Sonish Sivarajkumar

Email: sos86@pitt.edu **Phone:** +1 412 478 8959

Website | LinkedIn | GitHub | Google Scholar

RESEARCH INTERESTS

Natural Language Processing, Information Retrieval, Information Extraction, Few/Zero-shot Learning, Machine Learning, Deep Learning, Biomedical Informatics, Electronic Health Records(EHR), Real World Evidence(RWE), Explainable AI

EDUCATION

PhD in Intelligent Systems
University of Pittsburgh | 2021-Present — Pittsburgh, PA
Major: Informatics

Master's in Intelligent Systems
University of Pittsburgh | 2021-2022 — Pittsburgh, PA
Major: Informatics

Bachelor's in Electrical Engineering
APJ Abdul Kalam Technological University | 2016-20 — India
Government Engineering College - Thrissur

EXPERIENCE

Graduate Student Researcher | ***August 2021 – Present***
University of Pittsburgh, Intelligent Systems Program

- Areas: ML / AI, Information Retrieval, Information Extraction, few-shot/zero-shot learning, Healthcare NLP, Interpretable AI, Clinical Decision Making, Cancer Genomics
- Developing next-generation information retrieval (IR) frameworks to facilitate clinical and translational research
- Developing NLP algorithms for clinical knowledge discovery based on regular expressions, machine learning and deep learning to identify relevant information from 100,000+ clinical EHRs(Electronic health records) and patient notes.
- Working on Zero/Few-shot learning techniques such as prompt learning, Siamese Neural Networks(SNN) for Healthcare NLP applications to change the paradigm from "Deep learning" to "Deep thinking".
- End-to-end disease study from EHRs using information extraction and clinical NLP.
- Designed clinical machine learning systems using K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), Logistic Regression, Recurrent neural network(RNN), LSTM, Generative adversarial networks(GAN), AutoEncoders(AE), and fine-tuned Deep learning PLMs (Pre-trained Language Models) like BERT, BioBERT, Clinical BERT, etc.
- Worked on building AI tools for Cancer Genomics and Spatial Transcriptomics on Visium breast cancer data.

gRED AI Predictive Analytics Research Intern | ***May – July 2022***
Genentech (Roche), San Francisco

- Part of the Early Clinical Development Informatics(ECDi) team working on Clinical Operations in trial design, building predictive tools, and improving the drug and target/biomarker discoveries.

- Developed predictive clinical trials site recommendation tool, using advanced AI and NLP techniques.
- Responsible for developing a vector space model for Roche internal clinical trials sites and PIs (datasets: Citeline, AACT, ClincalTrials.gov, Roche Internal data)
- Implemented and tested this clinical trial site embedding based Information Retrieval system, with primary focus on Diversity and Inclusion.

Data Scientist | ***May 2020 – August 2021***
IQVIA, IQVIA AI Center of Excellence(CoE), India

- Areas: Machine Learning, Big Data, Time Series Analysis, Health Care Analytics, NLP
- Worked on "Country Patient Analytics" project, which is a big data tool for doing custom analytics on clinical and patient level RWE and EHR data.
- Worked on building Clinical AI and analytics systems using Real World Evidence(RWE) and Electronic Health Records(EHR) data
- Led a small team of 4 for completing an end-to-end Clinical trials pipeline automation project using NLP and deployed the application in IQVIA's private cloud.
- Worked on a POC for segmentation and targeting of Healthcare providers (HCPs)
- Led Apache airflow migration of the big data scheduler and cloud integration and deployment of an AI and Analytics platform
- Was awarded as the Best Employee (IQVIA Impact Award) in 2021 by a Senior Director of IQVIA

Data Science Intern | ***August 2019 – May 2020***
Fractal, India

- Areas: Risk Analytics, Customer Analytics, Data Analytics, Deep Learning
- Worked on Credit risk modeling and risk analytics for one of the largest financial companies in India.

Co-Founder and Head of Data Science | ***August 2017 – May 2019***
StartUp India

- Areas: Intelligent Systems, Machine Learning, Deep Learning, Robotics
- Head of the AIML team of 4 members. Developed a set of AI and Robotics solutions on Healthcare, which was acquired by Rajagiri Medical College, Kochi, India.

