جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Course Syllabus: Application of AI in Bioinfomatics - CS 321

Division	Computer, Electrical and Mathematical Sciences & Engineering
Course Number	CS 321
Course Title	Application of AI in Bioinfomatics
Academic Semester	Fall
Academic Year	2018/2019
Semester Start Date	08/26/2018
Semester End Date	12/11/2018
Class Schedule (Days & Time)	09:00 AM - 12:00 PM | Sun

Instructor(s)

Name	Email	Phone	Office Location	Office Hours
Vladimir Bajic	vladimir.bajic@kaust.edu.sa	+966128082386	4219, 3, Ibn Sina (bldg. 3)	Tuesday 10:00-11:00, Bld. 3, R4219

Teaching Assistant(s)

Name	Email
Dr Christophe Van Neste	christophe.vanneste@kaust.edu.sa

Course Information

Comprehensive Course Description	**Summary** Course consists of selected projects. The projects may change each year. These projects cover application of artificial intelligence (AI) to some of the relevant problems of analysis of large biological data and generally deal with complex information. Each year, the targeted problems change. Students get assigned one project and they work either alone or in groups of two. Students, in the interactive discussions with the whole class and the instructor, attempt to solve the project problems. Students regularly present their progress and defend their approach and results in front of the whole class. During one semester, several types of topics are dealt with (e.g., data integration; knowledge-, text- and data-mining of big biomedical data, genomic signal recongition, etc.). Students get direct experience in research methodology, report writing, presentations, and, most importantly, different ways of approaching solving AI applications for different bioinformatics problems. **Projects for 2018 year** **Project 1: Genomic signal recognition** There are several types of marker genomic signals that are useful in demarkating genes and their models/structure. In this topic, there ater four main signals we will target. In all four cases, the goal is to predict as accurately as possible each of the signal. Any AI/machine learning model can be used. Then, the models for the one type of signals will be compared regarding accuracy on genomes of several eukaryotic organisms. The signals are: a) Translation Initiation Site (TIS) in eukaryotic organisms. b) Poly(A) signal in mammalian organisms. c/ Donor splice site in eukaryotic organisms. d) Acceptor splice site in eukaryotic organisms. **Project 2: Non-synonymous mutations and splice sites** Sometimes there are mutations in the genome that do not alter the gene product but introduce previously non-existing signals in the genome. These signals may disrupt normal genome functioning potentially leading to diseases. The goal of this project is to identify such non-synonymous mutation in the 1000 genome project data. **Project 3: Applications of word2vec type methods for molecular function prediction** Data in many types of problems are described by very large number of descriptors, so called features. This causes the problem in the downstream analysis of the data. There are many ways how to compress information that encodes the data. One of latest approaches is based on the word2vec and similar types of methods. These methods are capable to compress information coming from several thousands of several tens of thousands of descriptors into several hundred new descriptors captured in the new feature vectors that describe original data items. This allows for significantly more efficient analysis of the original data that can be subjected further to various machine learning and AI processing. **Project 1. Apply above to analysis of function of different transcripts.** **Project 2. Apply above to the analysis of different data describing cancers.**
Course Description from Program Guide	These projects cover application of AI to some of the relevant problems of analysis of large biological data and generally deal with complex information. Each year problems change. Students get assigned one (1) project and they work either alone or in groups of 2. Students in the interactive discussions with the whole class and the instructor solve the project problems. Students regularly present their progress and defend their approach and results in front of the whole class. During one (1) semester several types of topics are dealt with. Students get direct experience in research methodology, report writing, presentations and, most importantly, different ways of approaching solving AI problems
Goals and Objectives	Course consists of selected projects. The projects may change each year. These projects cover application of artificial intelligence (AI) to some of the relevant problems of analysis of large biological data and generally deal with complex information. Each year, the targeted problems change. Students get assigned one project and they work either alone or in groups of two. Students, in the interactive discussions with the whole class and the instructor, solve the project problems. Students regularly present their progress and defend their approach and results in front of the whole class. During one semester, several types of topics are dealt with (e.g., data integration; knowledge-, text- and data-mining of big biomedical data). Students get direct experience in research methodology, report writing, presentations, and, most importantly, different ways of approaching solving AI applications for different bioinformatics problems.
Required Knowledge	C/C++, Java, Python, HPC (parallel computing) programming experience (desirable)
Reference Texts	Magana-Mora A, et al., Omni-PolyA: a method and tool for accurate recognition of Poly(A) signals in human genomic DNA. BMC Genomics. 2017 Aug 15;18(1):620. doi: 10.1186/s12864-017-4033-7. Word2Vec, http://www-personal.umich.edu/~ronxin/pdf/w2vexp.pdf Magana-Mora A et al., Dragon **TIS** Spotter: an Arabidopsis-derived predictor of translation initiation sites in plants. Bioinformatics. 2013 Jan 1;29(1):117-8. doi: 10.1093/bioinformatics/bts638. Malousi A, et al., SpliceIT: a hybrid method for **splice** signal identification based on probabilistic and biological inference. J Biomed Inform. 2010 Apr;43(2):208-17. doi: 10.1016/j.jbi.2009.09.004. Schweikert G, et al., mGene.web: a web service for accurate **computational** gene finding. Nucleic Acids Res. 2009 Jul;37(Web Server issue):W312-6. doi: 10.1093/nar/gkp479. Baten AK, et al., Fast **splice site** detection using information content and feature reduction. BMC Bioinformatics. 2008 Dec 12;9 Suppl 12:S8. doi: 10.1186/1471-2105-9-S12-S8.

Method of evaluation	**20.00**% - Presentation **10.00**% - Written report **50.00**% - Group Project(s) **20.00**% - Active participation
Nature of the assignments	Research project as defiend in the course description complemented by presentation of results, discussions on the methods of solution, written mid-term and final report.
Course Policies	Student absence of more than three times without justifiable reason will lead to failing the course.
Additional Information	Assessment of students is continuous.

Tentative Course Schedule

(Time, topic/emphasis & resources)

Week	Lectures	Topic
1	Sun 08/26/2018	Course introduction Projects introduction
2	Sun 09/02/2018	Different topics from data analysis and machine learning
3	Sun 09/09/2018	Students present progress in their projects. Discussions by all students.
4	Sun 09/16/2018	Students present progress in their projects. Discussions by all students.
5	Sun 09/23/2018	Students present progress in their projects. Discussions by all students.
6	Sun 09/30/2018	Students present progress in their projects. Discussions by all students.
7	Sun 10/07/2018	Students present progress in their projects. Discussions by all students.
8	Sun 10/14/2018	Students present progress in their projects. Discussions by all students.
9	Sun 10/21/2018	Writing reports
10	Sun 10/28/2018	Students present progress in their projects. Discussions by all students.
11	Sun 11/04/2018	Students present progress in their projects. Discussions by all students.
12	Sun 11/11/2018	Students present progress in their projects. Discussions by all students.
13	Sun 11/18/2018	Students present progress in their projects. Discussions by all students.
14	Sun 11/25/2018	Students present progress in their projects. Discussions by all students.
15	Sun 12/02/2018	Students present progress in their projects. Discussions by all students.
16	Sun 12/09/2018	Final reports discussion

Note

The instructor reserves the right to make changes to this syllabus as necessary.