ATTILA KERTÉSZ-FARKAS, PHD

LABORATORY HEAD –TENURED ASSOCIATE PROFESSOR

FACULTY OF COMPUTER SCIENCE,
HSE UNIVERSITY, MOSCOW, RUSSIA

http://www.hse.ru/en/staff/akerteszfarkas
attila.kerteszfarkas@gmail.com

PERSONAL

Hungarian (EU) citizen – Married – Parent of two kids

AREA OF EXPERTISE

Application of Machine Learning, Data Mining, Bioinformatics

ACADEMIC CAREER

2021	**Tenured Associate Professor** at HSE University, Moscow, Russia,
2021-present	**Laboratory Head**, Laboratory on AI for Computational Biology, HSE University, Moscow, Russia,
2015-present	**Assistant Professor (Docent)**, HSE University, Moscow, Russia
2013-2015	Postdoctoral Fellow, **University of Washington, Seattle WA, USA**
2010	**Ph.D.** in Computer Science, University of Szeged, Hungary,
2009-2013	Postdoctoral Fellow, Bioinformatics Group, International Centre of Genetic Engineering and Biotechnology (ICGEB), Trieste, Italy
2008-2009	Research Fellow, Division of Imaging and Applied Mathematics, CDRH**, U.S. Food and Drug Administration (U.S. FDA)**, Silver Spring MD, USA. Joint affiliation with Department of Biology, University of Maryland Baltimore County (UMBC), Catonsville MD, USA
2004-2008	Ph.D. Student, University of Szeged, Hungary
2004	**M.Sc.** in Computer Science and Mathematics, University of Szeged, Hungary,
2000-2004	Undergraduate Research Assistant, Research Group on Artificial Intelligence, Hungarian Academy of Sciences, Szeged, Hungary

SCHOLARSHIPS AND AWARDS

2017,2018	Best paper awards at AIST (Audio, Image, Speech, Text) intl. conference.
2007-2008	Predoctoral scholarship granted by the Foundation for Research and Teaching of Informatics, Hungary
2006-2007	Erasmus scholarship, **Technische Universität Dresden, Germany**
2003-2008	Honors College (Eötvös Collegium) for outstanding students at University of Szeged, Hungary
2004	1st prize at a Scientific Conference of Students, University of Szeged, Hungary
2002	1st prize at a Scientific Conference of Students, University of Szeged, Hungary

LABORATORY HEAD

Organizing research project for PhD, MSc, BSc theses,

Writing reports, grants, and research articles, giving talks,

Promoting professional working ethics,

Coaching students in their scientific careers.

CODING

C/C++, Matlab, Python, PyTorch, bash, code reviewing, debugging my students' code,

ADVISING

MSc, BSc, Intl. Interns: many, approx. 45

PhD students:

- Roman Chereshnev (2015-2019): Thesis title: Human gait controlling system using machine learning methods suitable for robotic prostheses for patients suffering from double transfemoral amputation. The thesis bases on one article in Q2 journal, two articles in Q3 journals.

  Date of Defense: 10 October 2019, https://www.hse.ru/sci/diss/290226447
- Pavel Sulimov (2016-2020): Title: Learning generative probabilistic models for mass spectrometry data identification. The thesis bases on three articles in Q1 journals and one article in a Q3 venue.

  Date of Defense: 9 September 2020, https://www.hse.ru/sci/diss/377487862
- Nikita Moshkov (2017-2022): Title: Automated self-learning system for detecting intercellular interactions and phenotyping neurons using the method of multiple local potential fixations. Nikita has four articles in Q1 journals so far.

