

Using Clinical Expert Decision Support System BioAnalyst to Diagnose Colon Diseases

Martina Cesalova PhD, MSCS¹, Ing. Milan Cesal PhD^{2*}, Ing. Karol Bitto³, Jaroslav Jansa, Ph.D⁴

¹Vysoka skola manazmentu, College of Management, Department of Mathematics and ICT, Bezrucova 64 91101 Trencin, Slovakia

²Biocons, Smidkeho 3 91108 Trencin, Slovakia

³KB-Soft S.R.O., Palackeho 85/5, 91101 Trencin, Slovakia

⁴Immobilisier Central Europe, Ltd., Evropská 52 Prague 6, 160 00, Czech Republic

DOI: [10.36348/SJMPS.2019.v05i12.003](https://doi.org/10.36348/SJMPS.2019.v05i12.003)

| Received: 20.11.2019 | Accepted: 27.11.2019 | Published: 11.12.2019

*Corresponding author: Milan Cesal

Abstract

The main objective of this contribution is to present a new alternative to effectively diagnose colon diseases in a timely, accurate and cost-effective manner, thanks to using technology. To be more specific, it introduces usage of an expert system that is based on using cluster, discriminant and multidimensional analyses and four biochemical parameters to effectively diagnose m. Crohn, colitis ulcerosa and large intestine cancer.

Keywords: Expert Systems, Diagnostics, BioAnalyst, Multivariate Analysis, Colon Diseases

Copyright @ 2019: This is an open-access article distributed under the terms of the Creative Commons Attribution license which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use (NonCommercial, or CC-BY-NC) provided the original author and source are credited.

INTRODUCTION

BioAnalyst is an adaptive expert system, which is based on using discriminant and cluster analyses, supported by some, specifically designed, discriminant criteria that are based on multi-dimensional analysis of biochemical and other para-clinical data. Diagnostic in BioAnalyst is based on using a priori knowledge of chosen biochemical parameters that were extracted from related groups of probands (so-called learning files). Then, thanks to using the learning files, the expert system creates a numerical model to describe so-called clusters, which represent individual groups of diseases that had been specified in advance. Finally, BioAnalyst analyzes specific biochemical parameters of the given patient and suggests his/her most probable diagnosis.

Although we have witnessed an exponential development in the field of information, expert or knowledge management systems and their application to the field of diagnostics, most of them are still based on using the nearest neighbor approach and therefore require an enormous amount of learning data [1]. Because this approach is based on using large amounts of data about specifically clinically diagnosed probands, it is significantly time consuming and also costly. Unlike the systems that are based on the nearest neighbor algorithm, BioAnalyst, thanks to using MONTE CARLO method, is able to validate success rates of diagnostics, while using data of just a few

probands. Because of this, BioAnalyst may be considered an effective tool to test potentially successful combinations of biochemical parameters, suitable for a cost-effective and fast diagnostics of diseases.

BioAnalyst encompasses the following objects:

- Client Manager – allows end users to create, preview and edit client's electronic health records.
- Diagnostics [of] Diseases – runs diagnostics based on choosing a desired diagnostic group and, at the same time, based on real patient's biochemical parameters.
- Diagnosis Manager – enables end users to create, edit, and view statistics of data that represents individual diagnoses.
- Diagnosis Group Manager – enables creation and management of diagnostic groups, supports collection and merging of diagnoses with identical parameters into common groups – clusters, and allows for visualization of diagnostics. The exploratory summarizing statistics feature also provides a quick reference to many frequently used statistics, such as the mean, standard deviation, skewness, and kurtosis. In addition, the summary statistics provides users with the correlation matrix and the variance/covariance matrix.
- Monte Carlo Modeling – performs data classification based on a set of artificially generated data and consequently statistically

processes the results of classification. This feature also provides end users with information about validity of values/parameters based on using a small amount of real data. Besides, Monte Carlo module visualizes results and presents the output in form of histograms.

