The Advanced Machine Learning for Innovative Drug Discovery (AIDD) project

Newsletter 5, February 2023

The Advanced Machine Learning for Innovative Drug Discovery (AIDD) project is a Marie-Skłodowska-Curie Innovative Training Network (ITN) for Early Stage Researchers (ESRs) funded by the European Commission under the Horizon 2020 Programme, <u>Marie Skłodowska-Curie grant agreement No 956832</u>. The project brings together fifteen academic and industry beneficiaries from ten European countries and the University of British Columbia (Canada) to train sixteen PhD students in close collaboration with partners from the USA, Australia, China, Israel, and other countries.

Project development

The AIDD project started on 01.01.2022 and will finish at the end of March 2025.

Third AIDD School - Autumn School on Machine learning in chemogenomics

The 3rd AIDD school was organised in Leuven, Belgium. The city has been a university city since 1425, and has beautiful architecture and vibrant student life. The event took place in the building of the <u>Irish College Leuven</u>, the campus of which dates back to the 17th century. The students were provided with an option to book their accommodation in the venue to encourage and facilitate interactions between the ESRs.

We were happy to welcome in Leuven physically nearly all of our speakers including five our PIs as well as invited speakers from Belgium, the Netherlands, UK, Israel and Hungary. All lectures except for those on the day of the company visit were also streamed via Zoom, and were available for the general public.



Figure 1. The Irish College, Town Hall and Leuven Fonske.

The School started on the 17th October. After a welcome from the local organisers Adam Arany and Yves Moreau, the scientific program was started with a two hour lecture on sampling methods in Bayesian inference by <u>Christel Faes</u> from the University of Hasselt, followed by an introduction to Variational Inference by <u>Adam Arany</u>. In the afternoon <u>Jaak</u> <u>Simm</u>, quantitative scientist of G-Research, London gave an interesting talk about pitfalls of different time-to-event modelling approaches. We closed the afternoon with a welcome reception to give space for the first social interactions, and provide an opportunity to get a sample from the famously diverse beer scene of Belgium.

The second day of the school (19th October) was opened with a talk by <u>Sara Magliacane</u> from the University of Amsterdam about causality-inspired machine learning focused on unsupervised domain adaptation problems. Subsequently, <u>Timur Madzhidov</u> from Reaxis presented about Artificial Intelligence and big data in synthetic chemistry, followed by an introductory explanation by Adam Arany about the topic and general task of the team assignment which took place the following week (see below). After the lunch break, <u>Yves</u> <u>Moreau</u> (KU Leuven) addressed, in his insightful talk about Ethics in AI, the potential beneficial and detrimental impacts of research on society and politics as well as the responsibility of scientists for their research developments by means of historic and more recent showcases. The day was closed by <u>Johan Suykens</u> from KU Leuven with a presentation about neural networks and kernel machines.

On the 20th, the fellows and present PIs were invited to visit the site of Janssen Pharmaceutica in Beerse, in northern Belgium. After a 90 minute bus trip, we attended an introductory presentation about the company's history followed by a private tour of the facilities, including the <u>Johnson & Johnson Innovation – JLABS</u>.

The following afternoon was packed with lectures by scientists from Janssen, giving insights into research conducted at the company. First, <u>Wouter Heyndrickx</u> talked about the recently finalised <u>EU Project Melloddy</u>, which involved various partners from industry focusing on

developing enhanced predictive deep learning models on decentralised data for federated and privacy-preserving ML. <u>Seong Joo Koo</u> continued with a lecture on High Content Imaging in drug discovery, followed by a talk by <u>Steffen Jaensch</u> about the new JUMP-Cell Painting dataset which is the result of a joint effort of various industrial and academic patterns to publish a large dataset of cell painting assay data. Due to the intention of several AIDD fellows to use this dataset in their projects, this talk was of particular interest to the fellows. <u>Ramil Nugmanov</u>'s talk about Bidirectional Graphormer for Reactivity Understanding Neural Network was the last presentation on Wednesday before the fellows and present PIs headed back to Leuven.

