Sumukh **Vasisht Shankar**

New Haven, Connecticut, USA

📧 sumukhmys1999@gmail.com | 🏠 www.svs.codes | 💻 sumukh-vasisht | 🔗 sumukh-vasisht-s | 👤 Sumukh VS

Education

Master of Science in Data Scinece   *Aug 2021 - May 2023*

NORTHEASTERN UNIVERSITY - GPA: 3.92/4.00   *Boston, MA*

Relevant Courses: Data Management and Processing, Supervised and Unsupervised ML, Geometric Deep Learning, Natural Language Processing

Bachelor of Engineering in Information Science   *Aug. 2017 - Aug 2021*

NATIONAL INSTITUTE OF ENGINEERING - GPA: 9.05/10.00   *Mysore, India*

Relevant Courses: Big Data Analytics, Recommender Systems, Artificial Intelligence, Graph Theory, Neural Networks

Research Experience

Yale University - Cardiovascular Data Science Lab   *New Haven, CT*

DATA SCIENTIST   *May 2023 - Present*

As a Data Scientist at the Cardiovascular Data Science Lab, Yale University, I leverage machine learning and natural language processing to advance cardiovascular data science research. My work encompasses diverse research projects, including Generative Adversarial Networks (GANs) for synthetic data generation, Retrieval-Augmented Generation (RAG) methods, and machine learning models for ECG image analysis. In addition to research, I build robust data infrastructures, including data architecture, mobile applications, and websites, to recruit research participants and deploy models on publicly accessible platforms. ***PI: Dr. Rohan Khera***

Northeastern University - Geometric Learning Lab   *Boston, MA*

RESEARCH FELLOW   *Aug 2022 - Sep 2023*

As a part of the Khoury Research Apprenticeship Program in Fall 2022 and Spring 2023, I focused on applying Geometric Deep Learning and Equivariant Neural Networks to fluid mechanics and material sciences. My research involved analyzing the symmetries inherent in fluid dynamics and optics, and developing scale, translation, and rotation symmetric equivariant neural networks that integrates physical priors, enhancing physical realism and data efficiency. ***PI: Dr. Robin Walters***

Northeastern University - Cybersecurity and Privacy Institute   *Boston, MA*

RESEARCH ASSISTANT   *Jan 2022 - Aug 2022*

Designed, collected, and analyzed Internet measurement campaigns to illuminate data localization practices in the European Union. Also implemented proof-of-concept models for various research papers and developed machine learning models to predict the physical locations of IP addresses. ***PI: Dr. David Choffness, Advisor: Dr. Alexander Gamero Garrido***

Network Science Institute - Lazer Lab   *Boston, MA*

DATA SCIENTIST   *Nov 2021 - May 2023*

Implemented and monitored scalable data pipelines to identify and characterize social behavior from social media data, primarily from Twitter. Developed and refined machine learning models to classify tweets as Misinformation or Disinformation. Created comprehensive Misinformation Dashboards for various events, providing real-time insights and visualizations. This work contributed to understanding the spread of mis-information on social media platforms. Collaborated with cross-functional teams to ensure the robustness and accuracy of the pipelines and models, enhancing the overall efficacy of the project. ***PI: Dr. David Lazer***

Publications

2024	**Development and Multinational Validation of a Novel Algorithmic Strategy for High Lp (a) Screening** Arya Aminorroaya, Lovedeep S Dhingra, Evangelos K Oikonomou, Seyedmohammad Saadatagah, Phyllis Thangaraj, <u>Sumukh Vasisht Shankar</u>, Erica S Spatz, Rohan Khera	*Nature Cardiovascular Research*
2023	**Controlling Dynamic Spatial Light Modulators using Equivariant Neural Networks** <u>Sumukh Vasisht Shakar</u>, Darrel D'Souza, Jonathan P Singer, Robin Walters. ICLR 2023	*ML for Materials - ICLR*

