

SARWAN ALI

Department of Computer Science
Georgia State University
[[Github Profile](#)] , [[Google Scholar Profile](#)]

Tel: +1 (404)-510-5915
Email: sali85@student.gsu.edu
[[Website](#)] , [[ORCID ID](#)]

RESEARCH INTERESTS

- Machine Learning, Artificial Intelligence, Data Mining, Algorithms, Bioinformatics, Combinatorial Optimization

EDUCATION

- Georgia State University, Atlanta January 2021 - June 2026
Ph.D (Computer Science) , **Advisor:** Murray Patterson
CGPA: 3.84/4.0
- Lahore University of Management Sciences (LUMS), Lahore August 2016 - June 2018
M.S (Computer Science)
 - MS Thesis Title:** Predicting Attributes of Nodes using Network Structure
Status: Published in ACM Transactions on Intelligent Systems and Technology (TIST), (2020), [[PDF](#)]
- University of Engineering and Technology (UET), Peshawar August 2012 - June 2016
B.S (Computer Science)
 - BS Thesis Title:** Cache Replacement Algorithm [[PDF](#)]

RESEARCH EXPERIENCE

- Georgia State University, Atlanta, GA, USA January 2021 - Present
 - Graduate Research Assistant, Advisor:** [Murray Patterson](#), [Alexander Zelikovsky](#)
 - Understanding biological processes in protein/DNA sequences using A.I.
- Boston College, Newton, Massachusetts, USA August 2022 - December 2022
 - Visiting Research Scholar, Advisor:** [José Bento](#)
 - Understanding the relationship between proteins and aptamers in 3D geometric machine learning.
- Robert Bosch LLC, Sunnyvale, CA, USA May 2022 - August 2022
 - Knowledge Engineering Intern (Summer Internship), Advisor:** [HyeongSik Kim](#)
 - Understanding sequence and interrupt patterns generated from manufacturing facilities using ML methods.
- IBM T. J. Watson Research Center Yorktown Heights (AI Foundations Group), NY, USA September 2021 - May 2022
 - Research Collaborator, Advisor:** [Pin-Yu Chen](#)
 - Benchmarking machine learning robustness for noisy biological sequences.
- University of Newcastle, NSW, Australia March 2020 - Dec. 2021
 - Research Collaboration, Advisor:** [Pablo Moscato](#) and [Luke Mathieson](#)
 - Designing efficient Memetic Algorithm To Find a Hamiltonian Cycle in a Hamiltonian Graph
- Lahore University of Management Sciences (LUMS), Lahore, Pakistan July 2018 - Dec. 2020
 - Research Associate (Data Analytics Lab), Advisor:** [Imdadullah Khan](#)
 - Understanding and designing big data solutions to deal with graphs- and time series-based data analytics.
- CECOS University, Peshawar, Pakistan July 2020 - Sept. 2020
 - Research Collaboration**
 - Sarwan Ali**, Abdul Haseeb. "Universities Ranking System Using Self and Cross Proclivity". Preprint (2020)
- Higher Education Commission (HEC), Islamabad, Pakistan March 2020 - June 2020
 - Data Analytics Project**
Project of HEC to predict the final marks of higher secondary school students using their previous years performance. This prediction is necessary because of cancellation of final exams due to Covid-19 pandemic.
- Pakistan Space & Upper Atmosphere Research Commission (SUPARCO), Karachi, Pakistan July 2014 - Jan 2015
 - Student Project**
National Student Satellite (PNSS-1) mission
- Conference/Journal Reviewer
 - NeurIPS (2019, 2021, 2022, 2023), ICML (2019), ECAI (2023), AusDM (2018), AAAI (2019, 2020, 2021), AECT (2019), TIST (2020), MBEC (2022)

HONOURS AND AWARDS

- *Individual*
 - **Molecular Basis of Disease (MBD)** Ph.D. Fellowship at Georgia State University
 - **Graduate Research Assistantship:** Fully funded scholarship for Ph.D. studies at Georgia State University
 - **NOP Scholarship:** Fully funded scholarship for MS computer science at LUMS (~ 1.6 Million PKR)
 - **Speed Programming:** Secured first position in Speed Programming competition held at UET Peshawar in 2015
- *As a member of a team*
 - **Grant from SUPARCO:** Won the grant to design "Telecommand Receiver Unit" for PNSS-1 Satellite which was part of Pakistan National Student Satellite Mission in 2014
 - **Best Android Application:** Awarded 1st prize by Government of KPK in 2015
 - **Best Business Model:** Awarded 1st prize by Tie Islamabad, Peshawar chapter in 2015
 - **Government Grant:** For Emergence Rescue Application (worth ~ 1 million PKR) from Government of Pakistan