In the field of clinical diagnostics, BioAnalyst may be used to:

- Quickly and effectively monitor population's health status, in order to focus on early detection of cardio-vascular and oncological diseases and recommend more specific examinations
- Monitor patients' health condition during the medical treatment or post-therapeutic period
- Enable or support fast diagnostics of several diseases, e.g. hepatopathy, diabetes mellitus, malignant diseases, atherosclerosis, Gilbert's disease, etc.

Among the most important features of the system undoubtedly belong the following:

- The learning process
- Processes related to classification of unknown objects, i.e. detection and diagnostics of diseases.

Outcomes of the above-mentioned fundamental processes/modules of the system are connected by an objective function that is based on the following rules:

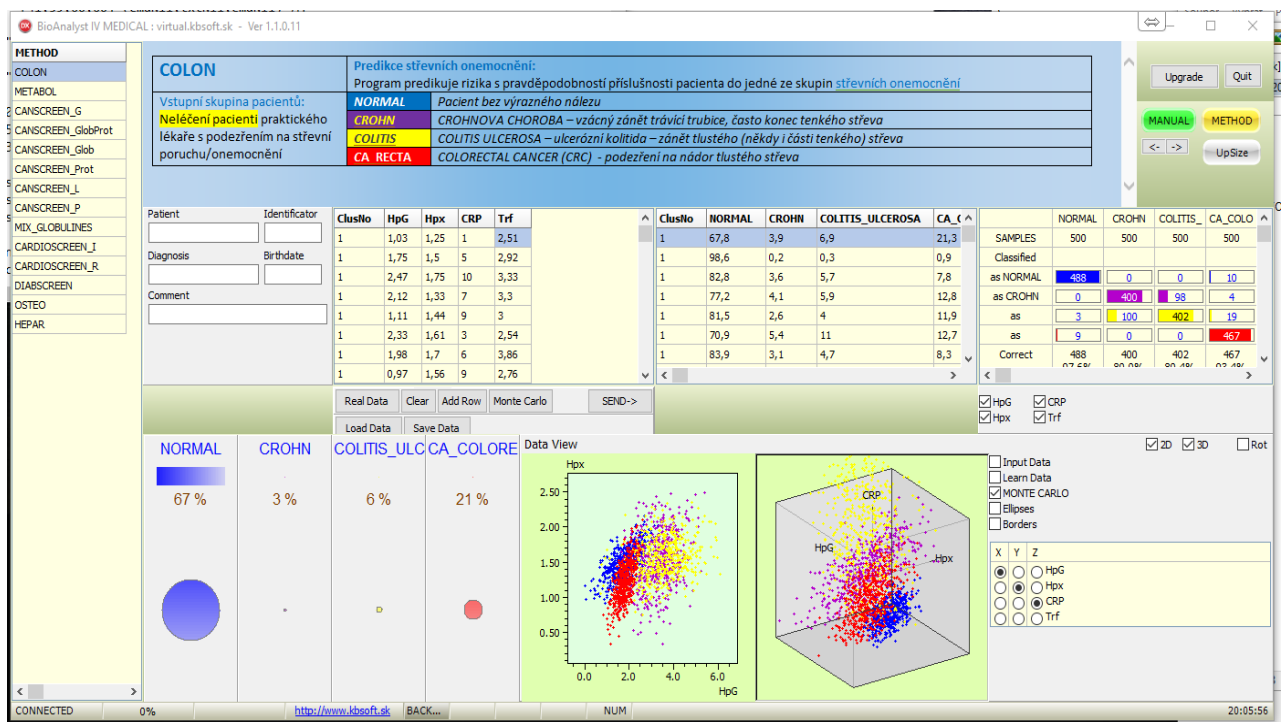
- If all the algorithms classify the given object into the same cluster, then the output will pinpoint the classification (so called fuzzy logic).
- If the algorithms classify the given object into several different clusters, then the output will question the dominant categorization, eventually, will organize the diagnosis under several clusters and show an uncertain result [2].