  Date of Defense: In progress. Expected in 2022
- Frank Acquaye (2021-2024): Title: Learning human-like reasoning for deep learning models for explaining and interpreting biomolecular data annotations. Expected in 2024

TEACHING EXPERIENCES

2017-present	Discriminative Methods in Machine Learning, (in English), HSE, Moscow, Russia
2017-present	Generative Methods in Machine Learning, (in English), HSE, Moscow, Russia
2016-present	Lecturer of Computational Mass Spectrometry (in English), HSE, Moscow, Russia
2015-2017	Lecturer of Advanced Topics in Machine Learning (Eng.), HSE, Moscow, Russia
2015	Lecturer at Bioinformatics Tools for Life Sciences Workshop (Eng.), Quito, Ecuador
2012	Lecturer at Theoretical and Practical Course on Advanced Bioinformatics (Eng.), ICGEB, Italy
2010-2013	Teaching assistant, Theoretical and Practical Course on Bioinformatics (Eng.), ICGEB, Italy
2007	Teaching assistant, Operational Research (in Hungarian), University of Szeged
2006	Teaching assistant, Algorithms and Data Structures (Hun), University of Szeged
2004-2005	Teaching assistant Introduction to Pascal, C and Java Programming (Hun), University of Szeged

MEMBERSHIPS

2018,2019	PC member, Intl' Conference on Analysis of Images, Social networks and Texts (AIST'18)
2017	PC member, Bioinformatics and Artificial Intelligence at International Joint Conference on Artificial Intelligence, (bai@IJCAI2017)
2014-2018	Industrial/Research Chair, Biotechno conference series
2012-2019	Editorial board of International Journal on Advances in Life Sciences
2011-2015	Consultant at ManuSTAT central (a statistical consulting service for life sciences)
2006-2014	John von Neumann Computer Science Society, Hungary
2005-2007	PhD student representative on the Board of Doctoral Schools, Faculty of Sci., Univ of Szeged
2005-2007	PhD student representative on the Board of the Institute of Informatics, University of Szeged

LIST OF PUBLICATIONS

THESES

1. Ph.D.: “*Protein Classification in a Machine Learning Framework*” 2008
2. M.Sc.: “*Compact Representation of Finite Languages with Nondeterministic Finite Automata*” 2004

PEER-REVIEWED JOURNAL PAPERS (*equally contributed authors; †corresponding author; **Q1 journal**)

