

SHUBHRO DEV

 github.com/spearb0lt  [linkedin.com/in/shubhro-dev](https://www.linkedin.com/in/shubhro-dev)  shubhro2004@gmail.com  [+918272979689](tel:+918272979689)

EDUCATION

- Rajiv Gandhi Institute of Petroleum Technology** 7.82 (CPI upto 7th semester)
(An Institute of National Importance along the lines of IITs)
Bachelor of Technology in Computer Science and Engineering 2022 - 2026
- Bharatiya Vidya Bhavans, Kolkata** 90.60%
Class 12th (CBSE Board) 2022

SKILLS

- Languages Used:** Python, C++, C, HTML, CSS, SQL
- Libraries:** LangChain, LangGraph, Huggingface, PyTorch, NLTK, NumPy, Pandas, Scikit-learn, Seaborn, Tensorflow, OpenCV, Matplotlib, SciPy
- Others:** AWS Bedrock, S3, Lambda, Sagemaker, Streamlit, NIM, Fine-tuning, Federated Learning, SQL, Power BI, Tableau
- Certifications:** Summer of Quant 2024(IIT-KGP), Web Developer Bootcamp 2024, Deep Learning Specialization

EXPERIENCE

- Siemens** January 2026 - June 2026
Technical Research Intern Global DAI R SDE-IN Team
 - Extensively researched on Validation and Verification of Output generated by LLMs.
 - Experimented with Deterministic ways to validate and verify Code and NLP use cases.
 - Worked on evaluating Bidify(existing RAG application) by inspecting all individual pipelines independently.
- VANET Lab, IIT Jodhpur** February 2025 - August 2025
Research Assistant Supervisor: Dr. Debasis Das, Professor IIT Jodhpur
 - Studied Time-Series Analysis, Transformers, Federated Learning, Membership Inference Attack(MIA), Encoders, Auto-encoders, VAEs, LSTM, GRU, RNN, Bi-RNN, Predictive Maintenance(PdM), Prognostics and Health Management (PHM) and Differential Privacy
 - Developed State of the Art architectures involving multimodal features for RUL prediction and Industrial PdM utilizing IDA-2024 Challenge SCANIA-X dataset
 - Proposed a novel hybrid architecture fusing a custom transformer-based time-series encoder with a sophisticated Transformer architecture, combined with differential privacy for added privacy.
- Core-AI Solutions** January 2025 - May 2025
Data Science Intern Remote
 - Engineered state-of-the-art multimodal AI models that integrate non-invasive MRI neuroimaging, blood plasma biomarkers (A β 42/40 ratio, p-tau181, NfL, GFAP), and cerebrospinal fluid parameters (A β 1-42, total tau, phosphorylated tau) to predict Alzheimer's disease onset and progression with high accuracy (AUC >0.90)
 - Conducted longitudinal predictive modeling on cohorts with 4+ years follow-up data, developing prognostic models that identify mild cognitive impairment patients at highest risk of conversion to Alzheimer's dementia using baseline multimodal assessments.
 - Integrated explainable-AI tools (SHAP and LIME) to visualize individual biomarker contributions and generate clinician-friendly reports highlighting top-ranked risk factors
- Adani Power** November 2024 - January 2025
Machine Learning Intern On-site
 - Developed time-series forecasting models using LSTM, GRU, and Prophet algorithms to predict power generation patterns and load demand, while conducting comprehensive power plant efficiency analysis using advanced data analytics methodologies to evaluate electricity generation, transmission, and distribution performance at Dahanu Thermal Power Station (DTPS).
 - Applied statistical analysis and optimization techniques including regression analysis, correlation studies, and linear programming for power loss minimization, implementing feature engineering techniques for temporal data including lag features, rolling statistics, and seasonal decomposition to develop custom algorithms for transmission efficiency calculation and grid performance optimization.
 - Created comprehensive performance dashboards utilizing Power BI and advanced visualization techniques to track key performance indicators including plant availability, heat rate optimization, fuel efficiency metrics, and transmission loss patterns across different voltage levels (11kV, 415V distribution networks), achieving actionable insights into operational optimization and cost reduction opportunities.
- VCA Lab, IIT(BHU) Varanasi** June 2024 - October 2024
Research Intern Supervisor: Dr. Sanjay Kumar Singh, Professor & HoD IIT-BHU
 - Studied Convolutional Neural Networks, Vision Transformers(ViT), Federated Learning, Transfer Learning, Multimodal models and Knowledge Distillation.
 - Conceptualized a Federated Learning algorithm for early diagnosis of Alzheimer's disease in patients based on multimodal data in form of biological markers and image data.