The next day (21st October) continued with official meetings (Supervisory Board and General Assembly) of the AIDD consortium which took place in the morning. In parallel, Bence Bruncsics from Budapest University of Technology and Economics presented his work on models for multivalent proteins focusing on binding kinetics in a system of differential equations. Srijit Seal from University of Cambridge continued with a presentation about cell image data processing, followed by Günter Klambauer's talk about Bayesian Deep Learning that aligned nicely with the main topic of the school. After lunch, the lecture series was continued with a presentation from Lewis Mervin from AstraZeneca about how drug design can benefit from federated learning and uncertainty prediction as well as molecular properties prediction approaches within the Design-Make-Test-Analyse (DMTA) cycle at AstraZeneca. In the later afternoon the fellows presented their research within the framework of a poster session.. This event turned out to be a valuable opportunity for the fellows to collect and exchange ideas for their projects with colleagues and PIs, and to update each other on their research. The evening was completed with a formal dinner at the Faculty Club providing a wonderful opportunity to network with fellow students and PIs in a loose atmosphere.

<u>Hanoch Senderowitz</u> from Bar-Ilan University, the project's associated partner, opened Friday's lecture series with a talk about Machine Learning in materials informatics. Before the lunch break, <u>Hugo Ceulemans</u> from Janssen Pharmaceutica talked about new training data sources for Drug Discovery Machine Learning. Friday afternoon was fully dedicated to discussing responsible conduct of research. For this purpose, <u>Marcel van der Heyden</u> from Utrecht University was invited to talk about the importance of correct practices in science. The fellows were motivated to think and discuss correct and improper practices in the research process.

After an exciting but also exhausting week full of talks, we started into the well-deserved weekend with a trip to the beautiful city of Gent. On Saturday morning we took a train to Gent where we first visited the Stadsbrouwerij Gruut, a small, local Beer Brewery in the centre of Gent. After a short introduction to the brewery, the fellows and present PIs were invited for a beer tasting and to offered herb and spice extracts which they used to adjust the beer to their liking in an informative workshop. Subsequently, we admired the beautiful architecture of Gent exploring the city first by boat and then on foot.

On Monday, 25th a hackathon was organised, during which the fellows worked on an assignment in small teams. The assignment was a distribution shift problem, applied to bioactivity data. At the end of the day, the approaches chosen by the groups and their results were compared. During the hackathon, fellows with various scientific backgrounds worked

together, providing a great opportunity to benefit from each other's experience and to exchange knowledge.

On the morning of the last full day of the AIDD school (26th October), <u>Martijn Oldenhof</u> from KU Leuven gave a talk about the intersection of Optical Chemical Structure Recognition (OCSR) and object detection. After a short coffee break <u>Mark Embrechts</u> continued with a presentation on how to infer missing data with auto-associators. The last long talk of this school was dedicated to the problem of improving the effectiveness of compound design, which was given by Marie Ledecq from <u>UCB</u>. The AIDD school concluded with an afternoon filled with pitches from scientists from the bioinformatics group at KU Leuven giving insights into the various research fields addressed in this research group.

The detailed list of all lecturers, as well as lecture slides, can be found at: <u>https://ai-dd.eu/lectures</u>.



Figure 2. AIDD Fellows on their weekend trip to Gent including a brewery visit and a boat trip though the historic city centre.

Voices about the school:



"I was really satisfied by the 3rd AIDD School at the Irish College of Leuven both in terms of organisation and content of the lectures. As a doctoral researcher interested in Bayesian machine learning methods, I particularly liked to learn more about this field, both from the theoretical and practical point of views. The content of the lectures was still diverse as we also had the opportunity to learn about new interesting topics such as the ethics of doing research with biological data, as well as to stay informed about most recent research in the field of machine learning for drug discovery. As in previous schools, all lectures were followed by valuable discussions and networking opportunities with the invited speakers. This time we also had the chance, we ESRs, to discuss and share our research ideas and latest advances during a successful poster session, and to work collaboratively on a problem solving exercise during a hackathon. Overall, the organising committee did a great job and I think we all agreed that it was a great experience full of new learnings, new ideas and, potentially, new collaborations!" - Yasmine Nahal, AIDD fellow, AstraZeneca



"As a member of the KULeuven side organising research group, I considered the invitation to join the school an excellent networking opportunity and to get a deeper insight into the biochemical component of my PhD. research. During the week, I experienced a warm and welcoming milieu of young professionals, who are always ready for discussion and have an open mind to the world. The broad variety of topics covered during the professional program, the buzzing atmosphere of the poster session, and the determination of the participants during the hackathon astonished me. I'm grateful to have been able to be a part of this experience."

- Andras Formanek, PhD at KU Leuven

Next AIDD School in Helsinki, March 20-23 will be organised as a hybrid meeting. Zoom access of external participants to selected presentations is currently planned and will be announced on the web site of the project (see <u>https://ai-dd.eu/news</u>).

AIDD won a bid to ENNS (<u>https://e-nns.org</u>) and will co-organise ICANN2024 (ENNS) at Lugano, Switzerland, September 17-20 2024.

