



SACHIN MISHRA

Roll No: MS2025013

MS Research, Artificial Intelligence & Data Science

International Institute of Information Technology Bangalore

+91-9111113267

sachin.mishra@iiitb.ac.in

Github | Website | Leetcode

LinkedIn | Google Scholar

EDUCATION

Degree	Institute	CGPA	Year
MS Research	International Institute of Information Technology, Bangalore	3.52/4	2025-2027
B.Tech. (CSE)	Government Engineering College, Raipur	3.58/4	2021-2025

EXPERIENCE

- NIT, Raipur (Industrial project with BigMint Tech. Pvt. Ltd.)** *Feb 2024 - May 2024*
Research Intern *Raipur, C.G.*
 - Built an end-to-end steel price prediction system, from data pipeline to deployment.
 - Built and evaluated classical ML (ARIMA, RF, XGBoost) to deep learning and transformer-based neural models.
 - Automated data pipelines and integrated local server backend to fetch real-time price data for model retraining.
 - Optimized model inference latency and ensured reproducible results through robust version control, achieving a 2.8% prediction error on yearly forecasts.
- IIIT-Naya Raipur** *May 2023 - Nov 2023*
Machine Learning & Research Intern *Naya Raipur, India*
 - Built an ensemble CNN-SVM model for multi-class disease classification with 91.42% accuracy.
 - Developed a non-invasive fitness monitoring app achieving 96.67% prediction accuracy.
 - Led performance analysis and data interpretation for research reports and client insights.

PROJECTS

- Gumbel-Softmax Dynamic Inference Router** *Jul. 2025 - Oct. 2025*
Built a data routing system, end-to-end routing policy for managing model complexity and inference cost. *Visit*
 - Differentiable Dynamic Inference (DDI) system using the Gumbel-Softmax trick in Deep Neural Nets to optimize for accuracy during training and optimize for latency during deployment.
 - Achieved optimal speed-accuracy trade-off: Maintained a peak Test Accuracy of 89.0% (near ViT-B-16's 90.0%) while routing **50%** of images via the Fast Path (latency **6.60** ms).
 - Minimized production latency overhead by implementing Hard Inference and freezing classification sub-models, resulting in a Fast Path latency only 0.80 ms slower than the absolute fastest model (MobileNetV2).
 - Validated model serving efficiency by demonstrating a **60%** latency reduction compared to the highest-accuracy standalone model (ViT-B-16), making complex vision architectures viable for real-time inference.
- OorjaNest** *Feb. 2025 - Apr. 2025*
A deep learning framework for rooftop extraction and solar potential mapping. *Visit*
 - Tools & technologies used:** Python, PyTorch/ Lightning, React.js, Node.js, flask
 - Achieved an IoU of 0.9394 on a custom-annotated rooftop dataset from Google earth.
 - Built on ResNet-34 + U-Net architecture with attention mechanisms.
 - Integrated segmentation outputs with solar irradiance maps, improving energy estimation accuracy by 15-30%.
- Optimisation of Deep Generative Models** *Aug. 2025 - Present*
Developing optimized Latent Variable Models & enhancing the performance of diffusion models *GitHub*
 - Tools & Technologies Used:** Python, PyTorch, NumPy, Matplotlib, CUDA, Hugging Face Transformers
 - Studied and analyzed diffusion and latent variable models from an optimization perspective, including loss formulations, KL-divergence terms, and ELBO-based objectives.
 - Explored theoretical aspects of convergence, stability, and objective design in generative models, focusing on trade-offs between reconstruction quality and regularization.
 - Conducted analytical evaluation of generative quality using standard metrics (FID, IS) and qualitative inspection to understand model behavior and failure modes.

Coursework Selected Projects

- Obesity Category Prediction** – Built a multi-class classification model (using classical & Adv ML models) to predict obesity levels from lifestyle and demographic data, achieving 91.29% accuracy in the XGBoost-based model.
- AI Resume Parser** – NLP-powered document parser using spaCy
- SmartCrop** – Crop recommendation using ML on soil and weather data
- Finance Tracker** – Personal expense tracking dashboard (MERN stack)

PUBLICATIONS & PATENT

- 2025** Sachin Mishra, "Automated Assessment of Rooftop Solar Energy Potential Using Deep Learning-Driven Building Footprint Extraction Model." *ADCOM 2025*
- 2025** Sachin Mishra "Dynamic Query Handling with RAG Fusion for PDF-Based Knowledge Retrieval Systems."
- 2025** **Patent:** Sachin Mishra et al. AI ENABLED AUDIO PROCESSING PEN FOR TRANSCRIPTION. IN Patent.
- 2025** S Mishra et al. "TransDFD: A Deepfake Detection System of Mesoscopic level Deepfake-guard-AI" *IEEE IATMSI*.
- 2024** S Mishra, A Singh. "IoT ML Driven Holistic Health Monitoring and Fitness Assessment Empowering Proactive Wellbeing Management." *IEEE SCES*.
- 2023** **Book Chapter** S Mishra et al. "UDR Fused Multimodal Approach for Disease Classification in Large Scale Dataset with Advanced CNNs." *SPRINGER LNNS*.

TECHNICAL SKILLS

- **Languages:** Python, JavaScript, SQL, LaTeX, Html/Xml/Json/Yaml/Css
- **AI / Data Skills:** Machine Learning, Deep Neural Networks, Deep Generative Models, NLP, Foundation Models, Time-series Analysis, Data Visualization, Training concepts (batch size, epochs, learning rate, GPU utilization).
- **Frameworks:** TensorFlow, PyTorch & Lightning, scikit-learn, Hugging Face, OpenCV, NumPy, Pandas
- **Cloud Platforms for ML:** GCP (Vertex AI, AutoML), Azure ML, Firebase, AWS, Linux, Windows
- **Experiment Tracking Dev Tools:** MLFlow, Git & GitHub, Jupyter Notebook, Colab, VSC

KEY COURSES TAKEN

- **Masster's Coursework :** Optimization, Reinforcement Learning, Self Supervised Learning, Machine Learning, Maths for ML, Deep learning & Foundational Models , AI Framework Testing.
- **UnderGrad's Coursework** Data Science, Computer Vision, Artificial Intelligence, Object-Oriented Programming (OOP's), Data Structures, Algorithms, DBMS, Operating Systems, Cloud Computing, Computer Networks, Web Services, Software Engineering & Product Management, CL & IP.

CERTIFICATIONS

- Supervised Machine Learning: Regression and Classification, by Andrew NG, Coursera.
- Advanced Learning Algorithms, by Andrew NG, Coursera.

POSITIONS OF RESPONSIBILITY

- **Student Head Coordinator**, i-HUB Chhattisgarh *Jan 2025 - June 2025*
- **Microsoft learn student ambassador (MLSA)**, Microsoft *Feb 2023 - July 2025*
- **Google Developer Student Lead**, Google's program *June 2023 - Apr 2024*

ACHIEVEMENTS

- **Institute Topper** Securing Rank 1, since 1st semester. *2021 - 2025*
 - **Secured 1st Position** Smart India Hackathon'24 at Institute level. *2024*
 - **Winner** ECOHACK GEC Raipur organized by National Service Scheme *2023*
-