Shikhar Shukla

shikharshuklams@gmail.com • +1 937 782 8345 • USA • GitHub • ORCiD • Portfolio

SUMMARY

Biomedical researcher with 10+ years of combined clinical and data science experience. Skilled in analyzing EHR, imaging, and physiologic time-series data and building reproducible pipelines for risk prediction and clinical forecasting. Experienced in applying statistical and machine learning methods to large-scale biomedical datasets. Strong record of peer-reviewed publications and close collaboration with clinicians on translational research.

WORK EXPERIENCE

Indiana University Remote

Senior Data Analyst *Aug 2025 - Current*

Postoperative Delirium Prediction Study - Dept of Anesthesiology and Critical Care Medicine

- Developed a multi-domain machine learning framework for predicting postoperative delirium using perioperative EHR data from >70 million anesthesia records, integrating intraoperative MAP dynamics, frailty indices, anesthetic exposures, and surgical context. The stacked ensemble (LightGBM-XGBoost-logistic meta-learner) achieved high discrimination (ROC-AUC ≈ 0.87) with well-calibrated risk estimates and clinically interpretable feature effects.
- Engineered domain-level risk models (Hemodynamics, Frailty, Anesthesia, Surgical Context, and Comorbidities) with calibrated meta-fusion to enhance interpretability and align modeling structure with clinical reasoning. Established reproducible preprocessing and explainability workflows (SHAP-based analysis) to ensure methodological transparency and support translational use by anesthesiology researchers. Manuscript in preparation for peer-reviewed submission.

Voice-Based Parkinson's Detection Study - Bridge2AI

- Developed reproducible pipelines to evaluate voice-derived digital biomarkers for Parkinson's Disease using the NIH Bridge2AI Voice dataset. Preprocessed MFCC time-series from 442 participants into participant-level sequences, benchmarking traditional statistical models against temporal deep learning architectures (PatchTST Transformer, 1D ResNet). Manuscript is in progress.
- Designed and optimized a residual convolutional framework (ResNet-1D) achieving strong discrimination and improved recall for PD detection, demonstrating that local spectrotemporal patterns capture disease-related microprosody more efficiently than Transformer-based approaches. The model provides a scalable, interpretable foundation for non-invasive digital phenotyping.

US Centers for Disease Control and Prevention Atlanta, GA

Public Health Informatics Fellow *Aug 2024 - July 2025*

- Developed an AI-assisted schema mapping framework in Palantir Foundry (Python, GPT-4 integration) to automate ingestion and transformation of multi-state fungal disease surveillance data. This work reduced manual curation and improved data readiness for epidemiologic research and national reporting.
- Standardized and refactored transformation pipelines for Fungal pathogen datasets, improving consistency, linkage accuracy, and reproducibility across state-submitted surveillance data, essential for reliable trend and outbreak analyses.
- Implemented a facility-matching and geocoding workflow using Foundry's AI Model Catalog, integrating NLP-based entity resolution and spatial standardization to improve data quality for antibiotic resistance and fungal pathogen surveillance.
- Designed a multi-stage enrichment pipeline for the NNDSS Operational Data Store, recovering missing surveillance variables and preparing *Candida auris* data for HL7-compatible epidemiologic analysis.
- Developed automated transformation pipelines for state-level histoplasmosis data, normalizing raw REDCap inputs into harmonized, analysis-ready tables for integration within CDC's national fungal surveillance system.
- Contributed to CDC's Generative AI Tiger Team, evaluating the use of large language models for metadata extraction and data standardization; recognized with the *CDC AI Excellence in GenAI Adoption Award* for advancing AI applications at CDC.

Indiana University School of Medicine Indianapolis, IN

Junior Data Scientist, Dept of Anesthesiology and Critical Care Medicine *Jan 2024 - May 2024*

- Developed a postoperative mortality prediction model by identifying Mean Arterial Pressure (MAP) time-series features associated with 30-day mortality risk, not captured by traditional methods. Applied time-series analysis, change point detection using Pruned Exact Linear Time (PELT), and L1 regularization to extract physiologic signal patterns beyond static thresholds.
- Engineered a reproducible feature extraction and modeling pipeline in Python using *tsfresh* and *tsfel*, integrating derived MAP features into logistic regression and survival-focused machine learning models. Findings were accepted and presented at the AMIA Annual Symposium 2024, demonstrating the clinical value of high-resolution intraoperative hemodynamic analytics.

Purkayastha Lab for Health Innovation (PLHI), Indiana University Indianapolis, IN

Research Data Analyst *Aug 2022 - May 2024*

- **Chronic Wound Healing Prediction:** Developed predictive modeling pipelines for 14,500+ chronic wounds, applying advanced imputation (XGBoost), statistical forecasting (ARIMA, Holt-Winters, Prophet), and deep learning (LSTM, BiLSTM). Conducted subgroup fairness analysis across race and gender to ensure equitable performance. Results validated with rigorous metrics (RMSE, R^2) and published as a second-author IEEE paper.
- **Radiology AI Assistant Project**: Integrated a general-purpose AI assistant into the open-source LibreHealth Radiology Information System, linking DICOM-SR annotations with OHIF viewer workflows to allow radiologists to retrain models directly in clinical use. Pioneered novel human-in-the-loop methods (few-shot and swarm learning) for real-time diagnostic model refinement, enhancing accuracy and reproducibility. Co-authored results presented at AIME 2023 (NSF-supported).
- **DHIS2-OpenCPU Statistical Analysis Tool**: Built a secure integration of DHIS2 with OpenCPU to deliver on-demand statistical testing within interactive dashboards. Eliminated manual workflows while maintaining reproducibility and data security; accepted for presentation at the DHIS2 2024 Annual Conference.

Family Dental Clinic Indore, India

Head Dentist and Clinical Data Analyst *Dec 2013 - Jun 2022*

- Analyzed over 2,000 patient records using statistical methods (t-tests, ANOVA, Chi-square) to evaluate treatment outcomes and explore risk factors related to socioeconomic status, hygiene, and diet.
- Led community-based oral health studies, applying descriptive and inferential analyses to assess disease prevalence, identify at-risk populations, and guide preventive health strategies.
- Integrated clinical practice with data-driven research, demonstrating early experience in linking healthcare delivery with evidence-based and reproducible analytic methods.

EDUCATION

Master of Science, Health Informatics

Indiana University-Indianapolis, (GPA: 3.8/4.0) *Aug 2022 - May 2024*

Bachelor of Dental Surgery

Devi Ahilya University-Indore, India *Oct 2006 - Dec 2010*

SKILLS

Programming & Analysis: Python, R, SQL; data wrangling, database querying, reproducible pipeline design, and workflow automation.
Machine Learning & Modeling: Statistical, ensemble, and deep learning methods (XGBoost, LightGBM, CNNs, Transformers); time-series and survival modeling; causal inference and Bayesian analysis; model calibration, interpretability (SHAP), and fairness evaluation.
Clinical & Biomedical Data Expertise: EHR (perioperative, ICU), physiologic monitoring, imaging (DICOM-SR), HL7 v2 & FHIR standards, clinical ontologies (SNOMED-CT, LOINC, ICD-10, RxNorm), and public health surveillance datasets.
Data Engineering & Integration: Palantir Foundry, DHIS2, SQL Server (SSMS, Azure Data Studio), dynamic SQL, PySpark, Azure Databricks, ETL design, data quality control, and LLM-assisted automations for health data workflows.
Reproducible Research & Cloud Computing: Git/GitHub, Docker; scalable analytics in Azure and GCP environments; privacy-preserving and HIPAA-compliant data handling.
Visualization & Scientific Communication: Tableau, Power BI, Matplotlib, Seaborn; publication-quality figures, and peer-reviewed research manuscripts and conference presentations.
Healthcare Background: 10+ years of combined experience in clinical practice, biomedical informatics, and applied AI research.

PUBLICATIONS & OTHERS

- Purkayastha, S., Isaac, R., Anthony, S., Shukla, S., Krupinski, E. A., Danish, J. A., & Gichoya, J. W. (2023). A general-purpose AI assistant embedded in an open-source radiology information system. arXiv. https://arxiv.org/abs/2303.10338
- Paddo, A. R., Shukla, S., Steiner, S. S., Purkayastha, S., & Sen, C. K. (2024). Interpolating and forecasting wound trajectory using machine learning approaches. 2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS), 146–152. https://doi.org/10.1109/CBMS61543.2024.00032
- Analysis of Intraoperative Blood Pressure Time Series and Extraction of Features Predictive of 30-Day Mortality - 1st Author, AMIA Annual Symposium 2024 (Poster Presentation)
- Enabling Data-Driven Exploration: DHIS2-OpenCPU Integration for Real-Time Statistical Analysis - 1st Author, DHIS2 Annual Conference 2024