

I. IDENTIFICATION DATA

Thesis title:	Machine learning for ttH mechanism Higgsboson detection from CERN ATLAS data
Author's name:	Jan Presperin
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Open Informatics
Thesis supervisors:	Prof. Jan Kybic, doc. Andre Sopczak
Reviewer's department:	Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment	challenging
<i>How demanding was the assigned project?</i>	
<p>The assigned project has been challenging as the signal and background separation is particularly difficult. There are many features which only contribute with a little separation power. The provided simulated data is in a format used in High Energy Physics and had to be converted in a suitable format for the machine learning application. The data from several simulated processes had to be normalized according to the number of expected events in the detector, each simulated event obtained an individual weight. The statistical interpretation of the results with respect to previous results was challenging as the same methods had to be applied for consistency.</p>	
Fulfilment of assignment	fulfilled
<i>How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.</i>	
<p>All goals were achieved with an expected significant increase in signal sensitivity. The result on low-level and high-level feature importance is novel. The developed software can be used for further studies in this field of research. An application on recorded data will follow with a new data set of features.</p>	
Activity and independence when creating final thesis	A - excellent.
<i>Assess whether the student had a positive approach, whether the time limits were met, whether the conception was regularly consulted and whether the student was well prepared for the consultations. Assess the student's ability to work independently.</i>	
<p>Jan Presperin met very well the time limits and progressed continuously towards his final results. He reported more than weekly on progress and made several intermediate status reports. He worked very independently in solving upcoming challenges with new ideas. He also presented his results at CERN meetings and received positive feedback, and suggestion which he implemented.</p>	
Technical level	B - very good.
<i>Is the thesis technically sound? How well did the student employ expertise in his/her field of study? Does the student explain clearly what he/she has done?</i>	
<p>The technical realization is convincing and at the state of the art level. The modular structure that the code can be used for an updated analysis is an advantage. Each step of the analyses is explained well.</p>	
Formal level and language level, scope of thesis	A - excellent.
<i>Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is</i>	

the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?

The formal level and the notations are consistent. The thesis is organized logically and easy to follow. Sufficient details are presented to follow each step what was done. The English language level is very good.

Selection of sources, citation correctness

B - very good.

Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?

The thesis adequately refers to earlier work on the topic. The selection of sources is correct. As a particularity of this research topic, previous results in recent years were presented only in internal meetings and in internal notes. The student's work is clearly distinguished from previous results. The bibliographic citations meet the standards.

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

As this research is part of a larger collaboration, the input data sample are being updated with better calibrations, further optimized object definitions, and adjusted preselections. The possibility of an efficient application of the developed software on new input data was therefore important in the designing the software, thus the software is also useful in further analyses of this research. The student has been very skillful in implementing the tasks, and efficient in solving naturally upcoming problems and technical challenges.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading.

The grade that I award for the thesis is **A - excellent.**

The thesis represents a solid research in a challenging field. The technical challenges to separate the signal from the background were high, and several machine learning algorithms were implemented and their performance compared. The result that an increase of a simulated data set will lead to an increase of the sensitivity is important. The quality of the figures and the thesis presentation is very good.

Questions for the thesis defense:

- 1) How do correlations of the features affect the stability of the results?**
- 2) Can you estimate by extrapolation the sensitivity gain with more simulated data?**
- 3) Which outlook do you see to increase the significance further?**

Date: 23.5.2022