BioAnalyst expert system is available with several premade [diagnosing] data files – so called methods. Based on these methods, as well as based on (the panel of) examinations that are required for the given method, algorithms in this system will classify (diagnose) a patient. At the same time, thanks to the system learning capability, BioAnalyst also allows for definition of new methods, i.e. new diagnosing approaches. When creating new methods, system users must have sufficient knowledge about biochemical parameters that could, eventually, will or cannot be used to enable the given differential diagnostics. During this process, users can and, at the same time, are required to specify the number of clusters needed to classify the desired diagnosis. Plus, they are also required to define names of files that will hold the data [results] of probands and patients, who represent the given clusters. While designing a new method, it is also unavoidable to make sure that the panel of examinations is suitable for the given case, or if it is possible to skip some exams. Plus, the method designer must decide if it makes sense to continue the data collection itself. BioAnalyst system provides users assistance in the above-mentioned situations also thanks to Real Data or Monte Carlo modeling.

In this particular study, BioAnalyst has been used as a tool to enable/support screening diagnostic of some serious colon diseases. The sampling method in this experiment was as following:

42 probands (patients) were divided into four groups:

1. probands with no colon disease (12), picture 1 – blue color
2. patients diagnosed with m. Crohn (10), picture 1 – purple color
3. patients diagnosed with colitis ulcerosa (10), picture 1 – yellow color
4. patients diagnosed with some kind of colon (large intestine) cancer (10), picture 1 – red color



Picture-1: Expert System BioAnalyst – Colon Cancer Diagnosing User Interface

Groups 2 and 4 were created based on data of patients, who had been recently diagnosed, i.e. patients without/before a therapy. Thanks to using BioAnalyst's correlated data approach, the experiment was based on choosing the following parameters:

1. CRP – C reactive protein
2. Hpg- Haptoglobin
3. HpX- Hemopexin
4. Trf- Transferrin

The patient's data (blood test results) were analyzed by BioAnalyst, including the learning process and Monte Carlo modules and are shown above (see Picture-1).

So-called benzidine test for occult blood started to be used about 40 -50 years ago and is still used for fast and relatively cheap population health screening [3]. However, just like 40-50 years ago, this colon occult blood test is sensitive, and/but not specific enough, as it detects any bleeding, for example, even bleeding caused by hemorrhoids. Therefore, it's possible to say that, still, there does not exist any cost-effective, easy-to-use and more specific screening test for colon examination, while, at the same time, the amount of people diagnosed with colon cancer shows increasing tendency [for example in Slovakia]. Although it is clear that the amount of probands in the above-mentioned experiment is rather small and it would be good to enlarge the sample of probands (e.g. in cooperation with some gastroenterologists), this study (especially Monte Carlo modeling with more than

80% diagnosing accuracy) already pinpoints the possibility of effective and cost-effective use of BioAnalyst in the future.

Multifactorial analysis provided/supported by expert system BioAnalyst seems to suggest another significant possibility of biochemical and other medical data processing and diagnostic of different types of diseases. Thanks to using BioAnalyst, it is possible to remove the correlated data that do not bring any new information and consequently decrease time and costs of certain health examinations, and increase effectiveness of health screening.

All in all, this study suggests that because of the fact that medical sciences require and are based on using enormous amount of data, expert systems, such as BioAnalyst have a huge potential to help achieve new knowledge and results in a time-saving, cost-effective, and, at the same time, valid manner.

REFERENCES

1. Hastie T., Tibshirani R., & Friedman J. (2017). *The Elements of Statistical Learning*, 2nd Edition, Springer.
2. Gaško, R., Bitto, K., & Česal, M. (1995). *Interpretácia laboratórnych testov pomocou multivariantnej analýzy: počítačový program BioAnalyst*. 2. vydanie. Košice: Elfa.
3. Cícvárek, Z. (1970). *Biochemické vyšetřovací metody*. 2. vydanie. Praha: Avicenum 80.