3. Polina Kudryavtseva, Matvey Kashkinov, and Attila Kertesz-Farkas†: *Deep convolutional neural networks help scoring tandem mass spectrometry data in database-searching approaches*, **Journal of Proteome Research**, 2021, Q1 Journal, Imp. Factor: 4.074 (JCR SCI 2020) https://doi.org/10.1021/acs.jproteome.1c00315
4. Nikita Moshkov, Botond Mate, Attila Kertesz-Farkas, Reka Hollandi, and Peter Horvath, *Test-time augmentation for deep learning-based cell segmentation on microscopy images*, **Scientific Reports**, 10, 5068, 2020, Q1 Journal, Impact Factor: 3.998 (JCR SCI 2020), https://doi.org/10.1038/s41598-020-61808-3
5. Pavel Sulimov, Anastasia Voronkova, and Attila Kertesz-Farkas†: *Annotation of tandem mass spectrometry data using stochastic neural networks in shotgun proteomics*, **Bioinformatics**, 36(12), 3781-3787, 2020, Q1 Journal, Impact Factor: 5.61 (JCR SCI 2020) https://doi.org/10.1093/bioinformatics/btaa206
6. Pavel Sulimov and Attila Kertesz-Farkas†: *Tailor: A Nonparametric and Rapid Score Calibration Method for Database Search-Based Peptide Identification in Shotgun Proteomics*, **Journal of Proteome Research**, 19(4), 1481-1490, 2020, Q1 Journal, Imp. Factor: 4.074 (JCR SCI 2020) https://doi.org/10.1021/acs.jproteome.9b00736
7. Roman Chereshnev, Attila Kertesz-Farkas†: *GaIn: Human Gait Inference for Lower Limbic Prostheses for Patients Suffering from Double Trans-Femoral Amputation*, **Sensors,** 18(12), 4146, 2018, Q1 Journal, Impact Factor: 3.275(JCR SCI 2020) https://doi.org/10.1021/acs.jproteome.9b00736
8. Pavel Sulimov, Elena Sukmanova, Roman Chereshnev, and Attila Kertesz-Farkas†: *Guided Layer-wise Learning for Deep Models using Side Information*, Communication in Computer and Information Science,1086, 2020, Q3 Journal, SJR Score: 0.168 (SJR 2018), https://doi.org/10.1007/978-3-030-39575-9_6
9. Yulia Danilova, Anastasia Voronkova, Pavel Sulimov, Attila Kertesz-Farkas†: *Bias in false discovery rate estimation in mass-spectrometry-based peptide identification*, **Journal of Proteome Research**, 18(5), 2354– 2358, 2019, Q1 Journal, Impact Factor: 3.780 (JCR SCI 2019) https://doi.org/10.1021/acs.jproteome.8b00991
10. Gleb Filatov, Bruno B. Bauwens, and Attila Kertesz-Farkas†: *LZW-Kernel: fast kernel utilizing variable length code blocks from LZW compressors for protein sequence classification*, **Bioinformatics**, 34(19), 3281–3288, 2018, Q1 Journal, Impact Factor: 5.481 (JCR SCI 2018), https://doi.org/10.1093/bioinformatics/bty349
11. Roman Chereshnev, Attila Kertesz-Farkas†: *RapidHARe: An Energy-Efficient Method for Real-Time Human Activity Recognition from Wearable Sensors*; Journal of Ambient Intelligence and Smart Environments, 10(5), 377-391, 2018, Q3 Journal, SJR-Score: 0.273 (SJR), Impact Factor: 0.878 (JCR SCI 2018), https://doi.org/10.3233/AIS-180497
12. Andrey Shestakov, Danila Doroshin, Dmitri Shmelkin, and Attila Kertesz-Farkas†: *Lookup Lateration: Non-linear Received Signal Strength to Distance Mapping for Non-Line-of-Sight Geo-localization in Outdoor Urban Areas*; 7th International Conference on Analysis of Images, Social Networks and Texts (AIST) 2018, Lecture Notes in Computer Science, Revised Selected Papers, 2018 – **Best paper award,** Q2 Journal, SJR score: 0.283 (SJR 2018) https://doi.org/10.1007/978-3-030-11027-7_23
13. Marina L. Mokrishcheva, Attila Kertesz-Farkas, and Dmitri V. Nikitin: *New bifunctional restriction-modification enzyme AloI isoschizomer (PcoI): bioinformatics analysis, purification and activity confirmation*; Gene, 660, 6-12, 2018, Q2 Journal, Impact Factor: 2.497 (JCR SCI 2018), https://doi.org/10.1016/j.gene.2018.03.069
14. Roman Chereshnev, Attila Kertesz-Farkas†: *HuGaDB: Database for Human Gait Analysis from Wearable Inertial Sensor Networks*; 6th International Conference on Analysis of Images, Social Networks and Texts (AIST) 2017, Lecture Notes in Computer Science, Revised Selected Papers, 124-134, 2017, Q2 Journal, SJR score: 0.295 (SJR 2017) https://doi.org/10.1007/978-3-319-73013-4_12
15. Uri Keich, Attila Kertész-Farkas, and William Stafford Noble: Mix-Max: an improved false discovery rate estimation procedure for shotgun proteomics, **Journal of Proteome Research**, 14(8) 3148-3161, 2015
16. Attila Kertész-Farkas, Uri Keich, and William Stafford Noble: Tandem mass spectrum identification via Cascaded search; **Journal of Proteome Researc**h, 14(8) 3027-3038, 2015, Q1 journal, Impact Factor: 4.268
17. Bruna Marini, Attila Kertész-Farkas, Hashim Ali, Bojana Lucic, Kamil Lisek, Lara Manganaro, Sándor Pongor, Roberto Luzzati, Alessandra Recchia, Fulvio Mavilio, Mauro Giacca, and Marina Lusic: Nuclear architecture dictates HIV-1 integration site selection; ***Nature*** 521, 227-231, 2015, Q1, Impact Factor: 40.137 https://doi.org/10.1038/nature14226