Meet AIDD partners at (see continuously updated list at <u>AIDD web site</u>):

- <u>IPAM workshop "Learning and Emergence in Molecular Systems"</u>, 23-27 January, Los Angeles, USA, **AIDD participant:** <u>Alexandre Tkachenko</u>
- <u>Royal Society conference on Machine learning and AI in biological science, drug discovery</u> <u>and medicine</u>, 1st March, London, UK, AIDD participant: <u>Djork-Arné Clevert</u>
- APS March meeting, 5-10 March, AIDD participant: Leonardo Medrano
- <u>Bayes Comp 2023,</u> 5-17 March, Finland, Levi, AIDD participant: <u>AIDD partners</u>
- <u>AIDD School.</u> 20-23 March, Helsinki, Finland, AIDD participant: <u>AIDD partners</u>
- <u>ELRIG Research and Innovation meeting: The rise of AI driven drug discovery session</u>, 29-30 March, Genome Campus Conference Centre, Hinxton, UK, **AIDD participant:** <u>Lewis Mervin</u>
- <u>Munich Interact Conference</u>, 30-31 March, Munich, Gemany, AIDD participant: <u>Peter Hartog</u>, <u>Paula Torren</u>
- International Conference on Artificial Intelligence and Statistics, 25 27 April, Valencia, Spain
- <u>The Eleventh International Conference on Learning Representations</u>, 1-5 May, Kigali, Rwanda, AIDD participant: <u>Adam Aramy</u>, <u>Djork-Arné Clevert</u>, <u>Günter Klambauer</u>, <u>Ana</u> <u>Sanchez-Fernandez</u>, <u>Emma Svensson</u>, <u>Alan Kai Hassen</u>
- <u>Helmholtz conference in Drug Discovery</u>, 15-16 May, Braunschweig, Germany, AIDD participant: <u>Ola Engkvist</u>
- Pint of Science Munich, 22-24 May, Munich, Germany, AIDD participant: Paula Torren
- <u>QSAR 2023.</u> 5-9 June, Copenhagen, Denmark, **AIDD participant:** <u>Igor Tetko. Günter</u> <u>Klambauer, Ana Sanchez-Fernandez, Emma Svensson, Alan Kai Hassen, Varvara</u> <u>Voinarovska</u>
- Hybrid Human Artificial Intelligence (HHAI), 26-30 June, Munich, Germany
- <u>AIDD School,</u> 26-30 June, Mölndal, Sweden, **AIDD participant:** <u>AIDD partners</u>
- <u>Gordon Conference in Computer Aided Drug Design</u>, 16 21 July, Vermont, USA, **AIDD** participant: <u>Floriane Montanari</u>, <u>Ola Engkvist</u>
- Intelligent Systems for Molecular Biology / European Conference on Computational Biology, 23-27 July, Lyon, France, AIDD participant: Adam Aramy
- <u>The Fortieth International Conference on Machine Learning</u>, 23-29 July, Hawaii Convention Center, USA, <u>Adam Aramy</u>
- <u>ELLIS Machine Learning for Molecules Workshop. Summer.</u> on-line, AIDD participant: <u>Günter Klambauer. Ana Sanchez-Fernandez. Emma Svensson</u>
- JKU Public PhD Seminars. TBA. on-line, AIDD participant: Ana Sanchez-Fernandez. Emma Svensson
- <u>ACS Fall, Harnessing the Power of Data.</u> 13-17 August, San Francisco, CA, USA, **AIDD** participant: <u>Igor Tetko</u>
- <u>12th World Congress on Alternatives and Animal Use in the Life Sciences</u>, 27-31 August, Canada, AIDD participant: Igor Tetko
- <u>ICANN2023,</u> 23-26 September, Cyprus, <u>Igor Tetko</u>
- <u>17. International Conference on Learning Representations (ICLR)</u>, 2-3 December, Sydney, Australia, **AIDD participant:** <u>Jürgen Schmidhuber</u>
- <u>NIPS 2023,</u> 10-16 December, New Orleans, USA, **AIDD participant:** <u>Adam Aramy, Günter</u> <u>Klambauer, Alan Kai Hassen</u>

Project publications with participation of AIDD fellows:

• Simm, J.; Arany, A.; De Brouwer, E.; Moreau, Y. Expressive Graph Informer Networks. In Machine Learning, Optimization, and Data Science; Nicosia, G., Ojha,