Pre-Prints

2024	**Automated Transformation of Unstructured Cardiovascular Diagnostic Reports into Structured Datasets Using Sequentially Deployed Large Language Models** Sumukh Vasisht Shankar*, Lovedeep Dhingra, Arya Aminorroaya, Philip Adejumo, Girish N Nadkarni, Hua Xu, Cynthia Brandt, Evangelos K Oikonomou, Rohan Khera	*Under Review - European Heart Journal - Digital Health*
2024	**A Novel Digital Twin Strategy to Examine the Implications of Randomized Control Trials for Real-World Populations** Phyllis M Thangaraj*, Sumukh Vasisht Shankar*, Sicong Huang, Girish Nadkarni, Bobak J Mortazavi, Evangelos K Oikonomou, Rohan Khera	*Under Review - Nature Communications*
2024	**Automated Diagnostic Reports from Images of Electrocardiograms at the Point-of-Care** Akshay Khunte, Veer Sangha, Evangelos K Oikonomou, Lovedeep Dhingra, Arya Aminorroaya, Andreas Coppi, Sumukh Vasisht Shankar, Bobak J Mortazavi, Deepak L Bhatt, Harlan M Krumholz, Girish Nadkarni, Akhil Vaid, Rohan Khera	*Under Review*
2024	**Equivariant Neural Networks for Controlling Dynamic Spatial Light Modulators** Sumukh Vasisht Shakar*, Rui Wang*, Darrel D'Souza, Jonathan P Singer, Robin Walters	*Under Review - JIMMI*
2024	**An Ensemble Deep Learning Algorithm for Structural Heart Disease Screening Using Electrocardiographic Images: PRESENT SHD** Lovedeep Dhingra, Arya Aminorroaya, Veer Sangha, Aline Pedroso Camargos, Sumukh Vasisht Shankar, Andreas Coppi, Murilo Foppa, Luisa Campos Caldeira Brant, Sandhi M Barreto, Antônio LP Ribeiro, Harlan Krumholz, Evangelos K Oikonomou, Rohan Khera	*Pre-print*
2024	**Retrieval-Augmented Generation for Extracting CHA$_2$DS$_2$-VASc Risk Factors from Unstructured Clinical Notes in Patients with Atrial Fibrillation** Philip Adejumo, Phyllis Thangaraj, Sumukh Vasisht Shankar, Lovedeep Dhingra, Arya Aminorroaya, Rohan Khera	*Pre-print*
2024	**Tracking the pre-clinical progression of transthyretin amyloid cardiomyopathy using artificial intelligence-enabled electrocardiography and echocardiography** Evangelos K Oikonomou, Veer Sangha, Sumukh Vasisht Shakar*, Andreas Coppi, Harlan Krumholz, Khurram Nasir, Edward J Miller, Cesia Gallegos-Kattan, Sadeer G Al-Kindi, Rohan Khera	*Pre-print*
2023	**Deep Learning-enabled Detection of Aortic Stenosis from Noisy Single Lead Electrocardiograms** Arya Aminorroaya, Lovedeep Dhingra, Veer Sangha, Evangelos K Oikonomou, Akshay Khunte, Sumukh Vasisht Shankar, Aline Pedroso Camargos, Norrisa Haynes, Ira Hofer, David Ouyang, Girish Nadkarni, Rohan Khera	*Pre-print*
2023	**CarDS-Plus ECG Platform: Development and Feasibility Evaluation of a Multiplatform Artificial Intelligence Toolkit for Portable and Wearable Device Electrocardiograms** Sumukh Vasisht Shankar, Evangelos K Oikonomou, Rohan Khera	*Pre-print*

Posters and Presentations

2024	**American Heart Association, Chicago** Using LLMs to extract tabular EHR data of heart failure cohorts across multicenter, non-interoperable health systems	*Moderated Digital Poster*
2024	**TMS AI in Materials International Conference, Cleveland, Ohio** Equivariant Neural Networks for Controlling Dynamic Spatial Light Modulators	*Presentation*
2024	**AI in Medicine Symposium, Yale University** CarDSPlus ECG Platform	*Poster*
2023	**AI in Healthcare Conference, New Orleans, LA** A Novel Digital Twin Strategy to Examine the Implications of Randomized Control Trials for Real-World Populations	*Poster*
2023	**The International Conference on Learning Representations (ICLR), Remote** Controlling Dynamic Spatial Light Modulators using Equivariant Neural Networks	*Poster*
2022	**Khoury Research Apprenticeship Showcase, Northeastern University** Controlling Dynamic Spatial Light Modulators using Equivariant Neural Networks	*Presentation*

Honors and Awards

Graduate Student Award for Outstanding Research Khoury College of Computer Sciences, NEU

2023 Bagged the graduate student award for outstanding research amongst 700+ peers graduating in 2023. *Boston, MA*

Research Apprenticeship Khoury College of Computer Sciences, NEU

2022-23 Awarded with a fully funded research project (14,000 USD) for the Fall 2022 and Spring 2023 semesters. *Boston, MA*

Global Nomination NASA SpaceApps Challenge 2020

2020 One of the final 20 teams in the challenge and received a Global Nomination for building an automated dust storm detector using deep neural networks. *Mysore, India*

Professional Experience

Walmart Inc. *Dallas, TX*

DATA ENGINEER III INTERN *May 2022 - Aug 2022*

Worked on an initiative to generate real-time reports on Walmart+ customer data, focusing on store signups. Built near real-time data streaming pipelines using Apache Spark and Apache Kafka to stream petabytes of click data into Google Cloud Storage. Developed comprehensive reports and interactive dashboards using Looker, providing actionable insights, and enhancing decision-making processes for the marketing and operations teams. This project improved data accessibility and reporting efficiency, allowing for timely analysis and response to customer behavior trends.

