Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Satyarth Mishra Sharma

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Predicting phenotype from genotype with machine learning

Supervisor: Professor Dmitry Yarotsky

Name of the Reviewer: Prof Gajendra P. S. Raghava

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Date: 11-12-2024**

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

In this thesis, the candidate has made a systematic and comprehensive effort to predict human health phenotypes based on genomic information. This research addresses a critical challenge in leveraging the full potential of genetic data, which continues to grow exponentially due to rapid advancements in high-throughput sequencing technologies. The thesis specifically addresses four key challenges associated with phenotypes: (i) the prediction of complex traits, (ii) ancestry prediction, (iii) understanding changes in protein stability, and (iv) detecting and mitigating errors in the genotyping process. The thesis is thoughtfully organized into six chapters. Chapter one provides the background and foundation of the study, while chapter six explores the future prospects of genomic medicine. The intervening four chapters focus on the methodologies developed to address the four identified problems, offering clear and concise solutions to each challenge. This well-structured thesis not only articulates each problem with precision but also proposes practical and innovative solutions to address real-world challenges. The work exemplifies a significant contribution to the field of genome medicine, offering valuable insights. The ability to accurately link genomic variations to phenotypic outcomes is essential for advancing personalized medicine, improving disease prevention strategies, and developing targeted therapeutic interventions. Number of papers have been already published in internationally reputed journals based on the work performed in this thesis. This indicate the quality of research performed in this study, as most of work have been peer reviewed by international experts. Overall, it is an excellent scientific contribution. Candidate should provide all resources developed in this study to public via freely accessible link.

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*~~