Mahadev Sunil Kumar

mahadevsunilkumar03@gmail.com +91 97467 26074 LinkedIn www.mahadevsunilkumar.com

Education

Amrita Vishwa Vidyapeetham (Amrita University) Kollam, Kerala, India
B.Tech. in Computer Science, Specialization in Artificial Intelligence Jun 2025
Thesis: Deep ScaleNet: A Novel Framework for Deploying Large Distributed CNNs on Edge Devices CGPA: 6.64/10
Advisor: Prof. Amitava Mukherjee

Loyola School Trivandrum, Kerala, India
Indian School Certificate (ISC), Computer Science & Mathematics Jun 2021
Score: 96.7%

Loyola School Trivandrum, Kerala, India
Indian Certificate of Secondary Education (ICSE), Science Jun 2019
Score: 90%

Research Interests

Optimization, Neuromorphic Systems, Computer Vision, Edge Computing

Research Experience

Undergraduate Student Researcher August 2023 – June 2025
Project: SlimEdge: Lightweight Distributed DNN Deployment on Constrained Hardware
Supervisors: Dr. Amitava Mukherjee (BITS Dubai); Dr. Arnab Raha (Intel Corp); Dr. Debayan Das (IISc); Dr. Gopakumar G (Amrita University)

- Designed dynamic filter-pruning for CNNs under memory/latency constraints; extended to Multi-View CNNs with per-view importance estimation using XGBoost.
- Built **reward functions** combining accuracy preservation, model size, inference time, device memory caps, and per-view salience.
- Implemented single-view optimization with a **Genetic Algorithm**; scaled to all views using **NSGA-II** for multiobjective trade-offs across accuracy, size, and latency.
- Developed sampling strategies (importance- and device-aware) using Beta-distribution parameterizations for pruning vectors.
- Used root-finding **(Newton–Raphson, Brentq, TOMS748)** to solve budget-constraint equations and align pruning targets with device memory ceilings.
- Deployed and profiled on **heterogeneous nodes** (Raspberry Pi, NVIDIA Jetson Nano, Apple Silicon)
- Implemented **device-aware partitioning heuristics** for distributed inference; validated robustness via cross-device accuracy and latency variance analysis.

Research Intern, ISRO Inertial Systems Unit (IISU) Aug 2023 – Oct 2023
Supervisors: Dr. Ouseph P. (IISU, ISRO) Dr. Durairaj R. (IISU, ISRO)

- Built a **YOLOv7-based system** for humanoid analogue-gauge reading; curated and annotated a custom dataset with variable lighting, glare, and dial geometries.

- Engineered post-processing for pointer-angle regression and scale normalization; achieved $\sim$**98%** accuracy on held-out gauges; created an inference workflow for deployment.
- Deployed trained model on an NVIDIA Jetson as well as optimized the model using **ONNX** and **TensorRT**.

Publications

Preprints

1. Mahadev Sunil Kumar, Arnab Raha, Debayan Das, Gopakumar G, Amitava Mukherjee – *SlimEdge: Lightweight Distributed DNN Deployment on Constrained Hardware*, preprint, 2025. (ArXiv)

Research Projects

SlimEdge: Lightweight Distributed DNN Deployment on Constrained Hardware Sep 2023 – Present

- Built a pruning + fine-tuning pipeline for MVCNN under user-set accuracy and model-size thresholds; achieved **1.2–5.0**× faster inference on heterogeneous devices while meeting accuracy targets.
- Quantified per-view salience by evaluating **55,665** random 12-D pruning vectors and training an **XGBoost** regressor to predict accuracy; obtained view-importance scores with $R^2 = 0.96$ and used them to guide pruning.
- Defined device compute capacity as the **harmonic-normalized inverse** of per-device inference time; biased pruning toward slower devices to balance latency across the distributed system.
- Formulated a **multi-objective** search over accuracy, size, and latency with inequality constraints; solved with **NSGA-II** (200 generations) and selected the lowest-latency Pareto solution; included a **single-objective GA** fallback with a penalty function for infeasible regimes.

Speech Emotion Detection (LSTM) May 2024 – Jun 2024

- Built MFCC and spectrogram pipelines; trained LSTM classifier for multi-class emotions; performed augmentation and class-imbalance handling.
- Evaluated with stratified splits; reported precision/recall/F1 per class and confusion matrices.

Reinforced Chess (DQN) Feb 2024 – Jul 2024

- Implemented self-play DQN with episodic reward shaping and n-step returns; stabilized training with target networks and experience replay.
- Benchmarked against heuristic baselines; analyzed policy improvement curves and move-quality statistics.

SDN DDoS Detection (SVM) Dec 2023 – Feb 2024

- Engineered flow-level features; trained SVM to detect DDoS in SDN settings; validated on held-out traffic traces.

Skin Disease CNN Oct 2023 – Dec 2023

- Built a custom 18-layer CNN for dermatology classification; normalized inputs, used label smoothing and early stopping; analyzed failure modes via Grad-CAM.

Teaching Experience

Teaching Assistant May 2025 – Jun 2025

Corporate and Industrial Relations, Amrita University Amritapuri, India

- Delivered sessions in **competitive programming, algorithms, and efficient coding** to the B.Tech. 2022–2026 batch; emphasis on complexity analysis, data structures, and problem patterns.
- Designed and taught *Code Hour*, a placement-oriented module covering greedy, DP, graph algorithms, recursion/backtracking, bit manipulation, and code hygiene.

- Mentored **143** students; created graded practice sets, solution walkthroughs, and rubric-based feedback to improve correctness and speed under time constraints.

Work Experience

Advanced Applications Engineering Analyst — Bengaluru, India
Accenture PLC — Oct 2025 – Present

Skills

Programming & Scripting: Python, C++, C, Shell (Zsh/Bash), MATLAB
ML & Deep Learning: PyTorch, TensorFlow, Keras, scikit-learn, XGBoost
Optimization & Numerical Methods: CMA-ES, NSGA-II, Genetic Algorithms, Newton–Raphson, Brentq, TOMS748
Model Compression & Edge Deployment: MVCNN pruning (per-view), distributed DNN partitioning, resource-constrained inference
Distributed & Systems: Docker, OpenMPI, Linux server administration, SSH, SMB/NFS
Data & Scientific Computing: NumPy, Pandas, SciPy, Matplotlib, Jupyter
Experimentation & Reproducibility: structured logging , ablation studies, reproducible Docker/Conda environments
Tooling: LaTeX (Elsevier/IEEE/Nature), Markdown

Specified Coursework

Core: Data Structures & Algorithms; Design & Analysis of Algorithms; Machine Learning; Deep Learning; Probability & Statistics; Linear Algebra; Optimization; Computer Networks; Operating Systems; Distributed Systems; Databases.
Online: Intro to LLMs in Python; Deep Learning for Text (PyTorch); Deep Learning for Images (PyTorch); RNNs for Language Modeling (Keras); Intermediate/Advanced ML (Coursera/Datacamp, 2023–2024).

Languages

English; Malayalam; Hindi; Latin
