UPPSALA UNIVERSITET

# Molecular Biotechnology Programme

Uppsala University School of Engineering

| **UPTEC X 07 046** | **Date of issue 2007-09** |
|---|---|


Author

**Daniel Edsgärd**


Title (English)

**Implementation of an automated model selection and validation algorithm and its application to breast tumour and allergen classification**

Title (Swedish)


Abstract

This work involves implementation of an automated model selection procedure and implementation of a validation technique giving an unbiased performance estimate of the selected models. The implemented model selection and validation algorithm is applied to a breast tumour and an allergen classification problem. From these studies it can be concluded that proper validation of selected models is critical as to not report biased performance estimates. Second, for both of the applications the composition of the retrieved feature subsets is not stable. However, high performing feature subsets are generally extracted by applying a genetic algorithm implemented in this work. The model selection of breast tumour classifiers indicates that predictors performing better than van't Veer's 70-gene predictor[48] should be possible to construct. Additionally, TSPYL5 is identified as a putative oncogene. Concerning the allergen classification it is concluded that high performing predictors can be constructed using a considerably smaller subset of peptide fragments (FLAPs) than previously reported. However, reliable epitopes may not be directly extracted due to unstable feature lists.



Keywords

statistical learning, cancer classification, microarray, gene expression, allergenicity, epitopes



Supervisors

**Mats Gustafsson**

**Department of Medical Sciences and Department of Engineering Sciences, Uppsala University**

Scientific reviewer

**Anders Karlén**

**Department of Medicinal Chemistry, Uppsala University**


| Project name | Sponsors |
|---|---|
| Language **English** | Security 2008-09 |
| **ISSN 1401-2138** | Classification |
| Supplementary bibliographical information | Pages **68** |


**Biology Education Centre** Biomedical Center Husargatan 3 Uppsala

Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217