

My research objective is to develop novel AI and ML algorithms to advance scientific discovery and understanding of biological systems. I would largely characterize my current research as [AI for Molecular Biology](#). My recent interests hybridize concepts and ideas from [deep learning](#), [language models](#), [natural language processing](#), and [bioinformatics](#). I am currently working on language models situated in domain knowledge to leverage content, structure and organization in disparate sources of data.

EDUCATION

Ph.D. Candidate in Computer Science, George Mason University, Fairfax, VA, USA

Aug. 2018 - Current

- CGPA: [3.80/4.00](#)
- Advisor: Dr. Amarda Shehu

BS in Software Engineering, University of Dhaka, Dhaka, Bangladesh

Jan. 2012 - Dec. 2015

- CGPA: [3.76/4.00](#)
- Dissertation: Migrating Design to Object Oriented Paradigm from Structured Program by Clustering Weighted Data Call Graph.
- Advisor: Shah Mostafa Khaled

PUBLICATIONS

Peer-reviewed Journal Publications

- NAR'24** A. Kabir, M. Bhattarai, S. Peterson, Y. Najman-Licht, K. Ø Rasmussen, A. Shehu, A. R. Bishop, B. Alexandrov, A. Usheva. "DNA breathing integration with deep learning foundational model advances [genome-wide binding prediction of human transcription factors](#)," Nucleic Acids Research, gkae783, 2024. (Impact Factor 2023: 16.6)
- Bioinformatics Advances'24** A. Kabir, A. Moldwin, Y. Bromberg, and A. Shehu. "In the [Twilight Zone](#) of Protein Sequence Homology: Do [Protein Language Models](#) Learn Protein Structure?," Bioinformatics Advances, vbae119, 2024. (Impact Factor 2024: 2.4)
- CSHL'23** Y. Bromberg, A. Kabir, R. Prabakaran, and A. Shehu. "[Variant prediction](#) in the age of Machine Learning," Cold Spring Harbor Perspectives in Biology, Cold Spring Harbor Laboratory, 2023. (Impact Factor 2023: 6.9)
- Bioinformatics'23** A. Kabir, M. Bhattarai, K. Ø. Rasmussen, A. Shehu, A. Usheva, A. R. Bishop and B. Alexandrov, "Examining [DNA breathing](#) with pyDNA-EPBD," Bioinformatics, vol. 39, no. 11, Nov. 2023. (Impact Factor 2023: 4.4)
- Biomolecules'22** A. Kabir and A. Shehu, "GOProFormer: A [Multi-Modal Transformer](#) Method for [Gene Ontology](#) Protein Function Prediction," Biomolecules, vol. 12, no. 11, 2022. (Impact Factor 2022: 6.064)

Peer-reviewed Conference Publications

- ACM-BCB'23** A. Kabir, A. Moldwin, and A. Shehu, "A Comparative Analysis of Transformer-based Protein Language Models for [Remote Homology Prediction](#)," Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology (BCB), and Health Informatics, Association for Computing Machinery, 2023. Best Paper Award.
- IEEE-ICKG'22** A. Kabir and A. Shehu, "Sequence-Structure Embeddings via [Protein Language Models](#) Improve on Prediction Tasks," International Conference on Knowledge Graph (ICKG), IEEE, 2022.
- BICOB'22** A. Kabir, T. Inan, and A. Shehu, "[Analysis of AlphaFold2](#) for Modeling Structures of Wildtype and Variant Protein Sequences," in Proceedings of 14th International Conference on Bioinformatics and Computational Biology, 2022. Best Paper Award.
- ACM-CSBW'20** Y. Du, A. Kabir, L. Zhao, and A. Shehu, "From Interatomic Distances to [Protein Tertiary Structures with a Deep Convolutional Neural Network](#)," ACM Conference of Bioinformatics and Computational Biology (BCB) Workshops: Computational Structural Biology Workshop (CSBW), Virtual, 2020.
- ACM-SIGSPATIAL'19** T. S. Khan, A. Kabir, D. Pfoser, and A. Züfle, "CrowdZIP: A System to [Improve Reverse ZIP Code Geocoding](#) Using Spatial and Crowdsourced Data," in Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2019.

Book Chapters

- Springer'22** A. Kabir and A. Shehu, "[Graph Neural Networks](#) in Predicting Protein Function and Interactions," in Graph Neural Networks: Foundations, Frontiers, and Applications, L. Wu, P. Cui, J. Pei, and L. Zhao, Eds. Singapore: Springer Singapore, 2022, pp. 541–556.

Other Publications

- BioRxiv** A. Kabir, M. Bhattarai, K.Ø. Rasmussen, A. Shehu, A.R. Bishop, B. Alexandrov, A. Usheva, "Advancing [Transcription Factor Binding Site Prediction](#) Using DNA Breathing Dynamics and Sequence Transformers via Cross Attention," 2024, doi:10.1101/2024.01.16.575935. (Under review)
- arXiv** A. Kabir and A. Shehu, "Transformer Neural Networks Attending to Both Sequence and Structure for Protein Prediction Tasks." arXiv 2206.11057, 2022.

