Prof. dr hab. inż. Jarosław Polański
Centre for Materials and Drug Discovery
Institute of Chemistry, University of Silesia
ul. Szkolna 9, 40-006 Katowice

July 18, 2026

Review of the Doctoral Dissertation *Few-shot machine learning for multimolecular systems modeling* by Karol Baran

This work was carried out under the supervision of Prof. Adam Kloskowski, Ph.D., Eng., while the author was enrolled in the Doctoral School at Gdańsk University of Technology. The subject of the candidate's research is the modeling of biomolecular systems using artificial intelligence (AI) methods, in particular, the physicochemical properties of ionic liquids, including toxicity and sorption. As his main research problem, the candidate analyzes the quality and size of available datasets, especially those based on descriptors specific to a particular area of analysis. In this context, he compares various machine learning methods, including those he originally designed.

There is no doubt that artificial intelligence (AI) is transforming our perception of the role of computational approaches in science, technology, and society. Although the long-term implications of AI for everyday life remain to be fully evaluated, AI has already become firmly established within the scientific community as powerful and indispensable research tool. One need only mention the groundbreaking significance of the AlphaFold program, which solved the problem of protein folding. This is illustrated by two 2024 Nobel Prizes (in physics and chemistry). AI works well when we have large, high-quality datasets. Face recognition is one such example. There are over 8 billion people in the world. Digital representations of their faces are widely available. Similarly, we have large datasets for polypeptide protein structures. A good example of the limitations is the method of retrosynthesis, regarded as one of the most significant recent revolutions in chemistry (Seemann, "Revolutions in Chemistry: Assessment of Six 20th Century Candidates," JACS Au 2023, 3, 9, 2378–2401, https://doi.org/10.1021/jacsau.3c00278). Data describing the domain of chemical reactions is fraught with numerous errors, and the dataset is not large enough to compensate for them. It turns out that in such a case, the solution may lie in the preliminary preparation of data by a chemist (a human), who encodes so-called reaction templates, thereby eliminating errors. Only then is this data processed by a computer. This approach by the Polish group led by Prof. Grzybowski has proven more effective than fully autonomous data processing by a computer.

As the research goal, the doctoral student has set out to study AI learning schemes for which only a small number of examples are available (few-shot learning). As examples for such learning, he uses diverse datasets covering physicochemical properties, toxicity, and sorption in ionic liquids, defining these problems as the modeling of bimolecular systems.

On page 19, the author presents the objectives of the thesis, which are:

G1. To evaluate the limitations of both classical methods and deep learning methods in the context of predicting the properties of ionic liquids (ILs).

G2. To identify operational limitations associated with the implementation of "few-shot" learning techniques (a small number of examples).

G3. Indicating modifications to existing algorithms or learning protocols aimed at significantly improving model accuracy in specific problem domains.

The main hypothesis of this thesis is that both conventional and deep machine learning methods are useful for modeling multi-molecular systems with limited data availability, provided they are properly designed, especially when using meta-learning techniques.

It formulates the following research questions (p. 21):

(1) What is the effectiveness and significance of combining classical and deep learning methodologies in addressing contemporary challenges in chemical research, such as carbon dioxide absorption and lignin isolation? (2) To what extent does the incorporation of physics-based rule and descriptor exploration enhance the predictive power of machine learning models with respect to biomolecular systems? (3) What are optimal strategies for GNN (graph neural network) models to achieve maximum performance in predicting the properties of ionic liquids (ILs)? (4) How do the quantity and quality of available data affect the predictive accuracy and robustness of GNN models when predicting the properties of ionic liquids (ILs)? (5) How does "few-shot" meta-learning compare to traditional machine learning paradigms in the context of modeling complex bimolecular systems? (6) Is the use of low-quality simulated data for the initial training of meta-learning models a viable strategy to address data scarcity in low-data-volume learning scenarios? (7) What are the inherent limitations of meta-learning-based approaches with respect to task similarity? (8) In what way does the similarity of chemical spaces across tasks limit the applicability and performance of meta-learning models? (9) Can modifications to the adaptation phase (inner loop) of meta-learning serve as an effective strategy for improving performance and reducing the standard deviation of metrics, particularly in the selection of the training set?