V., La Malfa, E., La Malfa, G., Jansen, G., Pardalos, P. M., Giuffrida, G., Umeton, R., Eds.; Springer International Publishing: Cham, 2022; pp 198–212. <u>https://www.springerprofessional.de/expressive-graph-informer-networks/20089896</u>

- Roland, T.; Böck, C.; Tschoellitsch, T.; Maletzky, A.; Hochreiter, S.; Meier, J.; Klambauer, G. Domain Shifts in Machine Learning Based Covid-19 Diagnosis From Blood Tests. J. Med. Syst. 2022, 46 (5), 23. <u>https://doi.org/10.1007/s10916-022-01807-1.</u>
- Klotz, D.; Kratzert, F.; Gauch, M.; Keefe Sampson, A.; Brandstetter, J.; Klambauer, G.; Hochreiter, S.; Nearing, G. Uncertainty Estimation with Deep Learning for Rainfall–Runoff Modeling. Hydrol Earth Syst Sci 2022, 26 (6), 1673–1693. https://doi.org/10.5194/hess-26-1673-2022.
- Vall, A.; Sabnis, Y.; Shi, J.; Class, R.; Hochreiter, S.; Klambauer, G. The Promise of AI for DILI Prediction. Front. Artif. Intell. 2021, 4. <u>https://doi.org/10.3389/frai.2021.638410</u>
- Kratzert, F.; Klotz, D.; Hochreiter, S.; Nearing, G. S. A Note on Leveraging Synergy in Multiple Meteorological Data Sets with Deep Learning for Rainfall–Runoff Modeling. Hydrol Earth Syst Sci 2021, 25 (5), 2685–2703. <u>https://doi.org/10.5194/hess-25-2685-2021</u>.
- Kim, P. T.; Winter, R.; Clevert, D.-A. Unsupervised Representation Learning for Proteochemometric Modeling. Int. J. Mol. Sci. 2021, 22 (23). <u>https://doi.org/10.3390/ijms222312882</u>.
- Hoedt, P.-J.; Kratzert, F.; Klotz, D.; Halmich, C.; Holzleitner, M.; Nearing, G.; Hochreiter, S.; Klambauer, G. MC-LSTM: Mass-Conserving LSTM. ArXiv E-Prints 2021, <u>arXiv:2101.05186.</u>
- Gauch, M.; Kratzert, F.; Klotz, D.; Nearing, G.; Lin, J.; Hochreiter, S. Rainfall–Runoff Prediction at Multiple Timescales with a Single Long Short-Term Memory Network. Hydrol Earth Syst Sci 2021, 25 (4), 2045–2062. <u>https://doi.org/10.5194/hess-25-2045-2021</u>.
- Clevert, D.-A.; Le, T.; Winter, R.; Montanari, F. Img2Mol Accurate SMILES Recognition from Molecular Graphical Depictions. Chem. Sci. 2021, 12 (42), 14174–14181. <u>https://doi.org/10.1039/D1SC01839F</u>.
- Widrich, M.; Schäfl, B.; Pavlovic, M.; Ramsauer, H.; Gruber, L.; Holzleitner, M.; Brandstetter, J.; Sandve, G. K.; Greiff, V.; Hochreiter, S.; Klambauer, G. Modern Hopfield Networks and Attention for Immune Repertoire Classification. In NeurIPS; 2020. <u>https://arxiv.org/abs/2007.13505</u>
- <u>Svensson, E.</u>; Hoedt, P.-J.; Hochreiter, S.; Klambauer, G. Robust Task-Specific Adaption of Models for Drug-Target Interaction Prediction; 2022., <u>https://openreview.net/forum?id=dIX34JWnIAL</u>
- <u>Hassen, A. K.; Torren-Peraire, P</u>.; Genheden, S.; Verhoeven, J.; Preuss, M.; Tetko, I. V. Mind the Retrosynthesis Gap: Bridging the Divide between Single-Step and Multi-Step Retrosynthesis Prediction; 2022. <u>https://openreview.net/forum?id=LjdtY0hM7tf</u>
- <u>Andronov, M.</u>; <u>Voinarovska, V</u>.; Andronova, N.; Wand, M.; Clevert, D.-A.; Schmidhuber, J. Reagent Prediction with a Molecular Transformer Improves Reaction Data Quality. *Chemical Science*, 2023. <u>https://doi.org/10.1039/D2SC06798F</u>.

Additional information

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions, grant agreement No 956832.

Disclaimer: the newsletter reflects only the authors' view and neither the European Commission nor the Research Executive Agency are responsible for any use that may be made of the information it contains.

Follow us at Twitter: <u>https://twitter.com/AiddOne</u>