Abdurrahman Abul-Basher

Email: arbasher@student.ubc.ca
Website: https://arbasher.github.io
Linkedin: https://www.linkedin.com/in/meitantei
Citizenship: **Canadian**

Education

The University of British Columbia (UBC), Vancouver, BC, Canada
PhD, Bioinformatics, 2020
Thesis title: Machine Learning Methods for Metabolic Pathway Inference from Genomic Sequence Information
Adviser: Professor Steven J. Hallam
Area of Study: Data Science, Machine Learning, and Bioinformatics

Concordia University, Montreal, QC, Canada
MASc, Information Systems Security, 2011
Thesis title: Mining Chat Logs to Extract Information about Authors and Topics for Crime Investigation
Adviser: Professor Benjamin C. M. Fung
Area of Study: Data Mining, Machine Learning, and Information Systems Security

King Abdulaziz University, Jeddah, Saudi Arabia
BSc, Computer Science, 2008
Thesis title: University Courses Timetabling System using Genetic Algorithm.
Adviser: Professor Mahmoud Kamel.
Area of Study: Algorithms.

Honors and Awards

11. **Faculty of Science - BCB2 Fund** ($18,200 ∼ $27,000 per year), The University of British Columbia (UBC), Canada. *2017-2020.*
10. *President's Academic Excellence Initiative PhD Award*, The University of British Columbia (UBC), Canada. *2020.*
9. **Four Year Fellowships** (4YF) ($18,200 per year + tuition fee), The University of British Columbia (UBC), Canada. *2013-2017.*
8. **Faculty of Science - Graduate Support Initiative (GSI) Fund** ($8,500 per year), The University of British Columbia (UBC), Canada. *2013-2017.*
7. **Faculty of Science - Graduate Student Support Program (GSSP)** ($25,000 per year), McGill University, Canada. *2013-2018.* **(Declined)**.
6. **Faculty of Science and Engineering - Graduate Student Support Program (GSSP)** ($27,000 per year), York University, Canada. *2013-2019.* **(Declined)**.
5. **Power Corporation of Canada Graduate Fellowships** ($5,000), Concordia University, Canada. *2009-2010.*
4. **Concordia Graduate Student Support Program (GSSP)** ($15,000 per year), Concordia University, Canada. *2009-2011.*
3. First Honor Graduate for graduating with high GPA from King Abdulaziz University, Saudi Arabia. *2008.*
2. Faculty of Computing and Information Technology (FCIT) distinguished award, King Abdulaziz University (KAU), Saudi Arabia. *2008.*
1. **King Abdulaziz University merit award** ($2,000), Saudi Arabia. *2004-2008.*

Teaching Experience

4. DSCI 571 Supervised Learning I, **Master of Data Science**, The University of British Columbia (1 semester).
3. DSCI 573 Feature and Model Selection, **Master of Data Science**, The University of British Columbia (1 semester).
2. DSCI 575 Advanced Machine Learning, **Master of Data Science**, The University of British Columbia (1 semester).
1. CPIT 201 Introduction to Computing, Computer Science, King Abdulaziz University (3 semesters).

Research and Professional Experience

5. Postdoctoral researcher/Applied machine learning scientist, Hallam Lab, The University of British Columbia, Vancouver, BC, Canada (2020-2021).
4. Ph.D. research student, Hallam Lab, The University of British Columbia, Vancouver, BC, Canada (2016-2020).
3. Research assistant (as a rotation student), Genome Sciences Centre, Vancouver, BC, Canada (2013-2016).
2. Part-time data scientist, BigOui Marketing Inc., Montreal, QC, Canada (2012-2014).
1. Research assistant, Data Mining and Security Lab, Concordia University, Montreal, QC, Canada (2020-2021).

Mentoring Experience

3. Aditi Nagaraj, MSc, The University of British Columbia (2 months).
2. Vaibhav Grover, BSc student, Indian Institute of Technology (3 months).
2. Nathan Nastili, BSc student, Simon Fraser University (4 months).
1. Alex Purdy, BSc student, University of Victoria (12 months).