18. Sean McIlwain, Kaipo Tamura, Attila Kertész-Farkas, Charles E. Grant, Benjamin Diament, Barbara Frewen, J. Jeffry Howbert, Michael R. Hoopmann, Lukas Käll, Jimmy K. Eng, Michael J. MacCoss, and William Stafford Noble: Crux: rapid open source protein tandem mass spectrometry analysis; **Journal of Proteome Research** 13(10) 4488-4491, 2014

19. Attila Kertész-Farkas, Beáta Reiz, Roberto Vera, Michael P. Myers, and Sándor Pongor: *PTMTreeSearch: a Novel Two-Stage Tree Search Algorithm with Pruning Rules for the Identification of Post-Translational Modification of Proteins in MS/MS Spectra*; **Bioinformatics** 30(2): 234-241, 2014

20. János Juhász*****, Attila Kertész-Farkas*****, Dora Szabo, and Sándor Pongor: *Emergence of Collective Territorial Defense in Bacterial Communities: Horizontal Gene Transfer Can Stabilize Microbiomes*; **PLoS One** 9(4), 2014

21. Roberto Vera, Yasset Perez-Riverol, Sonia Perez, Balázs Ligeti, Attila Kertész-Farkas, and Sándor Pongor: *JBioWH: an open-source Java framework for bioinformatics data integration*; **Database**, 2013

22. Beáta Reiz, Michael P. Myers, Sándor Pongor, and Attila Kertész-Farkas†: *Precursor Mass Dependent filtering of Mass Spectra for Proteomics Analysis*; Protein and Peptide Letters 21(8) 858-863, 2013

23. Beáta Reiz, Attila Kertész-Farkas, Somdutta Dhir, Sándor Pongor, and Michael P. Myers: *Chemical rule-based filtering of MS/MS spectra*; **Bioinformatics** 29(7) 925-932, 2013

24. Attila Kertész-Farkas, Beáta Reiz, Michael P. Myers, and Sándor Pongor: *Database searching in mass spectrometry based proteomics*; Current Bioinformatics 7(2) 221-230, 2012

25. Dóra Bihary, Ádám Kerényi, Zsolt Gelencsér, Sergiu Netotea, Attila Kertész-Farkas, Vittorio Venturi, and Sándor Pongor: *Simulation of communication and cooperation in multispecies bacterial communities with an agent based model*; Scalable Computing: Practice and Experience 13(1) 21–28, 2012

26. Beáta Reiz, Attila Kertész-Farkas, Sándor Pongor, and Michael P. Myers: *Data preprocessing and filtering in mass spectrometry based proteomics*; Current Bioinformatics 7(2) 212-220, 2012

27. Attila Kertész-Farkas, Beáta Reiz, Michael P. Myres, and Sándor Pongor: *PTMSearch: A Greedy Tree Traversal Algorithm for Finding Protein Post-Translational Modifications in Tandem Mass Spectra*; In: European Conference on Machine Learning and Principles and Practical Knowledge Discovery in Databases 2, Vol. 6912 Springer, 162-176, 2011

28. Emily Doughty*****, Attila Kertész-Farkas*****, Olivier Bodenreider, Gary Thompson, Asa Adadey, Thomas Peterson, and Maricel G. Kann: *Toward an automatic method for extracting cancer- and other disease-related point mutations from the biomedical literature*; **Bioinformatics** 27(3) 408-415, 2011

29. Somdutta Dhir, Mircea Pacurar, Dino Franklin, Zoltán Gáspári, Attila Kertész-Farkas, András Kocsor, Frank Eisenhaber and Sándor Pongor: *Detecting atypical examples of known domain types by sequence similarity searching: The SBASE domain library approach*; Current Protein Peptide Science 11(7) 538-549, 2010

