THESIS SUPERVISOR'S REPORT

I. IDENTIFICATION DATA

Thesis title:	Application of Machine Learning for the Higgs Boson Mass Reconstruction Using ATLAS Data.
Author's name:	Adam Herold
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Cybernetics and Robotics
Thesis supervisors:	Andre Sopczak, Jan Kybic
Supervisors' departments:	IEAP, K333 FEL

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

How demanding was the assigned project?

The task has been challenging as it involved a complex input data structure. Several data sets where provided which required a weighting to match the expectations in the real experiment. Furthermore, two different networks had to be applied, one for the best possible assignment of the elementary particles, and the other one for the reconstructing the Higgs boson mass. The neural network tself has not been particularly challenging.

Fulfilment of assignment **fulfilled with minor objections**

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 provided code for Higgs boson mass reconstruction was well studied and understood. The existing neural network code was further developed on the generated truth level, and also with simulated data of the ATLAS detector. The attribution of the reconstructed objects to the Higgs boson and top quark decays were studied and the correct attribution was quantified. The expected mass accuracy was determined and optimized the with simulated data of the ATLAS detector. Finally, the developed algorithm was applied to the actual data recorded by the ATLAS experiment and determine the Higgs boson mass, however, not its uncertainty. A minor objection is that the ordering of the features was not performed. Also a scaling should have applied such that the Higgs boson mass matched the known value. The code developed by the student is not well documented.

Activity and independence when creating final thesis **B - very good.**

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 has been very active, approached the scientific work systematically. He also reported in internal meetings and demonstrated his progress. The student was well prepared for the consultations and clearly able to work independently.

Technical level **B - very good.**

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?

The contents of the thesis is sound and his expertise is employed. The student has explained clear the research he has done.

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?

Overall the thesis is well written, and the English is very good. Minor format inconsistencies are in the citations, and also the application on real data could have been explained further and more conclusions drawn, also a list for future developments would have been useful.

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?

Adequate reference are given, and the student own contribution is well documented. The bibliographic citations meet the standard.

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.

Overll the quality of the thesis is good, and has potential for further development, some aspects that the features were not order and their number reduced was missing, also for the real data the jet variables where not used and when it was identified there was not enough time to correct it before the submission date..

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 thesis fulfills all tasks with minor objections, they could have been solved with more time before the submission date. The thesis describes very well the work done, it is original and the outcome useful. The usefulness would be higher with better description of the code. Scientifically, the scaling of the reconstructed masses and the estimate of the uncertainty would be useful. Also, the comparison with the missing mass calculator result is well done, however, a standard code was applied and not particularly adapted to the specific case. The use of two different input data production resulted in not using the jet information for the assignment and thus reduced the correct assignment probability significantly. These items should have been listed in the thesis for future research. Overall, the student show much motivation, addressed the tasks systematicalls and work independently while regularly reporting the progress.

The grade that I award for the thesis is B - very good.

Date: 20.1.2022 Signature: Andre Sopczak, Jan Kybic

REVIEWER'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	Application of Machine Learning for the Higgs Boson Mass Reconstruction Using ATLAS Data
Author's name:	**Adam Herold**
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Babar Ali
Reviewer's department:	Institute of Experimental and Applied Physics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

Evaluation of thesis difficulty of assignment.

Reconstructing Higgs boson mass produced in the association of top anti-top quark pairs is a challenging task because of the very busy environment. In different final states, there are lot of physics objects such as leptons, taus, jets, and missing energy (neutrinos).

Satisfaction of assignment **fulfilled with minor objections**

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 assignment is fulfilled in all aspects. I have minor comments please see below.

Method of conception **correct**

Assess that student has chosen correct approach or solution methods.

A classification neural network technique is developed for training and testing using simulated data. The approach is correct since the task in hand required to reconstruct Higgs mass and separate ttZ background.

Technical level **B - very good.**

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

The data recorded by the ATLAS detector at LHC CERN consists of different physics objects e.g., light leptons, taus, jets etc. I positively evaluate author's ability to work with such complicated data.

Formal and language level, scope of thesis **C - good.**

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

The thesis is concise and well written. I have no objections about the structure if the individual chapters. Citations are used properly. There are quite few repetitions about selection. Section 4.2 is difficult to follow, and it could be more precise. I would have expected detailed but concise explanation about the machine learning and neural network.

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.

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Overall, I positively evaluate all aspects of this thesis. I consider the work done is of good quality and could be very useful for the ttH group within ATLAS.

I evaluate handed thesis with classification grade **B - very good.**

Minor comments:
Equation 1.10, the leading order Feynman diagram for ttZ production is gluon-gluon and not via quark-antiquark.
Page 22: The common selection requires at least three jets. It should require at least four jets and out of them at least one b-jet. In a typical ttH event, we expect at least four jets (see Figure 1.4)
Nomenclature used for the variables in the CNN is difficult to follow. It could be simplified.
Figure 4.1: The description of z-axis is not given in the figure caption

Major comments:
Section 3.4: The narrow selection is used for the particle assignment, and it is stated that the wider selections are not used as the decay channels of those events are different. This is not entirely true. A typical ttW can also give the same final state and mimic the ttH event. While for a ttbar event, one of either light lepton or tau is fake (misidentified) as they originated from another source such as in decays of b flavored hadrons.

Figure 4.2: Separation achieved in MMC inspired loss is comparable with MMC itself. The mean values are smaller. This could be improved by just focusing on Higgs to tau tau decay mode as it is done in MMC case (1.3.2). The additional neutrino originating from anti-top decay has no significant impact in reconstruing Higgs mass.

Question: A classification NN approach is used inspired by the paper ref. [36]. In that paper, they used a stricter requirements e.g number of jets up to 6. Did you try to use the number of jets up to 6?

Question: Did you test the separation between ttH and ttZ if you had use data selection same as of real data? (Figure 4.5).

Date: **24.1.2022** Signature: