



BONESHWAR V K

B.S and M.S in Biological Sciences

✉ boneshwar.vk20@gmail.com

in LinkedIn

Education

- 2023 – 2025 **M.S. in Biological Sciences (Specialization in Computational Biology)**, *Indian Institute of Technology Madras*, CGPA: 8.12/10
- 2020 – 2023 **B.S. in Biological Sciences**, *Indian Institute of Technology Madras (Integrated B.S. & M.S.)*

Professional Experience

Machine Learning Scientist

PopVax

Antigen Design Team, Computational R&D division

(Jun 2026 – Present)

- Developed generative models for protein structure generation using drifting and spherical latents.
- Finetuned RFDiffusion 3 and Proteinix V2 with DPO and GRPO.
- Built a novel confidence head for structure prediction metrics and developability metrics (PAE matrix, gibbs free energy, Isoelectric Point, and SAP score) to steer RFD3 & produce more plausible structures.
- Ordered several binders using the steering model to test in vitro validation.

Associate Data Scientist

Convozen AI (NoBroker Technologies Solutions)

Speech Team

(May 2025 – Jun 2026)

- Pretrained **XTTS & PiperTTS** for Indic voices with TTFB<300ms; RTF<0.15 (T4 & Raspberry Pi).
- Performed exploratory architectural changes for end of speech on **XTTS V2** and **Fish Audio S2 pro**.
- Deployed and finetuned Moshi and Personalex for Indic voices with conversation turn delay 200 ms
- Integrated DeepFilterNet v3 for denoising. Built **S2S⁷** and **IVC⁸** model collaborating with **Google**

AI Intern

Sarvam AI

Founding AI Team

(Nov 2023 – Jan 2024)

- Built an end-to-end QA agent for CSVs and OCR pipeline to unlock large scale tabular data.
- SOTA on internal and Yahoo Finance datasets; OCR at $0.62\text{-CER}/0.71\text{-WER}^5$ with 80% layout fidelity.
- Enables scalable digitization and QA of clinical registries and EHRs in low-resource scripts.
- Designed Indic OCR alongside RAG with hybrid retrieval and cross-encoder re-ranking.

Autonomous Systems Engineer

10X, IITM Research Park

Artificial Intelligence Team

(May 2023 – Jul 2023)

- Prototyped an autonomous wheelchair with SLAM navigation & multilingual conversational interface.
- Improved navigation with a success rate of 76% using Cartographer SLAM.
- Improves assistive mobility for neuro-rehabilitation and elder-care settings.
- Implemented mapping and navigation using IMU, encoders, LiDAR with ROS⁹ and Indic chatbot.

Research Experience

Project member CRIS Lab

Ontology-Grounded LKM for disease diagnosis

Principal Investigators: Prof. Venkat Venkatasubramanian (Columbia University) Prof. Bala Natarajan (Kansas State University) (May 2025 – Dec 2025)

- 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 **KG² distilled transformer encoders** for clinical decision support with transparent reasoning.
- Curated corpus via GA4GH Phenopackets and engineering weighted disease-phenotype.

Research Associate HILCPS Lab

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

Principal Investigators: Prof. Babji Srinivasan (IIT Madras), Prof. Bala Natarajan (K-State) (Sept 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 and inductive embeddings integrating neighborhood states.

Research Intern **Inductive scRNA Imputation GNN**
 CPSWin Lab *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% ARI⁴, 48% runtime, 4.5% L₁ median error vs baselines.
- Improves downstream cell-state clustering and strengthens scRNA pipelines for tumor biopsies.
- Developed inductive GNN for unsupervised clustering across cohorts to handle inherent data sparsity.

Publications

CACE, Elsevier **Boneshwar, V. K.**, Deepesh Agarwal, Bala Natarajan, and Babji Srinivasan. *"Inductive graph neural network framework for imputation of single-cell RNA sequencing data."* Computers & Chemical Engineering 195 (2025): 109031.

BSPC, Elsevier **Boneshwar V. K.**, Vijayavallabh Jayamanikandan, Deepesh Agarwal, Bala Natarajan, and Babji Srinivasan. *"PAIGT: Position aware graph transformer for single-cell RNA."* (under review)

IEEE SLT 2026 Workshop **Boneshwar V. K.**, Udeshya Raj, Ishan Grover, Zaher Abdul Azeez. *"Adapting Indic Text-to-Speech for the Real World: A Survey of Architectures and an Industrial-Grade Benchmark"* (accepted)

ICASSP 2026 conference Udeshya Raj, **Boneshwar V. K.**, Ishan Grover, Zaher Abdul Azeez. *"When the Stop Token Loses: Diagnosing and Mitigating End-of-Sequence Failures in Autoregressive Neural Text-to-Speech Systems"* (submitted)