PUBLICATIONS

- *Journals (13)*
 - 13 Zahra Tayebi*, **Sarwan Ali***, Taslim Murad, Imdadullah Khan, and Murray Patterson, "PseAAC Protein Encoding for TCR Protein Sequence Classification." Accepted at Elsevier Computers in Biology and Medicine Journal (CIBM) (2023)
Impact Factor: 7.7
* Equal Contribution
 - 12 **Sarwan Ali**, Murray Patterson, "Improving Isomap Efficiency with RKS: A Comparative Study with t-SNE on Bioinformatics-based Protein Sequences." Accepted at Multidisciplinary Scientific Journal (J) (2023)
 - 11 Taslim Murad*, **Sarwan Ali***, Imdadullah Khan, Murray Patterson, "Spike2CGR: An Efficient Method For Spike Sequence Classification Using Chaos Game Representation." Published at Springer Machine Learning Journal Track at ECML-PKDD (2023)
Impact Factor: 7.5
* Equal Contribution
 - 10 **Sarwan Ali**, Bikram Sahoo, Alexander Zelikovsky, Pin-Yu Chen, Murray Patterson, "Benchmarking Machine Learning Robustness in Covid-19 Genome Sequence Classification." Published at Nature Scientific Reports (2023) [PDF]
 - 9 Haris Mansoor*, **Sarwan Ali***, Imdadullah Khan, Naveed Arshad, Muhammad Asad, Safiullah Faizullah, "Short-Term Load Forecasting Using AMI Data." Published at IEEE Internet of Things (IoT) Journal (2023) [PDF]
Impact Factor: 10.6
* Equal Contribution
 - 8 Zohair Hassan, **Sarwan Ali**, Imdadullah Khan, Mudassir Shabbir, Waseem Abbas, "Computing Graph Descriptors on Edge Streams." Accepted at ACM Transactions on Knowledge Discovery from Data (TKDD) (2023) [PDF]
Impact Factor: 2.71
 - 7 **Sarwan Ali**, Bikram Sahoo, Muhammad Asad Khan, Alexander Zelikovsky, Imdad Ullah Khan, and Murray Patterson. "Efficient Approximate Kernel Based Spike Sequence Classification." To Appear at IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS (TCBB), (2022). [PDF]
Impact Factor: 3.71
 - 6 **Sarwan Ali**, Babatunde Bello, Prakash Chourasia, Ria Thazhe Punathil, Yijing Zhou, and Murray Patterson. "PWM2Vec: An Efficient Embedding Approach for Viral Host Specification from Coronavirus Spike Sequences." MDPI Biology (2022). [PDF]
Impact Factor: 5.079
 - 5 **Sarwan Ali**, Yijing Zhou, Murray Patterson. "Efficient Analysis of COVID-19 Clinical Data using Machine Learning Models", Medical & Biological Engineering & Computing (2022) [PDF]
Impact Factor: 2.602
 - 4 Zahra Tayebi, **Sarwan Ali**, Murray Patterson. "Robust Representation and Efficient Feature Selection Allows for Effective Clustering of SARS-CoV-2 Variants" Algorithms 14(12) (2021) [PDF]
Impact Factor: 2.267
 - 3 **Sarwan Ali**, Simone Ciccolella, Lorenzo Lucarella, Gianluca Della Vedova, Murray Patterson. "Simpler and Faster Development of Tumor Phylogeny Pipelines" Journal of Computational Biology [PDF]
 - 2 **Sarwan Ali**, Muhammad Haroon Shakeel, Imdadullah Khan, Safiullah Faizullah, and Muhammad Asad Khan. "Predicting Attributes of Nodes Using Network Structure." ACM Transactions on Intelligent Systems and Technology (TIST), (2020) [PDF]
 - 1 Muhammad Ahmad, **Sarwan Ali**, Juvaria Tariq, Imdadullah Khan, Mudassir Shabbir, and Arif Zaman. "Combinatorial trace method for network immunization." Information Sciences 519 (2020): 215-228. [PDF]
Acceptance rate: 22%

■ Conference Proceedings (33)

