

Balu Bhasuran, PhD

Tallahassee, FL 32306 • balubhasuran08@gmail.com

<https://balubhasuran.github.io/> • <https://github.com/balubhasuran/>

Experience

Florida State University Postdoctoral researcher

Tallahassee, FL
Jul '23 – present

Mentor: Zhe He, PhD (zhe.he@cci.fsu.edu)

- Supervised investigation of LLMs response to patient-centric lab test result-related questions.
- Investigation into Retrieval augmented Generation (RAG) based LLM systems for patient-centric questions
- Development of a machine learning prediction model for pediatric transplantation rejection using EHR and transplant registry data
- Investigation of LLMs responses to clinical case report-based differential diagnosis using lab test data.
- Investigation of the role of seasonality in lab test results of Alzheimer's and Dementia patients using EHR data.

University of California, San Francisco Postdoctoral researcher

San Francisco, CA
Jan '20 – Jul '23

Mentor: Vivek Rudrapatna, MD, PhD (Vivek.Rudrapatna@ucsf.edu)

- Trained and clinically tested an early diagnosis prediction model for a rare disease using multicenter electronic health records (UCSF and UCLA)
- Developed a time-to-event model for Non-alcoholic steatohepatitis (NASH) using EHR
- Developed a natural language processing algorithm for identifying NASH patients from clinical notes
- Developed a text classification algorithm for Mayo endoscopic sub-scores using clinical notes.
- Patent pending for rare disease early diagnosis using multicenter electronic health records
- Patent pending for FDA collaboration on the use of LLMs for adverse event detection from clinical notes.

DRDO-BU Center for Life Sciences Junior and Senior Research Fellow

India
Jul '14 – Jun '2018

Mentor: Jeyakumar Natarajan, PhD (n.jayakumar@yahoo.co.in)

- Machine learning models text mining discovery of High-Altitude Diseases
- Three webservers of biomedical information extraction (Indian Copy Right)

Education

Bharathiar University

PhD, Computational Biology

India
Sep '14 – Dec '20

Thesis: Biomedical Text Mining Approaches: Applications in Disease Entity Recognition, Gene-Disease Association Extraction and Knowledge Discovery

Topics: Information Extraction, Biomedical Informatics

Advisor: Prof. Jeyakumar Natarajan, PhD (n.jayakumar@yahoo.co.in)

Mahatma Gandhi University

MCA, Computer Applications

India
Oct '10 – Mar '13

Thesis: Data Mining to predict efficient allocations in electrical grids

Relevant Coursework: Data structures, Algorithm Analysis and Design

Skills

Frameworks: Pytorch, Tensorflow, Scikit-learn, Pandas, SQL, AutoML, PySpark, Numpy, Scipy

Infrastructure: Unix-based SGE high performance computing cluster, Standalone GPUs

Research and Travel Grants

National Institute of Health

Jul '24 – Dec '24

Principal Data Scientist, Prediction of Health Outcomes and Adverse Events in Pediatric Organ Transplantation in Florida
Grant number R21LM013911

AHRQ

Mar'24-Feb'29

Principal Data Scientist, LabGenie, a web-based tool intended to improve patients' engagement in managing and acting upon their lab test results.

Merck

Aug '22 – Jul '23

Principal Data Scientist Patient trajectories and risk prediction in patients with nonalcoholic steatohepatitis disease. Grant number UL1 TR001872

Alnylam

Jan '21 – Jul '22

Principal Data Scientist, Reducing diagnostic delays in Acute Hepatic Porphyria disease.
Grant number: K99LM014099, UL1 TR001872, UL1TR001881

US Food and Drug Administration (FDA)

Feb '21 – Aug '22

Collaborator, Improving adverse event detection related to biologic immunosuppressant use – a pilot study of the BERT deep learning model adapted to real-world clinical notes.

