

Abhilash Dhal

A-27, Dhanashtra RBI Quarters, Mumbai
Phone number: +91-826-053-9435
adhalthbiophysics@gmail.com

[@github](#)
[@linkedin](#)
[@abhilashdhal.com](#)

EDUCATION

University of California, Davis, CA

Master of Science, Biophysics,

Dec' 2019

(Deep learning, Population and Quant. Genetics, Comp. Drug Design, Algorithm Design)

- **Thesis Project:** Developed, applied and evaluated **bayesian regression models** for GWAS and Genomic prediction. [\(SSBR-JWAS\)](#)

Indian Institute of Technology, Varanasi, India

Master of Technology, Biochemical Engineering,

Aug' 2016

- Junior research fellow(JRF)(**top 0.01%**) in the Graduate aptitude test examination(GATE)

Indian Institute of Technology, Varanasi, India

Bachelor of Technology, Biochemical Engineering,

Aug' 2015

- secured (**top 0.1%**) of 450,000 students in the Joint Entrance Exam(JEE)

PUBLICATIONS AND POSTERS

- Profiling antibody epitopes induced by mRNA-1273 vaccines and boosters [Frontiers \(2024\)](#)
- Distinguishing features of long COVID identified through immune profiling [Nature \(2023\)](#)
- Utilizing the autoantibody immune response to tumor antigens for kidney cancer early detection [JCO\(2022\)](#)
- High-resolution epitope mapping and characterization of SARS-CoV-2 antibodies in subjects with COVID-19 [Commun Biol \(2021\)](#)
- Phase Ib study of patients with metastatic castrate-resistant prostate cancer treated with different sequencing regimens of atezolizumab and sipuleucel [JITC \(2021\)](#)
- Genome-Wide Association Studies Combining Genotyped and Non-Genotyped Relatives Using Bayesian Regression Methods with Mixture Priors [\(PAG XXVIII, 2020\)](#)
- Investigation of azo-dye degradation by bacteria [\(ICOBM, 2016\)](#)

PROFESSIONAL REFERENCES

Dr.Minlu Zhang, Serimmune

Associate Director

Minlu.Zhang@serimmune.com

Phone: +1-513-410-5458

Dr.Thomas Campbell, Mindset Medical

Senior Data Scientist

thomascampbell1@gmail.com

Phone: +1-720-933-6212

Dr.Hao Cheng, UC Davis

Associate Professor

reworkhow@gmail.com

Phone: +1-240-848-8528

Brian Martinez, Cedarsinai

Senior Data Analyst

martinez.brian93@yahoo.com

Phone: +1-661-491-3059

RESEARCH EXPERIENCE

Developer/Contributor

May 2024 - Present

[Google Summer of Code, cBioPortal](#)

Mumbai, India

- Building an automated clinical metadata harmonization tool using advanced NLP methods.
- Designed several experiments for evaluating different sentence transformers and language models for ontology mapping of clinical bioinformatics data.
- Worked on several strategies of fine-tuning sentence transformers improving generated embeddings for textual attributes.
- Evaluated interpretability of model parameters for textual attribute harmonization through statistical approaches.
- Built infrastructure for quality assessment, monitoring and integration into cBioPortal.

