

TONY KABILAN OKEKE

Philadelphia, PA

✉ tonykabilanokeke@gmail.com  linkedin.com/in/t-k-o  Kabilan108

EXPERIENCE

Bioinformatics Data Scientist

GSK | Collegeville, PA

Apr 2022 – Sep 2022 | Apr 2023 – Sep 2023

- Developed data integration pipelines to enable joint analysis of internally generated spatial transcriptomics (10X Visium, GeoMx) data with public scRNA-seq datasets, providing novel biological insights.
- Developed machine learning models to predict carcinogenicity-initiating events for target molecules using rat transcriptomics data and chemical structures. Identified compound subgroups using UMAP embeddings.
- Built a web-based tool for performing statistical analysis and biomarker discovery on high-content imaging datasets using *Dash* and *Flask*, with *Dask* to scale computations. Deployed on *R Studio Connect*.
- Trained and deployed machine learning models (Decision Trees, Random Forests, SVMs) to identify imaging biomarkers in safety toxicology studies.
- Implemented deep learning models (Noise2Void, Cellpose) to improve the performance of *CellProfiler* segmentation and feature extraction pipelines.

Computational Research Assistant

Penn Medicine | Philadelphia, PA

Mar 2021 – Aug 2022

- Developed Python and R scripts to automate data ingestion, quality control and analysis for ELISA and Next-Generation Sequencing (NGS) assays, resulting in a 4X speed up of existing workflows. Conducted differential methylation and pathway enrichment analyses on NGS datasets.

PROJECTS

AbokiCode | Transformers, FastAPI, Telegram Bot API | Hackathon Winner

Sep 2023 – Present

- Developed a Learning Management System to facilitate easy course delivery to students in low-income areas. Implemented a web app using Next.js to enable course instructors customize the interface for students.
- Built LLM-based study assistant to aid students with questions related to course material. Utilized LlamaIndex to build retrieval augmented generation pipeline using ColBERT embeddings stored in MongoDB. Served assistant via FastAPI.

Meddibia | Tensorflow, Flask, Flutter | Hackathon Winner

Mar 2023

- Developed a medical diagnostic tool with a Flask API serving a Flutter-based IOS app.
- Fine-tuned neural networks (VGG16, EfficientNet) for skin lesion detection, achieving 70% accuracy. Developed API endpoints that leverage GPT-3 to extract symptoms from user messages. Trained a Naive Bayes classifier to predict diseases given a list of symptoms with 80% accuracy.

ML for Predicting Gene Ontology Enrichment | Tensorflow, Flask, Flutter

Sep 2022

- Developed an end-to-end pipeline for processing and analyzing over 11,000 RNA-seq datasets, leveraging high-memory GCP VMs to process datasets in batches.
- Trained and optimized an autoencoder neural network to reconstruct log fold-change values and predict Gene Ontology (GO) enrichment using latent space features.

Identification of Connectomic Biomarkers for Autism using ML | bctpy, scikit-learn

Aug 2022

- Computed graph theory measures on structural connectomes from ASD patients and employed machine learning models (LASSO, SVM) for feature selection and diagnostic accuracy assessment.
- Conducted a comprehensive study correlating connectomic biomarkers with clinical outcomes, utilizing data visualization tools to represent intricate relationships within the multimodal dataset.

EDUCATION

Drexel University

Philadelphia, Pennsylvania

Master of Science in Biomedical Engineering

Graduation: June 2024

Bachelor of Science in Biomedical Engineering

GPA: 4.00

Concentrations: Biomedical informatics & Neuroengineering

TECHNICAL SKILLS

Programming: Python, R, C++, SQL, Js/Ts, Shell, MATLAB

Web: Flask, FastAPI, React.js, Next.js

Bioinformatics: Seurat, ScanPy, SquidPy, CellProfiler

AI/ML: PyTorch, TensorFlow, sklearn, WandB, LangChain

CLUBS & ORGANIZATIONS

- Vice President, *Drexel Computational Design*, March 2021 – Present
- Member, *Sigma Xi*, January 2024 – Present
- Member, *Tau Beta Pi*, December 2021 – Present