Correlation Analysis as a Basis for Disease Pattern Recognition

W. Petrich[1], B. Dolenko[2], D.J. Fink[3], J. Früh[4], H. Greger[5], S. Jacob[6], F. Keller[5], A.E. Nikulin[6], M. Otto[7], M.S. Pessin-Minsley[3], O. Quarder[1], R.L. Somorjai[2], A. Staib[1], U. Thienel[8], G. Werner[1], H. Wielinger[1]

[1] Roche Diagnostics GmbH, Mannheim, Germany
[2] Institute for Biodiagnostics, National Research Council of Canada, Winnipeg, Canada
[3] Columbia University, Department. of Pathology, C.A.L.M., New York, USA
[4] Ludwigshafen, Germany
[5] Medinische Universitätsklinik Würzburg, Germany
[6] Universitätsklinik Tübingen, Germany
[7] TU Bergakademie Freiberg, Institut für Analytische Chemie, Freiberg, Germany
[8] New York Presbyterian Hospital, Dep. of Medicine, New York, USA

We observed differences between the mid-infrared spectra of sera originating from healthy volunteers and from patients with diabetes mellitus or rheumatoid arthritis. These differences were found to be significant in terms of the χ^2 test and the Kolmogorov-Smirnov significance level Q_{KS} within particular regions of the spectrum. They allow for a classification of the spectra using discriminant analysis.

In the example of discriminating between "diabetes mellitus" and "healthy" by disease pattern recognition (DPR) we obtain a sensitivity of 80 % and a specificity of 86 % (leave-one-out validation).

Some significant contributions are located in the 1000 cm^{-1} region, where there is a known glucose absorption peak. In a subsequent evaluation we investigated the correlation between disease pattern recognition and the glucose signature in the infrared spectrum. Although correlation analysis is indicative for a correlation between the DPR-score and the glucose concentration (Q_{KS} (1020 cm^{-1}) < 10^{-5}), the glucose concentration alone does not suffice for the complete correlation of the DPR-score with the actual status of health, i.e. "healthy" or "diabetes".

We used 384 samples to distinguish between serum from rheumatoid arthritis patients and healthy volunteers. 258 of these samples were used to teach a classifier in a robust linear discriminant analysis. The remaining 126 samples were examined in an independent validation of the DPR classifier, yielding a sensitivity of 84% and a specificity of 88%. The DPR-score was found to correlate better with the status of health than the rheumatoid factor alone.

[1] W. Petrich, B. Dolenko, J. Früh, M. Ganz, H. Greger, S. Jacob, F. Keller, A.E. Nikulin, M. Otto, O. Quarder, R.L. Somorjai, A. Staib, G. Werner, H. Wielinger, „Disease pattern recognition in infrared spectra of human sera with diabetes mellitus as an example“, Appl. Opt. **39**, 3372-3379 (2000).

[2] A. Staib, B. Dolenko, D.J. Fink, J. Früh, A.E. Nikulin, M. Otto, M.S. Pessin-Minsley, O. Quarder, R.L. Somorjai, U. Thienel, W. Petrich, „Disease pattern recognition testing for rheumatoid arthritis using infrared spectra of human sera“, (submitted)

[3] W. Petrich, A. Staib, M. Otto, R. Somorjai, „ Correlation between the health status of blood donors and the corresponding mid-infrared spectra of the serum“, (submitted)