

Pronab Chandra Roy

✉ pronabroy360@gmail.com | 🌐 pronabroy360 | 📞 pronabroy360 | 📄 pronabroy360
☎ +880 1533834552

Research Profile

By profession, I am a **Big Data Engineer** at **ACI Limited**, one of the largest conglomerates in Bangladesh, where I work on reliable data pipelines, KPI datasets, and analytics-ready reporting systems. Before this, I worked in **US healthcare RCM** at Augmedix / Commure, where I handled claims, denials, reimbursements, billing operations, and EHR-driven patient encounter workflows. I am now building public, reproducible projects around **CMS DE-SynPUF**, **TCGA/GTEX**, and **MIMIC-IV** to prepare for PhD labs working on large healthcare datasets, clinical data engineering, healthcare AI, and biomedical data mining.

Education

National Institute of Technology, Silchar [Certificate]

Assam, India

B.Tech in Computer Science and Engineering

Dec 2020 – June 2024

CGPA: **8.71/10.00** (Top 5% in the Department)

Thesis: Advanced Image Denoising through Combined Directional Wavelet based NLM and BM3D Technique.

Publications

1. A. D. Emon, **P. C. Roy**, L. D. Singh, and A. K. Saha, "Combining Haar Transform and Row-Column Directional RLE in Image Compression," *IEEE ICIC 2023*, Manado, Indonesia. DOI: [10.1109/ICIC60109.2023.10382031](https://doi.org/10.1109/ICIC60109.2023.10382031)

Work Experience

Big Data Engineer, ACI Limited

Jan 2026 – Present

- Maintain and improve production data pipelines so business teams receive clean and reliable data on time.
- Build SQL- and Python-based KPI datasets, reporting tables, and documentation for repeatable analytics work.

Data Analyst – RCM, Augmedix / Commure US Healthcare

Nov 2024 – Jan 2026

Dhaka, Bangladesh [Certificate]

- Worked with US healthcare RCM data covering claims, reimbursements, denials, billing operations, and payer-side patterns.
- Built dashboards for billing quality, denial trends, reimbursement patterns, and team productivity; reduced manual QA effort by 70%.
- Cleaned, validated, transformed, and documented healthcare operational data for consistent downstream reporting.
- Used and documented EHR/practice-management workflows across WebPT, eClinicalWorks, Epic, AdvancedMD, and DrChrono.

Research Intern – SN Bose Summer Internship, NIT Silchar

Jun 2023 – Jul 2023

Silchar, India

- Worked on a research problem involving image data, compression methods, and pixel-level transformation analysis.
- Implemented and tested ideas in Python/NumPy and contributed to a paper later published at IEEE ICIC 2023.

Selected Healthcare Data Projects

CMS DE-SynPUF Medicare Claims Pipeline – Claims Analytics

2026

- Loaded CMS DE-SynPUF beneficiary, inpatient, outpatient, carrier, prescription, diagnosis, and procedure files into a local DuckDB warehouse.
- Cleaned claim dates, beneficiary fields, cost columns, chronic-condition flags, diagnosis codes, procedure codes, and prescription events for consistent analysis.
- Built patient-year outputs showing total cost, service use, prescription use, chronic-condition summaries, and next-year high-cost labels.
- Created a Streamlit dashboard, FastAPI lookup service, model-training workflow, quality reports, and safe synthetic explanation reports for professor review.

CancerOmicsLake – TCGA/GTEX Bioinformatics Data Engineering

2026

- Connected open-access TCGA cancer project metadata with GTEx normal tissue expression data for breast, lung, and colon cancer examples.
- Parsed expression files, mutation MAF files, patient/sample metadata, and file manifests into clean parquet tables.
- Generated tumor-vs-normal expression summaries, mutation-frequency tables, cohort summaries, and cancer-gene-tissue relationship outputs.
- Exported graph-ready CSV files for Graphify/Neo4j-style visualization and added quality reports, ingestion traceability, tests, CI checks, API endpoints, and dashboard pages.

MIMIC-IV Clinical Data Warehouse – ICU EHR Analytics

2026

- Built a reproducible DuckDB workflow for MIMIC-IV-style hospital and ICU tables using demo, synthetic, and credentialed local execution modes.
- Prepared cleaned tables for patients, admissions, ICU stays, diagnoses, procedures, labs, medications, microbiology, and ICU events.
- Created cohort summaries, admission-level outputs, ICU-stay outputs, first-24-hour clinical features, and baseline mortality-model datasets.
- Added aggregate dashboards, FastAPI endpoints, validation/profile reports, synthetic demo packs, and guarded natural-language-to-SQL querying that avoids patient-level exposure.

Skills

Languages	Python, SQL, Bash
Data Engineering	ETL/ELT, Data Warehousing, DuckDB, PostgreSQL, MySQL, MongoDB, Apache Spark, Airflow, Kafka
AI/ML & Analytics	Pandas, NumPy, Scikit-learn, Keras, Matplotlib, Feature Engineering, Model Evaluation
Apps & Tools	FastAPI, Streamlit, Git, Linux, AWS, Databricks, pytest, Makefile-based workflows
Healthcare Domains	EHR Analytics, Claims Analytics, RCM, Clinical Data Warehousing, Cohort Building, Biomedical Data Mining
Research Interests	Large Healthcare Datasets, Biomedical Informatics, Clinical NLP/LLMs, Reproducible EHR/Claims Research

Coding Profiles

- **LeetCode** — Knight | 1900+ rating | 4× Top 500 Global Rank
- **CodeChef** — 4-Star | 1800+ rating | 200+ Problems Solved | Global Rank 12 (1 contest)

Certifications

- Deep Learning Specialization — *DeepLearning.AI / Coursera* Credential ID: ECD39RU4D56N
- Machine Learning with Apache Spark — *IBM / Coursera* Apr–Dec 2023

Honors and Awards

- ICCR Scholarship (2020–2024 Session) — Indian Council for Cultural Relations, fully funded Indian Govt. merit scholarship.
2020

Extra Curriculum

Competitive Programmer, Active on LeetCode and CodeChef; achieved Knight ranking and 4-star rating, reflecting strong algorithmic problem-solving foundations.

Healthcare Data Portfolio, Maintains public, privacy-safe repositories for healthcare data engineering projects intended for academic review and PhD outreach.

References

[Adam Mohiuddin](#)
Senior Manager, Commure (USA)
Email: adam.mohiuddin@commure.com
LinkedIn: [linkedin.com/in/apmo](https://www.linkedin.com/in/apmo)

[Dr. Anish Kumar Saha](#)
Assistant Professor,
Dept. of CSE, NIT Silchar
Email: anishkumarsaha@cse.nits.ac.in
Mobile: +91-7005345047