

NEHA SIMON

 USA |  +703-965-2736 |  nehasimon67@gmail.com

BIOINFORMATICS AND DATA SCIENTIST

PROFESSIONAL PROFILE

Results-driven Bioinformatics professional with expertise in developing scalable machine learning solutions, automating data pipelines, and analyzing biological data. Skilled in deploying workflows, differential gene expression, and integrating AWS, Docker, and Kubernetes to enhance research efficiency. Experienced in RNA-seq analysis, document processing with LLMs, and fraud detection through predictive modeling. Proficient in Python, SQL, TensorFlow, and cloud data systems, with bioinformatics tools like Pathway Studio, Nextflow, and Gene Ontology. Known for providing insights into biological research and optimizing workflows for scientific and health projects. Strong in communication, teamwork, project management, and time management.

CORE COMPETENCIES

- Bioinformatics Analysis
- Predictive Modeling
- Pipeline Development
- Cloud Computing & Infrastructure
- Programming & Scripting
- System Integration Testing
- Database Management
- Neuroinformatics & Imaging Data Analysis
- Test Planning/Automation
- Defect Tracking
- Process Improvement
- Strategic Planning & Execution
- Project Management
- Clinical Outcome Modeling
- Comparative Genomics
- Research Management
- Digital Transformation
- Proposal Writing

TECHNICAL SKILLS

Next-Generation Sequencing (NGS) Data Analysis (RNA-seq, ChIP-seq) | Differential Gene Expression & Pathway Enrichment (GOseq, LIMMA, featureCounts) | Bioinformatics Pipeline Development (Snakemake, FASTA/FASTQ processing) | Functional Genomics & Gene Ontology (GO) Analysis | Biological Database Expertise (GenBank, UCSC Genome Browser, UniProt, RCSB PDB) | Sequence Alignment & Comparative Genomics (Multiple and Pairwise Alignment) | Biomarker Discovery & Disease Signature Identification | Programming: Python, MATLAB, R, Perl, Bash/Linux | Machine Learning: TensorFlow, Keras, scikit-learn, PyTorch | Image Processing & Analysis: Feature extraction, segmentation | Bioinformatics Tools & Technologies: Pathway Studio, Nextflow, Gene Ontology, GenBank | Database Management: SQL, NoSQL, AWS DynamoDB | Cloud & Containerization: AWS, Azure, Docker | Visualization: NumPy, Pandas, Matplotlib

RESEARCH PROJECTS

SARS-CoV-2 Differential Gene Expression Analysis 2024

- **RNA-seq Workflow Development:** Designed pipelines for RNA-seq data alignment, differential expression analysis, and polymorphism identification using featureCounts and GOseq.
- **Significant Pathway Insights:** Identified immune response pathways (e.g., CXCL5, SERPINA3) through overrepresented chemokine activity, aiding in understanding immune mechanisms.

Detecting Cardiac Alternans Using Machine Learning

GMU Bioinformatics Research | August 2022 – December 2022

- **ML Model Development:** Created machine learning models to classify cardiac alternans, employing Python and TensorFlow to analyze Ca²⁺ imaging data.
- **Workflow Optimization:** Built efficient data analysis workflows to extract patterns and link biological conditions to cardiac alternans.

Poster Presentation: "Interferon Response in Down Syndrome"

Big Data in Biomedicine Symposium, Georgetown University | October 2018

- Presented research using Pathway Studio to study gene response to heightened interferon response in Down syndrome.
- Analyzed immune pathways, contributing insights to the biological mechanisms behind interferon dysregulation.
- Effect of Doxorubicin Treatment: Performed an end-to-end differential expression analysis of the cancer drug Doxorubicin on different genetic backgrounds. RStudio, LIMMA, ChIP-seq
- Stroke Prediction Project: This project aims to develop and deploy a machine learning model for predicting stroke risk in patients. The project was built utilizing Rshiny programming to analyze medical data and build a robust prediction tool.
Website: <https://nehasimon.shinyapps.io/stroke-prediction>

PROFESSIONAL EXPERIENCE

AttainX

Bioinformatics: Data Scientist/Engineer

February 2023- present

- Build and deploy predictive models using Scikit-learn and XGBoost to uncover biomarkers and disease signatures, supporting biomedical research and decision-making.
- Optimize cloud-based workflows by leveraging AWS services, including S3 for scalable data storage and Lambda for efficient serverless preprocessing, significantly reducing pipeline runtimes.
- Create clear, insightful data visualizations using ggplot, Matplotlib, and Seaborn to effectively communicate research findings to cross-functional teams
- Drive a fraud detection initiative for the U.S. Department of Agriculture (USDA), applying anomaly detection techniques and developing high-accuracy classification models using Python and scikit-learn to prevent illegal imports of regulated goods.
- Spearhead AI/ML innovation efforts within the AttainX Innovation Lab, designing and deploying predictive models that accurately forecast hospital length of stay, improving healthcare planning and resource allocation.
- Implement a Robotic Process Automation (RPA) solution using UiPath to automate timesheet compliance for over 300 employees, incorporating complex date logic for varying team schedules; reduce administrative workload by 80% and enhance DCAA audit readiness.
- Maintain and enhance RPA bots for the USDA Farm Loans Program, ensuring reliable and efficient processing of loan approvals.
- Develop Python scripts to extract and transform structured data into .yaml format for ingestion into ETL pipelines; orchestrate real-time data streaming with Apache Kafka and monitor via a Kubernetes dashboard.
- Containerize machine learning functions using Docker to enhance portability and scalability across diverse computing environments.

Neurosurgery Department, INOVA

Data Science Intern-Bioinformatics

January 2023–October 2023

- Designed and implemented a machine learning algorithm to analyze hospital readmission trends among spinal and brain surgery patients.
- Applied predictive analytics to identify key clinical factors contributing to post-operative readmissions.

Department of Biology, George Mason University (GMU)

Graduate Teaching Assistant

January 2022–December 2022

- Supported instruction for the undergraduate course BIOL 102: A Survey of Biodiversity and Ecology.
- Developed lesson plans, facilitated laboratory experiments, and proctored assessments to enhance student learning outcomes.

Research Intern (Mentor: Dr. Saleet Jafri), GMU

Machine Learning in Cardiac Alternans Detection

August 2022–December 2022

- Led a bioinformatics project using machine learning to detect cardiac alternans through calcium spark analysis.
- Developed a robust data processing pipeline to automate the classification of cardiac Ca^{2+} events, improving detection efficiency.

Department of Bioengineering, GMU

Data Analyst Intern

January 2021 – July 2021

- Analyzed experimental data on synaptic plasticity related to corticosteroid exposure.
- Utilized Python libraries (NumPy, Matplotlib, Pandas) to perform statistical analyses and generate insightful visualizations.

EDUCATIONAL BACKGROUND

George Mason University

Master of Science in Bioinformatics and Computational Biology

Bachelor of Science in Neuroscience

Fairfax, VA 22030

December 2022

May 2021

REFERENCE AND OTHER RELEVANT DOCUMENTS AVAILABLE ON REQUEST