Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Bogdan Kirillov

PhD Program: Life Sciences

Title of Thesis: Uncertainty Quantification and Neural Network Interpretation for studying CRISPR mechanics

Supervisor: Assistant Professor Maxim Panov

Name of the Reviewer: Dr. Martin Takac

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Date: 13-09-2023**

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 doctoral dissertation titled is an exemplary piece of research at the intersection of biotechnology, genomics, and machine learning. The work demonstrates a comprehensive understanding of the CRISPR-Cas system and its challenges, as indicated by the comprehensive abstract.

The structure of the thesis appears to be well-conceived and coherent. Beginning with the foundational background on CRISPR-Cas mechanisms, it delves into the complexities of data noise introduced at various stages of the experimental pipeline. The author astutely recognizes the need for advanced mathematical methodologies in handling these challenges, suggesting a thorough command over the subject.

The inclusion of state-of-the-art models for estimating cleavage efficiency stands out, especially with the biological validation through the unsupervised rediscovery of the seed region importance. It's particularly noteworthy that the author's work not only confirms known knowledge but leverages novel methods like Explainable Machine Learning and Neural Network Interpretation to achieve these findings.

Furthermore, the innovation introduced through prediction uncertainty as a novel axis for off-target cleavage efficiency analysis is commendable. It highlights a new dimension to the field and suggests a more sophisticated and nuanced approach to gRNA selection for gene editing.

Lastly, the introduction and validation of mathematical methods for supervised anomaly detection demonstrate the author's ability to merge theoretical concepts with practical applications, a trait that is essential for translational research. The successful application and validation of these methods to CRISPR off-target recognition further testify to the thesis's significance and potential impact on the field.

Evaluation of the Quality of Publications

The candidate's publication record is commendable, with articles in reputable journals that testify to the rigor and significance of the research conducted.

The first publication in "Nucleic Acids Research" is noteworthy. NAR is a prestigious journal in the field of molecular biology, known for its stringent peer-review process and its emphasis on high-quality, innovative research. The topic of the publication, which deals with the evaluation of Cas9–gRNA and Cas12a–gRNA specificity using Deep Kernel Learning, is highly relevant to the current advancements in genome editing. The co-authorship with esteemed scientists such as E. V. Koonin further underscores the significance and quality of the research.

The second publication in "Scientific Reports" highlights the candidate's versatility and ability to tackle complex computational problems. "Scientific Reports" is a well-regarded, multidisciplinary journal that emphasizes sound research. The work on measuring internal inequality in capsule networks demonstrates the candidate's depth in machine learning and its applications to anomaly detection.

In summary, the quality of the candidate's publications is high, reflecting a commitment to impactful and rigorous research. Both journals are respected in their domains, and the range of topics covered showcases the candidate's multifaceted expertise in both molecular biology and computational research.

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

I have found some minor issues that I hope could be fixed or justified during defense.

- It appears that some captions (both tables and figures) are not finished with ".". I recommend to always cast them as a sentence and hence finish the sentence to avoid the inconsistency.

- On multiple places, after a bold words follows a new sentence that sounds uncomplete. E.g. Page 21. Last paragraph, sentence after "Inference".
- Page 26, 2nd sentence after the "Rule-based systems" should be reformulated
- Figure 2-6. This Figure is from [77]. I would suggest to clearly make a note of this by explicitly saying that this is not motivated by [77] but is the copy from it. E.g. write (image source [77]).
- Section 2.6. you accidently added an extra line in your LaTeX source that created an intent.
- Figure 2-7 – add (image source [4]) in caption.
- Page 32 – after stating Uncertainty Quantification add (UQ).
- Section 2.7 – rewrite (for example here [115], here [116] and here [117]).
- I believe (4.4) is incorrect.
- Page 50. Are in the definition of $P(Y=C_j)$ missing norms for v_j and v_ks'?
- Page 50. To which "Supplementary note" are you referring?
- Page 52. Change networks to layer in *LSTM layer (Long Short Term Memory networks).*
- Page 54. I feel you should remove the empty line in your LaTeX source above the equation for $O_i(X)$. Also, I would define Sigmoid, BatchNorm and Linear as operators in LaTeX.
- Page 55. Align equations.
- Page 62. Remove "below" from "is shown on the Fig 5-1 below".

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*