Curriculum Vitae

Muhammad Abiodun Sulaiman, B.Tech., M.Sc.

Nigeria (working remotely)

Email: abiodun.msulaiman@gmail.com | Mobile: +234 810 831 6393

ORCID: 0000-0001-9161-2608 | Google Scholar | ResearchGate | LinkedIn | GitHub | Research Portfolio

Research Interests / Summary

I am a statistician and machine-learning engineer interested in using data to answer practical health questions. My background combines statistical modelling with hands-on experience building healthcare-data systems. During my B.Tech., I studied neonatal mortality in Nigeria over a 28-year period. For my M.Sc., I trained and evaluated a range of existing machine-learning models for ovarian cancer prediction using a clinical registry. Since then, I have worked on FHIR-based EHR pipelines, DHIS2-integrated decision-support tooling, and vartriage, a Python library for clinical variant interpretation that can process whole-genome VCF files containing more than four million variants. I am now looking to build on this experience through research in computational genomics, statistical modelling and clinical AI, including longitudinal patient trajectories and risk stratification. I work mainly in Python and also use R, SQL, Bash, and AWS.

Education

M.Sc. Statistics 2022–2024

University of Ilorin, Nigeria

- **Score:** 58.79%
- **Thesis:** Ovarian Cancer Prediction Using Machine Learning Algorithms.

B.Tech. Mathematics (Statistics option) 2013–2018

Federal University of Technology Minna, Nigeria

- **CGPA:** 2.43 / 5.00
- **Thesis:** Time Series Analysis of Neonatal Mortality in Nigeria, 1990–2017.

Publications & Research Output

Peer-Reviewed Publications

1. Dere, I. G., Ahmad, G. B. M., Oduleye, O. O., Ojekunle, J. A., & **Sulaiman, M. A.** (2023). Assessment of Operational Performance of Inland Water Transport in Borgu Local Government Area of Niger State. *Journal of Environmental Studies*, 5(1), 213–230. doi.org/10.5281/zenodo.21630892 (archived copy)
2. Usman, A., **Sulaiman, M. A.**, & Abubakar, I. (2019). Trend of neonatal mortality in Nigeria from 1990 to 2017 using time series analysis. *Journal of Applied Sciences & Environmental Management*, 23(5), 865. doi.org/10.4314/jasem.v23i5.15

Preprints

1. **Sulaiman, M. A.**, & Oyeyemi, B. F. (2026). A Curated Pharmacogenomic Allele Catalog for Sub-Saharan African Populations. *medRxiv*. doi.org/10.64898/2026.08.25.26361354
2. **Sulaiman, M. A.**, & Oyeyemi, B. F. (2026). Concordance of Automated ACMG Variant Classification with Expert-Curated Assertions: A Systematic Evaluation Using the ClinGen Evidence Repository. *Research Square*. doi.org/10.21203/rs.3.rs-10615801/v1

3. **Sulaiman, M. A.**, Oyeyemi, B. F., & Sarafadeen, H. (2026). Architectural Safety Mechanisms for Multi-Agent Clinical LLM Systems Under Knowledge Base Distribution Shift. *medRxiv*. doi.org/10.64898/2026.07.31.26359439

Book Chapters

1. Oyeyemi, B. F., **Sulaiman, M. A.**, Dauda, S. O., Oyewusi, H. A., Oladipo, O. O., & Adekilekun, H. A. (2026). From Gene Discovery to Clinical Applications: The Journey of Genomic Research. In *Genomics and Precision Healthcare: Innovations, Applications and Challenges* (pp. 39–66). Wiley. doi.org/10.1002/9781394367559.ch2
2. Oyeyemi, B. F., **Sulaiman, M. A.**, Dauda, S. O., Oyewusi, H. A., Oladipo, O. O., & Adekilekun, H. A. (2026). From Targeted Quantification to Untargeted Metabolomics: Applications and Prospects in Clinical Diagnosis. In *Computational Systems Biology for New Chemical Entities Bioprospection* (pp. 22–41). CRC Press. doi.org/10.1201/9781003645382-3

Dataset

1. **Sulaiman, M. A.**, & Oyeyemi, B. F. (2026). AfriPharmaGen Catalog: A Curated Pharmacogenomic Dataset for Sub-Saharan African Populations [Data set]. *Zenodo*. doi.org/10.5281/zenodo.21910084

Conference Presentations

1. Adekanmbi, O., **Sulaiman, M. A.**, Olomu, G., Ezule, P., & Ahmed, O. (2025). Scenario Planning for Next Best Action in Edo State (Abstract ID: DAC2025149). DHIS2 Annual Conference, University of Oslo, Norway. Physical presentation.
2. Oyeyemi, B. F., & **Sulaiman, M. A.** (2025). Elucidating the Epigenetic Landscape of Breast Cancer: Insights from Gene Expression Profiling of GDS662 Dataset (BSN 099). Biotechnology Society of Nigeria 37th Annual International Conference, Umudike. Physical presentation.

Research Experience

vartriage: Clinical Variant Interpretation Library 2024–present
Lead Developer

- Lead the core software engineering and bioinformatics development of vartriage, a Python pipeline that takes VCF files from gene panels and whole genomes and produces ranked candidate variants and ACMG/AMP-classified reports.
- Implemented the main processing workflow: VCF parsing, quality filtering, functional consequence annotation, population-frequency lookup, pathogenicity scoring, gene–disease linkage, variant prioritisation, and clinical report generation.
- Designed the pipeline to stream whole-genome files containing more than four million variants while using less than 2 GB of memory on commodity hardware.
- Added cohort analysis, trio inheritance analysis, structural-variant triage, mitochondrial-variant analysis, remote tabix scoring, and API-based annotation.
- Worked with Dr. Oyeyemi, a geneticist and biologist, who provides the project's genetics and biology guidance while I contribute the core software-engineering and bioinformatics implementation.
- Built automated tests, caching, typed interfaces, and command-line and Python APIs; released the package through PyPI.
- Benchmarked against GIAB and the ClinVar evidence repository, with validation results documented in the public project documentation.

Ovarian Cancer Prediction using Machine Learning Algorithms 2023–2024
M.Sc. Research

- Cleaned and standardised a clinical registry of 349 patients and explored risk factors including age, menopausal status and the CA125 biomarker.
- Selected 20 of 49 predictors with mRMR feature selection.
- Compared baseline models (logistic regression, SVM, KNN, decision trees) with ensemble methods (XGBoost, GBM, stacking and voting classifiers) using cross-validation across three train/test split ratios.
- Tracked experiments, metrics and parameters in MLflow.

Time Series Analysis of Neonatal Mortality in Nigeria (1990–2017) 2017–2018
Undergraduate Research

- Modelled neonatal mortality trends with ARIMA using national health data.
- Assessed the relationship between government health expenditure and neonatal survival and produced forecasts of future trends.
- Published in the Journal of Applied Sciences & Environmental Management (2019).

Professional Experience

All roles below were contracts.

MLOps / LLMOps Engineer 03/2026–present
Dou Associates (Remote)

- Build and run the AI service behind Slice24, a live product: a FastAPI microservice that generates and improves user content, with model routing and provider failover, retrieval over a vector store, content-safety checks, and an async job processor with idempotent retries.

AI/ML Engineer 07/2025–06/2026
Crea Cubed Inc., USA (Remote)

- Built multi-agent systems with LangChain, LangGraph and LlamaIndex, a Model Context Protocol (MCP) server for agent access to databases and tools, and retrieval-augmented generation over Pinecone and FAISS with output validation in Pydantic and Guardrails AI; deployed the services to Kubernetes through CI/CD.

MLOps Engineer / DPI Solutions Architect 05/2024–09/2025
Data Science Nigeria (Remote)

- Built an agentic assistant that recommends technology stacks through structured conversations, and automated model training, versioning and deployment on AWS with MLflow, Docker, Kubernetes and GitHub Actions.
- Built custom apps integrated into DHIS2, and led a team of four that delivered five scenario analysis decision-support use cases in Streamlit. Separately led a team of five data and cloud DevOps engineers.

Product Engineer 04/2023–11/2023
ParallelScore, USA (Remote)

- Built FHIR-based healthcare data pipelines on AWS for exchanging records between healthcare systems; a lineage setup with Apache Atlas, dbt and OpenMetadata cut compliance audit time from three days to ten hours.

Data/Cloud DevOps Engineer 10/2021–04/2024
Stata Insight Technologies Limited, Nigeria

- Built and deployed models for customer segmentation, fraud detection and predictive analytics, built REST APIs with FastAPI on AWS, and ran monitoring and cloud security (IAM, KMS, Vault).

Data Scientist 06/2020–09/2021
Prince_Analyst Concept, Nigeria

- Exploratory analysis, automated data-quality checks, data-analysis training for the team, customer segmentation with the marketing team, and A/B testing.

Software Engineer / Data Scientist 11/2019–05/2020
Rhics Technology Limited, Nigeria

- Built data-collection and reporting tools, Flask REST APIs for mobile and web apps, and a recommendation engine that matched investment options to users' risk profiles.

Technical Skills

Bioinformatics & Genomics: VCF analysis, GENCODE, gnomAD, ACMG guidelines, vartriage, BLAST, R, Bioconductor, QIIME 2, MaxQuant, Proteome Discoverer.

Machine Learning & Data Science: Scikit-learn, TensorFlow, XGBoost, CatBoost, MLflow, Statistical Modelling, NLP, PySpark.

Data Engineering & DevOps: Python, SQL, AWS, Docker, Kubernetes, GitHub Actions, CI/CD, Apache Airflow, Apache Kafka.

Software Engineering: REST APIs (FastAPI, Flask), System Design, System Integration.

Business Intelligence & Visualization: Power BI, Tableau, Plotly, Looker.

Certifications & Awards

- Data Science and Machine Learning with R | Decode Life, 2023
- AWS Machine Learning Foundations | Udacity, 2022
- Microsoft Certified: Azure AI Fundamentals | Microsoft, 2021
- Applied Data Science II: Machine Learning & Statistical Analysis (with honors) | WorldQuant University, 2020
- Certified: Microsoft Professional in Data Science | Microsoft, 2020

Professional Memberships & Service

- Global Burden of Disease (GBD) Collaborator
- Professional Statisticians Society of Nigeria (PSSN), membership no. PSSN/S24/00204
- AWS Community Builders (Machine Learning)
- Reviewer, Cancer Treatment and Research Communications (Elsevier)