RESEARCH EXPERIENCE

Shehu Lab, George Mason University, Virginia, USA

Aug. 2021 – Current

- Position: [Graduate Research Assistant](#)
- Contributions: Contributing towards developing multi-modal deep learning algorithms by solving molecular biology problems, including protein function prediction, explainable large protein language models and mutation effect understanding.
- Principal Investigator: Dr. Amarda Shehu, Professor, GMU.

Los Alamos National Lab (LANL), New Mexico, USA

Feb. 2023 - Dec. 2023

- Position: [Graduate Student](#), Physics & Chemistry of Materials Group (T-1)
- Contributions: Researched on [Transcription Factor-DNA \(TF-DNA\)](#) binding prediction problems and published two consecutive articles ([pyDNAEPBD](#) and [EPBDxDNABERT-2](#)).
- Principal Investigator: Dr. Manish Bhattarai, Staff Scientist-II, LANL.

Shehu Lab, George Mason University, Virginia, USA

May 2020 – Aug. 2020

- Position: [Graduate Research Assistant](#)
- Contributions: Worked on designing and developing a language model based regressor for protein stability prediction.
- Principal Investigator: Dr. Amarda Shehu, Professor, GMU.

Evolutionary Computation Research System (ECJ)

May 2019 - Aug. 2019

- Contributions: Developing and testing general purpose [genetic algorithms](#) for the ECJ framework ([Access link](#)).
- Principal Investigator: Dr. Sean Luke, Professor, GMU.

Study on [Object Detection and Few-shot Learning](#)

Jan. 2019 - May. 2019

- Contributions: Extended a fine-grained dataset involving subordinate categories of two class of species, such as cats and dogs, and applied a prototypical network for few-show learning to understand the learned-prototype representation ([Github](#)).
- Principal Investigator: Dr. Jana Kosecka, Professor, GMU.

TEACHING EXPERIENCE

Artificial Intelligence (CS 480)

Spring 2021

- Programming languages: [Lisp](#), [Python](#).
- Course Instructor: Dr. Sean Luke, Professor, GMU.
- Throughout this experience, I conduct one-to-one meeting at the office hours to discuss students' queries and clarifications; grade assignments and projects; and discuss about project materials with the instructor.

Computer Systems and Programming (CS 367)

Fall 2018, Spring 2018, Fall 2019, Spring 2019, Fall 2020

- Programming languages: [C](#), [Assembly](#).
- Course Instructors: Dr. Hakan Aydin, Professor, GMU; Dr. Yutao Zhong, Associate Professor, GMU; Kevin Andrea, Instructor, GMU.
- We discuss class, assignment and project materials in a biweekly manner; lecture on laboratory items before recitations; discuss students' questions on online (piazza) and in-person; grade assignments and projects; and solving the related problems.

HIGHLIGHTED PROJECTS (BEYOND PUBLICATIONS)

[Mutation stability classification using PProBERTa](#)

Fall 2021

- Highlights: Analysis of learned high-dimensional representation of protein sequences using linear and non-linear dimensionality reduction techniques, such as [PCA](#) and [T-SNE](#), from a pretrained Transformer model called [PProBERTa](#) to classify protein mutation stability ([Github](#)).
- Principal Investigator: Dr. Amarda Shehu, Professor, GMU.

[DeepDDG paper reconstruction for protein mutation analysis](#)

Fall 2021

- Highlights: Implementation of a paper titled as "DeepDDG: Predicting the Stability Change of Protein Point Mutations Using Neural Networks". This paper applies a neural network using biophysical attributes such as backbone dihedral angles, solvent accessible surface area, secondary structure, hydrogen bond, distance and orientation of amino acids and position-specific scoring matrix ([Github](#)).
- Course: [Machine Learning](#), Dr. Carlotta Domeniconi, Professor, GMU.

[DNCON paper reconstruction for protein tertiary structure prediction](#)

Fall 2020

- Highlights: Several attributes from DNCON paper, such as position-specific scoring matrix using PSI-BLAST, secondary structure and accessible surface area using SCRATCH, amino-acid potential statistics i.e Brauns, Jernigans and Levitts, were implemented for future research study ([Github](#)).
- Principal Investigator: Dr. Amarda Shehu, Professor, GMU.

[Protein Contact-map Prediction using Variational-Autoencoder from Sequence](#)

Spring 2020

- Highlights: Application and evaluation of [Variational-Autoencoder](#) for predicting contacts from protein sequences ([Github](#)).
- Course: [Deep Learning](#), Dr. Jana Kosecka, Professor, GMU.

Human pose estimation using DeepPose

Spring 2020

- Highlights: Following [DeepPose](#) article, [AlexNet](#) is applied to solve the task by constructing a regression problem from human pose estimation ([Github](#)).
- Course: [Computer Vision](#), Dr. Joran Duric, Professor, GMU.

Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records

Fall 2019

- Highlights: Application and evaluation of [decision tree](#), [support vector machine](#) and [a linear neural network](#) to classify whether a patient being readmitted or not in future given diagnosis result ([Github](#)).
- Course: [Data Mining](#), Dr. Harry Wechler, Professor, GMU.

INDUSTRY EXPERIENCE

Streams Tech Inc., Dhaka, Bangladesh

May 2017 - Jun. 2018

- Position: [Software Engineer](#)
- Contributions: Worked as full-stack frontend and backend developer on a project titled as System for Tracking Results and Evidence for Adaptive Management (STREAM). I use [AngularJS 1](#) framework to develop user interface, [ASP.NET](#) for web-application and services, and [MySQL](#) as database.

Samsung R&D Institute Bangladesh (SRBD), Dhaka, Bangladesh

Jun. 2016 - Dec. 2016

- Position: [Software Engineer](#)
- Contributions: Worked on a large team environment to develop [Tizen](#) mobile and gear application for Sales Force Automation and Management team.

Jantrik Technologies Ltd., Dhaka, Bangladesh

Jan. 2015 - Jun. 2015

- Position: [Intern Software Engineer](#)
- Contributions: Build the core module of To-Let and Watershed Data Management (WDM) using [ASP.NET MVC 5](#), [Web API 2](#), [AngularJS 1](#) and [Leaflet](#).

HONORS AND AWARDS

ACM-BCB'23	The paper titled "A Comparative Analysis of Transformer-based Protein Language Models for Remote Homology Prediction " received Best Paper Award at the ACM-BCB, 2023.
BICOB'22	Received Best Paper Award for the paper titled as "Analysis of AlphaFold2 for Modeling Structures of Wildtype and Variant Protein Sequences" at the BICOB, 2022.
SUST'15	Secured 1st Prize at Shahjalal University of Science and Technology (SUST) Inter University Software Competition, Sylhet, Bangladesh, 2015.
BASIS'15	Received 1st Prize at Bangladesh Association of Software and Information Services (BASIS) Code Warriors Challenge, Dhaka, Bangladesh, 2015.

SERVICES

External Reviewer

2024	Large Language Models for Biological Discoveries (AAAI 2024: LLMs4Bio)
2024	GIGAScience
2022-2024	Association for the Advancement of Artificial Intelligence (AAAI)
2021-2024	IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)
2024	Biology Methods and Protocols (BMP)
2021-2024	Intelligent Systems for Molecular Biology (ISMB)
2023	Computational Structural Biology Workshop (CSBW)
2021-2023	Biomolecules
2022	AI4Science
2022	Journal of Biomedical and Health Informatics (JBHI)
2022	IEEE International Conference on Bioinformatics and Biomedicine (BIBM)
2022	International Conference on Machine Learning (ICML)
2022	Bioinformatics
2021	Business Intelligence Data Management (BIDM)
2021	International Conference on Artificial Neural Networks (ICANN)
2021	Neural Processing Letters (NEPL)
2021	Journal of Ambient Intelligence and Smart Environments
2021	BioData Mining (BIDM)

Volunteering

- May'24 I lecture on prompt engineering, focusing on how AI tools can assist in developing research ideas, as part of the EPIC SPICE Internship program under the U.S. Department of Defense (DoD) STEM Internship Program.
- May-Dec'23 I serve on the organizing committee as the Program Committee (PC) Engagement role for the [AAAI 2024: LLMs4Bio](#) conference ([Access link](#)).
- Aug-Dec'22 I manage the [Machine Learning Reading Group](#) at GMU during Fall 2022, where we present a scholarly article each week and discuss the latest developments in the machine learning community.
- Aug-Dec'22 I collaborate with two newly enrolled graduate students to help them develop their research experience. We meet at least once a week to discuss papers, explore potential ideas, and brainstorm solutions.

COMPUTER SKILLS

Programming languages	Python , C++, Java.
Machine learning libraries	Numpy , Pandas , Scikit-learn , SciPy , Matplotlib .
Deep learning libraries	PyTorch .
Bio-simulation tools	PyMOL , Rosetta , Biopython .
Natural language processing	Hugging Face .
Web-development	HTML, CSS, ASP.NET MVC 5, ASP.NET Web API 2, AngularJS, Leaflet.
Databases	MySQL, PostgreSQL.
Version control	GitHub.
High-performance Computing	Slurm .

REFERENCES

[Amarda Shehu, Ph.D.](#)

Association: Academic Advisor and [Research Supervisor](#)

Vice President and Chief AI Officer, Office of the President

Professor of Computer Science, College of Engineering and Computing

Associate Vice President of Research for the Institute of Digital Innovation (IDIA)

Co-Director of Center of Excellence for Government Cybersecurity Risk Management and Resilience

Principal Investigator, Computational Biology Lab ([Shehu Lab](#))

George Mason University, Fairfax, VA, USA

Phone: +1-(703)-993-4135

Email: ashehu@gmu.edu