

Shashank Sharma

Research & Innovation Engineer, Neuromorphic Computing, TCS Research, Kolkata, India

shashank.0901sharma@gmail.com | +91-8490849270 | [shashankgsharma.github.io](https://github.com/shashankgsharma)

Google Scholar | github.com/shashank23088 | linkedin.com/in/shashankgsharma | ORCID 0009-0001-8159-2516

PROFILE

Research engineer in neuromorphic computing at TCS Research: I train spiking neural networks on time-series, audio and vision tasks and help move them onto neuromorphic hardware through NIR, using hardware-aware training to reduce the gap between simulated and on-chip accuracy. Before this, my Masters at IIIT Delhi focused on language and multimodal neural nets, LLM fine-tuning and inference, and vision-language models on live video. Currently studying the state of language modelling with event-driven networks in neuromorphic computing.

EDUCATION

Indraprastha Institute of Information Technology Delhi (IIIT Delhi) New Delhi, India
M.Tech., Computer Science & Engineering (Artificial Intelligence specialisation); CGPA 8.07/10 2023 – 2025

- Coursework: Machine Learning, Deep Learning, Advanced Deep Learning, Reinforcement Learning, Natural Language Processing, Information Retrieval, Artificial Intelligence

Adani Institute of Infrastructure Engineering Ahmedabad, India
B.E., Information & Communication Technology; CGPA 9.22/10 2019 – 2023

RESEARCH EXPERIENCE

TCS Research, Neuromorphic Computing Kolkata, India
Research & Innovation Engineer — Supervisors: Dr. Barnali Basak, Sounak Dey Sep 2025 – Present

Language models on neuromorphic hardware: literature review Apr – Sep 2026

- Reviewed 26 recent papers (2023–2026) on spiking and MatMul-free language models and their deployment on neuromorphic chips and FPGAs, comparing architectures, target hardware, and whether efficiency claims were measured or estimated.
- Found that only 3 of them report measurements on an actual neuromorphic chip, and identified the main deployment blockers, chiefly operations that chips do not support natively (e.g. softmax, normalisation); presented the review to the lab with proposed next steps, including cross-chip deployment and a measured energy benchmark.

Hardware-aware training and portability of SNNs across neuromorphic platforms Dec 2025 – Present

- Co-developed (with Dr. Barnali Basak) a hardware-aware synaptic LIF neuron in `snnTorch` that emulates the neuron dynamics of a chosen target platform (Lava-DL, Rockpool or Sinabs) during training, so one training pipeline produces models matched to each target; it accounts for differences in reset behaviour, numerical precision, time-step and decay conventions, state-update order and spike output. Extended to recurrent and adaptive-threshold neurons.
- Built the training and conversion pipelines for time series (FordA), spoken digits (SHD) and event-camera vision (DVS Gesture, N-MNIST), and co-developed those for image classification (CIFAR-10) and bird localisation (CUB-200-2011); the full study covers seven tasks. Standard `snnTorch` models lost accuracy after export via NIR, in some cases collapsing (SHD: 78% fell to 63% across targets), while hardware-aware models stayed within 0.5 points of their training accuracy on every target.
- Writing the user guide and tutorials for TCS Research's neuromorphic middleware (in progress).

Membrane-potential-to-spike knowledge distillation for image tasks Jan 2026 – Present

- Spiking networks that pass continuous membrane potentials between layers are more accurate than those that pass only spikes, but neuromorphic chips can run only the spike-based version. We train the spike-only network to mimic the membrane-potential network layer by layer, so the deployable model recovers much of the lost accuracy.
- Raised spike-only accuracy by up to 12.7 points across six image classification and localisation datasets. I identified the accuracy gap, proposed the distillation approach and built the training pipelines. Journal manuscript in preparation with S. Pradhan and D. Paul.

- **Accident detection in live traffic video with vision-language models** (collaboration with TU Darmstadt, Dr. Manisha Luthra, May–Sep 2024)
 - Aimed to replace classical accident-detection pipelines (object detector, tracker and hand-set rules), which transferred poorly across dashcam, CCTV and drone video, with multimodal LLMs. Reproduced two published classical baselines, which utilized unsupervised learning and yolo-deepsort based object tracking.
 - Benchmarked GPT-4/4o and open-source LLaVA models as the detector, mapping the accuracy–latency trade-off across model size, quantisation and frames per clip; accuracy depended strongly on viewpoint (53% dashcam to 74% CCTV for LLaVA-NeXT-Video-7B, about 1.3 s per clip).
 - To cut the LLM's cost, built a lightweight frame filter (background subtraction, vehicle tracking, distance between vehicles) that forwards only frames where vehicles come close and samples more densely around them; on the classical baseline it cut latency by 8% at a 4-point accuracy cost.
- **Crowd energy quantification from audio and video** (IEEE MIPR 2024 [P1], Jan–May 2024)
 - No standard measure of crowd energy exists, so we built a score for it: a visual part combines crowd density (STEERER) with heatmaps of the active crowd (Grad-CAM on a CNN trained to separate energetic from neutral crowds), an audio part rates cheering and loudness, and the two are weighted into one score.
 - My part was the audio branch. With three other annotators, collected a 91-minute dataset of cheering vs. neutral crowd audio from online videos; a spectrogram CNN from prior work was rejected after it rated neutral audio above cheering, so the final branch classifies CLAP audio embeddings with a small network and combines the result with loudness and spectral features.

