Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Nikita Klyuchnikov

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Multi-fidelity Classification and Active Search

Supervisor: Associate Prof. Evgeny Burnaev

Name of the Reviewer: Maxim Panov, Assistant Professor, CDISE, Skoltech

I confirm the absence of any conflict of interest.	**Signature:**  **Date: 25-01-2021**

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 considered thesis targets a very specific but important area of multi-fidelity classification and active search. The author considers several important problems:

1. Construction of multi-fidelity classification method based on the ideas from co-criging.
2. Development of new active search method specially tailored to analysis of multi-fidelity data.
3. Application of the developed methods to several important real-world data analysis problems.

All these topics are relevant for modern machine learning and data analysis. The author shows fluency in state-of-the-art machine learning methods and their application to challenging problems involving molecular data. Moreover, some of the methods require non-trivial modifications and lead to the algorithms, which are interesting for machine learning in general. I especially appreciate the usage of ideas from co-kriging (usually applied to regression) in the case of classification. What worries me is the significance of the experimental results of section 2. The author never reports standard deviations or confidence intervals, while the margins look small at all the plots/tables. In addition, I really miss computational complexity analysis for the method developed in Section 2. The other concern is section 4 where it seems that rather standard method for the single fidelity data are used rather than the ones developed by the author. Is it so?

In my opinion, thesis results are very relevant from the application perspective. Importantly, the author provides all the details on the algorithm construction and training procedures, which allows to directly applying them.

Finally, the results of the thesis research were published in well-reputed journals and conference proceedings including one publication in CORE A* conference and four publications in Q1 journals. Thus, the quality of the publications well supports the overall good scientific quality of Nikita's thesis research.

While I have overall positive opinion about the research contents of the thesis I think that the text deserves some improvements. First, the author should unify notations throughout the text. It is clear that the formulas were taken from different publications and it is the reason. This should be fixed in my opinion. Also, I highly recommend to proof read the text thoroughly as currently many sentences look not entirely correct or/and can be better formulated.

To sum up, I think that the issues found do not decrease the scientific quality of the thesis and Nikita Klyuchnikov deserves to be awarded with Skoltech PhD degree.

The list of issues (in the order of appearance in the text):

1. [list of publications] Many publications have incorrect page numbers, or completely miss page numbers.
2. [Page 3, main defense statement 1] I don't think that the wording "co-kriging of latent Gaussian processes" makes a lot of sense to me.
3. [last formula before section 2.2.2.1] I don't think that the equality is valid (due to integration).
4. [formula 2.8] Index L seem to be missing for the arguments of kernel k_l
5. [page 9 and further] I find it very non-standard to write indexes in summations as i=1..n, recommend to change to standard _{I = 1}^n
6. [section 2.2.3] It is strange that matrices A and D are defined only in the next section. Why not define right here?
7. [page 11] Not clear, what indexing means in first two formulas, it is some very non-standard notation.

8. [Section 3 and on] The quality of formulas and punctuation for them slightly deteriorates starting from Section 3.
9. [Figure 3] What do dotted lines mean above and below solid lines?
10. [page 77] It seems that citation is missing for Vega library.

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*