THESIS SUPERVISOR'S REPORT

CZECH TECHNICAL UNIVERSITY IN PRAGUE

I. IDENTIFICATION DATA

Thesis title:	**Rectifying Probabilistic Predictions of Neural Networks**
Author's name:	**Tuan Anh Ho**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Oleksandr Shekhovtsov
Reviewer's department:	Department of Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **extraordinarily challenging**

How demanding was the assigned project?

The project required reading and analyzing research papers, working theoretically and performing comprehensive experimental evaluations. The theoretical part posed open research problems and required high skill in statistics and optimization. The scope of the project is also rather broad, calling for generalization to several problems and exploring beyond. Part 4 of the assignment was posed as optional directions to explore. A solid start with the semestral project made it possible to pose such assignment. The project was also organized as a collaboration, with the supervisor contributing to the theoretical research.

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.

Primarily goals were achieved. The main technical parts 2-3 were addressed in depth, which is excellent. Part 4 of the assignment was proposing to explore new challenging problem formulations. It was not addressed, but it is completely in order, given the research nature of the assignment and the focus on parts 2-3.

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.

Tuan Anh was continuously working hard during the semestral and the master project and preparing well for consultations both in understanding the theory and discussing implementation results. In the last two month of the project the schedule was very tight but was successfully met. Compared to other students in my experience, the ability to build own rationale, make a research proposal or conduct analysis independently were somewhat lacking. Nevertheless, during the project I observed that the progress achieved in theoretical and implementation subtasks in the independent mode was constantly increasing.

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS SUPERVISOR'S REPORT

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?

Tuan Ahn successfully employed expertise in statistics, optimization, data science, pattern recognition, machine learning and software engineering from his field of study. He further learned and applied theoretical concepts and methods from the specialized literature. The thesis clearly explains what is done and to my best knowledge is technically sound.

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 formalisms and notations are used properly. The English is generally very good, but sometimes could use less pronouns and be more specific, for example "Having a calibration that would map accordingly, we could achieve that, however decision calibration does not allow that."
The explanations are rather intuitive, however a reader not familiar with the topic may encounter difficulties understanding as not all concepts and thoughts are introduced gradually and sequentially enough.
The organization is somewhat suboptimal in that all the background is followed by all the methods and followed by all the experiments. In this way the reader has to keep in mind everything what was done and it is difficult to track the logic and purpose. This is compensated to some extend by the overview in the introduction and references to upcoming experiments.

Selection of sources, citation correctness **A - excellent.**

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 cites results from many very recent works in the field, in particular Zhao et al., developed contemporaneously with the master project. The thesis cites as well key works in the statistical forecasting, where reliability concepts have been proposed and studied much earlier. Citations are used with a purpose covering all relevant works known to us. Every result taken from an external source is appropriately cited.

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.

The thesis presents novel theory and methods for calibration of statistical predictors derived from a particular purpose – the known downstream statistical decision task. The thesis was fully written by the student, however it contains a joint contribution of the supervisor and the student with the supervisor proposing the theory (this is also declared in the thesis). Intermediate results were submitted as a paper to ICLR 2023 but were not accepted. Since then we devoted and improved method (included in the thesis).
We believe it has significance and might be impactful, however the experiments showed (rather surprisingly) that task-specific calibration is in many cases not superior to a general task-agnostic calibration. We might need to adjust the methods or experiment with their implementation in order to improve the practical gain. We plan to submit a new paper including extended theory and methods in the future.

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS SUPERVISOR'S REPORT

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

It was a great experience to work with Tuan Anh. Over the course of the semestral and the master project he worked hard in the connected topics. The calibration topic was new to the supervisor(s) as well and our knowledge and the tasks were evolving along the way. He learned a lot on all sides: understanding applying and designing the theory studied in the program and beyond, implementing and evaluating the results in a scientifically convincing way, structuring and writing technical material. I am very happy with the results that were achieved and I am convinced that Tuan Anh Ho has reached the master's excellence level.

Q1. I believe there are some mistakes/typos in the thesis, please answer by correcting:

"Prior shift adaptation and Bayesian decision making in a cost-sensitive environment, with the first issue being more general"

"At calibration time, prior shift is known for one scalar parameter".

Q2. The experimental evidence obtained appears somewhat counterintuitive: we systematically observe in the experiments that the more task-specific methods perform worse than the less task-specific ones (e.g. DirectLoss is worse than Integral loss and Integral loss is worse than NLL). How this can be explained?

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

Date: **27.1.2023** Signature: