

Abdurrahman Abul-Basher

Vancouver, BC – Canada

✉ abdurrahman.mohd@gmail.com • [📁 arbasher.github.io](https://github.com/arbasher)

My research focuses on designing mathematical models to address complex challenges associated with microorganisms and their interactions (w.r.t metabolic pathways) in an ecosystem. Towards achieving this goal, I incorporate machine learning algorithms to infer biochemical reactions and metabolic pathways for single and multi-organismal genomes. Many of the proposed methodologies were already seen to achieve the state-of-the-art results in the metabolic pathway prediction task. Although, I am highly competent to address the aforementioned problem, I am always open to extend my knowledge to other bioinformatics related topics. These include: building a community-driven integrated ecosystem encompassing genomics, proteomics, and metabolomics data to solve the pathway reconstruction problem or discovering novel pathways, differential pathway enrichment analysis, understanding disease mechanisms and searching for new therapeutic targets, elucidating the metabolic functions of an organism, and understanding microorganisms interactions and their impacts on hosts, and vice versa.

Education

The University of British Columbia

Vancouver, BC, Canada

PhD in Bioinformatics

2013–2020

Thesis title: Machine Learning Methods for Metabolic Pathway Inference from Genomic Sequence Information.

Concordia University

Montreal, QC, Canada

MASc in Information Systems Security

2009–2011

Thesis title: Mining Chat Logs to Extract Information about Authors and Topics for Crime Investigation.

King AbdulAziz University (KAU)

Jeddah, Saudi Arabia

BSc in Computer Science

2004–2008

Thesis title: University Courses Timetabling System using Genetic Algorithm.

Research Experience

Applied machine learning scientist

The University of British Columbia

[📁 Hallam Lab](#), Vancouver, BC, Canada

2020–Current

Description: At present, I am developing a tool (“mltS”) which is an ensemble-based metabolic pathway prediction to improve both the performance and robustness of pathway predictions. This is achieved using concepts borrowed from both theoretical machine learning and meta-learning approaches. The rational idea is to estimate a score (“reliability”) associated with each member in a multi-label ensemble, and then use these scores to optimize the inference of metabolic pathways from genomic sequence information. mltS is a general-purpose method that can be applied to any multi-label data. This work is currently in preparation for submission to a top tier journal.

Advisor: Dr. Steven J. Hallam.

PhD research student

The University of British Columbia

[📁 Hallam Lab](#), Vancouver, BC, Canada


2016–2020

Description: Proposed multiple solutions to predicting metabolic pathways from genomic sequence information (over 5000 organismal genomes) at different levels of complexity and completion using supervised (e.g. leADS and mILGPR) and unsupervised (e.g. triUMPF and pathway2vec) machine/deep learning algorithms. My work was the first to establish link between the pathway prediction framework with the downstream pathway enrichment analysis method using machine learning approaches. Furthermore, some of these proposed treatments are extensible to closely-associated studies in bioinformatics and machine learning (multi-label learning for tagging, ranking, and recommendation).

Advisors: Dr. Steven J. Hallam (primary advisor), Dr. Aria S. Hahn, and Dr. Kishori M. Konwar.

Research assistant

Genome Sciences Centre

 *Birol Lab, Vancouver, BC, Canada*

2013–2016

Description: Conducted large-scale data analysis from PubMed articles using cutting-edge developments in the domain of biomedical natural language processing (BioNLP). These include building a discourse processing framework using convolutional neural networks to summarize and retrieve articles related to patient genomic and mutation profiles, addressing an efficient clustering module based on Nyström algorithm with subsampling technique, and leveraging the idea of representation learning to characterize various types of relations (e.g. discourse).

Advisors: Dr. Inanç Birol (primary advisor), Dr. Victoria A. Stuart, and Dr. Djallel Bouneffouf.

Part-time data scientist

BigOui Marketing Inc.

Montreal, QC, Canada

2012–2014

Description: Proposed two predictive modeling techniques (topic and graph based) to connect subscribers (BigOui) with local (Montreal) merchants and to perform quantitative and qualitative assessment of BigOui.

Research assistant

Concordia University

Data Mining and Security ( DMaS) Lab, Montreal, QC, Canada

2009–2011

Description: I was the major contributor in a team of 4 members to building a framework to analyze inappropriate messages in chat logs (over 1000 logs) and to discover relevant subjects using a combination of rule-based and graphical modeling techniques. I specifically designed and proposed two models (LDA-TOT and A-TOT) which can compute the contribution of each participant for each discussed subject in a chat while discovering participants' roles in a chat. Moreover, the two models have built-in functionality to characterize the evolution of subjects along with participants over time.

Advisors: Dr. Benjamin C. M. Fung (primary advisor).

