# 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%*

## 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 — ***Guide: Prof. Venkat Venkatasubramanian, Columbia University***
*(May 2025 – Present)*

- Implementing an interpretable, ontology-grounded LLMs—**LKM (Large Knowledge Model)**—for hierarchical differential diagnosis, leveraging MONDO/ORDO disease taxonomies and HPO phenotypic abnormalities with MeSH-aligned crosswalks for literature traceability.
- Engineering a phenomics-centric training corpus via *GA4GH Phenopackets* and weighted disease–phenotype edges; optimizing QA policies via KG-derived disease–phenotype communities.
- Investigating KG-aware inference with LKMs with hierarchical losses; binding assays and clinical actions to *LOINC/SNOMED CT/ICD-11*, and emitting calibrated, reasoning COT and healthcare driven queries under Clinical Decision Support safe guardrails.

Research Associate, HILCPS Lab — ***Guides: Prof. Babji Srinivasan, IIT Madras  Prof. Bala Natarajan, Kansas State University***
*(Sep 2023 – May 2025)*

- Implemented a graph-transformer for scRNA-seq that inductively assembles cell–gene community subgraphs, capturing glioma-relevant transcriptomic programs and microenvironmental cues via position–structure encodings (PSE) to recover global topology with accurate lower order patterns.
- Synthesized multiscale embeddings that integrated neighborhood cellular states with cohort-level transcriptomic manifolds, enhancing **cell-type annotation** and neoplastic–non-neoplastic separation across sparse single-cell profiles spanning multiple gold-standard scRNA-seq atlases.
- Deployed structure-aware attention with degree & path encodings, preserving biological hierarchy such as lineage trajectories, cell–cell communication across diverse single-cell benchmarks.

Research Intern, CPSWin Lab — ***Guide: Prof. Bala Natarajan, Kansas State University*** *(May 2024 – Jul 2024)*

- Developed an inductive GNN-based framework for scRNA-seq *imputation* that learns from subgraph neighborhoods to reconstruct gene–gene co-expression under severe data level noise and sparsity.
- Preserved cellular manifold structure to improve downstream clustering across diverse single-cell cohorts uisng unsupervised learning.
- Metrics reported: 60% improvement in Silhouette score, 14.9% in Adjusted Rand Index, 48% in runtime, and 4.5% in $L_1$ median error over baseline models.

AI Researcher, Sarvam AI — ***Guides: Vivek Raghavan, Prof. Pratyush Kumar, Sarvam AI*** *(Nov 2023 – Jan 2024)*

- Designed an end-to-end **LLM QA agent for large CSVs**; evaluated across GPT-3.5, GPT-4, Llama 2, and Mistral-7B, achieving state-of-the-art results on our internal large-CSV benchmark and other datasets like Yahoo finance dataset.
- Built a **multilingual OCR pipeline** for non-UTF-8 Indic PDFs (Hindi, Tamil, Telugu, Malayalam, Kannada) using PyMuPDF, PyPDF2, and Unstructured; benchmarked with *0.62 CER/ 0.71 WER*, and *80% layout accuracy* for robust information extraction.
- Implemented RAG over FAISS & Pinecone with adaptive chunking, hybrid lexical–vector retrieval, and cross-encoder re-ranking; evaluated using **RAGAS** , *Answer Relevancy*, *latency* and *end-to-end QA accuracy*.

## Professional Experience

Associate Data Scientist — **Convozen AI (NoBroker.com)** *(May 2025 – Present)*

- Integrated **DeepFilterNet v3** for neural spectral denoising and enhancing human voice activity; refined VAD, reducing false positives by 13%
- Actively preview S2S (*GPT-4o Realtime*, *Gemini 2.5*) and TTS models (*XTTS v2*, *Orpheus TTS*) into a low-latency, non-cascaded voice-bot pipeline with TTFB of 900 ms and an average conversation turn of 1.6 seconds.

Autonomous Systems Engineer — **10X, IITM Research Park** *(May 2023 – Jul 2023)*

- Designed a wheelchair with mapping and navigation system using IMU, encoder, LiDAR with **ROS algorithms**- (Hector SLAM, Cartographer SLAM) and a multilingual chatbot supporting 5 Indic languages using RASA
- Demonstrated the working our product to **Google India's CEO and Chief Justice of India** featuring on Times of India for our innovative realtime autonomous navigation system in a wheelchair

## Research Projects

MaskedAlign: Multiple Sequence Alignment for large scale genomes — ***Guide: Prof. Manikanda Narayanan*** *(Jan 2024 – May 2024)*

- Developed **MaskedAlign, a novel seq-to-seq masked transformer to enhance Multiple sequence alignment** increasing generalization for large scale unseen gene sequences, masking
- Implemented the model using PyTorch achieving superior alignment accuracy and computational efficiency compared to SOTA methodologies like BetaAlign, ClustalW and MAFFT with reduction in runtime upto **13%**

Recombinant Plasmid Assembly — ***Guide: Prof. Venkitasamy Kesavan*** *(Jan 2024 – May 2024)*

- Constructed and verified recombinant plasmids by performing alkaline-lysis miniprep, PCR amplification with gel confirmation and extraction, restriction digestion with BamHI and NdeI followed by ligation, and heat-shock transformation into $CaCl_2$-competent *E. coli* (DH5$\alpha$) with antibiotic selection, colony PCR, and restriction mapping to confirm inserts while maintaining clonal lineage.
- Tracked propagation trajectory and plasmid retention across passages of confirmed clones under selection by periodic minipreps, analytical digests, and agarose-gel checks to document construct stability along the colony-to-construct lineage.

Latent code guided GANs for image generation

***Guides: Prof. Babji Srinivasan, Prof. Rajagopalan Srinivasan***
*(Dec 2023 – May 2024)*

- Fine-tuned GANs on a custom dataset to disentangle latent codes for shape, size, texture, and color, enabling precise control over image generation with clear foreground–background separation.
- Achieved a 6% accuracy gain by integrating RGB-to-thermal infrared translation through edge-guided GANs, enhancing fidelity of foreground objects while allowing flexible background modification.

Microbial Population Trajectory and Cell Dynamics

***Coursework: BT2112 Microbiology Laboratory***
*(Aug 2022 – Dec 2022)*

- Mapped growth trajectories and viability by standardized inocula, optical-density tracking, hemocytometer totals, serial-dilution CFU on countable plates, and Gram-stained morphology under oil immersion.
- Profiled passage-wise responses to controlled pH, salinity, and temperature, linking biochemical panels (IMViC, TSI, catalase, oxidase, urease), enzyme plate assays, and antibiotic susceptibility (Kirby–Bauer, MIC) to trajectory stability and functional state.

## Technical Skills

Wet lab skills

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

Programming skills

- **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

## 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**.

Students Election Commission, 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