Md Abid Hasan

Principal Scientist II Bioinformatics
Roche Sequencing Solutions, Inc.
Pleasanton, CA 94588

Email: abidhasaniut@gmail.com
Phone: +1 951 907 2393
Website: https://abidhasan.creationzone.space/

Work Experience

Roche Sequencing Solutions – Principal Scientist II Bioinformatics Present
- Algorithm development in bioinformatics pipeline

Roche Sequencing Solutions – Principal Scientist I Bioinformatics 12/2019 – 09/2024
- Algorithm development in bioinformatics pipeline

University of California Riverside – Graduate Student Researcher 09/2014 – 11/2019
- Implementation of machine learning model on biomolecular data

Islamic University of Technology – Lecturer 01/2011 – 08/2014
- Conducting undergraduate courses in Computer Science department

Education

University of California Riverside 09/2014–10/2019
Ph.D. in Computer Science. (Advisor: **Prof. Stefano Lonardi**)

Islamic University of Technology, Bangladesh 01/2011–12/2012
M.Sc. in Computer Science. (Advisor: **Prof. M. A. Mottalib**)

Islamic University of Technology, Bangladesh 01/2007–11/2010
B.Sc. in Computer Science

Skills

Programming Languages
- Python, C++, Java

Machine learning & Bioinformatics
- Deep Learning, Algorithm Development, NGS Data Analysis, Data Visualization

Tools & Technologies
- PyTorch, TensorFlow, scikit-learn, shell scripting, cloud computing, Git, AWS, Linux, Jupyter, Github Copilot, LLM

Research Experience

Minimal Residual Disease (MRD) Internal
- Deep learning based method for tumor informed and plasma only disease monitoring

Genotype imputation Internal
- Training autoencoder model on large population dataset to learn complex haplotype pattern and genotype for predicting missing genotypes on samples

Germline filtering & TMB Internal
- A tree-based machine learning approach using LightGBM for classifying germline variants, and calculate TMB

Consensus read generation of read sequences with SBX technology Internal
- Algorithm implementation, and analysis to verify clustering and consensus read generation

Deduplication of sequencing data Internal
- Deduplication with Unique molecular identifier (UMI), Sequencing error suppression, and variant calls

Deep learning model for gene essentiality prediction in microbes Github repository
- Using DNA and protein sequence feature to prediction essential genes with Deep neural network

Multiple cancer type classification with somatic point mutation data Github repository
- Classify 12 types of cancer using only somatic gene mutation frequency data

Nucleosome stability prediction in *Plasmodium Falciparum* Github repository

- Use sequence and chemical properties of nucleotides for prediction of nucleosome stability

The genome of Cowpea (Vigna unguiculata [L.] Walp.

- Performed variant call analysis in a collaboration work with UCR Plant department

Patents

- **System and method for multimodal prediction of pateint outcomes** (2023)
- **System and method for virtual pathology** (in progress 2024)
- **System and methods for calculating Tumor Mutational Burden using machine learning models** (in progress 2024)
- **Molecular Consensus** (in progress 2024)
- **Variant calling** (in progress 2023)

Publications (Google Scholar)

- **Md Abid Hasan**, Stefano Lonardi, “DeeplyEssential: a deep neural network for predicting essential genes in microbes”, BMC Bioinformatics 21, 367 (2020)
- Stefano Lonardi, María Muñoz-Amatriaín, Qihua Liang, Shengqiang Shu, Steve I Wanamaker, Sassoum Lo, Jaakko Tanskanen, Alan H Schulman, Tingting Zhu, Ming-Cheng Luo, Hind Alhakami, Rachid Ounit, **Abid Md Hasan**, Jerome Verdier, Philip A Roberts, Jansen RP Santos, Arsenio Ndeve, Jaroslav Doležel, Jan Vrána, Samuel A Hokin, Andrew D Farmer, Steven B Cannon, Timothy J Close “The genome of cowpea (*Vigna unguiculata* [L.] Walp.)”, The Plant Journal, 98 (5), 767-782, June 2019
- **Md Abid Hasan**, Stefano Lonardi, “mClass: Cancer type classification with somatic point mutation data”. In *16th International Conference RECOMB-CG, QC, Canada*, 2018
- Anton Polishko, **Md Hasan**, Weihua Pan, Evelien M Bunnik, Karine Le Roch, Stefano Lonardi, “ThIEF: Finding genome-wide trajectories of epigenetics marks”. In *17th Workshop of Algorithm and Bioinformatics (WABI), Boston, USA*, 2017
- **Md Abid Hasan**, Md Kamrul Hasan, M Abdul Mottalib, “Linear regression based feature selection for microarray data classification”. In *Int. Journal of Data Mining and Bioinformatics, 11(2), 167-179*, 2015
- Kazi Mahbub Mutakabbir, Shah S Mahin, **Md Abid Hasan**, “Mining frequent pattern within a genetic sequence using unique pattern indexing and mapping techniques”. In *The International Conference on Informatics, Electronics & Vision (ICIEV)*, Dhaka, Bangladesh 2014
- Rasif Ajwad, Syed Nayem Hossain, **Md Abid Hasan**, “Optimizing DNA sequences using Tetra-nucleotide RankList”. In *The International Conference on Electrical Engineering and Information & Communication Technology*, Dhaka, Bangladesh 2014
- SM Samiul Islam, Md Rashed Asger, **Md Abid Hasan**, M Abdul Mottalib, “A modified algorithm for variable length DNA motif discovery”. In *The IEEE International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)*, Kuala Lumpur, Malaysia, 2013
- **Abid Hasan**, Shaikh Jeeshan Kabeer, Kamrul Hasan, Md Abdul Mottalib, “Discriminative Gene Selection Employing Linear Regression Model”. *Global Journal of Computer Science and Technology*, Vol 13, No 4-C, 2013
- **Md Abid Hasan**, Moin Mahmud Tanvee, Md Kamrul Hasan, M Abdul Mottalib, “Selecting features from high dimensional datasets using regression analysis”. In *5th International Conference on Knowledge and Smart Technology (KST)*, Chonburi, Thailand, Thailand, 2013
- **Abid Hasan**, Md Akhtaruzzaman Adnan, “High dimensional microarray data classification using correlation based feature selection”. In *The International Conference on Biomedical Engineering (ICoBE)*, Penang, Malaysia, 2012
- **Abid Hasan**, Golam Morshed Maruf, Hawlader Abdullah Al Mamun, Paul Kawn, “Cancer Classification from Microarray Data using Gene Feature Ranking”. In *International Journal Of Data Mining And Emerging Technologies*, 1 (2), 2011