- 33 **Sarwan Ali**, Prakash Chourasia, Murray Patterson. "Expanding Chemical Representation with k-mers and Fragment-based Fingerprints for Molecular Fingerprinting." Accepted at International Conference on Information Management and Big Data (SimBig) (2023)
- 32 **Sarwan Ali**. "Beyond Accuracy: Measuring Representation Capacity of Embeddings to Preserve Structural and Contextual Information." Accepted at International Conference on Information Management and Big Data (SimBig) (2023)
- 31 Prakash Chourasia, Taslim Murad, Zahra Tayebi, **Sarwan Ali**, Imdad Ullah Khan, Murray Patterson. "Efficient Classification of SARS-CoV-2 Spike Sequences Using Federated Learning." Accepted at International Conference on Information Management and Big Data (SimBig) (2023)
- 30 Taslim Murad*, **Sarwan Ali***, Prakash Chourasia, Haris Mansoor, Murray Patterson. "Circular Arc Length-Based Kernel Matrix For Protein Sequence Classification." Accepted at IEEE International Conference on Big Data
Acceptance rate: 17%
*Equal Contribution
- 29 Usama Sardar, **Sarwan Ali**, Muhammad Sohaib Ayub, Muhammad Shoaib, Khurram Bashir, Imdad Ullah Khan, Murray Patterson, "Sequence-Based Nanobody-Antigen Binding Prediction." Published at *International Symposium on Bioinformatics Research and Applications (ISBRA)*, 2023
- 28 **Sarwan Ali**, Pin-Yu Chen, Murray Patterson, "Unveiling the Robustness of Machine Learning Models in Classifying COVID-19 Spike Sequences." Published at *International Symposium on Bioinformatics Research and Applications (ISBRA)*, 2023
- 27 **Sarwan Ali**, Prakash Chourasia, Murray Patterson, "PDB2Vec: Using 3D Structural Information For Improved Protein Analysis." Published at *International Symposium on Bioinformatics Research and Applications (ISBRA)*, 2023
- 26 Prakash Chourasia, Taslim Murad, **Sarwan Ali**, Murray Patterson, "Enhancing t-SNE Performance for Biological Sequencing Data through Kernel Selection." Published at *International Symposium on Bioinformatics Research and Applications (ISBRA)*, 2023
- 25 **Sarwan Ali**, Haris Mansoor, Prakash Chourasia, Murray Patterson, "Hist2Vec: Kernel-Based Embeddings for Biological Sequence Classification." Published at *International Symposium on Bioinformatics Research and Applications (ISBRA)*, 2023
- 24 **Sarwan Ali**, Usama Sardar, Imdadullah Khan, Murray Patterson, "Efficient Sequence Embedding For SARS-CoV-2 Variants Classification." Published at *International Symposium on Bioinformatics Research and Applications (ISBRA)*, 2023
- 23 Zahra Tayebi, **Sarwan Ali**, Prakash Chourasia, Taslim Murad, Murray Patterson, "T Cell Receptor Protein Sequences and Sparse Coding: A Novel Approach to Cancer Classification." Accepted at International Conference on Neural Information Processing (ICONIP) 2023
- 22 **Sarwan Ali**, Babatunde Bello, Prakash Chourasia, Ria Thazhe Punathil, Pin-Yu Chen, Imdad Ullah Khan, Murray Patterson, "Virus2Vec: Viral Sequence Classification Using Machine Learning." Accepted at Conference on Health, Inference, and Learning (CHIL) [PDF]
Acceptance rate: 33% (Selected for Oral presentation (12% acceptance rate))
- 21 **Sarwan Ali**, Taslim Murad, Murray Patterson, "PCD2Vec: A Poisson Correction Distance Based Approach for Viral Host Classification." Published at International Joint Conference on Neural Networks (IJCNN) [PDF] [Slides]
- 20 Prakash Chourasia, **Sarwan Ali**, Murray Patterson, "Empowering Pandemic Response with Federated Learning for Protein Sequence Data Analysis." Published at International Joint Conference on Neural Networks (IJCNN) [PDF]
- 19 Taslim Murad, **Sarwan Ali**, Murray Patterson, "A New Direction in Membranolytic Anticancer Peptides Classification: Combining Spaced k-Mers with Chaos Game Representation." Accepted at International Joint Conference on Neural Networks (IJCNN) [Slides]
- 18 **Sarwan Ali**, Usama Sardar, Murray Patterson, Imdad Ullah Khan, "BioSequence2Vec: Efficient Embedding Generation For Biological Sequences." Published at Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) [PDF]
Acceptance rate: 17%
- 17 **Sarwan Ali**, Muhammad Ahmad, Muhammad Haroon Shakeel, Imdadullah Khan, Arif Zaman, and Asim Karim. "Efficient Data Analytics on Augmented Similarity Triplets." Published at IEEE International Conference on Big Data [PDF]
Acceptance rate: 18.6%
- 16 **Sarwan Ali**. "Evaluating COVID-19 Sequence Data Using Nearest-Neighbors Based Network Model." Published at IEEE International Conference on Big Data [PDF]
Acceptance rate: 18.6%
- 15 Haris Mansoor, **Sarwan Ali**, Shafiq Alam, Muhammad Asad Khan, Imdadullah Khan. "Impact Of Missing Data Imputation On The Fairness And Accuracy Of Graph Node Classifiers." Published at IEEE International Conference on Big Data [PDF]
Acceptance rate: 18.6%
- 14 Prakash Chourasia, **Sarwan Ali**, Murray Patterson, "Informative Initialization and Kernel Selection Improves t-SNE for Biological Sequences." Published at IEEE International Conference on Big Data [PDF]
Acceptance rate: 18.6%