- **Biomedical projects:** Drove biomedical projects for healthcare business partnerships (Moderna, Genentech, Verily, Cedar-Sinai, LabCorp, Yale and Stanford). Some in-depth statistical analysis and tools developed by me were:
 - **Proteome Wide Identification of Epitopes(PIE):**
 - * Applied and deployed an outlier sum statistic based approach using asymptotic theory for estimating outlier sum Z-Scores and p-value for feature selection across high dimensional bacterial peptide sequence data.
 - * Leveraged PIE tool for investigating peptide sequence signatures for viral strain stratification, mRNA Vaccine design evaluation, Differentiation in antibody signatures of Long and Convalescent COVID and evaluation of antibody responses in clear cell Renal Cell Carcinoma(ccRCC), non-small cell lung(NSCLC), Lyme disease and autoimmune diseases (SLE, Sjogren's and RA).
 - **Drug response studies:**
 - * Evaluated combination of atezolizumab with sipuleucel-T monoclonal antibody treatment for Prostate cancer using statistical approaches.
 - **Cancer Studies:**
 - * **Early detection of clear cell Renal cell carcinoma:**
 - Using peptide sequence aggregated signal demonstrated significant differentiation between early stage and late stage of ccRCC using PIE.
 - Collaborated with Christian Hoerner's group at Stanford University to present a poster for the findings ([The Early Detection of Cancer Conference](#))
 - * **Prognosis of NSCLC:**
 - Identified cross-reactive atezolizumab treatment specific peptide signatures to correct for confounding effects across actual prognostic antibodies.
 - Presented potential non-atezolizumab peptide signatures concordant with previous literature.
- **Statistical analysis of mRNA vaccine designs:** Drove **\$500K in revenue** for characterization of best mRNA vaccine and booster designs through statistical approaches for feature discovery and paired analysis. ([Frontiers in Immunology\(2024\)](#))
- **Supervised ML for disease diagnostics:** Deployed a feature selection tool with LASSO classifier achieving sensitivity of 87%, 99.5% specificity for SARS-CoV-2 IgG diagnostic panel. (Patent link: [PCT/US2021/038960](#))
- **Semi-supervised ML for data mining:** Deployed a GMM and PCA based semi-iterative method for mining control samples for 300+ diseases, 50000+ samples.
- **Graph ML for signal enhancement:** Deployed a graph based clustering tool for signal enhancement in Kmer sequencing data, improving SNR by 75% (Highest for any model in production).
- **ML based QC tools:** Deployed batch correction and clustering approaches for normalization of technical variation, correction of signal signal in abnormal samples, enhancing accuracy of cancer subtype and autoimmune disease classification.
- **Serosurveillance, Immune monitoring and forecasting:** Led research and development efforts cross-functionally for monitoring and forecasting immune signal changes over time in response to drug therapy and boosters.
- **Software Dev/Automation:** Spearheaded the development of 10+ Python-based pipelines and ETL processes for optimizing data handling, analysis and visualization of sequencing data

MENTORING

Serimmune

[Serimmune](#)

- Closely mentored one data analyst. Most of my mentoring was focussed on robust statistical method development, machine learning for the scientist, and software best practices.

UC Davis

- Mentored 50+ students in end to end development of GWAS software for evaluation of genetic merit in animal science (2019)
- Mentored 100+ students in a computational drug design course. (2018)

PERSONAL PROJECTS

Biomedical Research

[Serimmune](#)

- Worked on Infectious, cancer and autoimmune diseases to identify and validate disease specific signatures.
- Patented a novel feature selection model to improve sensitivity of SARS-CoV-2 diagnostic panel ([PCT/US2021/038960](#))

Open source ML/DL

[\(OmixHub, 2023-Present\)](#)

- **Overview:** Streamlining commonly used machine learning and bioinformatics algorithms for different omics datasets
- **Python client data search, accession and retrieval(SAR):** Developed a module to facilitate python based SAR to enable ease of downstream ML and AI applications for disease detection.
- **Applications:**
 - **Bioinformatics and ML methods on cancer datasets:** Dimensionality reduction, optimized density based clustering, multi-label supervised classification and multi-factor differential gene expression on [Genomic Data Commons](#) datasets.
 - **Multi-label classification of 3 kidney cancer subtypes and 2 tumor grades:** ([Project Link, 2024](#))
 - * Applied set sampling, minority sampling and data augmentation to resolve patient overlap for unbiased training, testing and validation split.
 - * Evaluated Adaboost, Random Forest, KNN, Decision trees, Naive bayes, SVM, Logistic regression and ensemble models to evaluate macro and micro ROC's on validation data.
 - * Best performing ensemble models achieved atleast 90% sensitivity at 100% specificity across each of the target variables in test data.
 - **Cohort stratification using Dimensionality reduction and clustering:** Using UMAP with hyper-opt tuned clustering showcased better cohort stratification of RNA kidney cancer expression data by tumor stage over subtype.

Deep-learning

- Multi-label classification of chest X-rays using DenseNet ([Project Notebook, 2024](#))
- Intracranial hemorrhage detection using convolutional neural networks ([Final Report](#))
- Real time strategy agents for Starcraft-II using deep reinforcement learning ([Final Report](#)) ([SC2-GCP-CNN](#))

TECHNICAL STRENGTHS

Programming Languages: Python, Julia, Java, Bash, R, MySQL
Databases

- **Cloud Technologies:** BigQuery, BigTable, Data Procs, SQL, AWS Redshift, HPC, Google Colab
- **Technical documentation:** Git, Bitbucket, Confluence, Jira, Visual Studio Code

Methods: Genome wide association studies(GWAS), Outlier sum statistics, hypothesis testing, Deep learning (Image classification), Natural language processing (NLP), Clustering, Statistical modeling (bayesian and frequentist)

Domain: Omics, Diagnostics, Images, Texts/Documents

Certifications: [AI using Tensorflow](#), [Data Science](#), [Web App Development \(Streamlit\)](#)