Defense Research and Development Organization (DRDO)

Data Mining and Text mining to Identify Enzymes that are potential candidates for high altitude diseases. Jul '14-Jun' 18

Mentorship and Thesis Committee

Undergraduate Research Opportunity Program (UROP), FSU

Aug'23 – Apr '24

LLM evaluation in differential diagnosis and lab test result-related questions

Rotation and PhD students

June '19 – present

Deep Learning for MRI Based Alzheimer's Disease Dementia Classification

Comparative Effectiveness of Ustekinumab and Vedolizumab in TNF-exposed Pediatric Patients with Ulcerative Colitis

Web application development for visualization of lab test results

Activities

Review Board Member, Graduate Studies Academic Regulations and Procedures, University of Nizwa

2023 – 2025

Publication chair, The 12th IEEE International Conference on Healthcare Informatics (IEEE ICHI 2024) 2024

Certifications

Microsoft Certified Professional (MCP) in Microsoft .NET Framework Application Development Foundation (Certification Id: 7292559)

Developing Enterprise Applications using VC#.Net, NIIT, New Delhi, India (Certificate No: 10DMZZZZ8845)

National Eligibility Test (UGC .NET) in Computer Science and Applications for Assistant Professor, Govt.of India (23699/NET)

Introduction to Python for Data Science Course, DataCamp (Id: 3,014,861)

Introduction to R Course, DataCamp (Id: 3,009,267)

Journal Reviewing

Bioinformatics, Oxford
Journal of the American Medical Informatics Association, Oxford
Artificial Intelligence in Medicine, Elsevier
Knowledge Based Systems, Elsevier
Scientific Reports, Springer Nature
NAR Genomics and Bioinformatics, Oxford
Gene, Elsevier
PLoS ONE, PLOS
IEEE Journal of Biomedical and Health Informatics, IEEE
BMC Bioinformatics, BioMed Central
BMC Medical Informatics and Decision Making, BioMed Central
Medical Oncology (MO), Springer Nature
Journal of Cannabis Research, BioMed Central
Biotechnology and Applied Biochemistry, Wiley
Journal of Biomolecular Structure and Dynamics, Taylor & Francis
Frontiers in Artificial Intelligence, Frontiers
Frontiers in Psychiatry, Frontiers
Frontiers in Research Metric and analysis, Frontiers
Applied Sciences, MDPI
Big Data and Cognitive Computing, MDPI
Mathematics, MDPI
Data, MDPI
Informatics, MDPI
Information, MDPI
AI, MDPI
Healthcare, MDPI
Future Internet, MDPI
Forecasting, MDPI
Electronics, MDPI
Machine Learning and Knowledge Extraction, MDPI
Computers, MDPI
Journal of Applied Bioinformatics & Computational Biology, SciTechnol
Artificial Intelligence in Health, ACS

Conference and Workshop Reviewing

American Medical Informatics Association, Annual Symposium 2023
American Medical Informatics Association, Clinical Informatics 2024
The IEEE/ACM international conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)
9th International Conference on Smart Computing and Communication (ICSCC 2023)

Awards

Indian Copy Right, Computer Software, SW-15957/2023, D-NER: Disease Name Recognizer from Literature, First Author
Indian Copy Right, Computer Software, SW-15958/2023, DisGeReExt: Disease Gene Relation Extractor from Literature, First Author
Indian Copy Right, Computer Software, SW-15974/2023, GD Miner: Gene-Disease Association Mining From Literature, First Author Filed Under: The Chairman, Defence Research & Development Organization (DRDO), GOVT OF INDIA, DTE OF ER & IPR
Best paper award, Data Mining, International Conference on Innovative Computing and Communication (ICICC-2018), Springer, India
Best poster award, Text Mining, International Symposium on Computational Biology and Bioinformatics (BioIndica-2016), India
Best poster award, Text Mining, Fifth Edition of National Workshop on Computer Vision, Image Processing Techniques and Data Analytics, India, 2015