COMPETITIONS

THOR × NeuroBench Challenge 2026: motor-imagery EEG decoding

Mar – Aug 2026

- With S. Pradhan, built a compact SNN that classifies left vs. right-hand motor imagery from 62-channel EEG (OpenBMI, 54 subjects): a learned spatial filter and signal-power rectification feed a single layer of leaky integrate-and-fire neurons, trained with a spike-rate penalty to keep activity sparse.
- Joint 3rd place in the compute-efficiency track: 69.5% accuracy with 58.8k synaptic operations and a 3.1 kB memory footprint, with neurobench benchmarking.

Vehant Smart City Surveillance Challenge 2024: person attribute recognition [code]

2024

- **Task:** recognise several attributes per person (clothing, accessories, footwear, pose, viewing angle) in Indian street-camera images, from only about 600 labelled images. With S. Kale and A. Khuntia; placed 2nd, write-up on arXiv [P2].
- Expanded the data 12-fold with augmentation and compared three models: two transformers (BEiT, Swin) reached 91% mean accuracy but overfitted and took about 14.5 hours each to train, while a ResNet50 with a loss weighted by attribute frequency (so rare attributes are not ignored) reached 86%, generalised better and trained $51\times$ faster; deployed it behind a YOLOv10 person detector for a real-time video demo.

LANGUAGE AND MULTIMODAL MODEL WORK

Perspective-controlled summarisation of healthcare answers — NLP course, IIT Delhi [code]

Jan – Apr 2025

Supervisor: Dr. Shad Akhtar

- **Task** (team of three): summarize answers from healthcare forums from a chosen perspective. Built two baselines: a prompted LLaMA-3.2-3B-Instruct model and a re-implementation of PLASMA, the method from the original paper on this task.
- Developed one of the three solutions built by the team, which included finetuning of LLaMA-3.2-3B-Instruct (4-bit, LoRA) in two steps: first on medical question answering, so the model learns medical language, then on summarizing from a chosen perspective.
- Added a training loss that penalises summaries that drift away from the chosen perspective or lose the meaning of the original answer.

LLM-based query system for traffic analysis from CCTV footage — LLM course, IIT Delhi [code]

Autumn 2024

Supervisor: Dr. Rajiv Ratn Shah

- **Task** (team of six): answer plain-language questions about CCTV traffic footage with a structured report (vehicle counts by type, congestion level, accidents). My part: fine-tuning the video-language model LLaVA-OneVision for accident detection and vehicle counting.
- Selected informative frames by how much consecutive frames change, used YOLO vehicle counts as labels, and fine-tuned with QLoRA on 200 DoTA and HEV-I videos. On 100 test videos, accident-detection accuracy rose from 65% to 80% and car-count error (MAE) fell from 4.3 to 2.7; model released on Hugging Face.

PUBLICATIONS

[P1] A. Anand, S. Jain, **S. Sharma**, A. P. Dominic, A. Gupta, A. Verma, R. Jaiswal, N. Lal, R. R. Shah, R. Zimmermann. *Pulse of the Crowd: Quantifying Crowd Energy through Audio and Video Analysis*. IEEE MIPR 2024, pp. 308–314. doi:10.1109/MIPR62202.2024.00056

[P2] S. Kale, **S. Sharma**, A. Khuntia. *Advanced Smart City Monitoring: Real-Time Identification of Indian Citizen Attributes*. Preprint, arXiv:2407.03305, 2024.

TECHNICAL SKILLS

Programming	Python, C++ , ML $\LaTeX$
Neuromorphic	snnTorch, spikingjelly, Rockpool, Lava, Sinabs, NIR, NeuroBench
ML	PyTorch, Hugging Face, Transformers, bitsandbytes, OpenCV
Systems	Linux, Git, Docker

TEACHING

- Teaching Assistant, IIT Delhi (2023–2025), for five courses: Introduction to Programming, Natural Language Processing, Data Structures & Algorithms, Artificial Intelligence, and Intelligent Systems.
- Ran tutorial sessions and mentored 25+ students, and occasionally gave lectures under the course instructors.
- Designed and graded assignments and tests, and helped students during lab sessions.

REFERENCES

Dr. Arijit Mukherjee

Principal Scientist, TCS Research
Manager at TCS Research
mukherjee.arijit@tcs.com

Sounak Dey

Senior Scientist, TCS Research
Team Lead at TCS Research
sounak.d@tcs.com