Shikhar Shukla

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

SUMMARY

Health informatics researcher with 10+ years of combined clinical and data science experience, including dental practice and biomedical informatics. Skilled in designing reproducible analyses of large-scale datasets (EHR, imaging, physiologic sensors, surveillance, and speech). Experienced in developing and validating statistical and machine learning models for risk prediction, clinical forecasting, and decision support. Strong record of translating methodological innovations into peer-reviewed publications, conference presentations, and collaborative research with clinicians and public health experts.

WORK EXPERIENCE

Indiana University Remote

Senior Data Analyst *Aug 2025 - Current*

NIH CIMDAR-HIVE Project & Bridge2AI Voice Projects

- Working on NIH-funded research initiative to design training resources and datathon workflows, equipping faculty and students at HBCUs and MSIs with applied skills in multimodal health data science, fairness in AI, and clinical time-series modeling.
- Developed and benchmarked sepsis prediction pipelines on the MIMIC-IV ICU dataset (>70,000 admissions), comparing statistical forecasting methods (ARIMA, Holt-Winters) with deep learning (Temporal Fusion Transformer) to evaluate trade-offs in recall, calibration, and interpretability for clinical forecasting.
- Built reproducible benchmarking pipelines for speech-derived clinical features (MFCCs) using the NIH Bridge2AI Voice Dataset; demonstrated the utility of LSTM and Transformer architectures for multivariate temporal modeling and established workflows for future disease-detection studies.

Postoperative Delirium Prediction Study

- Designed and validated predictive modeling pipelines on perioperative EHR data (>70M anesthesia records), engineering novel anesthesia-derived features (e.g., MAC exposure metrics, MAP thresholds) and optimizing recall-weighted models (AUROC 0.83, recall ~99%) for delirium risk detection.
- Partnered with anesthesiologists to balance methodological rigor with clinical interpretability, benchmarking machine learning models (XGBoost, Random Forest, logistic regression) alongside statistical odds ratio models for translational use.

US Centers for Disease Control and Prevention Atlanta, GA

Public Health Informatics Fellow *Aug 2024 - July 2025*

FungiSurv Informatics Infrastructure

- Built an AI-assisted schema mapping system in Palantir Foundry (Python, GPT-4 integration), enabling automated ingestion and transformation of multi-state fungal surveillance data; reduced manual mapping effort by ~60% while ensuring reproducibility through version-controlled human-in-the-loop validation.
- Refactored and standardized Case, Lab, Outpatient, and Hospitalization pipelines, strengthening integrity, linkage readiness, and governance across jurisdictional datasets, thereby supporting reliable national surveillance of fungal diseases.

NNDSS ODSE Enrichment Pipeline

- Designed and implemented a multi-stage enrichment pipeline (SQL + Palantir Foundry) across >200 relational tables in the NNDSS Operational Data Store; leveraged ERD mapping and dynamic SQL to efficiently extract and recover missing priority surveillance variables.
- Enhanced completeness of Candida auris monitoring datasets and prepared outputs for transition to HL7 v2-compatible standards, directly improving epidemiological analyses and interoperability.

Indiana University School of Medicine Indianapolis, IN

Research Assistant, Department of Anesthesiology *Jan 2024 - May 2024*

- Engineered and validated novel hemodynamic features from intraoperative Mean Arterial Pressure (MAP) time-series (>10 years of noncardiac surgeries), applying change point detection (PELT), dynamic time warping, and entropy-based complexity metrics to characterize patterns of instability associated with 30-day mortality.
- Identified 20 statistically significant predictors of adverse outcomes, integrating them into logistic regression and survival-focused ML models; findings highlighted clinical markers of risk and were accepted for presentation at the AMIA Annual Symposium 2024.

Indiana University - Health Informatics Lab Indianapolis, IN

Research Assistant *Aug 2022 - May 2024*

- **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).
- **Chronic Wound Healing Prediction:** Built a predictive modeling framework for >14,500 chronic wounds, combining advanced imputation (XGBoost) with forecasting methods (ARIMA, Holt-Winters, Prophet) and deep learning (LSTM, BiLSTM). Demonstrated novelty in subgroup fairness analysis across race and gender, ensuring equity in predictive accuracy. Established reproducible pipelines and validated findings with rigorous metrics (RMSE, R^2). Second-author paper accepted at IEEE.
- **DHIS2-OpenCPU Statistical Analysis Tool**: Architected a secure integration of DHIS2 with OpenCPU to deliver real-time statistical testing within an interactive web interface, emulating RStudio capabilities. This was the first tool to combine health system dashboards with on-demand statistical analysis, eliminating manual workflows while maintaining strict data security. First-author abstract accepted at DHIS2 2024 Annual Conference.

Family Dental Clinic Indore, India

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

- Directed operations across two clinics, performing complex oral surgeries, training junior clinicians, and managing patient workflows while integrating data analytics into routine care.
- Conducted statistical analyses (T-tests, ANOVA, Chi-square) on >2,000 patient records to evaluate treatment outcomes and risk factors such as socioeconomic status, hygiene, and diet.
- Designed and led community-based oral health screenings, applying descriptive and inferential methods to assess disease prevalence, identify at-risk groups, and shape preventive interventions.
- Initiated data-driven community health programs that improved treatment adherence and expanded access to preventive dental care in underserved populations.

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 & Data Science: Python, R, SQL; data wrangling, statistical analysis, reproducible workflows

Machine Learning & AI: Logistic regression, Random Forest, XGBoost, deep learning (LSTM, BiLSTM, Transformers, TFT), time-series forecasting, survival analysis, model interpretability (SHAP, ORs, CIs)

Clinical Informatics: EHR and perioperative workflows, HL7 v2, Epic (familiarity), decision support systems, disease surveillance platforms

Databases & Systems: SQL databases, database design, ETL pipelines, data integration (Palantir Foundry, DHIS2, OpenCPU)

Visualization & Reporting: Tableau, Power BI, Matplotlib, Seaborn; dashboards, statistical tables, manuscript-ready figures

Healthcare & Research: 10+ years in clinical care and biomedical informatics; advanced analytics, and translational 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
- AI Tiger Team Evaluation of OpenAI's Deep Research for Public Health - Contributor & Manuscript in Preparation, CDC AI Tiger Team