Shreya Patil

Mobile +1-443-453-8281 | shreya.patil.u@gmail.com | Baltimore, MD

[LinkedIn](#) | [Website](#) | Work Authorization: Lawful Permanent Resident (Green Card)

PROFESSIONAL SUMMARY

I thrive in understanding business challenges and delivering impactful results using research and data driven applications. With a strong background in Machine Learning, Deep Learning, Gen AI and High-Performance Computing (GPU) along with healthcare, pharmacovigilance and biomedical data domain expertise, I am dedicated to developing innovative solutions that address complex challenges.

CORE PROFICIENCIES

Relationship Building	Data Science	Pharmacovigilance Data
Communication & Collaboration	Neural Networks & Generative AI	Electronic Health Records
Technical Writing	Healthcare Data Analytics	Biomedical Data

PROFESSIONAL EXPERIENCE

Research Fellow, U.S. Food and Drug Administration (FDA)   Jul 2023- Mar 2025
- Applying Deep Learning techniques to discover patterns in FDA Adverse Event Reporting System (FAERS) database using node classification.
- Conducted comparative analysis between traditional statistical models (preferential attachment, node fitness) and ML models for inference and prediction.
- Evaluated neural network algorithms (CNN, GCN, GAT) for node connection prediction, demonstrating accuracy comparable to traditional statistical models.
- Analyzed healthcare data relationships by constructing Knowledge Graph of drug and adverse effect using biomedical LLM BERN2.
- Implemented data preprocessing and feature extraction on 2 TB of Electronic Health Records data.

Data Scientist, Redapt Inc   May 2022- Aug 2022
- Implemented NLP-based medical entity extraction system using BERT to analyze unstructured data and extract structured information to reduce manual review time.
- Built predictive model using Decision Trees and XGboost to optimize resource management, resulting in 25% reduction in processing time for the clients admin team.
- Created custom analytics pipeline integrating Azure Blob Storage and PowerBI for tracking resource allocation.

Research Assistant, UMBC, MD (PI: Dr. Milton Halem)   Aug 2021- May 2023
- Created a forecasting algorithm for a Wildfire Digital Twin to forecast the spread and the smoke impacts on air quality.
- Designed a Deep Neural Network architecture for predicting time series microphysics parameters, to replace components in the NASA Unified-Weather Research Forecasting (NU-WRF) model.
- Trained and tested Auto-Keras Neural Net Architecture Search algorithm on NU-WRF models microphysics output over US region achieving RMSE of 0.015.
- Evaluated Global Tree Cover Loss to predict the impact of mega wildfires and deforestation on the irreversible climate change in tropical forests.
- Acquired and analyzed ~30 TB data from the Hadoop distributed file system to retrieve essential parameters of geospatial data.

Software Engineer, Capgemini Technology Service LTD, INDIA   Mar 2018- Nov 2019

- Designed and implemented automation pipelines for database management and data processing workflows in Azure DevOps environment for enterprise clients.
- Mentored 10+ junior developers on software development best practices, version control workflows, and cloud infrastructure strategies.
- Created documentation frameworks and technical specifications ensuring knowledge transfer and code maintainability.

EDUCATION

Master of Professional Science, Data Science *University of Maryland, Baltimore County (UMBC), Baltimore, MD*
Master of Technology, Computer Science and Engineering *Shivaji University, MH, India*
Bachelor of Engineering, **Computer Science and Engineering** *Shivaji University, MH, India*

TOOLING & TECHNICAL PROFICIENCIES

- **Programming**: Python (Pandas, NumPy, NLTK, Pytesseract, spaCy, Scikit-learn, PyTorch, Tensorflow, AutoKeras, DGL, Neo4j, NetworkX, Seaborn), Linux, C++ (CUDA), SQL.
- **Big Data**: Spark, Distributed Computing, Azure Cloud Platforms.
- **Tools**: Jupyter, GitHub, VS code, Databricks, Azure ML Studio, Tableau.

PUBLICATION / ARTICLES

- Accelerating the WRF-CHEM Model Using a Machine Learning Emulation *Poster*, 2022 American Meteorological Society Conference.
- Patil S U (2019), Optimal Number of Cluster Identification using Robust K-means Algorithm. International Research Journal of Engineering and Technology.
- Patil S U (2018), A Review of Clustering and Clustering Quality Measurement. International Research Computer Engineering in Research Trends.
- Technical Content Creation *Medium*

PROJECTS

Prediction of Drug Binding Affinity of Protein Using Spark ML *View Repo*
Extracted biological data for a single protein target. Used PySpark classification models to predict binding affinity between sample drug compounds and target protein. Compared multiple models to find optimum solution for prediction.

Fabric Defect Detection Using CUDA Parallel Processing *View Repo*
Implemented GPU-accelerated image processing pipelines for textile defect detection algorithms using CUDA. Evaluated multiple detection methods including Gabor filtering, image difference, and Canny edge detection to optimize detection accuracy across various fabric types.

Optimal Number of Cluster Identification using Robust K-means Algorithm *View Repo*
Developed a Robust K-means Algorithm to identify optimal numbers of clusters in categorical data by removing noise clusters. Measured goodness of clusters using Silhouette Coefficient.

Computer Vision for Histopathological Cancer Detection *View Repo*
Developed CNN-based image processing models (VGG, AlexNet) to automatically detect cancer patterns in medical tissue images, improving diagnostic efficiency and accuracy. Enhanced model performance through data augmentation and preprocessing techniques that improved detection of subtle patterns in diverse histopathological samples.