instances. Thirty cases of three types of inborn errors of metabolism were studied with the programme and it could correctly identify the 10 cases with galactosemia, another 10 cases with fructosemia and the remaining 10 with propionic acidemia. There was no wrong identification of the normal cases as disordered cases.

The results are summarized in Table 1:

OJHAS Vol 5 Issue 3(1): Automated Screening for Three Inborn Metabolic Disorders: A Pilot Study www.ojhas.org/issue19/2006-3-1.htm

4

Table 1: Performance of the Program for Screening Inborn Metabolic Disorders

	Inborn Metabolic Disorder	No. Of Cases	Recog-nition %	False +ve	False -ve
1.	Galactosemia	10	100%	-	-
2.	Fructosemia	10	100%	-	-
3.	Propionic Acidemia	10	100%	-	-

Discussion:

The decision tree programme identified the cases of inborn errors of metabolism correctly. The decision tree pointed towards that particular provisional diagnosis utilising a certain combination of all the seven (nonspecific) attributes mentioned in the methodology. A clinician will not be able to "analytically decide" all the permutation and combinations possible with 7 clinical features (each having 4 grades of severity), whereas for a decision support system, this is a rather simple, quick and easy job. Strengths of the study therefore lies in the quick and correct identification of 3-5 close differential diagnoses and laboratory investigations can be performed only in such direction. Clinical symptoms in addition to output from the classifier may be a quick way of probably initiating general management to be followed by disorder specific management in all suspected cases of inborn errors of metabolism. This shall avoid unnecessary expenditure on performing the tests for the whole spectrum of disorders. As an output, the classifier shall also help stratify which patient should receive expensive laboratory testing.

However one of the limitations of this study was the small sample size; 30 cases was too small a number to test for the inborn errors of metabolism. But when more data on inborn errors of metabolism is available, the decision tree can be expanded and the validation of the program can be extended to many such cases. This was only a pilot study and further study is in progress. Presently more samples are being collected and other types of classifiers viz., Logistic Regression Analysis, Artificial neural networks and Support Vector Machines (SVM) are being tried out.

The program in the present study differs from the other computer programs for the diagnosis of inherited metabolic diseases in various ways. The program developed for this study was a decision tree based computer program which took as input the ordinal valued data of patient's medical conditions (whether the symptom is absent/low/moderate/severe) and the output provided the provisional diagnosis. The present study program differed from Metanet, a neural network program, in that it took as input the amino acid results and the common clinical abnormalities and the output provided the provisional diagnosis accompanied by a numerical belief vector, which indicated the degree of confidence of the program in the diagnosis. Information on diseases and their conditions, treatment of such diseases and integration of that information with other databases and the inclusion of statistical knowledge of such diseases were the typical features of the other above described programs. Though the program in the present study did not include such additional information, it can be used as a tool for the diagnosis of IEMs.

Other investigators (8) had described sequential testing for clinical trials. The sequential nature of the data was not from additional patients, but rather from longer follow-up which we shall try to do.

Artificial neural networks had been used (9) for the parallel evaluation of all fine-motoric data, leading to a reclassification of patients suffering from Wilson's disease, based on actual fine-motoric abilities but not reflecting the clinical classification at the time of

OJHAS Vol 5 Issue 3(1): Automated Screening for Three Inborn Metabolic Disorders: A Pilot Study www.ojhas.org/issue19/2006-3-1.htm

5

manifestation. Presently we are not in a position to try such things at our Institute.

The pioneers of machine learning applications to IMDs (10) endorse that the mined data confirm the known and indicate some novel metabolic patterns that may contribute to a better understanding of newborn metabolism. We have been presently using some of their techniques but applying to our own growing databases – which are very distinct from the Western population features.

Conclusion:

Decision support systems that help physicians are becoming very important part of medical decision making, especially in medical diagnostic processes and particularly in those where decision must be made effectively and reliably.

The aim of this work had been to study and apply a suitable technique for mining medical data for prediction of or screening for the inborn metabolic disorders namely galactosemia, fructosemia and propionic acidemia. This approach was more suitable for the given set of minimal data on IMDs. The ability to handle situations robustly and the ability to classify samples made this approach attractive for healthcare applications. This kind of decision support systems could help the healthcare delivery personnel immensely for early screening of IMDs.

In continuation of the pilot study presented here, we are now trying various other data mining methods on an extended database of clinical features of IMDs.

Acknowledgements:

References:

1. Fernandez J, Saudubray J-M, Van Bergeh G. Eds, Inborn Metabolic Diseases: Diagnosis and Treatment. 4th ed, New York, Springer-Verlag, 2006

2. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Chapman & Hall / CRC Press, 1984

3. Hofestadt R, Mischke U, Scholz U. Knowledge Acquisition, Management and Representation for the Diagnostic Support in Human Inborn Errors of Metabolism, Stud Health Technol Inform. 2000; 77:857-62

4. Kauert R, Topel T, Scholz U, Hofestadt R. Information System for the Support of Research, Diagnosis and Therapy of Inborn Metabolic Diseases, Medinfo. 2001; 10(Pt 1): 353-6

5. Pince H, Cobbaert C, van de Woestijne M, Lissens W, Willems JL. Computer Aided Phenotyping of Dyslipoproteinemia, Int J Biomed Comput. 1988 Dec; 23(3-4):251-63

6. Wyett CE. "MetaNet" http://medexpert.imc.akh-wien.ac.at/metanet_info.html (Accessed July 2006).

7. Smith S, The Classification Algorithm http://www.cs.mdx.ac.uk/staffpages/serengul/The.Classification.algorithm.htm (Accessed July 2006).

8. Troendle JF, Liu A, Wu C, Yu KF. Sequential testing for efficacy in clinical trials with non-transient effects. Stat Med. 2005 Nov 15; 24(21): 3239-50

9. Hermann W, Wagner A, Kuhn HJ, Grahmann F, Villmann T. Classification of fine-motoric disturbances in Wilson's disease using artificial neural networks. Acta Neurol Scand. 2005 Jun;111(6):400-6

10. Baumgartner C, Bohm C, Baumgartner D, Marini G, Weinberger K, Olgemoller B, Liebl B, Roscher AA. Supervised machine learning techniques for the classification of metabolic disorders in newborns. Bioinformatics 2004 Nov 22; 20(17):2985-96. Epub 2004 Jun 4

OJHAS Vol 5 Issue 3(1): Automated Screening for Three Inborn Metabolic Disorders: A Pilot Study www.ojhas.org/issue19/2006-3-1.htm

6