GDAŃSKI
UNIWERSYTET
MEDYCZNY

www.gumed.edu.pl
58 349 11 11
info@gumed.edu.pl
ul. M. Skłodowskiej-Curie 3a,
80-210 Gdańsk

Gdańsk, 24 August 2026

Krzesimir Ciura, PhD, DSc
Department of Physical Chemistry
Medical University of Gdańsk
Faculty of Pharmacy
Al. Gen. J. Hallera 107
80-416 Gdańsk

Review of the doctoral dissertation of Karol Baran
entitled "Few-shot machine learning for multimolecular systems modeling"

The doctoral dissertation submitted for review, prepared by Karol Baran, MSc Eng, originated at the Department of Physical Chemistry of the Gdańsk University of Technology, and was supervised by Professor Adam Kloskowski. In the research carried out, the candidate undertook the development and application of machine learning methods, with particular emphasis on *few-shot learning* techniques, for the modeling and prediction of the properties of ionic liquids.

I regard the subject matter as valuable and timely on two independent levels. The first is set by the object of the research itself. Ionic liquids (ILs) are a promising group of chemical compounds owing to their negligible vapor pressure, wide liquid range, thermal stability, and two-component, ionic nature, which allows properties to be designed deliberately through the choice of the cation–anion pair. The same structural flexibility, however, is the principal obstacle to the rational selection of an ionic liquid for a specific application. The number of possible ion pairs far exceeds what can be covered experimentally, and structure–property relationships in multicomponent systems, additionally modified by temperature and pressure, do not lend themselves to simple additive descriptions. Reliable predictive models are therefore a prerequisite for the rational exploration of this chemical space.

The second level is methodological and extends beyond the chemistry of ionic liquids. The candidate addressed the problem of modeling low-count data, that is, a situation in which the number of available observations is orders of magnitude smaller than that required by typical deep learning architectures. This difficulty is common wherever data are produced by costly and time-consuming experiments, for example in medicinal chemistry, chemical safety assessment, or materials engineering.

Karol Baran, MSc Eng, set out to solve the following three research problems: (i) assessment of the limitations of classical machine learning and deep learning methods in

GDAŃSKI
UNIWERSYTET
MEDYCZNY

www.gumed.edu.pl
58 349 11 11
info@gumed.edu.pl
ul. M. Skłodowskiej-Curie 3a,
80-210 Gdańsk

predicting the properties of ionic liquids; (ii) determination of the actual usefulness and limitations of *few-shot learning* techniques in the modeling of multimolecular systems; (iii) development of modifications to existing algorithms and training protocols that improve model accuracy in selected applications.

The dissertation was prepared as a series of five thematically related publications, accompanied by an extensive commentary on the results obtained. The whole forms a coherent work of 172 pages, written in English. The theoretical introduction covers artificial intelligence and machine learning, the fundamentals of chemoinformatics, and the ways of representing molecular structure for modeling purposes. This is followed by a literature review, the aims and research questions, and the methodology used, together with the data sets, validation protocols, molecular representations, and algorithms applied. The next chapter synthesizes the results obtained in the individual publications of the series. The dissertation is completed by a list of 175 references and an appendix containing the full texts of the publications forming the series together with the author contribution statements.

The series comprises papers published in reputable journals: *The Journal of Physical Chemistry B*, *Science of the Total Environment*, *ACS Sustainable Chemistry & Engineering*, *Chemical Research in Toxicology*, and *Journal of Chemical Information and Modeling*. The leading role of Karol Baran is immediately apparent, since in each of the papers he is the first author and the corresponding author, and three of the five papers have only two authors. This is also confirmed by the author contribution statements, according to which the candidate was responsible for the conception, methodology, software, data analysis, visualization, and preparation of the original version of the manuscript.

In my view, the candidate's independence in obtaining funding deserves emphasis. Karol Baran, MSc Eng, is the principal investigator of a National Science Centre research project awarded under the PRELUDIUM 22 call (no. UMO-2023/49/N/ST5/01043, 2024–2027), and also the principal investigator of two computational grants (no. PLG/2024/017245 and no. PT00996). Securing his own funding at the doctoral stage testifies to research maturity and to the ability to run a scientific project independently.

Substantive assessment:

The first issue examined by the candidate was the assessment of the usefulness of classical machine learning methods for modeling systems involving ionic liquids when the amount of data is very limited. The analysis was carried out on two data sets of considerable

GDAŃSKI
UNIWERSYTET
MEDYCZNY

www.gumed.edu.pl
58 349 11 11
info@gumed.edu.pl
ul. M. Skłodowskiej-Curie 3a,
80-210 Gdańsk

practical importance: carbon dioxide solubility expressed as the Henry's law constant (76 records) and the yield of lignin isolation from biomass (about 110 records). In the first case, it was shown that including the fractional free volume (FFV) as a molecular descriptor markedly improves the quality of nonlinear models; for the MLP network, the coefficient of determination on the test set increased from 0.738 to 0.833. In the case of lignin, the candidate combined biomass composition, process parameters, and ionic liquid structure in a single model, obtaining R^2 in the range 0.63–0.77. The models were interpreted using Shapley values, the LIME method, and CN2 association rules, which made it possible to identify the factors governing both processes.

