Balu Bhasuran, Ph.D

Bakar Computational Health Sciences Institute (BCHSI)
University of California, San Francisco (UCSF)

BCHSI, UCSF
490 Illinois St, Floor 2, #2114
San Francisco, CA 94143
Ph: 415-688-9702
Email: balu.bhasuran@ucsf.edu

Education

2014-2020 **Ph.D.** Computer Applications with Bioinformatics: Bharathiar University, India|2020

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

Advisor: Prof. Jeyakumar Natarajan

2010-2013 **MCA** Computer Applications: Mahatma Gandhi University, Kerala, India. | 8.19| 2013

2007-2010 **B.Sc.** Computer Science: University of Kerala, Kerala, India. |8.24|2010

Professional Experience

August 2021 – Till date: Post Doctoral Fellow, BCHSI at University of California, San Francisco (UCSF), USA
January 2021– August 2021: Affiliate, BCHSI at University of California, San Francisco (UCSF), USA

July 2014 – June 2018: Junior Research Fellow (JRF) , Major Research Project (MRP) at DRDO-BU Center for Life Sciences, Coimbatore, India
September 2014- January 2020: Full time Ph.D. research scholar in the field of Biomedical Literature Mining

Publications (Selected)

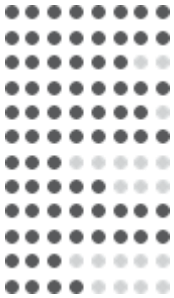
1. **Balu Bhasuran**,Gurusamy Murugesan, Sabenabanu Abdulkadhar, and Jeyakumar Natarajan. "Stacked ensemble combined with fuzzy matching for biomedical named entity recognition of diseases."***Journal of biomedical informatics*** 64 (2016): 1-9.**[Q1, IF 8.0]**
2. **Balu Bhasuran,** and Jeyakumar Natarajan. Automatic extraction of gene-disease associations from literature using joint ensemble learning**.** ***PloS one*.** 2018 Jul 26;13(7):e0200699.**[Q1, IF 3.24]**
3. **Balu Bhasuran,** Devika Subramanian, and Jeyakumar Natarajan. "Text Mining and Network Analysis to Find Functional Associations of Genes in High Altitude Diseases."***Computational Biology and Chemistry*** (2018).[**Q2**, **IF 3.73]**
4. **Bhasuran, Balu, and Jeyakumar Natarajan.** "DisGeReExT: a knowledge discovery system for exploration of disease–gene associations through large-scaleliterature-wide analysis study." **Knowledge and Information Systems** (2023): 1-25.[**Q1, IF: 2.53**]
5. 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 **[Q1, IF: 2.53]**
6. Subramanian, Devika, **Balu Bhasuran**, and Jeyakumar Natarajan. "Genomic analysis of RNA-Seq and sRNA-Seq data identifies potential regulatory sRNAs and their functional roles in Staphylococcus aureus." ***Genomics*** 111.6 (2019): 1431-1446. **[Q2, IF 4.31]**
7. Maroli, Nikhil, Naveen Kumar Kalagatur, **Balu Bhasuran**, Achuth Jayakrishnan, Renuka Ramalingam Manoharan, Ponmalai Kolandaivel, Jeyakumar Natarajan, and Krishna Kadirvelu. "Molecular Mechanism of T-2 Toxin-Induced Cerebral Edema by Aquaporin-4 Blocking and Permeation." ***Journal of chemical information and modeling*** 59, no. 11 (2019): 4942-4958. **[Q1, IF 6.162]**
8. Gurusamy Murugesan, Sabenabanu Abdulkadhar, **Balu Bhasuran**, and Jeyakumar Natarajan. "BCC-NER: bidirectional, contextual clues named entity tagger for gene/protein mention recognition." ***EURASIP Journal on Bioinformatics and Systems Biology*** 2017, no. 1 (2017): 7.
9. **Balu Bhasuran**, and Jeyakumar Natarajan. Distant supervision for large scale extraction of gene-disease associations from literature using DeepDive. International Conference on Innovative Computing and Communications, ***Lecture Notes in Networks and Systems*** 56, doi 10.1007/978-981-13-2354-6
10. Maroli, Nikhil, **Balu Bhasuran**, Jeyakumar Natarajan, and Ponmalai Kolandaivel. "The Potential role of Procyanidin as a Therapeutic Agent against SARS-CoV-2: A Text Mining, Molecular Docking and Molecular Dynamics Simulation Approach." (2020) ***Journal Of Biomolecular Structure And Dynamics***, 1-16 [**Q2, IF 3.392**]

IF as reported by Clarivate in its 2022 Journal Citation Reports

Q value based on the SCImago Journal Rank (SJR) indicator

Skills & Abilities

Programming & Software



- Python
- Java
- C
- C++
- Shell Script
- HTML &CSS
- Asp.NET
- SQL
- Cytoscape, Gephi
- Java Script
- R

Methodology

- Machine Learning
- Information Extraction
- Classification
- Prediction
- Distant Supervised Learning
- Named Entity Recognition
- Relation Extraction
- Knowledge Grpahs
- Network Analysis

Data Science

- Pytorch, TensorFlow,Keras
- DeepCheck,SHAP,mlextend
- AutoML,AutoGluon
- Matplotlib,Seaborn
- NLTK, SciSpacy,MedSpacy
- ScikitLearn
- cTAKES,metamap,Stanza
- MLjar,FlaML,CausalML
- Word2Vec,FastText

Web Servers

http: //www.biominingbu.org:8080/dner/
http: //www.biominingbu.org:8080/gdminer/

http: //www.biominingbu.org/DisGeReExT/
http: //www.biominingbu.org:8080/bccner/

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-June 2014)
- **Introduction to Python for Data Science Course**, DataCamp (Id: 3,014,861)
- **Introduction to R Course**, DataCamp (Id: 3,009,267)

Accomplishments

- **Junior Research Fellow**-JRF at DRDO -BU Center for Life Sciences 2014-2016
- Extended JRF at DRDO -BU Center for Life Sciences 2016-2017
- **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
- Developed innovative software packages that automated the Biomedical Text Mining Available at ***http://www.biominingbu.org***.
- **University Grand Commission National Eligibility** Test (**UGC .NET**) in the field of Computer Science for Assistant Professor in November 2014.

Experience Snapshot

Data Science and Machine Learning

- 7+ years of experience in text mining, natural language processing, machine learning applied to text data, information extraction and information retrieval
- Experience in working with Deep Learning platforms TensorFlow, Keras, PyTorch etc.
- Experience in working with Automated Machine Learning platforms AutoGluon, H2O, TPOT etc.
- Excellent understanding of machine learning techniques and algorithms, such as CRF, SVM, Graph Kernels, Decision Tree and Random Forests, etc. with ensemble learning using MALLET, LibSVM, EnsembleSVM etc.
- Experience in Python libraries for text data analyses and machine learning such as NLTK, Spacy, ScikitLearn, Word2Vec etc.
- Experience with common data science toolkits, such as R, Weka, NumPy, MATLABetc.
- Experience with data visualization tools, such as Cytoscape, Gephi, Matplotlib,Seaborn, Ploty, D3.js, GGplot
- Proficiency in using query languages such as SQL, PL/SQL etc.
- Good applied statistics skills, such as distributions, statistical testing, regression etc.
- Good scripting and programming skills in Python, C++, JAVA, R and C

Biomedical Domain

- Developed Named Entity Recognition Methodologies/Tools for Disease,Gene/Protein entities
- Developed Relation Extraction Methodologies/Tools for Gene-Disease Associations and Biomolecular Events
- Developed a disease centric methodology by finding functional associations of genes in high altitude diseases
- Developed a data mining pipeline for finding potential regulatory sRNAs and their functional roles in Staphylococcus aureus
- Experience with the following NLP tasks: tokenization, part of speech tagging, morphological decomposition, chunking, segmentation, regular expressions using OpenNLP, Stanford CoreNLP, BioLemmatizer etc.
- Experience working with biomedical open-source ontologies and terminologies such as MeSH, UMLS, SNOMED, PharmGKB, OMIM etc.
- Experience with publication databases such as PubMed, PubMed Central, ClinicalTrials.govetc.
- Experience working with open-source lexicons such as Dbpedia, WordNet etc.

Poster Presentations

a) **[Distant supervision for large scale extraction of gene-disease associations from literature using DeepDive**], International Conference on Innovative Computing and Communication (ICICC-2018), Springer,Guru Nanak Institute of Management, West Punjabi Bagh, New Delhi, **(Best paper award)**

b) [**Text Mining to Identify Gene-Gene Interactions of High Altitude Diseases**], International Symposium on Computational Biology and Bioinformatics, Kerala university (2016) (**Best poster award**)

c) [**Text Mining and Network Analysis to Identify Genes Related To High Altitude Diseases**], Fifth Edition of National Workshop on Computer Vision, Image Processing Techniques and Data Analytics, Amrita School of Engineering Amrita Vishwa Vidyapeetham, (2015) (**Best poster award**)

d) [**Deploying Biomedical Text Mining for Knowledge Discovery in Precision medicine**]International Conference on Bioscience and Bioinformatics (ICBB-2017), Bharathiar University, Coimbatore

Conferences & Workshops

1. Podium Presentation, AMIA 2023 Informatics Summit,March 13 - 16,Seattle, WA,#IS23
2. 17th International Conference on Bioinformatics (INCOB), Jawaharlal Nehru University, New Delhi, 26th – 28th September, 2018
3. Two-day Workshop on Big Data in Life Sciences, Department of Environmental Sciences, Bharathiar University, Coimbatore, Tamilnadu, 05th -06th September, 2018
4. International Conference on Phytomedicine, Department of Botany, Bharathiar University, Coimbatore, Tamilnadu, 29th - 31st August, 2018
5. National Conference on Computational Biology, DRDO-BU Center for Life Sciences, Bharathiar University Campus, Coimbatore, Tamilnadu, 26th June, 2018
6. International Conference on Innovative Computing and Communication (ICICC'18), Springer,Guru Nanak Institute of Management, West Punjabi Bagh, New Delhi, 05th -06th May, 2018
7. MOTIF, Department of Bioinformatics, Bharathiar University, Coimbatore, Tamilnadu, 2nd March, 2018
8. International Conference on Biosciences and Bioinformatics (ICBB'17), Bharathiar University, Coimbatore, Tamilnadu, 14th - 16th December, 2017
9. Third National Conference on Computational Biology, DRDO-BU Center for Life Sciences, Bharathiar University Campus, Coimbatore, Tamilnadu, 27th and 28th June, 2016
10. Science Communication Workshop, The Wellcome Trust/DBT India Alliance, Vinayaka Missions University, Salem, Tamilnadu, 21st September, 2015
11. National Seminar on Bioinformatics and Bio-Pharmaceuticals (NSBB'15), Department of Bioinformatics, Bharathiar University, Coimbatore, Tamilnadu, 3rd March, 2015
12. Chemical Weapons Convention (CWC) Awareness Program, department of chemicals and pharmaceuticals and Indian Chemical Council, Southern Region, 04th December, 2014

Personal Information

- Sex & Marital Status: Male & Single
- Age & DOB: 30 years, 07th Feb 1990.
- Nationality / Citizen: Indian
- Relocation: Willingness to relocate both Domestic and International

Profiles

Google Scholar ResearchGate LinkedIn

References

Kalpana Raja, PhD, MRSB, CSci
Faculty,
School of Medicine
Yale University
New Haven, CT, USA
kalpana.raja@yale.edu

Zeeshan Ahmed, Ph.D.
Assistant Professor of Medicine
Department of Medicine
Rutgers Robert Wood Johnson Medical School.
Rutgers Biomedical and Health Sciences.
Rutgers, The State University of New Jersey.
zahmed@ifh.rutgers.edu, (848) 932-5866

Jeyakumar Natarajan, Ph.D
Professor & Head
Department of Bioinformatics,
Bharathiar University
Coimbatore, Tamil Nadu, India – 641046
Mobile: +91 -9443494336
E-mail: n.jeyakumar@yahoo.co.in