Skoltech

Skolkovo Institute of Science and Technology
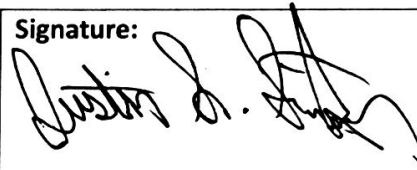
Jury Member Report – Doctor of Philosophy thesis.

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: Justin S Smith

I confirm the absence of any conflict of interest	Signature:
Yes.	
	Date: 20-09-2020

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

The thesis presents the development of a neural network-based model for the prediction of band structures, the evaluation and modification of dropout-based uncertainty estimation methods, and active/transfer learning improvements of the model. This is all combined into one framework for elastic strain engineering applications. Overall, the thesis is well structured and high quality. All content and methods employed within the thesis fit well within the topic of the dissertation. To the best of my knowledge the results are novel, and provide a powerful tool for exploring materials space. The publications that resulted from this work are novel and important additions to the field.

List of comments for consideration:

1) The initial random data set seemed very large compared the data generated in the very few active learning iterations. This made the active learning work seem to be tacked on, rather than an ultimate effort to produce the best model with minimal data. Perhaps a justification on why one wouldn't start with a much smaller initial data set, then active learn to error convergence, is warranted. Could it be possible to build the model with much less PBE data? This would be important information for future studies wanting to use these methods for other materials.

2) Perhaps I overlooked it, but I could not find how the transfer learning data (GW) was selected. Was this data simply a random subsample of the larger PBE data?

3) Ensemble-based UE is said to "increase training time". This is not necessarily true since models in the ensemble can be trained (and evaluated) in an embarrassingly parallel manner. It does, however, increase the overall computational resources required.

4) Are there any major disadvantages to using the dropout-based UE method? Dropout tends to require a very large number of model parameters compared to non-dropout models, does this slow the model training and evaluation down significantly? How much larger did you need to make the dropout-based models compared to a normal model?

5) How (if possible) could these methods be extended to have a single model predict the band structure on multiple crystal phases and/or elements? This could make an interesting discussion in the future work section.

Provisional Recommendation

☑ *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*