Manuscripts in Preparation

Survey: Single-Cell Landscapes **Boneshwar, V. K.**, Rakhul Kumar Babushankar, Aabila Tharzeen, Deepesh Agarwal, and Bala Natarajan. *"Mapping the Landscape of RNA-Seq Computational Tasks: A Review of Algorithms, Model Architectures, Learning Strategies and Evidence Harness."*

Uncertainty Quantification and noise modeling **Boneshwar, V. K.**, Rakhul Kumar Babushankar, Deepesh Agarwal, and Bala Natarajan. *"Sequencing-Agnostic Uncertainty-Aware Graph Modeling for scRNA-seq Cell-Type Annotation"*

Select Deep Learning Projects

PathoMorph Disease Progression Visualiser **PathoMorph 3D: Evidence-Based 3D Disease Progression Visualiser**
Personal Project (September 2026 – Present)

- PathoMorph: 3D interactive web app animating disease progression from healthy to end stage.
- Implemented a custom three.js shader rendering multiple diseases from their documented origin sites.
- Grounded every visual parameter in 200+ peer-reviewed papers with an independent fact-check.
- **Impact:** Evidence-traceable educational tool for visualising diseases as they appear in real life.

Personalised intelligent AI companion **Offline AI companion on edge device**
Personal Project (July 2026 – Present)

- Building an AI companion with more focus on support towards research, job and personal development.
- Supports very fast end to end conversational latency (less than 1.4 seconds) using LiteRT.
- Implemented EdgeVDB based retrieval framework for personalized memory accessibility on device.
- **Impact:** Open source project paving way to future edge device deployments of AI agents.

MaskedAlign: Seq-to-Seq MSA Transformer **Masked Language Modeling for large scale genomes**
Principal Investigator: Prof. Manikanda Narayanan (Jan 2024 – May 2024)

- Enhance generalization of MSA¹⁰ on unseen gene families using a masked transformer using PyTorch.
- Higher alignment accuracy vs BetaAlign/ClustalW/MAFFT with runtime reduction up to 11%.
- Built seq-to-seq model with masked language modeling and robust training for long genomic data.
- **Impact:** Speeds comparative genomics and variant discovery relevant to cancer cohort analyses.

Latent code guided GANs¹¹ for image generation **Enforcing control over latent space in GANs**
Principal Investigators: Prof. Babji Srinivasan, Prof. Rajagopalan Srinivasan (Nov 2023 – Mar 2024)

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

- Realtime Traffic Bot ***Realtime monitoring of traffic violations: Helmet, speed and triple riding***
CVI Club, CFI, IITM (Jun 2022 – Jan 2023)
- Implemented end-to-end Python pipeline for detecting traffic violations in realtime.
 - Achieved **F1 score 0.79** and accuracy **89%** leveraging **Yolov5** on Stanford Cars.
 - Deployed on **Jetson Nano** & Incubated in **Nirmaan Incubation Cell, IIT Madras**.
 - **Impact:** Aids traffic police and institutions to monitor any violation in realtime and penalize the riders.

Coursework

- Core CS & Math Machine Learning, Deep Learning, Linear Algebra, Series and Matrices, NLP¹
- Core biology Cell Biology, Molecular Biology, Genomics and Proteomics, Biophysics, Genetics
- Other Computational Courses Data Structures and Algorithms in Biology, Algorithmic Approaches in Computational Biology, Bioinformatics, Computational Neuroscience

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

Leadership and Community Service

- 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 Machine Learning based pipeline for analyzing workouts (pushups, situps, pullups)
 - **Suvidha Foundation (NGO):** Led Machine Learning team and designed framework for teaching AI to support weaker economic sections

Achievements and Extra-Curricular Activities

- Academic Honors & Achievements
- Featured in **Times of India**. Live demo to Google India's CEO and the Chief Justice of India.
 - 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 Advanced 2020 (success rate: 4%)
- Cricket
- Lead Narmada hostel, IIT Madras in Schroeter 2024
 - Lead the high school cricket team in Inter-District school tournament 2018, Tamil Nadu, secured 3rd 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.
- 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**

*- Equal contributions, 1- Natural Language Processing 2- Knowledge Graphs, 3- Human Phenotype Ontology, 4- Adjusted Rand Index, 5- Character Error Rate & Word Error Rate, 6- Time To first byte, 7- Speech to speech (Model with no separate STT-LLM-TTS modules: Gemini Native, GPT 4o Realtime), 8- Instant Voice Cloning, 9- Robot Operating System, 10- Multiple Sequence Alignment, 11- Generative Adversarial Networks