PUBLICATIONS

- **Sivarajkumar, Sonish**, and Yanshan Wang. "HealthPrompt: A Zero-shot Learning Paradigm for Clinical Natural Language Processing." *arXiv preprint arXiv:2203.05061* (2022). Accepted for AMIA Symposium 2022, as one of 8 finalists in Best Paper competition 2022.
- **Sivarajkumar, S.,** Viggiano, S., Oniani, D., Visweswaran, S., & Wang, Y. (2022). Extraction of Sleep Information from Clinical Notes of Alzheimer's Disease Patients Using Natural Language Processing.
- Oniani, David, **Sonish Sivarajkumar**, and Yanshan Wang. "Few-Shot Learning for Clinical Natural Language Processing Using Siamese Neural Networks." *arXiv preprint arXiv:2208.14923* (2022).
- Koyilot, Mufeeda C., Priyadarshini Natarajan, Clayton R. Hunt, **Sonish Sivarajkumar**, Romy Roy, Shreeram Joglekar, Shruti Pandita et al. "Breakthroughs and Applications of Organ-on-a-Chip Technology." *Cells* 11, no. 11 (2022): 1828.

- April Sagan, **Sonish Sivarajkumar**, Hatice Osmanbeyoglu "Computational methods for delineating spatially informed cell context-specific regulatory programs." *UPMC Cancer Retreat 2021.*

CONFERENCE PRESENTATIONS

- **Sivarajkumar, Sonish**, and Yanshan Wang. "HealthPrompt: A Zero-shot Learning Paradigm for Clinical Natural Language Processing. *Paper Presentation, AMIA Symposium 2022.*
- April Sagan, **Sonish Sivarajkumar**, Hatice Osmanbeyoglu "Computational methods for delineating spatially informed cell context-specific regulatory programs." *UPMC Cancer Retreat 2021*

TALKS

- Clinical NLP and Few/Zero-shot learning for Clinical Text Extraction. *Presented at: University of California-San Francisco (UCSF) Seminar series, August 2022*
- Few-shot and zero-shot Learning for Clinical Information Extraction. *Presented at: Merck Text Mining Task Force Seminar series, August 2022*
- Guest lecture on 'Programming in R' in Foundations of Health Informatics, University of Pittsburgh; *June 2022.*
- Explainable Natural Language Processing(NLP). *Presented at: Department of Biomedical Informatics, University of Pittsburgh; April, 2022.*
- Zero-Shot Learning for Clinical Natural Language Processing. *Presented at: Intelligent Systems Program AI Forum, University of Pittsburgh; February,2022.*
- Guest lecture on 'Natural Language Processing' in Foundations of Health Informatics, University of Pittsburgh; *February 2022.*

SKILLS AND INTERESTS

Skills and Interests: Machine Learning, Deep Learning, Natural Language Processing, Information Retrieval, Information Extraction, ETL of Electronic Health Records data, Clinical Decision Support, Knowledge Graph, Explainable AI, Knowledge Graphs
Languages: Python, R, Java, C, SQL, Git
Technologies: Docker, AWS, Joyent Triton Cloud, CI/CD, Data Engineering(ETL-Spark-Hive)
Libraries: Transformers, NLTK, Spacy, Pandas, Scikit-learn, Jupyter, Keras, Networkx, Tensorflow, Pytorch, Stellargraph, OpenPrompt, Pyspark

OTHER PROFESSIONAL ACTIVITIES

Editorial Activities

Journal of the American Medical Informatics Association (JAMIA) Student EditorialBoard Member | *2022-2023*

Peer Review

ICHI 2022 (IEEE International Conference on Healthcare Informatics)| *2022*
ICML 2022 (International Conference on Machine Learning)| *2022*
LREC 2022 (Conference on Language Resources and Evaluation) | *2022*

Workshops

Co- organizing the AMIA 2022 NLP Working Group Pre-Symposium | *Washington,DC, Nov 5, 2022*

Volunteering

Translational Bioinformatics Year-in-Review team, AMIA Informatics Summit | *2021 &2022*
Student Volunteer, AMIA Annual Symposium *2022*

Memberships

American Medical Informatics Association (AMIA) | *2020-Present*
International Society of Computational Biology (ISCB) | *2021-Present*
Institute of Electrical and Electronics Engineers (IEEE)| *2019-Present*

AWARDS

- Fellowship – School of Computing and Information, University of Pittsburgh | *2021-2022*
- Fellowship – School of Computing and Information, University of Pittsburgh | *2022-2023*
- AMIA 2022 Best Paper Award Finalist | *2022*