- Studied Supervised Learning Techniques, Data Preprocessing, Attention Mechanisms, Deep Neural Networks, and Transfer Learning.
- Experimented with various algorithms and techniques such as Attention, PCA, Color channel splitting analysis and Image partitioning analysis for better lung cancer classification and recognition.
- Proposed a novel method for lung cancer diagnosis of multi-source medical images from different datasets. Built a custom architecture based on transfer learning and attention mechanisms, which could detect and classify lung cancer from CT scans as well as histopathological images.

PROJECTS

- **Cross-Industry Privacy-Preserving Framework for PdM based on IDA 24 Industrial Challenge:** 
 - Preprocessed the [SCANIA-X](#) dataset and engineered a hybrid architecture that combines numerical and categorical features of the dataset into transformer embeddings and a TabTransformer for least information loss. Also implemented various differential privacy(DP) algorithms(Spectral-DP, DP-SGD) to protect training data. Designed a global model architecture that would perform well for other PdM datasets that involve numerical, categorical, or any multimodal features in their datasets.
 - Infused a Federated training approach based on flwr library, allowing global model aggregation across heterogeneous clients having different computing power, parameters, and hyperparameters; mirroring real-life industrial equipment.
 - Presently it is the best RUL prediction model on the dataset with MSE of 2725 whilst keeping training data privacy. Developed an advanced MIA that considering white-box, gray-box, and black-box features, along with time-series specific seasonality/trend features; achieved a MIA success rate(AUC) of 49.12% and MIA accuracy of 49.59%, thus solidifying the claim.
- **SENTRAL Multi Spectrum Stock Analysis using custom LLMs:** 
 - Conducted a fundamental analysis of target companies using 14 metrics (P/E, P/B, Debt, ROE, etc.) plus Piotroski F-Score and Altman Z-Score via custom functions. Executed technical analysis using 18 indicators (EMA20/50, SMA20/50, RSI14, MACD, ATR, etc.) and performed peer comparisons across sector companies.
 - Collected news via seven APIs and web scraping, filtered items by relevance score, then fed the curated corpus to 10 LLMs (including finance-tuned and SOTA reasoning models) for sentiment extraction and signal generation. Visualized correlations and feature importance to support decision making.
 - Applied Transformer and LSTM models to forecast price trends for reference and ensembled sentiment with fundamental and technical indicators to compute buy/hold/sell probabilities. Boosted newbie portfolio worth by +45% over five months.
- **Context-Aware Multimodal Knowledge Retrieval System:** 
 - Built a context-aware parser to automatically extract and separately process multimodal content (images, tables, equations, graphs, text) and summarize each modality into vector embeddings. Employed modality-specific pipelines and selected appropriate LLMs per content type to improve summary quality; stored embeddings and summaries in ChromaDB using Hugging Face embeddings.
 - Designed a multi-vector retrieval strategy that links document summaries to original content for improved context preservation. On query, the system retrieves relevant multimodal content and generates comprehensive, source-cited answers that reference text, table data, and image insights.

PUBLICATIONS

1. S. Dev, P. Roy, N. Chakraborty, R. Sarkar, “**Lung Cancer Identification from CT Scans using a Soft-attention enabled Deep Transfer Learning Model**” (Accepted at IEEE ISACC 2025) 
2. S. Dev, A. Ghosh, D. Das, “**A Differentially Private Hybrid Transformer-Mamba Architecture for Industrial Predictive Maintenance on Multivariate Time Series Data**” (In submission: IEEE Transactions on Intelligent Transportation Systems)
3. S. Dev, J. Konwar, M. Aggarwal, S. Mishra “**A Multiview Multiobjective Clustering Approach to Unsupervised Extractive Summarization**” (Manuscript in preparation; Target: AACL 2026)

ACHIEVEMENTS AND EXTRA-CURRICULAR

- Global Rank **10** at [Meta Hacker Cup 2025](#) (AI Track)
- Reviewer at International Conference on Cyber Security and Artificial Intelligence(ICCSAI 2025)
- **Finalist** of DPBH 2023 conducted by Ministry of Consumer Affairs
- Codechef Max Rating **1622**(3*)
- Secretary of [OWASP](#) RGIPT Student Chapter
- Qualified JEE-Advanced 2022 with AIR 17078
- Secured AIR 1859 in WBJEE 2022