- 13 **Sarwan Ali**, Taslim Murad, Murray Patterson. "PSSM2Vec: A Compact Alignment-Free Embedding Approach for Coronavirus Spike Sequence Classification." Published at International Conference on Neural Information Processing (ICONIP) 2022 [[PDF](#)]
 - 12 Taslim Murad*, Prakash Chourasia*, **Sarwan Ali***, Murray Patterson, "Hashing2Vec: Fast Embedding Generation for SARS-CoV-2 Spike Sequence Classification." Published at Asian Conference on Machine Learning (ACML), 2022 [[PDF](#)]
* Equal Contribution
Acceptance Rate: 32%
 - 11 **Sarwan Ali**, Taslim Murad, Prakash Chourasia and Murray Patterson, "Spike2Signal: Classifying Coronavirus Spike Sequences with Deep Learning" Published at IEEE International Conference on Big Data Computing Service and Applications (BigDataService), 2022 [[PDF](#)]
 - 10 Prakash Chourasia, **Sarwan Ali**, Simone Ciccolella, Gianluca Della Vedova and Murray Patterson. "Clustering SARS-CoV-2 Variants from Raw High-Throughput Sequencing Reads Data." In International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) (2022).
 - 9 Bikram Sahoo, **Sarwan Ali**, Alex Zelikovskiy, Pin-Yu Chen, Murray Patterson "Evaluating the Robustness of ML Models in SARS-CoV-2 Genome Sequences Generated Using TGS Technology" Published at *International Symposium on Bioinformatics Research and Applications (ISBRA)*, 2022
 - 8 **Sarwan Ali**, Murray Patterson, "Spike2Vec: An Efficient and Scalable Embedding Approach for COVID-19 Spike Sequences" in IEEE International Conference on Big Data (IEEE BigData), 2021 [[PDF](#)]
Acceptance Rate: 19.6 %
 - 7 Inaam Ul Hassan, Abdul Haseeb, **Sarwan Ali**. "Locally Weighted Mean Phase Angle (LWMPA) Based Tone Mapping Quality Index (TMQI-3)." In International Conference on Intelligent Vision and Computing (ICIVC), 2021 (**Best Paper Award**). [[PDF](#)]
 - 6 **Sarwan Ali**, Tamkanat-E-Ali, Muhammad Asad Khan, Imdadullah Khan, Murray Patterson. "Effective and scalable clustering of SARS-CoV-2 sequences." In International Conference on Big Data Research (ICBDR), 2021. [[PDF](#)]
 - 5 **Sarwan Ali**, Bikram Sahoo, Naimat Ullah, Alexander Zelikovskiy, Murray Patterson, Imdadullah Khan. "A k-mer Based Approach for SARS-CoV-2 Variant Identification." In International Symposium on Bioinformatics Research and Applications (ISBRA), 2021. [[PDF](#)]
 - 4 **Sarwan Ali**, Haris Mansoor, Naveed Arshad, and Imdadullah Khan. "Short Term Load Forecasting using Smart Meter Data." In Proceedings of Tenth ACM International Conference on Future Energy Systems, pp. 419-421. 2019. [[PDF](#)]
 - 3 **Sarwan Ali**, Maria Khalid Alvi, Safi Faizullah, Muhammad Asad Khan, Abdullah Alshangiti, and Imdadullah Khan. "Detecting DDoS Attack on SDN Due to Vulnerabilities in OpenFlow." International Conference on Advances in the Emerging Computing Technologies (AECT), pp. 1-6, 2019. [[PDF](#)]
 - 2 **Sarwan Ali**, Haris Mansoor, Imdadullah Khan, Naveed Arshad, Muhammad Asad Khan, and Safiullah Faizullah. "Fair allocation based soft load shedding." In Intelligent Systems Conference (IntelliSys), pp. 407-424, 2020. [[PDF](#)]
 - 1 Asad Ullah, **Sarwan Ali**, Imdadullah Khan, Muhammad Asad Khan, and Safiullah Faizullah. "Effect of Analysis Window and Feature Selection on Classification of Hand Movements Using EMG Signal." In Intelligent Systems Conference (IntelliSys), pp. 400-415, 2020. [[PDF](#)]
- *Posters and Workshops (4)*
 - 4 **Sarwan Ali**, Prakash Chourasia, Murray Patterson. "Anderson Acceleration For Bioinformatics-Based Machine Learning" in Knowledge Discovery in Healthcare Data (KDH@IJCAI) (2023)
 - 3 **Sarwan Ali**, "Information We Can Extract About a User From 'One Minute Mobile Application Usage'" in IEEE International Workshop on the Security, Privacy, and Digital Forensics of Mobile Systems and Networks (MobiSec 2023) under 2023 IEEE INFOCOM
 - 2 **Sarwan Ali**, Babatunde Bello, Murray Patterson "Host Specificity of the Coronaviridae through the Lens of Information Gain" in Annual Satellite Conference of RECOMB on Comparative Genomics (2022)
 - 1 **Sarwan Ali**, Bikram Sahoo, Pin-Yu Chen, Murray Patterson "Benchmarking Machine Learning Robustness in Covid-19 Genome Sequence Classification" in Robustness in Sequential Data, CVPR workshop (2022).