The second research problem was to establish how two-molecule systems should be represented in graph neural networks. The candidate compared three variants of graph construction and five partial charge models, showing that the best and most reproducible results are obtained when the ions are optimized separately and then combined into a single graph structure ($R^2 = 0.79 \pm 0.05$), and that the approach to geometry optimization has a greater influence on model quality than the choice of the method for determining charges. A non-obvious result was also obtained: for density and surface tension, models trained on raw data outperformed models based on data cleaned by the candidate, which the author attributes to the broader coverage of chemical space by the raw data set. This part of the work also revealed a fundamental limitation of deep learning, namely the size of the training set required, which justified turning to meta-learning methods.

The third issue was the assessment of the usefulness of few-shot learning and the development of a modification of the MAML algorithm. The candidate showed that with a small pool of tasks meta-initialization does not in practice differ from random initialization, and proposed a two-stage training protocol using data simulated by the COSMO-RS method, which raised R^2 from 0.518 to 0.643. The real advantage of meta-learning was demonstrated in the modeling of ionic liquid toxicity, where MAML retained its predictive ability with adaptation sets of 4–16 compounds, that is, in a range in which tree-based methods fail. On this basis, an original modification, Att-MAML, was proposed, in which only the layer scaling feature importance is updated in the inner loop. This modification clearly reduced the standard deviations of the metrics, decreasing the dependence of the result on the random choice of the adaptation set, which is of fundamental practical importance when only a few measurements are available.

The last challenge taken up by the candidate was to transfer the conclusions concerning task similarity to the modeling of sorption. Using a set of activity coefficients at infinite dilution, it was shown that model performance depends in a statistically significant way (Kruskal–Wallis test, $p = 0.04$) on the similarity of the test task to the meta-training tasks,

GDAŃSKI
UNIWERSYTET
MEDYCZNY

www.gumed.edu.pl
58 349 11 11
info@gumed.edu.pl
ul. M. Skłodowskiej-Curie 3a,
80-210 Gdańsk

approximated by the Tanimoto coefficient between solute molecules. In response, a second original algorithm, TSA-Reptile, was proposed, in which the loss function is scaled inversely to task similarity, so that atypical tasks are given greater weight during meta-training. The method proved more effective than MAML for five of the six distant tasks, which leads to the practical conclusion that the algorithm should be selected according to the position of the task relative to the training data.

I rate the results obtained highly, and all of them have already been verified by independent reviewers in reputable scientific journals. I consider the two original modifications of meta-learning algorithms and the systematic resolution of the question of the graph representation of two-molecule systems to be particularly valuable. It is worth emphasizing that the candidate has made the source code he developed available in a public GitHub repository, in accordance with the principles of open science, which allows the analyses carried out to be verified and reproduced. The English used by the candidate is of a high standard, and the figures prepared are clear and legible.

My reading of the dissertation prompted two questions, which I present below.

1. In Chapter 5, the author draws attention to the considerable variability of prediction results in the *few-shot* regime, expressed by high standard deviations of the metrics for adaptation sets of 2–8 compounds, and attempts to limit it through the original Att-MAML modification. At the same time, the results presented indicate that the advantage of a given method depends on the size of the adaptation set: meta-learning dominates below a certain threshold, above which Gradient Boosting proves more effective. This raises the question of whether the author considered the use of consensus modeling, combining the predictions of the meta-learning algorithms developed here with classical ensemble models such as Gradient Boosting. Averaging or weighting the predictions of models with a different error character could provide an alternative and computationally cheaper route to variance reduction than modifying the meta-learning algorithm itself.

2. Chapter 5.2 shows that graph neural network models trained on raw data outperform models based on data subjected to expert cleaning, which was explained by the broader coverage of chemical space by the raw data set. An alternative explanation of this observation, however, remains the greater size of that set. Was a comparison carried out with the number of records equalized, which would make it possible to separate the influence of structural diversity from the influence of training set size?

In summary, I assess the doctoral dissertation of Karol Baran, MSc Eng, very positively. The subject he has taken up is demanding, situated at the interface of ionic liquid chemistry and advanced machine learning methods, and the results obtained show that he has an excellent

GDAŃSKI
UNIWERSYTET
MEDYCZNY

www.gumed.edu.pl
58 349 11 11
info@gumed.edu.pl
ul. M. Skłodowskiej-Curie 3a,
80-210 Gdańsk

command of these areas. The candidate's overall output, five publications in reputable journals, in all of which he is the first author, together with research funding obtained independently from the National Science Centre, indicates that he has a very good command of research methodology and is prepared to conduct scientific work independently.

I state unequivocally that the doctoral dissertation of Karol Baran, MSc Eng, submitted to me for assessment, meets the requirements set for doctoral dissertations under Article 187 of the Act of 20 July 2018, Law on Higher Education and Science (Journal of Laws 2023, item 742), and I request that Karol Baran, MSc Eng, be admitted to the further stages of the procedure for the award of the doctoral degree. At the same time, in view of the high standard of the research carried out, I request that this doctoral dissertation be awarded a distinction.

Signature Not Verified
Dokument podpisany przez
Krzesimir Ciura
Data: 2026.08.24 14:15:26 CEST

KATEDRA I ZAKŁAD CHEMII FIZYCZNEJ
prof. dr hab. n. farm. Wiesław Sawicki | wieslaw.sawicki@gumed.edu.pl | al. gen. J. Hallera 107, 80-416 Gdańsk
kierownik - 58 349 12 79 , sekretariat (tel./faks) – 58 349 16 52 | chemfiz@gumed.edu.pl