FICO *Bangalore, India*

SOFTWARE DEVELOPER INTERN *Jan 2021 - Jun 2021*

Implemented and integrated a Regular Expression Validation feature for a Mortgage Decision Management Platform tool, enhancing data accuracy and validation efficiency. Developed and deployed multiple APIs to facilitate the conversion of data into various formats for optimal storage across different database systems. Successfully increased the project's code coverage by over 11% through comprehensive regression testing, improving the reliability and maintainability of the codebase. This work significantly contributed to the robustness and functionality of the platform.

Antriksh Labs Pvt. Ltd. *Mysore, India*

RESEARCH INTERN *Jan 2020 - Dec 2020*

Initiated active methodological research practices to identify optimal visualization techniques for both structured and unstructured data. Conducted qualitative research and developed intelligent data pipelines for an AutoML SaaS product. Leveraged distributed systems such as Spark and Hadoop to accelerate data processing and statistical analysis, significantly enhancing the efficiency and effectiveness of data workflows. This work contributed to advancing the capabilities of the AutoML platform and improving its user experience.

Academic Projects

Classification of Movies into Genres through Analysis of Posters using CNN (DS5220 - NEU) Developed well performing custom CNN models that identifies the movie genre conveyed by its poster. The CNNs were built with sigmoid activation function along with RMSProp and ADAM optimizers.

Automated Dust Storm Detector using Deep Neural Networks Used Image Processing and Deep Neural Network model to detect dust storms using satellite image data through use of Stacking Ensemble by implementing Integrated Stacking methods with multilayer perceptron and Binary Classification. Predicted the impact of dust storm on population and vegetation of the region on a web interface which sends emergency warning to users along with respective news and tweets using NLP by using reverse geo-encoding technique.

Facial Recognition Based Smart Attendance System (Major Project - NIE) Developed a facial recognition algorithm using Deep Learning model and Extended Binary Patterns Histogram algorithm. The model developed achieved high accuracy (>95%) for both facial detection and recognition aspects of the software. Built an end-to-end software which maintained attendance in a central database along with the application. This reduced down the time taken by professors to mark attendance in a typical classroom setting by more than 50%.

Teaching

National Institute of Engineering *Mysore, India*

Data Analytics, IS0450 *Fall 2019*

Taught my peers the different types of Recommender Systems and helped them develop recommender systems over the course of two weeks under the guidance of Professor C K Vanamala.

NIE IEEE Computer Society *Mysore, India*

Game Development with Python - Resource Person *Fall 2019*

Rresource person for a 1-day workshop for a technical fest - Rubix 2019. Ttaught the ways of game development using Python and helped the attendees develop a total of six games.

NIE Summer of Code (NSOC 6.0) *Mysore, India*

Lead Resource Person *Summer 2019*

Taught Progressive Web Application development and Data Mining to over 50 students over the course 7 sessions.

Community

2022-23	**Career Peer Advisor** Khoury College of Computer Sciences, NEU Career peer advisor for MSDS-Align students where I mentored DS students about different career paths in the data field.	*Boston, MA*
2020	**Organizer** Biblus - 2020 One of the organizing members of Biblus 2020 - A national level paper presentation competitions held at NIE, Mysore.	*Mysore, India*
2020	**Organizer** IEEE Women in Engineering - Panel Discussion One of the organizing members of a panel discussion on the topic - 'Courage to Lead: Inner Dimensions of Leadership' as a part of Women's Day Celebrations of 2020.	*Mysore, India*
2019-20	**Student Activities Committee Coordinator** NIE IEEE Student Branch Scouted for and invited Resource Persons for various technical and social activities of the organization involving over 300 members. Oversaw various activities and co-ordinated with companies and arranged Industrial Visits to DELL EMC and CISCO.	*Mysore, India*
2019-20	**Teaching Volunteer** U&I Volunteerd at an NGO for almost an year to teach Mathematics and English to underprivileged children on the weekends.	*Mysore, India*
2019-20	**Website Committee Convener** TechNIEks 2019 & 2020 Was a part of the committee that developed websites for the college's techn-cultural fest techNIEks in 2019 and 2020.	*Mysore, India*
2018-19	**Student Branch Networking Coordinator & Digital Design Lead** NIE IEEE Student Branch Designed posters, banners, certificates, videos, magazines, and newsletters for the student club. Networked with other IEEE Student branches and reported all activities done by NISB to the IEEE Bangalore Section.	*Mysore, India*

Technical Skills

Languages	Python, R, SQL, C++, C, HTML, CSS, Javascript, Angular, React, Node.js, Vue.js
Database	MySQL, PostgreSQL, Google Firebase, Mongo DB, AWS
Data Science	Supervised and Unsupervised Machine Learning, Deep Learning, Neural Networks, Natural Language Processing
Data Engineering Tools	Spark, Hadoop, Kafka, Looker, Tableau, HDFS, Zookeeper