30. József Dombi and Attila Kertész-Farkas†: *Applying Fuzzy Technologies to Equivalence Learning in Protein Classification*; **Journal of Computational Biology** 16(4) 611-623, 2009

31. Róbert Busa-Fekete, Attila Kertész-Farkas, András Kocsor, and Sándor Pongor: *Balanced ROC analysis (BAROC) protocol for the evaluation of protein similarities*; Journal of Biochemical and Biophysical Methods 70(6) 1210-1214, 2008

32. Attila Kertész-Farkas, Somdutta Dhir, Paolo Sonego, Mircea Pacurar, Sergiu Netoteia, Harm Nijveen, Arnold Kuzniar, Jack A.M. Leunissen, András Kocsor, and Sándor Pongor: *Benchmarking protein classification algorithms via supervised cross-validation*; Journal of Biochemical and Biophysical Methods 70(6) 1215-1223, 2008

33. Attila Kertész-Farkas, András Kocsor, and Sándor Pongor: *Equivalence Learning in Protein Classification*; In: P. Perner (Ed.) Machine Learning and Data Mining in Pattern Recognition, Springer Verlag, Heidelberg, LNAI (4571) 824-837, 2007

34. János Z. Kelemen, Attila Kertész-Farkas, András Kocsor, and László G. Puskás: *Kalman Filtering for Disease-State Estimation from Microarray Data*; **Bioinformatics** 22(24) 3047-3053, 2006

35. László Kaján, Attila Kertész-Farkas, Dino Franklin, Nelly Ivanova, András Kocsor, and Sándor Pongor: *Application of a simple log likelihood ratio approximant to protein sequence classification*; **Bioinformatics** 22(23) 2865-2869, 2006

36. Paolo Sonego, Mircea Pacurar, Somdutta Dhir, Attila Kertész-Farkas, András Kocsor, Zoltán Gáspári, Jack A.M. Leunissen, and Sándor Pongor: *A Protein Classification Benchmark collection for Machine Learning*; **Nucleic Acids Research** 35 232-236, 2006

37. András Kocsor, Attila Kertész-Farkas, László Kaján, and Sándor Pongor: *Application of compression-based distance measures to protein sequence-classification: a methodological study*; **Bioinformatics** 22(4) 407-412, 2006
38. Attila Kertész-Farkas and András Kocsor: *Kernel-based Classification of Tissues using Feature Weightings*; Applied Ecology and Environmental Research 4(2) 63-71, 2006
39. Attila Kertész-Farkas, Zoltán Fülöp, and András Kocsor: *Compact Representation of Hungarian Corpora*; (in Hungarian), Hungarian Journal of Applied Linguistics (1-2) 63-70, 2005

BOOK CHAPTERS

40. Attila Kertész-Farkas, András Kocsor, and Sándor Pongor: *The application of Data Compression-based Distance to Biological Sequences*; In: Frank Emmert-Streib. (Ed.) Information Theory and Statistical Learning, 73-88, 2008
41. Dmitry V. Nikitin, Attila Kertész-Farkas, Alexander S. Solonin, and Marina L. Mokrishcheva: *Bifunctional prokaryotic DNA-methyltransferases*; In: Dr. Anica Dricu (Ed.) Methylation (ISBN 980-953-307-146-1), 2012