Formally, the dissertation is an extensive commentary on the series of publications listed on page 61:

A1. **Baran, K.***, & Kloskowski, A. (2023). Graph Neural Networks and Structural Information on Ionic Liquids: A Cheminformatics Study on Molecular Physico-chemical Property Prediction. Journal of Physical Chemistry B, 127, 10542-10555. doi.org/10.1021/acs.jpcb.3c05521
A2. **Baran, K.***, Barczak, B., & Kloskowski, A. (2024). Modeling lignin extraction with ionic liquids using machine learning approach. Science of the Total Environment, 935, 173234-173248. doi.org/10.1016/j.scitotenv.2024.173234
A3. **Baran, K.***, & Kloskowski, A. (2025). Unlocking Structural Insights into CO2 Absorption with Ionic Liquids: A Comparative Study of Machine Learning, Association Rules, and Meta-Learning for Modeling Henry's Law Constant. ACS Sustainable Chemistry & Engineering, 13, 12805-12817. doi.org/10.1021/acssuschemeng.5c06122

A4. **Baran, K.***, Derron, T., Eichenlaub, J., & Kloskowski, A. (2026). Few-Shot Prediction of Toxicity of Ionic Liquids Supported by Attentive Model-Agnostic Meta-Learning. Chemical Research in Toxicology. doi.org/10.1021/acs.chemrestox.6c00022
A5. **Baran, K.***, & Kloskowski, A. (2026). Exploring Chemoinformatics Aspects of Few-Shot Meta-Learning by Example of an Infinite Dilution Activity Coeffcient in Ionic Liquid Prediction. Journal of Chemical Information and Modeling. doi.org/10.1021/acs.jcim.6c00067

Chapter 1 (Introduction) provides a theoretical introduction to chemoinformatics and AI methods, helping the reader understand the issues addressed in this thesis. Chapter 2 (Related Works) situates the thesis's subject matter within the context of the literature. Chapter 3 (Scope and Goals of the Thesis) presents the objectives, hypotheses, and research questions. In Chapter 4, the doctoral candidate discusses the methods and datasets used in conducting his research.

Chapter 5 presents a discussion of the results in relation to subsequent research, in which the doctoral candidate formulates problems by relating them to a series of studies that do not discuss the A1-A5 publications directly, but rather to an arbitrarily selected set of issues. In Chapter 5.1, for example, the doctoral candidate discusses the prediction of properties of datasets with a small number of objects using various machine learning methods. The data analyzed are the DS1 dataset (CO_2 absorption in ionic liquids) and the DS2 dataset (lignin isolation from biomass using ionic liquids). Both datasets are relatively small data reservoirs. For DS1, the molecular descriptor used is the Fractional Free Volume (FFV), which can also be estimated using molecular dynamics simulations. Various classical machine learning methods are employed to model the D1 and D2 datasets. Tables 5.1 and 5.2 present cross-validated statistics, including CV R^2, CV, RMSE, and R^2 test. Table 5.3 shows, for example, the prediction statistics for Henry's law constant. Generally, the cross-validated constants range from approximately 0.4 to 0.8. The results presented in Chapter 5.1 are referred to publications A2 and A3.

Chapter 5.2 specifically discusses the challenges of applying graph neural networks to the optimization of bimolecular systems, including, in particular, the impact of methods for encoding bimolecular ionic liquid systems as graphs, with three scenarios (A, B, C) involving separate or joint optimization.

Chapter 5.3 is devoted to meta-learning applications. Meta-learning is the concept "learning how to learn." The model learns based on a wide variety of tasks. We believe this will allow it to adapt quickly to different tasks. The author takes a similar approach by presenting a two-stage meta-learning protocol for similar tasks (e.g., dissolved substances vs. toxicity indices) and adapting it to the task of interest (the test task). This method was applied in preliminary studies on the DS1 dataset. The task was defined as predicting the solubility of a specific solute (in the case of the DS1 dataset—gases) in ionic liquids. The dataset for meta-learning contained 84 data points across 15 different gases. The results are presented in Table 5.6 (publication A3). An interesting discovery is that good models can be obtained even though the meta-learning dataset lacked data points for the molar fraction of CO_2 in ionic liquids. This result represents an intriguing discovery, suggesting that the solubility of new gases can be successfully modeled even with limited data by using meta-learning, i.e., learning from other data, known as Agnostic Meta-Learning (MAML).