Peer-reviewed Publications

1. Anna L Silverman*, Madhumita Sushil*, **Balu Bhasuran***, Dana Ludwig, James Buchanan, Rebecca Racz, Mahalakshmi Parakala, Samer El-Kamary, Ohenewaa Ahima, Artur Belov, Lauren Choi, Monisha Billings, Yan Li, Nadia Habal, Qi Liu, Jawahar Tiwari, Atul J. Butte, and Vivek A. Rudrapatna. Algorithmic identification of treatment-emergent adverse events from clinical notes using large language models: a pilot study in inflammatory bowel disease. *Journal of Clinical Pharmacology and Therapeutics*, 2024.
2. Perseus V. Patel, **Balu Bhasuran**, Vignesh G. Ravindranath, Melvin B. Heyman, Sofia G. Verstraete, Atul J. Butte, Vivek A. Rudrapatna. Real World Effectiveness of Ustekinumab and Vedolizumab in TNF-exposed Pediatric Patients with Ulcerative Colitis, *Journal of Pediatric Gastroenterology and Nutrition*, 2024
3. **Balu Bhasuran**, and Jeyakumar Natarajan. "DisGeReExT: a knowledge discovery system for exploration of disease-gene associations through large-scale literature-wide analysis study." *Knowledge and Information Systems* (2023): 1-25
4. Yadav, A., **Balu Bhasuran** & Oviya, I. R. (2023, November). A Novel Approach for Classifying DNA Barcodes Using Ensemble NLP Models. In 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE) (pp. 1-5). IEEE.
5. **Balu Bhasuran**. BioBERT and Similar Approaches for Relation Extraction. *Methods Mol Biol.* 2022;2496:221-235. doi: 10.1007/978-1-0716-2305-3_12. PMID: 35713867.
6. **Balu Bhasuran**. Combining Literature Mining and Machine Learning for Predicting Biomedical Discoveries. *Methods Mol Biol.* 2022;2496:123-140. doi: 10.1007/978-1-0716-2305-3_7. PMID: 35713862.
7. Natarajan, Jeyakumar, **Balu Bhasuran**, and Gurusamy Murugesan. "Big Data Analytics: A Text Mining Perspective and Applications in Biomedicine and Healthcare." In *Big Data Applications in Industry 4.0*, pp. 367-408. Auerbach Publications, 2022.
8. Maroli N, **Balu Bhasuran**, Natarajan J, Kolandaivel P. The potential role of procyanidin as a therapeutic agent against SARS-CoV-2: a text mining, molecular docking and molecular dynamics simulation approach. *J Biomol Struct Dyn.* 2022 Feb;40(3):1230-1245. doi: 10.1080/07391102.2020.1823887. Epub 2020 Sep 22. PMID: 32960159; PMCID: PMC7544928.
9. Sabenabanu Abdulkadhar, **Balu Bhasuran**, and Jeyakumar Natarajan. "Multiscale laplacian graph kernel combined with lexico-syntactic patterns for biomedical event extraction from literature." *Knowledge and Information Systems* (2020) 1-31
10. Maroli N, Kalagatur NK, **Balu Bhasuran**, Jayakrishnan A, Manoharan RR, Kolandaivel P, Natarajan J, Kadirvelu K. Molecular Mechanism of T-2 Toxin-Induced Cerebral Edema by Aquaporin-4 Blocking and Permeation. *J Chem Inf Model.* 2019 Nov 25;59(11):4942-4958. doi: 10.1021/acs.jcim.9b00711. Epub 2019 Nov 5. PMID: 31644276.
11. **Balu Bhasuran** & Natarajan, J. (2019). Distant supervision for large-scale extraction of gene-disease associations from literature using DeepDive. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 2* (pp. 367-374). Springer Singapore.
12. Subramanian D, **Balu Bhasuran**, Natarajan J. Genomic analysis of RNA-Seq and sRNA-Seq data identifies potential regulatory sRNAs and their functional roles in *Staphylococcus aureus*. *Genomics.* 2019 Dec;111(6):1431-1446. doi: 10.1016/j.ygeno.2018.09.016. Epub 2018 Oct 7. PMID: 30304708.
13. **Balu Bhasuran**, Natarajan J. Automatic extraction of gene-disease associations from literature using joint ensemble learning. *PLoS One.* 2018 Jul 26;13(7):e0200699. doi: 10.1371/journal.pone.0200699. PMID: 30048465; PMCID: PMC6061985.
14. **Balu Bhasuran**, Subramanian D, Natarajan J. Text mining and network analysis to find functional associations of genes in high altitude diseases. *Comput Biol Chem.* 2018 Aug;75:101-110. doi: 10.1016/j.compbiolchem.2018.05.002. Epub 2018 May 2. PMID: 29763853.
15. Murugesan G, Abdulkadhar S, **Balu Bhasuran**, Natarajan J. BCC-NER: bidirectional, contextual clues named entity tagger for gene/protein mention recognition. *EURASIP J Bioinform Syst Biol.* 2017 Dec;2017(1):7. doi: 10.1186/s13637-017-0060-6. Epub 2017 May 5. PMID: 28477208; PMCID: PMC5419958.
16. **Balu Bhasuran**, Murugesan G, Abdulkadhar S, Natarajan J. Stacked ensemble combined with fuzzy matching for biomedical named entity recognition of diseases. *J Biomed Inform.* 2016 Dec;64:1-9. doi: 10.1016/j.jbi.2016.09.009. Epub 2016 Sep 12. PMID: 27634494.

