CDDM: Clinical database for data mining

Evangelista D[1,2], Cutugno F[1], Castello G[2], Colonna G[3], Desantis S[4], Costantini S[2,3]

Motivation

Clinical laboratory databases contain usually large collections of structured medical data related to diseases progression and therapy response. The progressive increase of clinical data electronically stored is opening the possibility to carry out large-scale studies aimed to discovery correlations between new research data and correlated diseases. For these reasons many databases have implemented data mining techniques to extract unknown and potentially useful information and to discover new associations and correlations between a large data collection. We have developed a clinical database, named CDDM, to collect clinical data and evaluations related to cytokines, chemokines and growth factors concentrations and immunophenotypes data on healthy subjects and patients affected from different diseases. We have also implemented some statistical tools to correlate significatively clinical data and experimental results. At present we have focused our attention on auto-immune diseases and cancers.

Methods

The database was developed using: MySQL version 5.0.75-0 ubuntu 10.2; Web server Apache/2.2.11 (Ubuntu) PHP/5.2.6-3 ubuntu version 9.04; phpMyAdmin version 3.1.2 deb1 ubuntu 0.2. The software was realized with the scripting language PHP, the Javascript technology for dynamic contents, the markup language HTML and with style sheet CSS 2.0. For the statistical analysis, the t-test formula, correlation formula and related graphs we developed specific PHP scripts whereas for the other statistical analysis we used the aggregate functions of MySQL.

Results

The user may register to the CDDM portal and access to some services on the basis of privileges related to its role (i.e. physician, administrative personal, PhD, etc). The available functions are subdivided in two sections: patient history and statistical tools. In the first section there are case histories of patients with information related to their diagnosis, biological analyses as well as clinical data, cytokine and immunophenotypes data evaluations. In the second section, being

[1] Dipartimento di Scienze Fisiche, Università Degli Studi di Napoli Federico II, Italy [2] Centro Ricerche Oncologiche Mercogliano, "Fiorentino Lo Vuolo", Mercogliano, Italy [3] Dipartimento di Biochimica e Biofisica & Centro di Ricerca Interdipartimentale di Scienze Computazionali e Biotecnologiche, Seconda Università di Napoli, Italy [4] Altravia srl, Roma, Italy

publicly accessible, the user can choose the disease, filter the patients on the basis of gender, age and experimental data and select the best tool to perform a statistical analysis. In particular, we have implemented: i) median, mean, variance, standard deviation, min and max value for the selected protein; ii) t-test value related to the comparison between cytokine concentrations in control group and patients; iii) Pearson correlation between different cytokines with related graph; iv) Pearson correlation between each cytokine and some clinical data with related graph. Moreover we are planning to open the data set to other diseases and implement other statistical tools and classification methods to improve or discover new predictive relationships among data groups.

Contact e-mail
susan.costantini@unina2.it