# Ashutosh Roy

**Deep Learning Researcher | Speech AI & Affective Computing**

+91-9810599837 | ashu2003roy@gmail.com | linkedin.com/in/ashutosh-roy | github.com/ashutosh-roy | Gurugram, Haryana

## Summary

Deep Learning researcher specializing in hybrid architectures (Wav2Vec2, Transformers, BiLSTMs) for high-accuracy signal classification and affective computing. Proven track record in developing robust preprocessing pipelines for emotional and physiological stress detection in audio data. Published researcher (RECCAP 2026, IIT Palakkad; co-author, IIT Delhi) with hands-on experience in speech processing, multimodal learning, and self-supervised audio-visual representations. Proficient in PyTorch, Hugging Face, and transformer-based sequence models.

## Experience

**Research Intern** **Jan 2025 – Jun 2025**
*Indian Institute of Technology (IIT) Delhi* *New Delhi, India*

- Evaluated and benchmarked 3 OCR systems (Tesseract, Marker, Gemini OCR) on diverse historical document images — handwritten manuscripts, printed text, mixed layouts — designing a per-class selection strategy to maximize extraction fidelity.
- Integrated image understanding into a multimodal retrieval pipeline capable of cross-modal querying over OCR-extracted text and visual document features.
- Fine-tuned **Surya OCR** model to convert scanned books with varied layouts into structured Markdown output; extracted geographic and historical semantic data from historical maps.
- Co-authored **SARCH: Multimodal Search for Archaeological Archives** — combining OCR, image classification, and hybrid search for large-scale archival data. *arxiv.org/abs/2511.05667*

**AI/ML Engineer Intern** **Oct 2024 – Nov 2024**
*MitoVoid AI* *Remote*

- Built an OCR-based PDF analysis module to extract and interpret medical reports — combining **Tesseract** with structured text parsing for downstream NLP inference.
- Designed end-to-end multimodal pipelines integrating visual document inputs with language model outputs for healthcare applications.

## Projects

**Speech Emotion & Stress Detection** | *Python, PyTorch, Wav2Vec2, BiLSTM, Hugging Face* | ***RECCAP 2026*** Aug 2024 – Present

- Designed audio preprocessing pipeline: VAD, silence removal, log-Mel spectrogram generation, and MFCC feature extraction across 3 corpora (RAVDESS, TESS, SAVEE).
- Hybrid **Wav2Vec2 + BiLSTM** architecture achieved **85.9% accuracy**, **0.856 F1**, RMSE **0.1268** on continuous stress regression; outperformed SVM, Random Forest, and CNN baselines. Paper accepted at RECCAP 2026, IIT Palakkad.

**Medical Chatbot with OCR Pipeline** | *Python, Mistral 7B, QLoRA, Tesseract, OpenCV* Mar 2024 – Apr 2024

- Built a document vision pipeline: Tesseract OCR + **OpenCV** preprocessing (denoising, binarization, deskewing) to extract structured text from clinical report images; output fed to fine-tuned Mistral 7B for medical NLU.

**Exam-Helper RAG System** | *Python, LLaMA 3, FAISS, LangChain, Streamlit* Dec 2024 – Apr 2025

- Implemented multimodal document ingestion supporting scanned PDFs and image-based Q&A — combining vision-based document parsing with dense vector retrieval for context-accurate LLM responses.

**Music Identification App** | *Python, ACRCloud API, Streamlit, NumPy* Sep 2024 – Dec 2024

- Built an audio signal processing pipeline: waveform capture → feature extraction → acoustic fingerprinting via ACRCloud API; designed cross-platform UI with Flutter.

**BulkyMail** | *Python, Streamlit, SMTP, Jinja2* Dec 2024

- Built a Streamlit-based bulk email automation tool with dynamic Jinja2 templating for personalized large-scale outreach; integrated SMTP APIs for reliable delivery of 100+ emails per run.

## Technical Skills

**Speech & Audio Processing**: Wav2Vec2, Audio Feature Extraction (MFCC, log-Mel, VAD), Silence Removal, Acoustic Fingerprinting, Emotion & Stress Detection, Affective Computing
**Deep Learning & Research**: PyTorch, Hugging Face, Transformers, BiLSTM, Self-Supervised Learning, CNN, QLoRA/LoRA, Evaluation & Benchmarking
**Computer Vision & Multimodal**: OCR (Tesseract, Surya, Gemini OCR), Document Understanding, Image Preprocessing, Multimodal Retrieval, OpenCV
**Libraries**: OpenCV, NumPy, Pandas, Scikit-learn, Matplotlib, LangChain, FastAPI
**Databases & Search**: FAISS, Pinecone, Apache Solr
**Languages**: Python, C++, SQL
**Tools**: Git, Docker, Jupyter Notebook, Google Colab, Linux

## Education

**Chhattisgarh Swami Vivekananda Technical University (CSVTU)** **2022 – 2026**
*B.Tech (Hons) in Computer Science and Engineering (Artificial Intelligence)* *Bhilai, Chhattisgarh*

**St. Columbus School** **2020 – 2021**
*Higher Secondary (Class XII - Science PCMB) - 87%* *Faridabad, Haryana*