17. **Balu Bhasuran**, Schmolly K, Kapoor Y, Jayakumar NL, Doan R, Amin J, Meninger S, Cheng N, Deering R, Anderson K, Beaven SW, Wang B, Rudrapatna VA. Reducing diagnostic delays in Acute Hepatic Porphyria using electronic health records data and machine learning: a multicenter development and validation study. medRxiv [Preprint]. 2023 Aug 31:2023.08.30.23293130. doi: 10.1101/2023.08.30.23293130. PMID: 37693437; PMCID: PMC10491361. (Under review at JAMIA), 2024.
18. He, Z., **Balu Bhasuran.**, Jin, Q., Tian, S., Hanna, K., Shavor, C., ... & Lu, Z. (2024). Quality of Answers of Generative Large Language Models vs Peer Patients for Interpreting Lab Test Results for Lay Patients: Evaluation Study. *arXiv preprint arXiv:2402.01693*. (Under review at JMIR), 2024.
19. Silverman, A. L., **Balu Bhasuran**, Mosenia, A., Yasini, F., Gupta, S., Mardirossian, T., ... & Rudrapatna, V. A. (2022). Accurate, Robust, and Scalable Machine Abstraction of Mayo Endoscopic Subscores from Colonoscopy Reports. medRxiv, 2022-06. (Under review at IBD), 2024
20. **Balu Bhasuran**, Gurusamy Murugesan, and Jeyakumar Natarajan. "Literature Based Discovery (LBD): Towards Hypothesis Generation and Knowledge Discovery in Biomedical Text Mining." arXiv preprint arXiv:2310.03766 (2023). (Under review at ACM Computing Surveys), 2024

Conference Abstracts

21. **Balu Bhasuran**, Shadera Slatter, Gail Fernandes, Boshu Ru, Joe Yang, Xiao Zhang, Ravi Shankar, Jin Ge, and Vivek Rudrapatna. NASHDetection: A Natural Language Processing Method for Identifying Patients With Non-Alcoholic Steatohepatitis Using Clinical Notes. *Official journal of the American College of Gastroenterology | ACG*, 2023.
22. **Balu Bhasuran**, Shadera Slatter, Gail Fernandes, Boshu Ru, Joe Yang, Xiao Zhang, Ravi Shankar, Jin Ge, and Vivek Rudrapatna. Uncontrolled Diabetes and Hypertension Are Associated With the Risk of New-Onset Cirrhosis in Patients With Nonalcoholic Steatohepatitis *Official journal of the American College of Gastroenterology | ACG*, 2023.
23. Anna L Silverman*, Madhumita Sushil*, **Balu Bhasuran***, Dana Ludwig, James Buchanan, Rebecca Racz, Mahalakshmi Parakala, Samer El-Kamary, Ohenewaa Ahima, Artur Belov, Lauren Choi, Monisha Billings, Yan Li, Nadia Habal, Qi Liu, Jawahar Tiwari, Atul Butte, and Vivek Rudrapatna. Algorithmic identification of treatment-emergent adverse events from clinical notes using large language models: A pilot study in inflammatory bowel disease. *Official journal of the American College of Gastroenterology | ACG*, 2023.
24. Silverman, Anna L., **Balu Bhasuran**, Arman Mosenia, Fatema Yasini, Saransh Gupta, Narbe Mardirossian, Rohan Narain, Justin L. Sewell, Atul Butte, and Vivek A. Rudrapatna. "Accurate, Robust, And Scalable Abstraction Of Mayo Endoscopic Subscores From Colonoscopy Reports." In GASTROENTEROLOGY, vol. 162, no. 7, pp. S618-S619. 1600 John F Kennedy Boulevard, Ste 1800, Philadelphia, Pa 19103-2899 Usa: Wb Saunders Co-Elsevier Inc, 2022.