

Gliwice, 20.06.2023

Dr hab. Monika Pietrowska

Narodowy Instytut Onkologii im. Marii Skłodowskiej - Curie - Państwowy Instytut Badawczy

Oddział w Gliwicach

ul. Wybrzeże Armii Krajowej 15

44-102 Gliwice

tel. (32) 278 96 72

e-mail: monika.pietrowska@gliwice.nio.gov.pl

Review Report on PhD Dissertation of Aleksandra Suwalska

entitled "Developing a system of automatic identification of cellular subpopulation in data from single-cell mass cytometry with the use of algorithms for grouping of high dimensional data"

completed in the Faculty of Automatic Control, Electronics and Computer Science at the Silesian University of Technology,

under the supervision of **Joanna Polańska, PhD, DSc**

The development of new technologies especially those generating large amounts of data creates a natural need to develop new methods for analyzing them or to implement existing solutions for analyzing this data. Mass cytometry represents one of the most advanced techniques for analyzing complex biological systems at the level of a single cell. Initially, studies using the mass cytometer were mainly concerned with analyzing the expression of antigens of cellular differentiation. With the development of this technology, it also began to be used to the study of the proteome, which allowed in-depth study of many complex biological systems, such as cancer tissues. Like flow cytometry, mass cytometry involves specific labeling of cellular antigens with monoclonal antibodies . However, unlike flow cytometry, these antibodies are conjugated to non-radioactive heavy metal isotopes (instead of fluorochromes), which are detected by mass spectrometry (instead of photomultipliers). The use of isotopes therefore makes it possible to overcome the limitations of autofluorescence or spectral overlap, and thus to detect more antigens than in flow cytometry. Due to the fact that mass cytometry is based on the use of stable metal isotopes as markers of individual biological individuals, today it is possible to simultaneously analyze more than 40 parameters at the level of a single cell. The dissertation submitted for evaluation proposes new solutions for technology that is not yet routinely available. Thus, the PhD student also used already published and available data (the works were published between 2015 and 2020) for her analysis.

The topic of the doctoral dissertation is up-to-date, highly relevant to contemporary medical and biological research, and relates to the analysis of data obtained with modern, developing technology with great potential.

In general, the dissertation is built on six main chapters: Introduction, Pre-gating of mass cytometry data, Batch effect correction analysis, Identification of cell subpopulations, Summary of the doctoral thesis, Supplementary materials (the whole length is 109 pages). The description of the thesis is detailed enough for a reviewer to be able to assess selection of methods for the assumed goals of the work, the obtained results and the conclusions drawn. The selection of references (59 items) is appropriate for the content of the work and is not limited only to the most recent works, but also includes previous source works. The doctoral dissertation is written in English. The layout of the work in this form is much more individual than traditional. The reviewer's attention was drawn to the very precisely executed figures. In my opinion, it contains all information necessary for preparation of this review report.

I have a few questions related to the content of the dissertation:

- (1) What error removal and lab artifacts does the PhD student have in mind in Section 1.4.7 Data Preprocessing? In particular, which ones are specific to data obtained by mass cytometry technique?
- (2) What is the superiority of the GMM method over commercial solutions? Has the PhD student performed such a comparison for any data set?
- (3) In addition to the number of cells in the data set, what parameters differed between the sets used to evaluate the algorithm for batch effect correction? What are the limitations of batch effect correction?
- (4) What was the variance of expression values for the same marker in each group for Tuberculosis dataset?
- (5) Which indexes and why does Ms. Suwalska consider the most important for evaluating the results of unsupervised analyses?
- (6) During the analyses, did the PhD student notice an effect of the number of cells on the results obtained?

In the opinion of the reviewer the Author's work contributed to the development of mass cytometry data analyses and is assessed very highly. A list of the PhD student's research works was attached to the dissertation and includes five publications. The PhD student participated in several projects (8 to be exact) and went on research fellowships (4 times). Achievements presented in this part of the work are related to the topic of the doctoral dissertation. It would be much more readable and consistent if the publications were presented independently without conference reports. Stand-alone presentations (lectures) have a different scientific value in the reviewer's opinion than poster presentations. In presenting her achievements, Ms. Suwalska could divide her work into such categories. The results have been published in good scientific journals. In the case of a PhD student, such achievements should be considered significant. I consider the achievements of the PhD student as very good at this stage of her academic career.

General conclusion

To sum up, Aleksandra Suwalska obtained results of significant value, the chosen direction is undoubtedly developed. In the opinion of the reviewer, the Author clearly described the obtained results in the presented dissertation, although quite briefly. Perhaps the short description of the results is due to the fact that the Author believed that she would repeat the information contained in the publications. Nevertheless, additional information, especially regarding the work done by the Author in the results section, would make the reviewer's work easier and leave no room for guesswork. The above minor remarks absolutely do not affect my high rating of the doctoral dissertation by Ms. Aleksandra Suwalska. The content of the dissertation proves the scientific maturity of Ms. Suwalska, her data analysis skills and fully justifies applying for a doctoral degree. Moreover, the dissertation represents an original solution to a scientific problem, demonstrates the candidate's general theoretical knowledge of the discipline and the ability to carry out scientific work independently.

To conclude, in my opinion the content and quality of the presented thesis fully match the requirement for a PhD dissertation, and indicate scientific maturity of the Candidate. Hence, I recommend admission of the dissertation to the public defense and I apply for a distinction for this doctoral thesis.

The doctoral dissertation meets the conditions specified in Article 13 of the Act of 14 March 2003 on academic degrees and academic title and on degrees and title in art (Journal of Laws No. 65, item 595 with later amendments).

Monika Pietrowska, PhD, DSc