SARWAN ALI

Department of Computer Science
Georgia State University
Github Profile [1] , Google Scholar Profile [2]

Tel: +1 (404)-510-5915
Email: sali85@student.gsu.edu
Website [3] , ORCID ID [4]

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)
    Thesis URL: https://dl.acm.org/doi/abs/10.1145/3442390

- University of Engineering and Technology (UET), Peshawar   August 2012 - June 2016
  B.S (Computer Science)

  - **BS Thesis Title:** Cache Replacement Algorithm
    Thesis URL: https://arxiv.org/abs/2107.14646

RESEARCH EXPERIENCE

- Georgia State University, Atlanta, GA, USA   January 2021 - Present

  - ***Graduate Research Assistant, Advisor: Murray Patterson*** [5]***, Alexander Zelikovsky*** [6]
    * 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*** [7]
    * 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*** [8]
    * 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*** [9]
    * Benchmarking machine learning robustness for noisy biological sequences.

- University of Newcastle, NSW, Australia   March 2020 - Dec. 2021
  - ***Research Collaboration, Advisor: Pablo Moscato*** [10] ***and Luke Mathieson*** [11]
    * 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** [12]
    * 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***

[1] www.github.com/sarwanpasha
[2] https://scholar.google.com/citations?user=9dtXSoAAAAAJ&hl=en
[3] https://sarwanpasha.github.io/
[4] www.orcid.org/0000-0001-8121-2168
[5] https://scholar.google.com/citations?user=49aQ7scAAAAJ
[6] https://scholar.google.com/citations?user=UzCRNJwAAAAJ&hl=en
[7] https://scholar.google.com/citations?user=KZL-4GEAAAAJ
[8] https://scholar.google.com/citations?user=_lxx4ycAAAAJ&hl=en
[9] https://scholar.google.com/citations?user=jxwlCUUAAAAJ&hl=en
[10] https://www.newcastle.edu.au/profile/pablo-moscato
[11] https://www.uts.edu.au/staff/luke.mathieson
[12] https://scholar.google.com/citations?user=aMwi6UwAAAAJ&hl=en

* ✱ **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), 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 ($\sim$ 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 $\sim$ 1 million PKR) from Government of Pakistan

PUBLICATIONS

- *Journals (10)*

  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
  URL: https://www.nature.com/articles/s41598-023-31368-3

  9 Haris Mansoor, **Sarwan Ali**, Imdadullah Khan, Naveed Arshad, Muhammad Asad, Safiullah Faizullah, "Short-Term Load Forecasting Using AMI Data." Accepted at IEEE Internet of Things (IoT) Journal
  Impact Factor: 9.936

  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)
  Impact Factor: 2.713

  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).
  Impact Factor: 3.71
  URL: https://ieeexplore.ieee.org/abstract/document/9891853

  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).
  Impact Factor: 5.079
  URL: https://www.mdpi.com/2079-7737/11/3/418

  5 **Sarwan Ali**, Yijing Zhou, Murray Patterson. "Efficient Analysis of COVID-19 Clinical Data using Machine Learning Models", Medical & Biological Engineering & Computing (2022)
  Impact Factor: 2.602
  URL: https://link.springer.com/article/10.1007/s11517-022-02570-8

  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)
  Impact Factor: 2.267
  URL: https://doi.org/10.3390/a14120348

  3 **Sarwan Ali**, Simone Ciccolella, Lorenzo Lucarella, Gianluca Della Vedova, Murray Patterson. "Simpler and Faster Development of Tumor Phylogeny Pipelines" Journal of Computational Biology
  URL: https://doi.org/10.1089/cmb.2021.0271

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) . **URL:** `http://dx.doi.org/10.1145/3442390`

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.
URL: `https://www.sciencedirect.com/science/article/pii/S0020025520300396?via%3Dihub`
Acceptance rate: 22%

- *Conference Proceedings (21)*

21 **Sarwan Ali**, Taslim Murad, Murray Patterson, "PCD2Vec: A Poisson Correction Distance Based Approach for Viral Host Classification." Accepted at International Joint Conference on Neural Networks (IJCNN)

20 Prakash Chourasia, **Sarwan Ali**, Murray Patterson, "Empowering Pandemic Response with Federated Learning for Protein Sequence Data Analysis." Accepted at International Joint Conference on Neural Networks (IJCNN)

