

SHUBHAN MITAL

shubhanmital@gmail.com | +1 (858) 214-7189 | github.com/Shubhanflash22 | linkedin.com/in/shubhan-mital

RESEARCH AT YUANYUAN LAB, UNIVERSITY OF CALIFORNIA SAN DIEGO

- Built a multi-stage CV pipeline chaining YOLOv8 detection, DeepSORT multi-object tracking, a physics-constrained finite state machine, and transfer-learned 3D ResNet-18 spatiotemporal classification with automated CVAT annotation to convert **34+ hours** of raw construction site footage into per-subactivity energy profiles for mobile electric excavator charging dispatch
- Achieved **88.81%** recognition accuracy across 5 activity classes, surpassing a prior 3-class benchmark of 86.7%; applied Bayesian regression for hidden-state power estimation and built automated tooling for class-balanced clip extraction
- Benchmarking deterministic certainty-equivalent MPC against scenario-based stochastic MPC over a 24-hour dispatch window to produce real-time mobile charging schedules that curb on-site battery depletion, demand charges, and grid carbon impact

EDUCATION

University of California, San Diego MS in Machine Learning and Data Science (ECE)	<i>Sept 2025 - Mar 2027</i> CGPA: 3.83 / 4.00
BITS Pilani B.E. in Electrical and Electronics Engineering and M.Sc. in Mathematics	<i>July 2019 - July 2024</i> CGPA: 8.48 / 10.00

SKILLS

Languages: Python, SQL, C++, MATLAB, Java
ML/DL Frameworks: PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face
ML/AI Techniques: Deep Learning, CV, NLP, Bayesian Inference, Optimization (MILP/MPC), Predictive Modeling
Data & Cloud: Pandas, NumPy, AWS, Azure Databricks, Snowflake, Git

EXPERIENCE

- | | |
|--|----------------------------------|
| Data Scientist at Piramal Pharma Ltd., Mumbai | <i>Sept 2024 - Sept 2025</i> |
| <ul style="list-style-type: none">• Owned end-to-end ML lifecycle (HLD/LLD, feature engineering, model selection, cross-validation, deployment) for a forecasting model predicting inventory stock-out/expiry risk 24 months ahead, improving planning visibility• Engineered 4+ scalable ML data pipelines from LLD to production using Azure Databricks, ADF, Snowflake, and Azure Function Apps, boosting processing efficiency by 60% and eliminating manual accrual consolidation• Translated ML pipeline outputs into Qlik Sense dashboards, driving data-informed decision-making across 20+ teams• Automated pricing workflows & deployed NLP-driven chatbots, reducing turnaround time for business processes by 70% | |
| Data Science Intern at Piramal Pharma Ltd., Mumbai | <i>Jan 2024 - June 2024</i> |
| <ul style="list-style-type: none">• Led full-cycle delivery of 3 analytics and reporting solutions – from stakeholder ideation and HLD through LLD, development, and production deployment – driving a 20% improvement in operational efficiency across business units and key stakeholders• Streamlined price scrapers, self-ship trackers, & offtake reports for e-Commerce platforms, cutting manual reporting overhead and improving operational efficiency by 50% and data accuracy by 90% across 3+ marketplaces | |
| SDE Intern at Amazon Development Centre, Bangalore | <i>July 2023 - December 2023</i> |
| <ul style="list-style-type: none">• Developed and integrated automated data validation and pipeline auditing tools into production UIs, detecting discrepancies across 5+ downstream data pipelines and preventing financial losses of up to \$5 million• Implemented CI/CD for 2 critical data pipelines with comprehensive functional and integration test coverage, cutting release validation cycle time by 15 hours per sprint and improving team operational excellence metrics | |

PROJECTS

- SolarInvestAgent: Agentic AI for Residential PV & BESS Sizing - LangGraph, HiGHS, Python**
- Architected a hybrid decision-support system coupling a HiGHS-backed MILP optimizer with an LLM-orchestrated multi-agent reasoning loop (LangGraph) for feasibility-validated residential PV & BESS sizing recommendations
 - Reduced panel-count RMSE by **13.8×** and CAPEX MAE by **2.3×** relative to a single-shot LLM baseline across 20 household scenarios, while maintaining **100%** constraint-feasible outputs across all test cases
- StayTuned: Multimodal Attention Detection - MediaPipe, TensorFlow, PyTorch, Python**
- Engineered a real-time facial attention classifier using a 10-feature geometric BiLSTM (EAR, MAR, 3D PnP head pose), achieving a macro-F1 of **0.61** on the DAiSEE benchmark across 4 engagement-level classes
 - Fused facial predictions with a Whisper-based 75-feature audio model; deployed live inference to an ESP32 OLED display for real-time hardware-in-the-loop classroom attention monitoring with multimodal integration
- Spatial-Aware Noise for Text-to-Image Diffusion - SDXL, PyTorch**
- Designed NPNet, a lightweight SVD-based noise transformation network with a prompt-conditioned residual branch, improving SDXL spatial accuracy by **+10.6%** and compositional fidelity by **+10.1%** over baseline
 - Confirmed seed-dependent spatial variability via chi-squared tests ($\chi^2=249.89$, $p=4.3\times10^{-7}$) across 6,900 evaluation images on the Comp90 benchmark, establishing statistically significant noise-conditioned controllability