

Mandavi Singh

s.mandavi@op.iitg.ac.in | [LinkedIn](#) | github.com/mandavi-singh | [Portfolio](#)

SUMMARY

Final-year B.S. (Honors) in Data Science & AI at IIT Guwahati, with a prior B.Sc. in Mathematics. Research internships at IIT Guwahati: **sleep-apnea detection** under Dr. Ritika Jain (ongoing) and **computational phonetics** under Dr. Neeraj Sharma. Industry experience: deep-learning fall detection at Hybionics, and SNS-based agriculture forecasting and AI agents at LTID, Tokyo. Has delivered multiple end-to-end ML applications deployed on the cloud.

EDUCATION

Indian Institute of Technology Guwahati

B.S. (Honors) in Data Science and Artificial Intelligence

Guwahati, India

2023 – 2027

Patna University

B.Sc. in Mathematics

Patna, India

2020 – 2023

EXPERIENCE

Research Intern

Aug 2026 – Present

Dr. Ritika Jain, IIT Guwahati

Guwahati, India

- Classifying **sleep apnea (OSA vs. CSA)** on three benchmark datasets – PhysioNet, UCDDDB, and MIT-BIH – using SpO₂, nasal airflow, respiratory signals, and ECG.
- Building a 1D CNN + **Transformer** pipeline in PyTorch with bandpass filtering and sliding-window training for robust apnea-event detection.

Research Intern

Jan – May 2026 | [Code](#)

Dr. Neeraj Sharma, IIT Guwahati

Guwahati, India

- Computational phonetics study on PHOIBLE 2.0 (2,716 languages) merged with Glottolog and Köppen climate data; replicated Everett (2017) *languages in drier climates use fewer vowels* (humidity–vowel slope +5.41 vs. paper's +5.37, $p < 0.001$).
- Extended the analysis to 112 Indian languages across four language families; found a significant N→S vowel gradient ($\rho = -0.233$, $p = 0.013$) and a strong tonal–humidity association ($p = 1.6 \times 10^{-5}$) using OLS family-controlled regression.

Machine Learning Intern

Feb – Jul 2026 | [Code](#)

Hybionics

Remote

- Built deep-learning **fall detection** on the SisFall wearable-sensor dataset: trained CNN and LSTM models on accelerometer/gyroscope time series and optimized for **recall** so missed falls – the costly failure mode – are minimized.

AI Engineer Intern

Oct 2025 – Mar 2026 | [Code](#)

Long Term Industrial Development Co. Ltd. (LTID)

Tokyo, Japan (Remote)

- Built ML models forecasting agricultural trends and outcomes from social-media (SNS) and environmental signals; developed **AgriIntel**, a Streamlit platform combining crop-yield prediction, disease detection, and live weather/market intelligence for 12 countries.
- Designed and implemented **AI agents** for automation and decision support; owned research, testing, and technical documentation of the AI algorithms.

PROJECTS

TokenLab – How LLMs See Text | Python, Streamlit, Hugging Face transformers, tiktoken

[GitHub](#) | [Live](#)

- Runs real tokenizers locally (Qwen BPE, GPT-style tiktoken, BERT WordPiece) to reveal exact tokens, token IDs, and vocabulary statistics, with side-by-side tokenizer comparison, token/cost analytics, and a Text→Tokens→Embeddings learning module.

Smart Parking Pricing | Python, Pandas, NumPy, Matplotlib, Streamlit

[GitHub](#) | [Live](#)

- Dynamic pricing models for smart-city parking: baseline linear, demand-based, and location-aware competitive pricing using Haversine distance between lots; live Streamlit dashboard with interactive map and per-lot/timestamp analysis.

AI Research Paper Assistant | Python, Streamlit, TF-IDF (scikit-learn), LLM API

[GitHub](#) | [Live](#)

- Upload-any-PDF tool that generates structured summaries, study notes, and MCQs, plus retrieval-augmented (RAG) Q&A over a local TF-IDF vector store; deployed on Streamlit Cloud.

Arogya Mitra – Healthcare AI Assistant | Python, Scikit-learn, Google Gemini Pro, Streamlit

[GitHub](#) | [Live](#)

- Predicts likely diseases from symptoms using Random Forest and SVM on the Kaggle Disease–Symptom dataset, then generates guidance with Gemini Pro; deployed with PDF export.

TECHNICAL SKILLS

Languages: Python, MATLAB

ML & DL: Scikit-learn, XGBoost, LightGBM, PyTorch, TensorFlow/Keras, CNNs, LSTMs, Signal Processing (bandpass filtering, sliding windows), Time-Series Modeling, Class-Imbalance Handling, Hyperparameter Tuning (GridSearchCV), Evaluation (F1, Recall, ROC)

LLMs & NLP: Hugging Face Transformers, tiktoken, Tokenization (BPE/WordPiece), RAG, LLM Tool Agents, OpenRouter / Gemini APIs

Data & Tools: Pandas, NumPy, SciPy, Statsmodels, Streamlit, Flask, Power BI, Git/GitHub, Jupyter/Colab, ~~TeX~~

Achievements: Convolve 4.0 (Pan-IIT AI/ML Hackathon) – Semi-Finalist