

As a data scientist at UCSF with a focus on deep learning, I have experience in building out biology-based deep learning projects. My work includes digitizing medical records in Rwanda and developing biology-based foundational models. Deeply passionate about machine learning, I am eager to contribute, grow, and make a positive impact.

SKILLS

Tools and Languages	Statistics, Probability, Bayesian Statistics, Python, Git, Numpy, Pandas, AWS (Sagemaker, EC2, S3, ECS, Lambda), Scikit-learn, Pytorch, Keras, Tensorflow, Jax, Docker
Supervised Learning	Regression, Decision Trees, Support Vector Machines, Boosting
Unsupervised Learning	Clustering, Principal Component Analysis
Deep Learning	Feed Forward Network, CNNs, RNNs, Transformer-based Models, VAE's, Diffusion Models

TECHNICAL EXPERIENCE

Data Scientist: Assistant Specialist Step 3 **Oct 2022 — Present**
UCSF *San Francisco, CA*

- Led the end-to-end development of pretraining a transformer-based deep learning model to learn the relationship between DNA sequence and chromatin confirmation (ATAC-seq) utilizing PyTorch and AWS SageMaker.
- Achieved nearly 90% accuracy by finetuning my model in identifying transcription factors' binding sites on DNA, crucial for understanding gene regulation.
- Enhanced the understanding of gene regulation and its impact on diseases through precise identification of transcription factor binding sites

Data Science Intern **Jun 2021 — Aug 2021**
University of Virginia *Remote (originally Rwanda)*

- Created an advanced technique to extract key details such as drug names from Rwandan hospital paper medical records, using YOLO for segmenting text and LSTMs to analyze the sequence of drugs.
- Offering a way to enhance medical documentation and patient care.

Devops Intern **May 2020 — Aug 2020**
ResMed (Propeller Health) *Madison, WI*

- I improved our company's systems by managing various infrastructure projects on Amazon Web Services (AWS), including Lambda, EC2, S3, IAM, Cloudwatch, Cloudtrail, and ECS, enhancing my expertise in AWS products.
- Created an automated alert system that notifies us via Slack of AWS service failures, enhancing real-time system monitoring and uptime

EDUCATION

Master of Science in Data Science, *University of Virginia* 2022
Masters of Science in Systems Engineering, *University of Virginia* 2021

PROJECTS AND PUBLICATIONS

UTX condensation underlies its tumour-suppressive activity (paper)

Nature paper that discovered how a protein called UTX fights cancer by forming special clusters inside cells, a process unrelated to its known chemical activities. This finding opens new avenues for understanding how our cells prevent cancer. Unveiled that UTX's ability to cluster together plays a crucial role in controlling gene activity, essential for keeping cells healthy and preventing cancer, providing a fresh perspective on disease prevention and treatment strategies.

UVa Building Classifier (github)

Classified buildings at UVa using transfer learning.

Comparing different ML Methods for Song Classification (github)

This project focuses on classifying songs into genres using only their lyrics, evaluating the effectiveness of both traditional machine learning algorithms and neural networks in accomplishing this task.

Portfolio Website

My portfolio is a website written in Sveltekit using Tailwind for styling. It is hosted on Vercel and is a showcase of my ability to learn and use new technologies on my own.