Pattern Recognition System Development for the Classification of Prostate Mass Spectrometry Data

Spiros KOSTOPOULOS[1], Dimitris GLOTSOS[1], Konstantinos SIDIROPOULOS[2], Pantelis ASVESTAS[1], Evripidis MIKOS[1], George SAKELLAROPOULOS[1], and Ioannis KALATZIS[1*]

[1]Department of Biomedical Engineering, Technological Educational Institute of Athens, Greece, *ikalatzis@teiath.gr
[2]EMBL-EBI, European Molecular Biology Laboratory, Welcome Trust Genome Campus, Hinxton, Cambridge, UK

Aim: the development of a pattern recognition system for the classification of prostate Mass Spectrometry (MS) signals.

Material & Methods: Data were retrieved from the National Cancer Institute (USA) Clinical Proteomics Database. *Class I* included the feature vectors characterizing **63** normal cases with prostate specific antigen *(PSA) <1* (**HL**) and **190** benign cases with *PSA>4* (**BE**). *Class II* included feature vectors characterizing **26** malignant cases with *PSA from 4 to 10* and **43** cases with *PSA>10*.
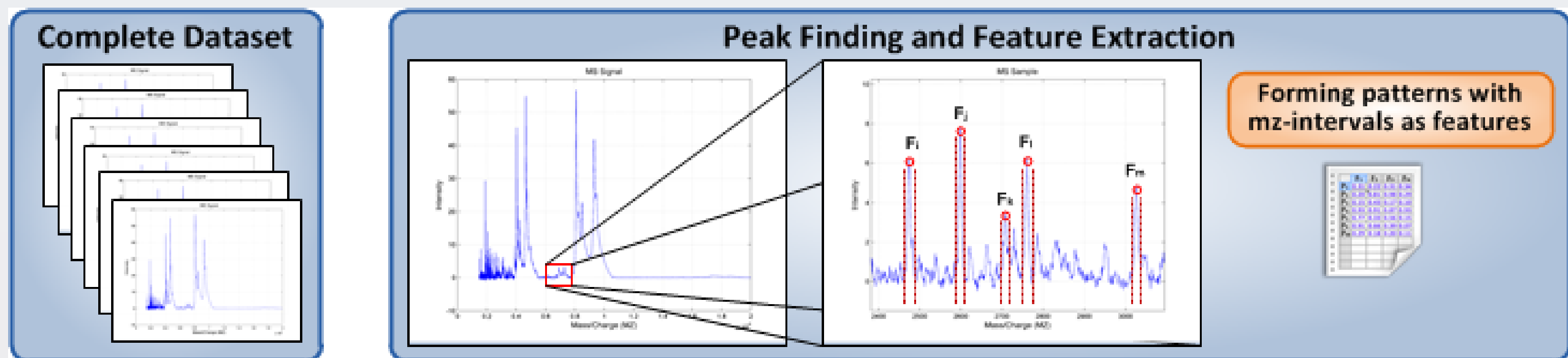
Classification: employing the Probabilistic Neural Network (PNN).

System evaluation: using the External Cross Validation (ECV). The training group comprised the 2/3 of the total number of patterns, while the rest 1/3 consisted the test group. The mean overall accuracy of the test group classification, over the 10 repetitions of the whole procedure, is considered the expected accuracy of the system when generalized on unknown data.

Training phase: using the exhaustive search (EXS), the sequential forward selection (SFS) and the sequential floating forward selection (SFFS) techniques. Feature combinations were evaluated using resubstitution (RESUB) and leave-one-out (LOO) methods. The PNN overall accuracy (OA) and the J3 class separability criterion were used for the optimal feature combination selection.

Test phase: The optimal feature combinations are characterized by the highest value of the criterion used (OA, J3) with the lowest feature vector dimensionality. The overall accuracies were calculated from the corresponding confusion matrices.

Results: The optimal feature combinations at each repetition were achieved with the SFFS technique combined with the LOO method, giving an average overall accuracy equal to 80.8%.

PNN accuracies at test phase