Soham Pahari

Linkedin: https://www.linkedin.com/in/sohampahari/
Github: https://github.com/suhanpahari

Email : paharisoham@gmail.com
Mobile : +91-90645-20673

Profile Summary

Third-year student with strong expertise in Python, statistics, and machine learning, currently working on computer vision research in IoT and image processing. Proficient in TensorFlow, Hugging Face, and Flask-Streamlit for building scalable ML systems, with a focus on performance optimization. Actively engaged in university research, exploring innovative solutions in deep learning and vision-based applications.

Education

- **University of Petroleum and Energy Studies** — Dehradun, India
  Bachelor of Computer Science (Data Science major); — *Aug 2022 - May 2026*

Skills Summary

- **Languages**: Python, Java, C
- **Libraries**: Scikit-learn, NLTK, OpenCV Flask, Hugging Face, TensorFlow, Langchain, PyTorch
- **Vector Database**: Chroma, Pinecone, Drant
- **Framework**: LangChain, LlamaIndex

Internship Experience

- **Bahas Pvt Ltd** — Remote
  ML Development Intern — *May 2024 – July 2024*
  - **Regional Language Emotion Classification**: Developed multi-model system (**Custom finetuned BERT**, SVM, Random Forest) for emotion classification in regional languages. Optimized models for accuracy.
  - **Deployed Interactive App**: Built a Streamlit app for real-time emotion prediction with model selection.
- **NIT Warangal dept of Artificial Intelligence** — Warangal, India
  Research Intern – Big Data Lab — *May 2025 – Present*
  - **Novel Drug Innovation Model**: Researching a deep learning-based framework for drug discovery using a modified encoder-decoder architecture with custom gating mechanisms and domain-specific embeddings.

Projects

- **Cricket Event Detection using Transfer Learning and Voting**: Developed a deep learning-based event detection model for cricket, integrating **VGG-16, VGG-19, and ResNet50** with a voting mechanism for accurate classification. The model detects key events such as six, four, out, wide, no ball, drs, and no action, achieving 94.66% accuracy. This system automates highlight generation, enhances real-time decision-making for analysts and umpires, and improves fan engagement through instant updates.

  Technologies used: TensorFlow, Keras, OpenCV, scikit-learn.

- **Query-Based Event Searching in Cricket**: Developed an automated video analysis system for cricket event detection using signal processing and computer vision techniques. **The method integrates intensity-based segmentation, optical character recognition (OCR), and text-based querying to identify key events** such as "four", "six", and "wicket" in long-form video recordings. This multi-modal approach enhances event retrieval efficiency and improves sports analytics for automated highlight generation.

  Technologies used: OpenCV, Tesseract OCR, scikit-learn, TensorFlow.

- **Delhi Pollution Prediction in Time Series with Sequential Models)**: Developed a model addressing Delhi's pollution using **ARIMA, LSTM, and a custom hybrid metaheuristic algorithm combining Dung Beetle, Quantum Swarm, Hybrid Genetic, Red Deer, and Gravitational Algorithms**. Achieved a 15% increase in efficiency and a 9-10% boost in prediction accuracy through advanced optimization techniques. Focused on minimizing errors and enhancing performance for reliable, long-term pollution forecasting. Currently in its final phase, the project offers scalable solutions for future environmental challenges.

  Technologies used: TensorFlow, Keras, scikit-learn, pmdarima (for ARIMA, SARIMAX).

Honors and Awards

- Selected among the top 200 out of 17,000 participants in the **Hackahazard 2025** AI Hackathon. Certificate Link
- Scored 99.28 percentile in Mathematics in CUET 2022.

Certifications

- **Decision Making and Reinforcement Learning** — April 2025
  Coursera – Columbia University — *Certificate Link*