CLASSIFICATION OF BIOMEDICAL DATA

Basic Information

This is a course, which contributes to MSc award in Biology

Course period	From February 1st till June 1st, and from September 1st till February 1st, 2 semesters
Study credits	8 ECTS credits
Duration	288 hours
Language of instruction	English
Academic requirements	- BSc degree in Biology, Chemistry, Physics or Environmental Sciences or equivalent (transcript of records), - Good command of English (certificate or other official document)

Course Description

«Classification of biomedical data» is an extensive course, which is created to provide a student the ability to perform the classification analysis using machine learning for the various medical problems. This course will also provide the skill of the imbalanced data processing in order to ameliorate the further classification analysis.

For the successful graduation of the course, the students are expected to know the hyperplane algebra and hypercomplex numbers algebra, which are obligatory to construct the algorithms for the Support Vector Machines (SVM).

The most important part of this course is the practical part, which includes the classification of the features and anomalies in the health electronic records – ECG and EEG and in the medical images from the different scan sources, using both intensity- and geometry based models.

The course is designed to cover a wide range of possible classification algorithm structures, which are special for a given area or a given organ.

Course Aims

- To help students study new discipline or to ameliorate the existing knowledge in classification analysis.
- To assist students to construct their own strategy of medical data classification algorithms.
- To give students the understanding of how to estimate the errors of the given classification algorithm.
- To introduce the basic principles of the feature and anomaly classification of medical data.

Course Objectives

The course has been designed to:

- ensure that students are familiar with a basic classification routines.
- ensure that students know the main real- and complex-valued classification algorithms, which are suitable for the most medical data analysis.
- give students the optimal structures of the classification algorithms for the medical data analysis.
- give students an appreciation of the feature and anomaly detection routines.
- provide students with the principles of the imbalanced data processing.
- provide students with the basic classification routines on medical images.

Learning Outcomes of the Course

By the end of the course, the students will be able to:

- ameliorate the imbalanced data for further classification analysis,
- estimate the error of the given classification prediction,
- classify the features and anomalies on the medical images and health electronic records.

Course (module) Structure

Learning Activities	Hours

Lectures	32
Practice sessions / Seminars,	50
Self-study Assignments	170
Final Exam (including preparation)	36
Total study hours	**288**

Course Outline

Week	Lectures	Practice session / Assignments	Hours
Semester 2			
1-6	Preclassification Routines	• Prior knowledge in classification • Imbalanced Data classification • Rough set based classification • Home assignment No 1	(7/7/40)
7-14	Types of Classification	• Model-Based Classification • Prototype-Based Classification • Random Over-Sampling Technique • Case-Based Classifiers • Home assignment No 2	(7/7/40)
	Pass/Fail Exam		
Semester 3			
1-18	Classification of Patterns and Anomalies	• Support Vector Machines in Classification • Machine-Learning Classifier • Complex Valued ELMs • Feature Extraction from Electronic Health Records • Home assignment No 3	(18/36/90)
	Final Exam		36

Course Instructors and Tutors, Contact Information

Instructor	Contact Information
Andrey SHUVAEV Ph.D. in Biophysics, Associate professor, School of Biology and Biotechnology Siberian Federal University, Krasnoyarsk	Svobodny, 79 Room 32-13 Tel +7 391 206 2165 Email: AShuvaev@sfu-kras.ru, andrey.n.shuvaev@gmail.com

Assessment

The final exam is an individual problem in the form of a simulated experimental dataset. The student must demonstrate the abilities in:

• use the prior knowledge for ameliorate the rough sets and imbalanced data (20 points maximum),

• construct the classification algorithm – real- or complex-valued for a given problem (40 points maximum),

• perform the classification of the features or anomalies in the given health electronic record (40 point maximum).

Grade policy:

A (excellent work)	91-100 %
B (above average work)	81-90 %
C (average work)	71-80 %
D (below average work)	50-70 %
F (failed work)	<50 %

Core reading

1. Abbate, D., Asmundis, R. and Guarracino, M.R. (2010) Chapter: *Prior Knowledge in the Classification of Biomedical Data*. Book: Combining Soft Computing and Statistical Methods in Data Analysis. Springer, Berlin, HeidelbergCleophas, T. and Zwinderman, A. (2014) *Machine Learning in Medicine – Cookbook* (in 3 books). Springer, Cham.
2. Fernandez, A. et al. (2018) Chapter: *Imbalanced Classification for Big Data*. Book: Learning from Imbalanced Data Sets. Springer, Cham.
3. Colantonio, S. et al. (2010) Chapter: *Prototype–Based Classification in Unbalanced Biomedical Problems*. Book: Successful Case-based Reasoning Applications - I. Springer, Berlin, Heidelberg.

4. Yin, X.-X., Hadjiloucas, S. and Zhang, Y. (2017) Chapter: *Pattern Classification*. Book: Pattern Classification of Medical Images: Computer Aided Diagnosis. Springer, Cham.
5. Leemput, K. van et al. (2005) Chapter: *Model-Based Brain Tissue Classification*. Book: Handbook of Biomedical Image Analysis. Springer, Boston, MA.
6. Kumar, S.S. and Inbarani, H.H. (2015) Chapter: *Modified Soft Rough Set Based ECG Signal Classification for Cardiac Arrhythmias*. Book: Big Data in Complex Systems. Springer, Cham.
7. Kumar, S.S. and Inbarani, H.H. (2015) Chapter: *Classification of ECG Cardiac Arrhythmias Using Bijective Soft Set*. Book: Big Data in Complex Systems. Springer, Cham.