Chapter 5.4 describes meta-learning problems using an attention mechanism for toxicity prediction, and Chapter 5.5 discusses meta-learning that accounts for task similarity (Task Similarity Aware Meta-Learning).

The analysis presented above provides only a brief overview of the problems examined by the doctoral candidate. In general, the dissertation addresses a significant problem in computational chemistry: optimizing machine learning and deep learning models under limited data availability. On the one hand, the author uses ionic liquids as a model for testing AI methods. On the other hand, he refines and investigates chemoinformatics and the physical chemistry of ionic liquids. By analyzing deep learning architectures and meta-learning strategies, he demonstrates that good learning results can be achieved even with a limited amount of training data. One of the key findings is the demonstration that it is possible to propose a molecular descriptor that improves the quality of nonlinear models used to predict the solubility of carbon dioxide, and through the use of association rules, it was revealed that there are specific combinations of functional groups that favor both microscopic (free volume) and macroscopic (Henry's constant) properties describing sorption. Importantly, the Ph.D. candidate's research confirms the concept that association rules provide reliable guidance regarding favorable ion pairs that promote solubility. This is one of the conclusions that can be interpreted directly in the field of chemistry. It is an important conclusion because the methods presented may influence future research in areas where the interpretation of models will be an effective way to design new ionic liquids.

I hold this dissertation in very high regard. It is a well-structured and engaging work that makes for a compelling read. This study addresses the most current issues in computational chemistry as they relate to ionic liquids, which are currently of great experimental interest. The author demonstrates a deep understanding of the methodologies of AI and chemoinformatics. All analyses are supported by correctly applied statistics, cross-validation, etc. The work required conducting a vast number of computer simulations, programming procedures, and data analysis. Also of great significance are the author's scientific maturity and creativity in selecting AI methods, as well as the ability to adapt them and develop innovative approaches. The publication of the obtained results in renowned journals also deserves high praise here.

One drawback of the paper is its complexity. It contains a very large amount of data and is difficult for the reader to comprehend. Structuring the discussion directly in parallel with the publications would make it easier for the reader. However, this is merely a subjective impression. From a reviewer's perspective, it is not possible to analyze all the presented analyses in detail. Therefore, I would like the author to consider how the various methods used in the paper fit into the context of the problem defined as feature learning versus feature engineering. This is particularly interesting in the context of meta-learning. Would it be possible to apply such methods, for example, in retrosynthesis, which is still not sufficiently effective when using purely feature learning methods? I look forward to an interesting discussion during the defense.

The dissertation unquestionably merits an excellence award. The author applies the most up-to-date AI methods to predict the properties of ionic liquids, adapting them to modern computational chemistry of these biomolecular molecules. Innovatively, he describes how the bimolecular structure of an ionic liquid can be encoded in such systems. The work contains many interesting results. Of particular interest is the intriguing discovery that meta-learning—where a model learns from multiple tasks and adapts to other tasks—works for different models of ionic liquid behavior, e.g., dissolved substances versus toxicity indicators. These are extremely interesting problems, the understanding of which still lies ahead of us. The results presented by Mr. Baran significantly shape the landscape of computational methods available for the design of ionic liquids in modern chemistry. The doctoral candidate's research confirms that association rules provide reliable guidance on favorable ion pairs that

promote solubility. This is one of the conclusions that can be interpreted directly within the field of chemistry. This conclusion is important because the methods presented can significantly influence the design of new ionic liquids.

In conclusion, the results presented in this dissertation constitute a significant contribution to the development of AI protocols for modeling ionic liquid properties. I confirm the dissertation demonstrates a clear correspondence between the stated research objectives, the methodology employed, and the conclusions drawn. I conclude that Mr. Karol Baran meets all the conditions specified in the paragraph 187 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2020 r. poz. 85 z późniejszymi zmianami). I therefore recommend that Mr. Karol Baran be admitted to the subsequent stages of the doctoral degree proceedings.[1]

Dokument
podpisany przez
Jarosław Polański
Data: 2026.07.20
07:32:43 CEST

[1] The original review was written in Polish. DeepL was used for translation, and Grammarly was subsequently applied to refine the translated text.