

# Mehak Gurnani

+44 7464197660 | [mehak.gurnani2000@gmail.com](mailto:mehak.gurnani2000@gmail.com) | [LinkedIn](#) | [GitHub](#) | [Personal website](#)

## TECHNICAL SKILLS

---

**Programming Languages:** Python, R, SQL (incl. MySQL)

**Libraries & Frameworks:** (Python) PyData stack, HuggingFace, PyTorch, TensorFlow, (R) tidyverse, caret, Shiny

**ML & Statistical Methods:** Classical ML (regression, classification, ensemble models, clustering, time-series analysis, NLP, deep learning, hypothesis testing)

**Tools & Platforms:** Git, Jupyter, RMarkdown, Bash, Tableau, Microsoft Office, Agile/Scrum

## RELEVANT WORK EXPERIENCE

---

### Imperial College London

**December 2024 - Present**

*Honorary Research Associate, Cardiac Electrophysiology and AI Research Group*

*London, UK*

- Granted the title in recognition of continued contributions to AI-ECG and cardiology research beyond the formal role
- Leading first-author manuscripts and contributing to co-authored publications within the group

### Owlstone Medical

**December 2024 - February 2025**

*Data Scientist*

*Cambridge, UK*

- Built machine learning and signal processing models in Python to classify disease from multi-sensor breath data, focusing on modularity and scalability
- Engineered features and developed regression analyses to quantify system variability, informing experimental design and lab workflows
- Communicated technical findings using Python-based visualizations to support project and business decisions
- **Tech Stack:** Python (NumPy, pandas, scikit-learn, Plotly), Git, CI/CD, object-oriented design, unit testing

### Imperial College London

**Oct 2023 – Nov 2024**

*Data Science Research Assistant, Cardiac Electrophysiology and AI Research Group*

*London, UK*

- Developed and validated end-to-end ML and statistical models for AI-ECG analysis using multimodal clinical data from transnational cohorts (1M+ records)
- Led projects on unsupervised subgroup discovery (tree-based models), CNN-based mortality prediction, and time-series DL for ICD therapy optimization
- Collaborated with clinicians and computational scientists to ensure clinical relevance and technical robustness
- Presented at top cardiology AI conferences; work led to 2 first-author and 3+ co-authored publications
- **Tech Stack:** Python (PyData stack, TensorFlow), R (tidyverse, survival, caret); Models: statistical, unsupervised/clustering (kNN, DDTTree, VAE), supervised (ensemble-based), deep learning (CNNs, LSTMs, transformers); Git, HPC, GPU computing

### elmeas

**March 2023 - June 2023**

*Data Analyst*

*London, UK*

- Built clinical data models and derived insights from health data designed for integration with openEHR systems
- Wrote advanced SQL queries to support backend development of a medical compound database and address user data needs
- Leveraged Python and RESTful APIs to interface with PostgreSQL databases for clinical data aggregation and retrieval
- **Tools and Standards:** Python, SQL (PostgreSQL), REST APIs, openEHR, ICD-9/10, SNOMED, ATC

## EDUCATION

---

### Imperial College London

*MSc Health Data Analytics and Machine Learning, Distinction*

**2022 – 2023**

- **Key modules:** Machine Learning, Statistics, Advanced Analytics, Clinical Data Management, Computational Epidemiology, Translational Data Science
- Leveraged real-world datasets (UK Biobank, OMICS data) using **R, Python and SQL**
- Proficient with predictive and network modeling, Machine Learning methodologies and Bayesian methods
- **Thesis:** Extending the Applications of AI-ECG Models to Uncover and Predict Clinically Relevant Phenogroups

*BSc Medical Biosciences, First Class Honours*

**2019 – 2022**

- Achieved Distinction in all written scientific pieces based on independent research projects
- Achieved **80.5%** in the Biomedical Data Science using Machine Learning models for biomarker identification

## PROJECTS

---

**Sequential Sentence Classification from RCT Abstracts** – using Word2Vec embedding with a bidirectional LSTM

**Transfer Learning on Colorectal Cancer Images** – combined fine-tuned pre-trained ResNet50 with a Stacked Classifier

**Causal Modeling for CVD Risk Prediction** – used Cox regression, graphical models and structured causal modeling

**Geospatial Modeling of Crimes Against Women** – implemented Spatial-temporal Bayesian Modeling on INLA

**Breast Cancer Survival Analysis** – using clinical and genetic predictors for survival analysis modelling using R

**Facebook Data Challenge** – analyzed consumer trends and demographics for market segmentation insights on Python