Publications

Published

10. **Abdurrahman Abul-Basher**, Ryan J. McLaughlin, and Steven Hallam. "Metabolic pathway inference using non-negative matrix factorization with community detection." *Journal of Computational Biology, 2021.* [impact factor: 1.479]
9. **Abdurrahman Abul-Basher**, Ryan J. McLaughlin, and Steven Hallam. "Metabolic pathway inference using non-negative matrix factorization with community detection." *ICCABS, 2020.*
8. **Abdurrahman Abul-Basher** and Steven Hallam. "Leveraging heterogeneous network embedding for metabolic pathway prediction." *Bioinformatics, 2020.* [impact factor: 6.937]
7. **Abdurrahman Abul-Basher**, Ryan J. McLaughlin, and Steven Hallam. "Metabolic pathway inference using multi-label classification with rich pathway features." *PLOS Computational Biology, 2020.* [impact factor: 4.475]
6. **Abdurrahman Abul-Basher** and Benjamin CM Fung. "Analyzing topics and authors in chat logs for crime investigation." *Knowledge and information systems, 2014.* [impact factor: 2.822].

Under Review

5. **Abdurrahman Abul-Basher**, Aditi N. Nagaraj, Ryan J. McLaughlin, Julia Anstett, and Steven Hallam. "leADS: improved metabolic pathway inference based on active dataset subsampling." *Bioinformatics, 2022.* [impact factor: 6.937].

In Submission

4. **Abdurrahman Abul-Basher**, Ryan J. McLaughlin, Julia Anstett, Aditi N. Nagaraj, and Steven Hallam. "Leveraging multiple (less-trusted) sources to improve metabolic pathway prediction." *2022*.
3. **Abdurrahman Abul-Basher** and Steven Hallam. "reMap: relabeling multi-label pathway data with groups to enhance predictive performance." *2022*.

PrePrints

2. **Abdurrahman Abul-Basher**. "Modeling metabolic pathways as bags (with augmentation)." *arXiv, 2019*.
1. **Abdurrahman Abul-Basher**, Alex Purdy, and Inanç Birol. "Event extraction from biomedical literature." p. 1-13. *bioRxiv, 2015*.

Poster Presentation

3. Abdurrahman Abul-Basher and Steven Hallam. "Leveraging heterogeneous network embedding for metabolic pathway prediction", BIOF, IOP and GSAT programs (B.I.G.) retreat, Vancouver, BC, *2019*.
2. Abdurrahman Abul-Basher and Steven Hallam. "Metabolic pathway inference using multi-label classification with rich pathway features", BIOF, IOP and GSAT programs (B.I.G.) retreat, Vancouver, BC, *2018*.
1. Abdurrahman Abul-Basher, Connor Morgan-Lang, and Steven Hallam. "Machine learning approach to recovering metabolic pathways from metagenomics sequences", Centre for Microbial Diversity and Evolution (CMDE) retreat, Victoria, BC, *2016*.

Affiliation

5. Hallam Lab, Microbiology & Immunology, The University of British Columbia, Vancouver, BC, Canada (2016–Current).
4. Graduate Student Society, The University of British Columbia, Vancouver, BC, Canada (2013 - 2020).
3. BC Cancer Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada (2013 - 2016).
2. The National Cyber-Forensics and Training Alliance, Concordia University, Montreal, QC, Canada (2009 - 2011).
1. Graduate Students' Association, Concordia University, Montreal, QC, Canada (2009 - 2011).

Research and Methodology Developments

5. Developed an efficient **active subsampling algorithm** to address the class imbalance problem for the multi-label data, which is a challenging task in the **data science** and machine learning fields.
4. Developed multiple solutions using **novel artificial intelligence algorithms** to predicting metabolic pathways from genomic sequence information (over 8000 organismal genomes) at different levels of complexity and completion using supervised and unsupervised machine/deep learning algorithms.
3. Designed a discourse processing framework using **convolutional neural networks** to summarize and retrieve articles related to patient genomic and mutation profiles.
2. Collaborated in developing an efficient clustering module based on **Nyström algorithm** with subsampling technique, and leveraging the idea of representation learning to characterize various types of discourse relations.
1. Proposed a framework to analyze inappropriate messages in chat logs and discover relevant subjects using a combination of **heuristic rule-based and graphical modeling techniques**.

Projects

2. I am the core developer of mltS, leADS, reMap, triUMPF, pathway2vec, and mlLGPR tools which are applied for metabolic pathway prediction, and are in the process to be integrated to MetaPathways v3 that is a popular tool to generate environmental pathway genome database from multi-organismal genome sequence information.
1. I developed LDA-TOT and A-TOT that compute the contribution of each participant for each discussed topic in a chat while discovering participants' roles in a chat.

Invited Talks

2. Center for Protein Degradation at Dana-Farber Cancer Institute, Harvard Medical School, Harvard University, 2021, Virtual talk.
1. International Conference on Computational Advances in Bio and Medical Sciences, 2020, Virtual talk.