Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Dmitry Shadrin

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Data-driven modeling of plant growth dynamics in controlled environments

Supervisor: Professor Maxim Fedorov

Name of the Reviewer:

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** **Date: 26-09-2020**

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

- The relevance of the topic of dissertation work to its actual content

The thesis of Dmitry Shadrin on Data-driven modeling of plant growth dynamics in controlled environments addresses topics with high relevance for both – scientific progress as well as societal challenges. The thesis picks up a global trend to utilize mathematical methods of computer vision, machine learning and statistical analysis to gain more quantitative approaches to plant/ crop production. The overall target to improve crop production for a growing population with less resource input and lower environmental footprint and risk from the environment (e.g. toxicity form soils and soil additives) is of utmost importance. The technology component is an essential one in a holistic view, which should however also include responsible use (e.g. avoiding waste) as well as addressing consumption patterns in production, processing and with consumers.

- Brief evaluation of the thesis quality and overall structure of the dissertation.

The thesis is clearly on the technological side of the solution of the challenges – targeting methods development and testing relevant fort digital agriculture/ farming as well as precision phenotyping for breeding with some extension towards high-capacity phenotyping.

The thesis is clustered in several sections:

The introduction, first, provides valuable information on the objectives, the structure and important co-authorship statements. I would like to thank the candidate for this very appropriate and thorough statements. I am reporting in many interdisciplinary PhD theses: acknowledging in such detail and soundness the own contribution and the contribution of others should become standard: Especially in highly interdisciplinary work it is absolutely essential to frame own contribution in the team effort – in this context it is not a weakness, but a strength to show that one has not done everything alone, which the thesis reports on. This also gives the opportunity to the reviewers to lower the strictness in those parts, where the candidate is not the central expert in a study.

The introduction also contains an interesting and swell researched review on global trends in agriculture. The quality of this review is very good and the depth of the analysis indicates that the candidate has deeply emerged himself into a field, which is not his home turf, but providing a highly relevant opportunity to use CV, ML and IT-technologies at large. This review already includes a quantitative analysis of publication and funding activities and shows nicely the growing importance of the topics. The candidate tries to bring these numbers into an economic frame by indicating financial values. From my own experience this is a sensitive issue, as financial scores are often not easily comparable and numbers often not available. Nevertheless, the approach is valid and delivers – within the indicated limitation – quantitative data for a process that I can validate from working in tis field in the last 20 years.

The background section focusses rather on the IT-related aspects – namely a section in control systems ("internet of things", computer vison and ML and on modelling of plant growth. The first section quickly converges to artificial growing systems (IoT-approaches seem to be a bit artificial terminology here). These technologies today often include state-of-the-art controllers and concepts. The much longer and more in-depth description of the development and role of CV and ML to analyze plant growth contains aa good overview of technologies available in literature to analyze growth of plant parts/ organs and entire plants. It is applaudable that the candidate here covers a wide range of methods, but always targets for useability, which is important to achieve the goals set at the beginning of the thesis. The overview also addresses the robustness/ useability of models to describe plant growth. The candidate also reports about publications that promised improved performance due to ML and introduces different ML approaches and their benefits and disadvantages. In each section the candidate identified major limitations – obviously "promising" that he wants to address these in this thesis. The section on germination covers specifically the high-tech parts of analysis of early root development (remark: the wide range of germination assays, which would also urgently need improvement are only addressed superficially. However, I do not want to count this to negatively, as the width of the approach is already very wide. The same applies to the coverage of environmental

monitoring systems. I like the development of the text, always leading back to ask for the simplest solution needed and not going for the most high-tech one first. Finally, the text addresses pathogen detection. Here again the text can only be (a bit) superficial since the wealth of methods parallels the diversity of pests and pathogens in plants. The last section addresses deterministic models of growth and biomass dynamics. This section could also be massively extended if the candidate would target for a deep review of models. However, given the target of an engineering approach, the depth is sufficient to deduce relevant understanding of the growth processes.

The first experimental chapter addresses the experimental setup for growth plants and to collect data, Kalman – filters for simplification, various segmentation modes and integrated approaches to deliver minimal approaches to achieve usability. The growth setup is a classical approach to follow leaf and plant growth over time, which delivers the required parameters. Models were developed and tested to improve predictive power in an application-oriented mode. Again, I like the reductionistic approach to reduce complexity to the parameters needed instead of achieving a full understanding. However, it is important to realize that this includes the risk that the range to which predictions can be applied is limited.

The 4th chapter address plant growth models fed by (mainly temporal data) using RNNs for this the experimental setup was extended to also deliver nutritional data of the nutrient solution. Here a flaw is that the figures indicate "leaf area", while this seems to be projected leaf area, since leaves do not shrink that massively in the diurnal cycle. Here clearly movements affect the measurement, but at the same time enrich the data set. Here I would have hoped for a more in-depth analysis of the results. The text indicates that the candidate understands the restrictions of the system (p 107 middle), but a more thorough terminology throughout the text would have been favorable. The predictive power obviously also depends on the richness of the signal. The title of the section "CV in industrial scale experiments" seems to be a bit exaggerated; however, it is applaudable that the candidate moves into setups that are relevant for industry. Predications of early phases of growth are well targeted. Seed germination assay adds yet another dimension to the thesis. To evaluate the performance of CNNs and set up another experimental system. The results are promising, but useability needs to be shown in real high-throughput approaches. The candidate then extends the analysis into early detection of pathological responses of plants using spectral analysis again the results are promising, but the applicability for the extremely wide range of plant disease would need to be addressed. Chapter 5 adds another application targeted on phytotoxicity (termed "environmental parameters"), spatially explicit water quality and on testing phytotoxicology in petroleum -contaminated soils as well as the analysis of phosphogypsium on soil and plant performance. While each of these issues/ questions has a high relevance, chapter 5 gives the impression of a collection and less of targeted approach (I would have recommended to lower the number of practical examples and instead deepen some analysis of e.g. plant growth. This would have strengthend the case of the thesis).

The conclusion section is rather a summary of the most important results and indicates the value of using data-intensive approaches.

I have covered most of the following aspects already in the above evaluation. This I add only some short comments on these specific issues.

- The relevance of the methods used in the dissertation

The thesis showcases the power of CV, ML and NN for analysis of data-intensive applications in plant sciences and digital agriculture. The diversity of the methods is large. Not knowing about the habits in the Russian systems, I would have reduced the number of application examples, and rather had focused on one plant-based application system, where all the methods can be addressed.

- The scientific significance of the results obtained and their compliance with the international level and current state of the art

The candidate ´has a very thorough understanding and mathematical toolset, which he can develop, adapt and more relevant use in various applications. This is impressive. The candidate has shown the capability to perform at international level. The thesis is well placed and targets internationally relevant approaches. On an international scale there is a urgent need to scientists with the capability to bridge between sound mathematics and application. The candidate has shown his ability to contribute significantly.

- The relevance of the obtained results to applications (if applicable)

See comments above,

- The quality of publications

It is hard for me, being a plant scientist, to judge the publication quality in engineering-oriented journals. However, the Journals are international standard. Co-authored publications addressing applications in plant sciences are also in IEEE journals. However, there are two publications in J Soils and sediments and Plat Methods, respectively, which are well received in the application fields.

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

The only issue that I consider worth correcting is the terminology with respect to leaf area and projected leaf area. As mentioned above, the candidate addresses this issue in the text, but an earlier indication to the reader and correct naming of the parameters in the figures would be approprioate.

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*