CONFERENCES

42. *Generative probabilistic modelling of peptide-spectrum matching in tandem mass spectrometry*; International Congress on Biotechnology: State and Prospects of Development, Moscow Russia, February 25-27, 2019
43. *Bias in false discovery rate estimation in mass-spectrometry-based peptide identification*; International Congress on Biotechnology: State and Prospects of Development, Moscow Russia, February 25-27, 2019
44. *Lookup Lateration: Non-linear Received Signal Strength to Distance Mapping for Non-Line-of-Sight Geo-localization in Outdoor Urban Areas*; 7th International Conference on Analysis of Images, Social Networks and Texts (AIST), Moscow Russia, 5-7 July 2018 – **Best Talk Award**
45. *High-dimensional generative probabilistic models for peptide-spectrum-matching in tandem mass spectrometry*; Proteomics-2017, Valencia Spain, 8-11 November 2017
46. *PTMTreeSearch: a new algorithm for post-translational modification identification in tandem mass spectrometry data;* Proteomics-2017, Valencia Spain, 8-11 November 2017
47. *Cascaded false discovery rate control tandem mass spectrometry (MS/MS) data for peptide identification*; Proteomics-2017, Valencia Spain, 8-11 November 2017
48. *HuGaDB: Database for Human Gait Analysis from Wearable Inertial Sensor Networks*; International Conference on Analysis of Images, Social Networks, and Texts (AIST), Moscow Russia, 2017 – **Best Talk Award**
49. *Peptide identification in tandem mass spectrometry data via cascade search*; US HUPO; Seattle USA, 6-9 April 2014
50. *PTMSearch: A Greedy Tree Traversal Algorithm for Finding Protein Post-Translational Modifications in Tandem Mass Spectra*; European Conference on Machine Learning and Principles and Practice of knowledge discovery in databases; Athens Greece, 5-9 September 2011
51. *Equivalence Learning in Protein Classification*; International Conference on Machine Learning and Data Mining in Pattern Recognition; Leipzig Germany, 18-20 July 2007
52. *Classification of Tissues using Feature Weightings*; VII. Hungarian Conference on Biometrics and Biomathematics; Budapest Hungary, 5-6 July 2005.
53. *Kernel-based learning with dimension weighting*; Scientific Conference of Students, University of Szeged; April 2004,
54. *Compact Representation of Hungarian Vocabulary with Nondeterministic Finite Automata*; I. Conference on Hungarian Computational Linguistic; Szeged Hungary, 10-11 December 2003
55. *Compact representation of finite languages with nondeterministic automata*; Scientific Conference of Students, University of Szeged; October 2002,
56. *EMU: A tool for the Extraction of MUtations with disease associations from literature*; Growth Factor and Signal Transduction Conferences: System Biology, Integrative, Comparative and Multi-scale Modeling, Iowa USA, 11-14 June 2009

TALKS

57. *Large-scale localization method for urban area*, The 3rd Professor Day Academic Conference by Huawei Russian Research Centre, Moscow Russia, December 2016
58. *Cascade search for identification of heterogeneous tandem mass spectrometry data*, HSE International Laboratory of Algorithms and Technologies for Network Analysis, Nizhniy Novgorod Russia, November 2016
59. *False discovery rate control for database search methods over heterogeneous biological data*, Keynote talk at Analysis of Images, Social Networks, and Texts (AIST'2016), Yekaterinburg Russia, April 2016
60. *Statistical methods in MS/MS spectrum identification,* Universidad San Francisco de Quito, Ecuador, Oct. 2015
61. *Algorithmic identification of post-translational modifications of proteins in tandem mass spectra*, Luxembourg Centre for Systems Biomedicine, Université du Luxemburg, Luxemburg, April 2013
62. *Noise filtering and compression of mass spectra for proteomics*; Symposium at International Centre for Genetic Engineering and Biotechnology; Trieste Italy, June 2012
63. *A new algorithm for identification of post-translational modification in tandem mass spectrometry data*; Symposium at International Centre for Genetic Engineering and Biotechnology; Trieste Italy, June 2011
64. *Equivalence Learning in Protein Classification*; Young Researcher Symposium on Intelligent Systems at the John von Neumann Computer Society; Budapest Hungary, November 2007
65. *Protein databases and similarity measures for protein classification*; Bioinformatics seminar at Rényi Mathematical Institute; Budapest Hungary, October 2007