Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Evgenii Tsymbalov

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Machine Learning for Elastic Strain Engineering

Supervisor: Associate Professor Alexander Shapeev

Name of the Reviewer: Alexey Zaytsev

I confirm the absence of any conflict of interest	**Signature:**  **Date: 19-09-2020**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications

The summary of issues to be addressed before/during the thesis defense

Provisional Recommendation

The presented thesis combines the surrogate modeling approaches to the problem of crystal deformation. Author develops a complex framework for the prediction of crystal properties based on surrogate modeling approach. He proves that the model based on convolutional neural

network model tailored to the problem specifics can predict the desired properties with a relative error about 1%. On top of this, Evgenii provides a recipe for the surrogate model improvement based on an adaptive selection of the design of experiments (aDoE). The uncertainty estimates required for aDoE is based on Gaussian Processes (GP) regression model distilled from a neural network's output.

The structure of the thesis looks good and provides the reader with all the necessary components for understanding. Both background introduction and methodology description are oriented on the computer science community as physics involved is explained from the basics. In addition, Evgenii provides a deep overview of used machine learning approaches suitable for non-specialist.

The developed framework is thoroughly tested in a numerous numerical experiments, related both to elastic strain engineering and fully-connected neural networks. The results were presented in top-tier international conference on machine learning (IJCAI, A* CORE rating) and high-impact interdisciplinary journal, which demonstrates both quality of publications and international acceptance of the works. The scientific footprint produced in the process is twofold: the author contributes to the state of the art in the uncertainty estimation and active learning for neural network models and explores demandable properties of the deformed diamond and silicon crystals.

I recommend that the candidate should defend the thesis by means of a formal thesis defense, yet would like him to comment more on the following matters:

- Can the author provide an evidence that considered GP models provide results of reasonable quality for both their mean prediction and uncertainty estimates? How to take advantage of the structure of multidimensional output while constructing GP model?
- It would be interesting to compare other uncertainty estimation approaches in the scope of the main problem of the work to provide practical recipes on how to construct and use surrogate models in this area.
- Why the author use Bayesian Neural Networks? Now GP model is based on mean predictions by NN model and don't require multiple runs of NN with different dropout masks.
- The author doesn't use specific multifidelity approaches to construct surrogate models, while there is a number of GP regression models designed to work with multifidelity data e.g. cokriging and derivative works in this area

+ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*