

## BIOINFORMATICS METHODS IN ANALYSIS OF MEDICAL DATABASES

**Supervisor: Jacek Waniewski, Prof.**  
**Co-supervisor: Małgorzata Dębowska, PhD**

*Institute of Biocybernetics and Biomedical Engineering PAS*

*Department IV*

*Laboratory of Mathematical Modeling of Physiological Processes*

Nowadays, there is a very rapid increase in medical data worldwide. Patients are described by a set of even several hundreds, or in the case of genetic data - several thousands/millions features. Statistical models are mostly used if the number of patients significantly exceeds the number of features to obtain statistical significance. In addition, in practice, the importance of individual feature is often assessed. This approach ignores the possibility of interaction of features emphasized by clinicians. The aim of the planned PhD thesis is the use of various bioinformatics and statistical methods for the analysis of large and complex medical databases. In particular, anthropometric, demographic, biomarkers determined in blood, the results of therapeutic interventions, genetic and other data in patients with various diseases will be analyzed. In addition to the typical statistical analysis, it is planned to use bioinformatics methods (as *lasso* - least absolute shrinkage selection operator, SVM - support vector machine and others) in order to select a set of features related to a specific clinical problem and to perform patients classification. Interaction of features and the role of the selected set of features need to be analyzed in context of clinical symptom. The appropriate methods of feature selection should allow to reduce the number of features by omitting irrelevant or redundant information. The selection of an appropriate method of analysis for a given data set is crucial. *Deep learning* method can be beneficial for some datasets whereas *linear regression* will be the most suitable for others. Systematic and comprehensive approach to medical databases, appropriate visualization, allow for proper clinical interpretation. Potential, practical applications of the obtained information include accurate diagnosis and optimization of therapy for a single patient. The project will be carried out in the international cooperation with Karolinska Institute (Stockholm).