INVITED TALKS

- **Title:** Efficient Sequence Embedding For SARS-CoV-2 Variants Classification
Organization: Boston College, MA, USA

TEACHING AND PROFESSIONAL EXPERIENCE

- Georgia State University, Atlanta, USA
 - **Teaching Assistant**
 - * Fundamentals of Bioinformatics (Instructor: [Lee S. Katz](#))

Fall 2022, Spring 2023

- * CSC 2720 Data Structures (Instructor: [Shiraj Pokharel](#)) Spring 2022
- * CSC 4760 Big Data Programming (Instructor: [Peng Wang](#)) Summer 2021
- * CSC 2720 Data Structures (Instructor: [Jun Yi](#)) Fall 2021

- Lahore University of Management Sciences, Lahore, Pakistan

- **Teaching Assistant**

- * CS 210 Discrete Mathematics Fall 2017
 - * CS 510 Design & Analysis of Algorithm Fall 2018
 - * CS 5312 Big Data Analytics Spring 2019
 - * CS 5312 Big Data Analytics Spring 2020

- **Web Developer**

- * High Performance Computing Lab July 2017 - September 2017

- University of Engineering and Technology, Peshawar, Pakistan

- **Android Instructor**

June 2016 - July 2016

- Freelancing

- **Software Development**

January 2014 - Present

COURSES

References

- | | | | |
|---------------------------------------|-----------------------|------------------------|--------------------------------|
| ▪ Advanced Machine Learning | ▪ Data Mining | ▪ Dr. Pablo Moscato | pablo.moscato@newcastle.edu.au |
| ▪ Image and Video Encoding | ▪ Advance OS | ▪ Dr. Murray Patterson | mpatterson30@gsu.edu |
| ▪ Design and Analysis of Algorithms | ▪ Applied Probability | ▪ Dr. Imdadullah Khan | imdad.khan@lums.edu.pk |
| ▪ Advance Computer Architecture | ▪ Software Reuse | ▪ Dr. Iftikhar Ahmad | ia@uetpeshawar.edu.pk |
| ▪ Software Engineering for Smart Grid | ▪ Big Data Analytics | ▪ Dr. Alex Zelikovskiy | alexz@gsu.edu |

ADDITIONAL SKILLS AND EXPERIENCE

- **IT SKILLS:** R, Matlab, Java, Weka, C++, Python
- **SPORTS:** Basketball (won inter-school basketball championship), Cricket (won inter-departmental cricket championship)
- **VOLUNTEER WORK:** Worked as instructor at Special school involving children with disabilities.