Selected Publications

Accepted/Under Review

6: **Abdur Rahman M. A. Basher** and Steven Hallam. "Multi-label Pathway Prediction based on Active Dataset Subsampling." *under review*.

5: **Abdur Rahman M. A. Basher** and Steven Hallam. "reMap: Relabeling Metabolic Pathway Dataset with Bags to Enhance Predictive Performance." *under review*.

4: **Abdur Rahman M. A. Basher**, Ryan J. McLaughlin, and Steven Hallam. "Metabolic Pathway Inference using Non-negative Matrix Factorization with Community Detection." *ICCABS, 2020*.

3: **Abdur Rahman M. A. Basher** and Steven Hallam. "Leveraging Heterogeneous Network Embedding for Metabolic Pathway Prediction." *Bioinformatics, 2020*.

2: **Abdur Rahman M. A. Basher**, Ryan J. McLaughlin, and Steven Hallam. "Metabolic Pathway Inference using Multi-Label Classification with Rich Pathway Features." *PLOS Computational Biology*,

2020.

1: **Abdur Rahman M. A. Basher** and Benjamin CM Fung. "Analyzing Topics and Authors in Chat Logs for Crime Investigation." *Knowledge and information systems*, 2014.

In Preparation/Non-peer Reviewed

3: **Abdur Rahman M. A. Basher** and Steven Hallam. "Leveraging Multiple (Less-Trusted) Sources to Improve Metabolic Pathway Prediction." *in preparation*.

2: **Abdur Rahman M. A. Basher** and Steven Hallam. "Modeling Metabolic Pathways as Bags (with Augmentation)." 2020.

1: **Abdur Rahman M. A. Basher**, Alex Purdy, and Inanç Birol. (2015). "Event Extraction from Biomedical Literature." *biorxiv*. 1-13.

Presentations

4: Abdur Rahman M. A. Basher, Ryan J. McLaughlin, and Steven Hallam. "Metabolic Pathway Inference using Non-negative Matrix Factorization with Community Detection.", International Conference on Computational Advances in Bio and medical Sciences (ICCABS), Virtual Conference, 2020.

3: Abdur Rahman M. A. 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: Abdur Rahman M. A. 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: Abdur Rahman M. A. 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.

Conference Participations

Participated in more than 3 conferences for collaborations, presenting, exchange of ideas, demonstration of tools, etc.

Fellowships and Awards

10: Faculty of Science - BCB2 Fund (\$18,200 ~ \$27,000 per year), The University of British Columbia (UBC), Canada. 2017-2020.

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

Teaching Assistant.....

The University of British Columbia


Vancouver, BC, Canada
2016–2017

I was a TA for the following Master of Data Science (MDS) courses:

- o DSCI 571 Supervised Learning I
- o DSCI 573 Feature and Model Selection
- o DSCI 575 Advanced Machine Learning

Affiliations

Microbiology & Immunology

 Hallam Lab, Vancouver, BC, Canada


University of British Columbia
2016–Current

Graduate Student Society

 GSS, Vancouver, BC, Canada

University of British Columbia
2013–2020

Canada's Michael Smith Genome Sciences Centre

 GSC, Vancouver, BC, Canada

BC Cancer Agency
2013–2016

Skills

Programming Languages.....

Python • R • MATLAB • L^AT_EX • C++ • C# • BASH • Shell script

Packages and Toolkits.....

- o I have more than 5 years experience working with various visualization tools and machine/deep learning packages such as: • tensorflow • Scikit-learn • pandas • Plotly • Altair • Matplotlib • MALLET • GMTK • Stanford CoreNLP • BANNER • Gensim • NLTK • LSTM • GRU • RNN • CNN • scikit-bio • Graph learning • Representation learning
- o I have more than 3 years experience working with various downstream genomics analysis tools such as: RDKit • GSEA • g:Profiler • Cytoscape • EnrichmentMap • Phred • RAST • GenSAS • MG-RAST • Pathway-tools • EggNOG • BLAST • KAAS • rapsearch 2 • MinPath • MetaBin • KEGG • MetaCyc • Pathway Commons • WikiPathways • ConsensusPathDB • Gene Ontology (GO) • Molecular Signatures Database (MSigDB)

Miscellaneous.....

- I proposed or currently addressing concerns about multi-label classification. These include: label correlations • representation learning of labels • noise in the instance/label spaces • inductive bias • active dataset subsampling • knowledge distillation • imbalanced and tail labels • partial multi-label problem • a combination of multi-instance and multi-view problem • bagging labels (grouping labels into non-disjoint sets) • high inference cost for the extreme multi-label case
- I have in-depth experience with pathway reconstruction problem from genomics data while also exploring solutions to extend pathway prediction and analysis from metatranscriptomics, proteomics, metabolomics data

References

Available upon request.