CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS SUPERVISOR'S REPORT

I. IDENTIFICATION DATA

Thesis title:	Application of Machine Learning for the Charged Higgs Boson Search Using ATLAS Data.
Author's name:	Jiří Pospíšil
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	André Sopczak
Reviewer's department:	Institute of Experimental and Applied Physics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

How demanding was the assigned project?

The thesis was very demanding as it applied Machine Learning to a challenging task of separating a specific signature (signal) and unwanted background events. In addition to the machine learning optimizations, the project also required to get familiar with data format used in the ATLAS experiment, and also using ATLAS software like root and trexfitter. Further challenges were the network optimization with the provided features, and the determination of the working point for highest signal sensitivity. The developed machine learning should also be flexible to cope with different signal masses, having very different signal kinematics.

Fulfilment of assignment **fulfilled**

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.

The thesis fulfilled all tasks. The primary goal to develop and test machine learning algorithms is fulfilled. As the simulation of the signal by the ATLAS collaboration took several weeks, the student developed the machine learning code first on a similar signal to separate the background. Important for the task was the statistical analysis. Over the assigned task, the student also learned to use the trexfitter package to calculate the significance (confidence level) in a separate way, which is compatible to the method used in other analyses by the CERN collaborations for a better compatibility in the comparison of the results. The trexfitter package was also used for producing plots of signal background processes with a high standard.

Activity and independence when creating final thesis **A - excellent.**

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.

The student met the time limits very well and reported more than weekly on the progress of his studies. Jiri was well prepared for the discussions on the progress. He also presented his results at a meeting at CERN and received positive feedback. He was also able to answer questions on specific analysis aspects during and following his presentation. He demonstrated very good research capabilities and solving problems independently.

Technical level **A - excellent.**

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?

Technical aspects are well explained and expertise from his field of studies on machine learning is applied. Jiri used several machine learning algorithms and compared their performance. Very good separation power was achieved between the signal and background.

CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS SUPERVISOR'S REPORT

Formal level and language level, scope of thesis **B - very good.**
Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?
The standard formalism of the thesis and the notations are correct. The thesis has sufficient details and is well presented. The English language is very good and clear.

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 lists previous work adequately. The reference list is rather short, but sufficient. The selection of sources is adequate. The student's work is very clear. The bibliography style in the reference list could be more unified.

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.
Overall, the student performed very well in this research project applying the machine learning techniques. The topic is novel and it extends previous research in a new analysis channel. The result is promising and shows for small signal masses significant improvement over previous results. The code is well written and documented which allows the future use to follow up on this analysis. The student has been very skillful in scientific research with the application of machine learning tools.

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.

In summary, the thesis contains a solid piece of research work with the application of machine learning. The student performed very well, worked independently and communicated his ideas and progress regularly.

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

Minor comments / questions:

How is made use of the GPU and how much does it speed up the training?

Date: **30.5.2022**

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

REVIEWER'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	Application of Machine Learning for the Charged Higgs Boson Search Using ATLAS Data
Author's name:	**Jiří Pospíšil**
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Karel Smolek
Reviewer's department:	Institute of Experimental and Applied Physics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

Evaluation of thesis difficulty of assignment.

The aim of the thesis is the application of machine-learning for the selection of events with a charged Higgs boson, produced in the ATLAS experiment at the LHC. The thesis focuses on final states with two same-sign charged light leptons, a hadronically decaying tau lepton and multiple quarks. The final state is rather complex. Therefore, it can be expected the procedure for distinguishing of signal and background events will be complicated. The data acquisition and processing within the ATLAS experiment is rather complex. The application of the data samples provided by the ATLAS collaboration requires mastering many technical details (e.g. related to the data format, SW used with the ATLAS collaboration). Furthermore, the topic of the thesis requires understanding of elements of particle physics terminology and phenomenology. It can be difficult for a student not studying high energy physics.

Satisfaction of assignment **fulfilled**

Assess that handed thesis meets assignment. Present points of assignment that fell short or were extended. Try to assess importance, impact or cause of each shortcoming.

The thesis fulfills all goals specified in the thesis assignment.

Method of conception **correct**

Assess that student has chosen correct approach or solution methods.

The main goal of the thesis is to develop and to optimize the procedure for distinguishing of signal and background events. Due to a high complexity of the analyzed final states with many parameters describing the events, the application of neural networks was an optimal approach to reach aims specified in the thesis assignment.

Technical level **A - excellent.**

Assess level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.

The student managed to use programs specific for machine learning and for data analysis used in the ATLAS collaboration. Two neural networks, MLP and TabNet, were implemented. The student successfully processed detailed tuning and optimizations to get optimal results in the signal events selection. He estimated corresponding cross section exclusion limits. I positively evaluate, that his results indicate a higher sensitivity for tbH^+ production than methods published by the CMS collaboration.

Formal and language level, scope of thesis **A - excellent.**

Assess correctness of usage of formal notation. Assess typographical and language arrangement of thesis.

The thesis is well written. The description of the used methods and obtained results is clear and comprehensible, the structure of the thesis is well arranged. There is almost no typo in the text.

I noticed a mistake at the end of the section 2.2: A charged Higgs boson H^+ can decay to an anti-tau lepton and a tau-neutrino, not to an anti-tau lepton and a muon. On the page 5, the correct decay channel of a W^+ boson is an anti-lepton and a neutrino (not an anti-lepton and a muon). In the introductory sections 2.2 and 2.3, it would be advisable to mark antiparticles with a bar above a symbol of a particle. Also, Feynman diagrams in Figures 2.1, 2.2 and 2.3 would be clearer,

if correct particle/antiparticle labels were chosen, as well as if arrows signing a particle momentum orientation in a spacetime were used.

Selection of sources, citation correctness **B - very good.**

Present your opinion to student's activity when obtaining and using study materials for thesis creation. Characterize selection of sources. Assess that student used all relevant sources. Verify that all used elements are correctly distinguished from own results and thoughts. Assess that citation ethics has not been breached and that all bibliographic citations are complete and in accordance with citation convention and standards.

The citations are used properly throughout the thesis. In the Bibliography section, the list of authors for each reference should have the same format (sometimes a first name is missing, sometimes it is fully written, and sometimes only a first character of a first name is written).

Additional commentary and evaluation

Present your opinion to achieved primary goals of thesis, e.g. level of theoretical results, level and functionality of technical or software conception, publication performance, experimental dexterity etc.

All goals of the thesis were achieved. The student obtained nice results, applicable in practice within the ATLAS experiment. I have no serious objection to the thesis.

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Summarize thesis aspects that swayed your final evaluation. Please present apt questions which student should answer during defense.

I positively evaluate all aspects of the thesis. The student demonstrated a high level of knowledge in the field of machine learning and the ability to apply his skills within the real experiment in the field of high energy physics.

Question: It is well known that a simulation of so complex object like the ATLAS detector may not reflect fully all real details. For example, I can imagine, that real measurement performance of the detector ATLAS can be slightly different than supposed in the simulation used for the generation of training data. Or, there are several models used for a creation of jets from primary quarks. What could you say about the robustness (sensitivity of neural network training on the quality of simulated input data) of the machine learning method presented in the thesis?

I evaluate handed thesis with classification grade **A - excellent.**

Date: **3.6.2022** Signature: