



Supervisor's statement of a final thesis

Student: Ján Jendrušák
Supervisor: Ing David Přihoda
Thesis title: Antibody sequence analysis using Deep Learning
Branch of the study: Computer Science

Date: 18. 8. 2020

<i>Evaluation criterion:</i>	<i>The evaluation scale: 1 to 4.</i>
1. Fulfilment of the assignment	<u>1 = assignment fulfilled,</u> 2 = assignment fulfilled with minor objections, 3 = assignment fulfilled with major objections, 4 = assignment not fulfilled
<i>Criteria description:</i> Assess whether the submitted FT defines the objectives sufficiently and in line with the assignment; whether the objectives are formulated correctly and fulfilled sufficiently. In the comment, specify the points of the assignment that have not been met, assess the severity, impact, and, if appropriate, also the cause of the deficiencies. If the assignment differs substantially from the standards for the FT or if the student has developed the FT beyond the assignment, describe the way it got reflected on the quality of the assignment's fulfilment and the way it affected your final evaluation.	
<i>Comments:</i> The assignment was fulfilled completely, all the objectives were achieved and reported.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
2. Main written part	90 (A)
<i>Criteria description:</i> Evaluate whether the extent of the FT is adequate to its content and scope: are all the parts of the FT contentful and necessary? Next, consider whether the submitted FT is actually correct – are there factual errors or inaccuracies? Evaluate the logical structure of the FT, the thematic flow between chapters and whether the text is comprehensible to the reader. Assess whether the formal notations in the FT are used correctly. Assess the typographic and language aspects of the FT, follow the Dean's Directive No. 26/2017, Art. 3. Evaluate whether the relevant sources are properly used, quoted and cited. Verify that all quotes are properly distinguished from the results achieved in the FT, thus, that the citation ethics has not been violated and that the citations are complete and in accordance with citation practices and standards. Finally, evaluate whether the software and other copyrighted works have been used in accordance with their license terms.	
<i>Comments:</i> The thesis is written in a very readable fashion without any noticeable grammatical errors. The structure corresponds to structure of scientific publications. The background knowledge needed to understand the thesis is explained extensively, although some paragraphs could have been revisited and organized more logically. All sources are properly cited, the thesis correctly distinguishes between contributions and previous results. Parts of the discussion and the case study chapter could have been written out more extensively in the text, as opposed to just referring the reader to the tables and figures.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
3. Non-written part, attachments	100 (A)
<i>Criteria description:</i> Depending on the nature of the FT, comment on the non-written part of the thesis. For example: SW work – the overall quality of the program. Is the technology used (from the development to deployment) suitable and adequate? HW – functional sample. Evaluate the technology and tools used. Research and experimental work – repeatability of the experiment.	
<i>Comments:</i> The analysis results are organized very well, the code is comprehensive and presented mostly using Jupyter Notebooks with visualization of the analysis process. Advanced machine learning libraries, bioinformatics tools, grid schedulers and data visualization libraries were used very skilfully, greatly exceeding my expectations for an undergraduate student. Beyond the scope of the assignment, Jan also managed very well to make the results fully reproducible, the whole analysis pipeline including preprocessing, model training and result analysis is written down using Makefile steps or Jupyter Notebooks, including necessary parallelization of time-intensive tasks.	
<i>Evaluation criterion:</i>	<i>The evaluation scale: 0 to 100 points (grade A to F).</i>
4. Evaluation of results, publication outputs and awards	95 (A)
<i>Criteria description:</i> Depending on the nature of the thesis, estimate whether the thesis results could be deployed in practice; alternatively, evaluate whether the results of the FT extend the already published/known results or whether they bring in completely new findings.	

Comments:

Results were extensively explored and visualized. The analysis is reproducible and applicable to any antibody-antigen binding prediction task. Beyond the scope of the assignment, baseline approaches were implemented and evaluated, providing a very useful comparison to the deep learning approach. All implemented approaches were able to classify antibodies from vaccinated subjects unseen during training, which is considered a safe validation split strategy to avoid information leakage. Contrastingly, applying the deep learning classifier to a separate vaccination study produces performance close to random choice. Nevertheless, this result is properly reported, discussed and compared to better performing baseline approaches implemented also by Jan.

Evaluation criterion:

The evaluation scale: 1 to 5.

5. Activity and self-reliance of the student

5a:
1 = excellent activity,
2 = very good activity,
3 = average activity,
4 = weaker, but still sufficient activity,
5 = insufficient activity
5b:
1 = excellent self-reliance,
2 = very good self-reliance,
3 = average self-reliance,
4 = weaker, but still sufficient self-reliance,
5 = insufficient self-reliance.

Criteria description:

From your experience with the course of the work on the thesis and its outcome, review the student's activity while working on the thesis, his/her punctuality when meeting the deadlines and whether he/she consulted you as he/she went along and also, whether he/she was well prepared for these consultations (5a). Assess the student's ability to develop independent creative work (5b).

Comments:

Jan was not afraid to pursue hard scientific questions and report the results thoroughly in clear visual form. He was able to both follow my suggestions as well as come up with his own creative solutions to problems. My experience as a supervisor felt more like supervising a master's student, if not better.

Evaluation criterion:

The evaluation scale: 0 to 100 points (grade A to F).

6. The overall evaluation

95 (A)

Criteria description:

Summarize which of the aspects of the FT affected your grading process the most. The overall grade does not need to be an arithmetic mean (or other value) calculated from the evaluation in the previous criteria. Generally, a well-fulfilled assignment is assessed by grade A.

Comments:

The thesis provides an in-depth look on predicting binding of antibodies to a specific antigen, demonstrated on Hepatitis B. The deep learning method was able to classify antibodies from vaccinated subjects unseen during training. Although applying the deep learning classifier performed poorly when applied to a separate vaccination study, this result is properly reported, discussed and compared to implemented baseline approaches.

Signature of the supervisor: