BONESHWAR

B.S and M.S in Biological Sciences

boneshwar.vk20@gmail.com
Academic portfolio
LinkedIn

Summary

My primary research interests are applications of deep learning particularly developing architectures like graph neural networks, large language models, large reasoning models or large knowledge models, and generative pipelines involving GANs and diffusion models with a focus on applied research, especially in multi-omics data focused on personalized diagnosis and medicine for brain cancer. In the long term, I aspire to contribute to the open source research community by developing foundational interpretable models that are robust and can generalize across multiple domains and modalities within healthcare, progressing towards AGI.

Education

2020–2025 **B.S. in Biological Sciences**, *Indian Institute of Technology Madras*

M.S. in Biological Sciences (Specialization in Computational Biology), *Indian Institute of Technology Madras*, *CGPA: 8.12/10.0*

2018–2020 **Higher Secondary Education**, *Velammal Vidyalaya, Central Board of Secondary Education*, *Percentage: 95.0%*

Technical Skills

Programming
- **Programming Languages:** Python, C++, C#, R
- **Software and tools:** Gromacs, AutoDock, Weka, Pymol, VMD, Blender
- **Libraries and packages:** PyTorch, NetworkX, PyG, DGL, Gensim, Rpy2, Scikit-learn, Pandas, TensorFlow

Wet lab
- **Techniques:** Molecular cloning, Colony PCR, CFU quantification
- **Technologies & Instruments:** Thermocycler, Gel documentation, Spectrophotometer

Journals

Computers in Biology and Medicine — *PAIGT: Position aware graph transformer for single-cell RNA**. Authors: **Boneshwar V. K**, Vijayavallabh Jayamanikandan , Deepesh Agarwal, Bala Natarajan, and Babji Srinivasan. (Under review)

Computers & Chemical Engineering — *Inductive graph neural network framework for imputation of single-cell RNA sequencing data.* Authors: **Boneshwar, V. K.**, Deepesh Agarwal, Bala Natarajan, and Babji Srinivasan.

Research Experience

Project member, CRIS Lab

Ontology-Grounded LKM for disease diagnosis

Principal Investigators: Prof. Venkat Venkatasubramanian (Columbia University) (May 2025 – Present)

- Building interpretable, ontology-grounded Large Knowledge Model (LKM) for hierarchical diagnosis.
- Prototyping LKM with distilled chain-of-thoughts with rare disease-symptoms and neuro-oncology.
- Leverages **Llama** for safe, local & interpretable clinical decision support with transparent reasoning.
- Curated phenomics corpus via ***GA4GH Phenopackets***; engineered weighted disease–phenotype edges and KG[1] communities using ***MONDO/ORDO and HPO***[2] ***attributes***; implementing KG-aware inference with hierarchical reasoning and standards-aligned actions (LOINC/SNOMED CT/ICD-11).

Research Associate
HILCPS Lab

Foundational models using multi-omics based cell trajectory inference for Gliomas

Principal Investigators: Prof. Babji Srinivasan, Prof. Bala Natarajan (K-State) (Dec 2023 – May 2025)

- Built inductive frameworks for scRNA-seq to recover global topology and glioma-relevant patterns.
- Enhanced **cell-type annotation** & neoplastic–non-neoplastic separation across sparse scRNA atlases.
- Supports early, pre-clinical risk stratification & prevention by unifying glioma microenvironment profiles.
- Designed position–structure encodings (PSE) and multiscale embeddings integrating neighborhood states with cohort manifolds while preserving lineage and cell–cell communication patterns leveraging graph transformer aimed for recovering trajectories and pseudotime.

Research Intern
CPSWin Lab

Inductive scRNA Imputation GNN

Principal Investigator: Prof. Bala Natarajan (K-State) *(May 2024 – July 2024)*

- Impute scRNA-seq dropouts by reconstructing subgraph neighborhoods gene–gene co-expression.
- SOTA Metrics: *+60%* Silhouette, *+14.9%* ARI[3], *48%* runtime, *4.5%* L_1 median error vs baselines.
- Improves downstream cell-state clustering and strengthens scRNA pipelines for tumor biopsies.
- Developed inductive graph neural network for unsupervised clustering across cohorts to handle the inherent data sparsity associated with single-cell technologies.

AI Researcher
Sarvam AI

LLM QA and Indic OCR

Principal Investigators: Prof. Pratyush Kumar, Vivek Raghavan *(Nov 2023 – Jan 2024)*

- Built an end-to-end agent for large CSV QA and multilingual OCR to unlock large scale tabular data.
- SOTA on internal and Yahoo Finance datasets; OCR at *0.62-CER/0.71-WER*[4] with *80%* layout fidelity.
- Enables scalable digitization and QA of clinical registries and EHRs in low-resource scripts.
- Designed CSV QA agent; built Indic OCR (Hindi/Tamil/Telugu/Malayalam/Kannada); implemented RAG with hybrid retrieval and cross-encoder re-ranking.

Professional Experience

Associate Data Scientist

Convozen AI (NoBroker.com)

Data Science Team *(May 2025 – Present)*

- Delivered a low-latency speech-to-speech voice-bot with robust VAD in noisy, real-world call flows.
- VAD false positives reduced by *13%*; *900 ms* TTFB[5]; *1.6 s* average conversation turn.
- Enables AI based realtime conversational voice bots in healthcare and other customer care operations.
- Integrated DeepFilterNet v3 for neural spectral denoising and non-cascaded[6] Speech-to-speech collaborating with Google and Azure; exploring TTS for human vocal mimics—SOTA commercial voice bots mirroring human vocals and prosody with biomimetic vocal-tract cues (timbre, breath, coarticulation).

Autonomous Systems Engineer

10X, IITM Research Park

Artificial Intelligence Team *(May 2023 – Jul 2023)*

- Prototyped an autonomous wheelchair with SLAM navigation & multilingual conversational interface.
- Live demo to Google India's CEO and the Chief Justice of India; featured in ***Times of India***.
- Improves assistive mobility for neuro-rehabilitation and elder-care settings.
- Implemented mapping and navigation using IMU, encoders, LiDAR with ROS[7] (Hector SLAM, Cartographer) and a RASA-based chatbot for 5 major indic languages.

Research Projects

MaskedAlign: Seq-to-Seq MSA Transformer

Multiple Sequence Alignment for large scale genomes

Principal Investigator: Prof. Manikanda Narayanan *(Jan 2024 – May 2024)*

- Enhance generalization of MSA[8] on unseen gene families using a masked transformer using PyTorch.
- Higher alignment accuracy vs BetaAlign/ClustalW/MAFFT with runtime reduction up to *11%*.
- Speeds comparative genomics and variant discovery relevant to cancer cohort analyses.
- Built seq-to-seq model with masked language modeling and robust training for long genomic data.

Recombinant Plasmid Assembly

Recombinant Plasmid Engineering

Principal Investigator: Prof. Venkitasamy Kesavan *(Jan 2024 – May 2024)*

- Construct and verify recombinant plasmids for reliable downstream expression studies.
- Insert verification via restriction mapping; maintained clonal lineage with stable plasmid retention.
- Establishes a robust cloning backbone for functional assays like oncogene perturbation models.
- Executed miniprep, PCR and gel, BamHI/NdeI digests, ligation, and DH5α transformation with antibiotic selection and colony PCR monitoring clonal lineage across passages.

Latent code guided GANs[9] for image generation

Enforcing control over latent space in GANs

Principal Investigators: Prof. Babji Srinivasan, Prof. Rajagopalan Srinivasan (Dec 2023 – Mar 2024)

- Disentangle latent factors for controllable image synthesis and RGB→thermal translation.
- Metrics: *+6%* accuracy and consistent foreground–background separation over baselines.
- Enables privacy-preserving synthetic data for controlled image generation.
- Fine-tuned GANs to separate shape, size, texture & color with edge-guided translation for RGB to thermal.

Microbial Population Trajectory and Cell Dynamics

Microbial Dynamics Profiling

Coursework: BT2112 Microbiology Laboratory (Jan 2022 – May 2022)

- Quantify growth trajectories and stress responses across controlled environmental gradients.
- Derived stable phenotype panels and MIC[10] profiles across pH/salinity/temperature conditions.
- Reinforces reproducible phenotyping workflows transferable to microbiome QC in oncology pipelines.
- Performed OD tracking, hemocytometer counts, serial-dilution CFU[11], Gram staining, biochemical panels, enzyme assays, and antibiotic susceptibility (Kirby–Bauer, MIC).

Leadership and Outreach

CVI Club, CFI, IIT Madras

- Project member 2022-2023, developed an end-to-end traffic violation detection system (helmet, triple riding, over-speeding) using **YOLOv5**, achieving F1: 0.79 and 89% accuracy on Jetson Nano, deployed via CCTV infrastructure and incubated at **Nirmaan, IIT Madras**.

SEC, IIT Madras

- Actively served on IIT Madras election body, overseeing the Student General Elections 2021-2022 with 10,000+ electorate and upholding integrity in the Academic Affairs Secretary race.

Social Impact & NGOs

- **Omdena AI Singapore:** Implemented ML based pipeline for analysing workouts (pushups, situps, pullups)
- **Suvidha Foundation (NGO):** Designed framework for teaching ML to support weaker economic sections

Achievements and Extra-Curricular Activities

Martial Arts

- Secured **1st** place and **2nd** place at **World Unified Martial Arts Federation 2011 and 2012**
- Active participant of **Ryu Kyu Kobudo & Shonrinryu Karate International seminar and Okinawa karate and Kobudo Invitation championship 2013**

Cricket

- Lead Narmada hostel, IIT Madras in Schroeter 2024
- Lead the high school cricket team in Inter-District school tournament 2018, Tamil Nadu, secured 3^{rd} place in the tournament and being the highest run scorer among the team
- Represented Tamil Nadu at **Nationals 2015** held by united cricket federation of India.

Academics

- Secured **All India Rank 9 (AIR 9)** in the JEE MAINS B.Arch conducted in 2020.
- Secured top 1 percentile among 1,30,000 candidates in JEE Mains & Advanced 2020

**- ongoing/submitted, 1- Knowledge Graphs, 2- Human Phenotype Ontology, 3- Adjusted Rand Index, 4- Character Error Rate & Word Error Rate, 5- Time to first byte, 6-Model with no separate STT-LLM-TTS modules, 7- Robot Operating System, 8- Multiple Sequence Alignment, 9- Generative Adversarial Networks, 10- Minimum Inhibitory Concentration, 11- Colony-Forming Units*