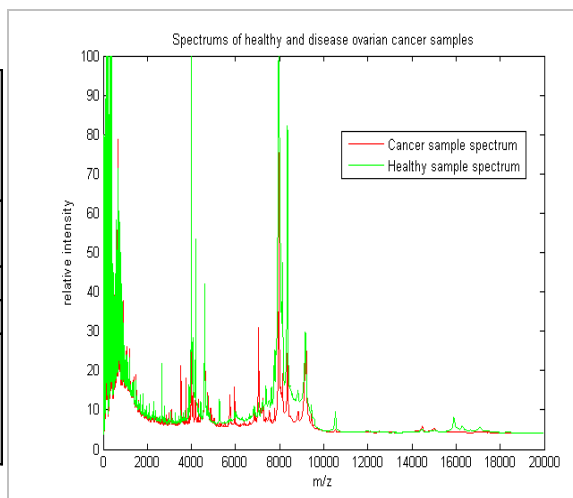




KATEDRA SYSTEMÓW AUTOMATKI

Zespół projektowy:	1. Jakub Piwowarski - kierownik 2. Maciej Rzeziński 3. Szymon Scharmach
Opiekun:	prof. dr hab. inż. Maciej Niedźwiecki
Klient:	MedicWave AB, Szwecja
Data zakończenia:	Styczeń 2010 r.
Słowa kluczowe:	Klasyfikacja widm proteinowych; Diagnostyka raka we wczesnym stadium;



TEMAT PROJEKTU:

Optymalizacja algorytmów Q5 oraz SVM, jako klasyfikatorów widm proteinowych

CELE I ZAKRES PROJEKTU:

1. Analiza, implementacja i ocena algorytmu Q5;
2. Analiza, implementacja i ocena algorytmu SVM;
3. Połączenie algorytmów Q5 i SVM oraz ocena hybrydowego algorytmu;
4. Implementacja ww. algorytmów w języku C++;
5. Porównanie powyższych algorytmów i wyłonienie najlepszego;

OSIĄGNIĘTE REZULTATY:

1. Zapoznanie się i opanowanie podstaw teoretycznych algorytmów Q5 (redukcja wymiarów PCA, maksymalizacja separowalności LDA) i SVM (w wersji Sequential Minimal Optimization);
2. Implementacja w języku Matlab algorytmów Q5 i SVM, a także testowanie i ocena ich działania;
3. Opracowanie algorytmu hybrydowego (jako połączenie PCA i SVM) oraz jego implementacja w języku Matlab i przeprowadzenie testów wydajności;
4. Implementacja wymienionych algorytmów w języku C++ oraz ocena ich działania;
5. Zestawienie wyników oraz wyłonienie najlepszego rozwiązania, którym jest algorytm Q5 ze wspierającym algorytmem hybrydowym;

CECHY CHARAKTERYSTYCZNE ROZWIĄZANIA, KIERUNKI DALSZYCH PRAC:

Zastosowane algorytmy klasyfikacji zostały przystosowane do problemu klasyfikacji widm proteinowych (duża liczba wymiarów sięgająca 45000, nierówne wartości widma, duże wahania wyników zależnie od warunków pomiaru). Wszystkie zaimplementowane rozwiązania osiągały bardzo wysokie wartości wskaźników klasyfikacji (sensitivity, specificity, PPV, NPV pomiędzy 95 a 100%), co dowodzi właściwego doboru metod do założonego celu oraz poprawnej implementacji.

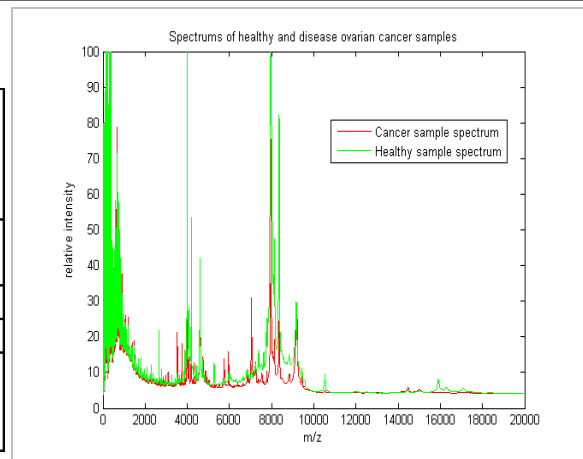
W ramach projektu rozważano zagadnienia zarówno dwu (rak, zdrowy) jak i trzyklasowe (rak złośliwy, rak łagodny, zdrowy).

Osiągnięte rezultaty są bardzo obiecujące. Opracowane klasyfikatory mogą skutecznie służyć do wczesnego wykrywania zmian rakowych, zakładając posiadanie bogatego zbioru danych uczących.



DEPARTMENT OF AUTOMATIC CONTROL

Project team:	1. Jakub Piwowarski - leader 2. Maciej Rzeziński 3. Szymon Scharmach
Supervisor:	prof. dr hab. inż. Maciej Niedźwiecki
Client:	MedicWave AB, Sweden
Date:	January 2010 yr.
Key words:	Classification of protein spectra; Early cancer diagnosis



PROJECT TITLE:

The optimal performance of Q5 and the favorable performance of SVM for classification of spectra from Lung Cancer patients

OBJECTIVES AND SCOPE:

1. Analysis, implementation and evaluation of classifier using Q5 algorithm;
2. Analysis, implementation and evaluation of classifier using SVM algorithm;
3. Combination of Q5 and SVM and evaluation of such complex classifier;
4. Implementation of all the classifiers in C++ programming language;
5. Comparison of classifiers mentioned above and selection of the best one;

RESULTS:

1. Authors became familiar with Q5 and SVM algorithms and related issues, like PCA dimensionality reduction, LDA separation maximalization and Sequential Minimal Optimization;
2. Implementation of Q5 and SVM algorithms in Matlab language, testing and assessing of their performance;
3. Elaboration of hybrid algorithm, utilizing both Q5 and SVM features, implementing it in Matlab and conducting tests and evaluation;
4. Development of C++ implementations of mentioned algorithms;
5. Algorithms and implementations comparison, choice of the best algorithm, which shows to be Q5 with supporting hybrid algorithm;

MAIN FEATURES, FUTURE WORKS:

Used algorithms have been optimized to fit the needs of protein spectra classification (that means handling vast dimensionality, reaching up to 45000, disturbances in spectra). All implemented methods achieved great performance, which is understood as reaching values of classification parameters (as sensitivity, specificity, PPV, NPV) greater than 95%, up to 100% in some cases.

Both two- (cancer and control) and three-class (cancer, benign cancer and control) cases were considered.

Results are very promising. The developed algorithms and their implementations may be successfully used in early cancer diagnosis, provided using ample training data set.