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)

18 **Sarwan Ali**, Usama Sardar, Murray Patterson, Imdad Ullah Khan, "BioSequence2Vec: Efficient Embedding Generation For Biological Sequences." Accepted at Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)
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." Accepted at IEEE International Conference on Big Data
Acceptance rate: 18.6%
URL: `https://ieeexplore.ieee.org/abstract/document/10021104`

16 **Sarwan Ali**. "Evaluating COVID-19 Sequence Data Using Nearest-Neighbors Based Network Model." Accepted at IEEE International Conference on Big Data
Acceptance rate: 18.6%
URL: `https://ieeexplore.ieee.org/document/10020653`

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." Accepted at IEEE International Conference on Big Data
Acceptance rate: 18.6%
URL: `https://ieeexplore.ieee.org/document/10020694`

14 Prakash Chourasia, **Sarwan Ali**, Murray Patterson, "Informative Initialization and Kernel Selection Improves t-SNE for Biological Sequences." Accepted at IEEE International Conference on Big Data
Acceptance rate: 18.6%
URL: `https://www.computer.org/csdl/proceedings-article/big-data/2022/10020217/1KfT8iuxRS0`

13 **Sarwan Ali**, Taslim Murad, Murray Patterson. "PSSM2Vec: A Compact Alignment-Free Embedding Approach for Coronavirus Spike Sequence Classification." Accepted at International Conference on Neural Information Processing (ICONIP) 2022

12 Taslim Murad*, Prakash Chourasia*, **Sarwan Ali***, Murray Patterson, "Hashing2Vec: Fast Embedding Generation for SARS-CoV-2 Spike Sequence Classification." Accepted at Asian Conference on Machine Learning (ACML), 2022
* Equal Contribution
Acceptance Rate: 32%

11 **Sarwan Ali**, Taslim Murad, Prakash Chourasia and Murray Patterson, "Spike2Signal: Classifying Coronavirus Spike Sequences with Deep Learning" Accepted at IEEE International Conference on Big Data Computing Service and Applications (BigDataService), 2022

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" Accepted 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
URL: `http://arxiv.org/abs/2109.05019`
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**).
URL: `https://arxiv.org/abs/2109.08774`

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.
URL: `https://arxiv.org/abs/2108.08143`

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.
URL: `https://arxiv.org/abs/2108.03465`

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.
URL: `https://dl.acm.org/doi/abs/10.1145/3307772.3330173`

3 **Sarwan Ali**, Maria Khalid Alvi, Safi Faizullah, Muhammad Asad Khan, Abdullah Alshanqiti, 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.
URL: `https://www.researchgate.net/publication/338987710_Detecting_DDoS_Attack_on_SDN_Due_to_Vulnerabilities_in_OpenFlow#fullTextFileContent`

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.
URL: `https://doi.org/10.1007/978-3-030-55187-2_32`

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.
URL: `https://doi.org/10.1007/978-3-030-55190-2_30`

- *Posters and Workshops (3)*

    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   August 2022 - December 2022
        * CSC 2720 Data Structures   January 2022 - June 2022
        * CSC 4760 Big Data Programming   June 2021 - August 2021
        * CSC 2720 Data Structures   August 2021 - December 2021

- Lahore University of Management Sciences, Lahore, Pakistan

    - *Teaching Assistant*
        * CS 210 Discrete Mathematics   August 2017 - January 2018
        * CS 510 Design & Analysis of Algorithm   August 2018 - January 2019
        * CS 5312 Big Data Analytics   January 2019 - June 2019
        * CS 5312 Big Data Analytics   January 2020 - June 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

- Advanced Machine Learning
- Image and Video Encoding
- Design and Analysis of Algorithms
- Advance Computer Architecture
- Software Engineering for Smart Grid
- Data Mining
- Advance OS
- Applied Probability
- Software Reuse
- Big Data Analytics

References

- Dr. Pablo Moscato — pablo.moscato@newcastle.edu.au
- Dr. Murray Patterson — mpatterson30@gsu.edu
- Dr. Imdadullah Khan — imdad.khan@lums.edu.pk
- Dr. Iftikhar Ahmad — ia@uetpeshawar.edu.pk
- Dr. Arif Zaman — arifzaman@gmail.com

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.