Aman Tripathi

+91-9005330740 | amantripathi3273@gmail.com | linkedin.com/in/aman-tripathi27 | github.com/in/Aman-Tripathi27

Summary

M.Sc. Data Science candidate focused on deep learning for biological and physiological data. Built GNN-based molecular interaction models, CNN-Transformer hybrid architectures for temporal signal decomposition, and biomedical ML systems deployed in production. Strong PyTorch foundation across graph, sequence, and time-series modalities. Published IEEE researcher. Looking to apply model development and signal processing skills to physiological metrics at the intersection of hardware and human performance.

Education

Manipal Academy of Higher Education — Manipal, Karnataka
Master of Science in Data Science — *2024 – 2026*

Chandigarh University — Kharar, Punjab
Bachelor of Science in Computer Science, Statistics, and Mathematics — *2021 – 2024*

Experience

Temple (Eternal) — Gurugram, Haryana
Deep Learning Engineer — *March 2026 – July 2026*

- Developing deep learning models for sleep onset and offset detection using physiological time-series signals (EEG, accelerometer, PPG), applying temporal segmentation and change-point detection techniques to accurately identify sleep/wake boundaries.
- Designed and trained CNN-LSTM hybrid architectures to capture both local morphological patterns and long-range temporal dependencies in multi-channel biosignal data for sleep stage characterization.
- Built preprocessing pipelines for raw physiological recordings including artifact rejection, bandpass filtering, and epoch-level feature extraction; applied VMD-based signal decomposition to isolate sleep-relevant frequency components.
- Implemented and benchmarked multiple model architectures (threshold-based baselines, XGBoost, LSTM, Transformer) for sleep onset/offset prediction, evaluating performance using epoch-level accuracy, F1-score, and temporal boundary error metrics.
- Collaborated on end-to-end ML pipeline design from raw wearable sensor data ingestion to model inference, targeting real-world deployment constraints in consumer sleep-tracking hardware.

Technical Skills

Languages: Python, SQL
Deep Learning: PyTorch, PyTorch Geometric, TensorFlow, CNN, LSTM, Transformer, Graph Autoencoder, VMD-CNN-Transformer hybrid architectures
Signal Processing: Variational Mode Decomposition (VMD), STL decomposition, ADF/KPSS stationarity testing, temporal feature engineering, frequency-domain analysis
Biomedical ML: Graph Neural Networks on molecular data, RDKit, DrugBank, biomedical feature engineering, link-prediction metrics (ROC-AUC, AP)
ML Engineering: scikit-learn, XGBoost, model benchmarking, SHAP interpretability, Diebold–Mariano significance testing
Deployment & Dev: Streamlit, Hugging Face Spaces, AWS (S3, SageMaker), Git, Jupyter Notebook

Projects

Drug–Drug Interaction Prediction using Graph Neural Networks | *PyTorch, PyTorch Geometric, RDKit* 2025

- Converted molecular SMILES structures from DrugBank into graph representations using RDKit; nodes encode atom features, edges encode chemical bonds — directly analogous to physiological sensor graph construction.
- Trained a Graph Autoencoder (GAE) in PyTorch Geometric to learn latent molecular embeddings and predict adverse drug–drug interactions via link prediction.
- Achieved strong link-prediction performance (ROC-AUC, Average Precision); model generalizes to unseen drug pairs, demonstrating robust biological pattern learning.

- Deployed real-time inference application on Hugging Face Spaces via Streamlit; end-to-end pipeline from raw molecular data to production prediction.

Temporal Signal Decomposition & Hybrid Deep Learning Forecasting | *PyTorch, VMD, CNN-Transformer* 2025

- Applied Variational Mode Decomposition (VMD) to decompose complex non-stationary time-series into intrinsic mode functions — a technique directly transferable to physiological signal preprocessing (ECG, PPG, EEG).
- Designed and trained a VMD-CNN-Transformer hybrid: CNN layers extract local temporal features, Transformer captures long-range dependencies; validated stationarity via ADF/KPSS before modeling.
- Benchmarked 8 model architectures (ARIMA, XGBoost, LSTM, CNN-Transformer); XGBoost achieved RMSE 56.43 (NSE 0.985) vs. ARIMA collapse (RMSE 555.33, negative NSE) — demonstrating architecture selection discipline.
- Engineered 50+ temporal features (lag windows 1–365, rolling statistics, harmonic encodings); Diebold–Mariano tests confirmed deep model superiority ($p<0.001$); produced SHAP-based model interpretability diagnostics across a multi-component fused signal.

Personalised Medicine Recommendation System | *Python, scikit-learn, TF-IDF, Streamlit* 2024

- Built a patient-facing medicine recommendation system over real-world clinical review data; engineered combined feature vectors from drug metadata, condition labels, and patient-reported outcomes.
- Applied TF-IDF vectorization and cosine similarity to identify clinically similar drug alternatives; system returns top-5 recommendations with condition context and patient ratings.
- Deployed live on Streamlit Cloud with model artifacts hosted on Hugging Face Datasets; demonstrates end-to-end biomedical ML deployment pipeline.

Leadership & Achievements

Novo Nordisk GBS Hackathon 2025 – Team Lead

Sep – Oct 2025

- Led team to deliver a working ML prototype under tight deadlines in a healthcare-adjacent problem domain; managed technical scoping, execution, and stakeholder communication.

IEEE CIEES 2025 – Published Research

2025

- Published peer-reviewed paper on a large-scale NLP pipeline for public health signal analysis; processed 700+ noisy Reddit documents through a full ML workflow.
- Benchmarked classical models (SVM, DT, NB, KNN, RF) against DistilBERT transformer; applied BERTopic and TF-IDF semantic clustering with Zipf filtering.
- Compared K-Means and BERTopic clustering approaches, translating findings into a peer-reviewed publication.