

SHADAB ANSARI

M.Tech (Computer Science & IT) | Machine Learning | Data Science | Explainable AI | MLOps

shadabahamed63@gmail.com | Hyderabad, Telangana, India

GitHub: github.com/shadabdlN | LinkedIn: [linkedin.com/in/shadab-ansari-dln](https://www.linkedin.com/in/shadab-ansari-dln) | Medium: shadabdlN.medium.com

PROFESSIONAL SUMMARY

M.Tech (Computer Science & IT) student at MANUU, Hyderabad, with hands-on experience building an end-to-end, explainable machine learning system for colorectal cancer survival prediction on 257,827 patient records from the SEER cancer registry. Skilled in Python, scikit-learn, XGBoost, LightGBM, SHAP and survival analysis, with MLOps and deployment experience in MLflow, DVC, GitHub Actions, Flask and Docker. Seeking a Data Scientist, Machine Learning Engineer or AI/ML research role.

TECHNICAL SKILLS

Programming & Databases: Python, SQL, MySQL

Machine Learning: scikit-learn, XGBoost, LightGBM, Gradient Boosting, Random Forest, Logistic Regression, SVM, Optuna (hyperparameter tuning), Feature Selection, Cross-Validation, Model Calibration, Imbalanced Data Handling

Survival Analysis & Statistics: lifelines, scikit-survival, Cox Proportional Hazards, Random Survival Forest, Kaplan-Meier, C-index, SciPy, Bootstrap Confidence Intervals, Decision Curve Analysis

Explainable AI (XAI): SHAP, LIME, Partial Dependence Plots, Permutation Importance

Data Analysis & Visualization: NumPy, Pandas, Matplotlib, Seaborn, Plotly, Exploratory Data Analysis (EDA)

MLOps & Deployment: MLflow (Experiment Tracking, Model Registry), DVC, DagsHub, pytest, GitHub Actions (CI/CD), Docker, Kubernetes (Minikube), Flask, Gunicorn, REST API

Web & Tools: HTML5, CSS3, Git, GitHub, Jupyter Notebook, VS Code

PROJECTS

Explainable ML for 5-Year Colorectal Cancer Survival Prediction

M.Tech Research

Python, scikit-learn, XGBoost, LightGBM, SHAP, lifelines, MLflow, DVC, Flask, Docker, Kubernetes, GitHub Actions

GitHub: github.com/shadabdlN/crc-seer-survival

- Built an end-to-end ML pipeline on 257,827 colorectal cancer patients from the SEER 17 registries (cohort selection, data cleaning, feature engineering) to predict five-year overall survival.
- Benchmarked 6 classifiers (Logistic Regression, SVM, Random Forest, Gradient Boosting, XGBoost, LightGBM) with stratified 5-fold cross-validation; the final Gradient Boosting model achieved ROC-AUC 0.877 (95% CI 0.874-0.880), recall 0.878 and calibration slope 1.02 on a held-out test set of 50,338 patients, beating the stage + age + sex clinical baseline (ROC-AUC 0.835).
- Validated robustness with temporal validation (trained on 2010-2015, tested on 2016-2018: ROC-AUC 0.878) and external validation of the frozen model on the AACR Project GENIE BPC CRC dataset (ROC-AUC 0.790).
- Compared 5 time-to-event survival models (Cox PH, CoxNet, Random Survival Forest, Gradient Boosting Survival, XGBoost-Cox; best C-index 0.809) and explained predictions with SHAP, LIME and partial dependence.
- Automated a reproducible MLOps workflow: DVC pipeline (prepare, train, evaluate), MLflow experiment tracking and Model Registry on DagsHub, pytest tests on synthetic data, and GitHub Actions CI with a no-patient-data check and Docker build.
- Deployed a Flask web app that returns a calibrated 5-year survival probability with the top SHAP risk factors for each patient, containerized with Docker and Gunicorn, with Kubernetes deployment manifests.

Google Drive Clone | HTML5, CSS3

Frontend Project

- Replicated the Google Drive user interface with file listing, folder navigation and a responsive layout.

RESEARCH & PUBLICATIONS

- S. Ansari, M. Omar, "Explainable, Calibrated and Externally Validated Machine Learning for Five-Year Colorectal Cancer Survival", manuscript prepared for submission to ICCIDSDT-2027 (Springer LNNS); extended journal version with GENIE external validation in preparation.

EDUCATION

M.Tech, Computer Science & Information Technology

Pursuing (2nd Year)

Maulana Azad National Urdu University (MANUU), Hyderabad | Research Guide: Dr. Mohd Omar

Master of Computer Applications (MCA)

2022 - 2024

Maulana Azad National Urdu University (MANUU), Hyderabad | CGPA: 8.4/10

ACHIEVEMENTS, LEADERSHIP & CERTIFICATIONS

- 3rd Rank, Coding Ninjas Coding Byte Contest.
- Coordinator of Cultural Activities, MANUU; Member, Food-cum-Audit Committee, Boys' Hostel, MANUU (2023-24).
- Certification: 7-Day Bootcamp, Google Drive Clone (Dev Town with Microsoft Learn and GDSC KIIT).