EPSRC CENTRE FOR DOCTORAL TRAINING IN DIGITAL HEALTH TECHNOLOGIES (TECH4HEALTH)

Shaun Yap

University College London (UCL)

Background

My background is in **data science** and **statistics**, with a strong focus in **statistical modelling, Bayesian inference, feature engineering** and **deep learning.**

BSc (Hons) Data Science | *2020–2023*

MSc Data Science & Statistics | *2023-2024*

Statistics Tutor (Generalised Linear Models & Bayesian Forecasting) | *2025*

Data Scientist (R&D) AI fraud analysis | *2025*

PhD in Digital Health Technologies *2025–Present*

Previous Research Work

M-Estimation: Estimation and inference from statistical models in Julia

Developed statistical computing tools for robust inference in Julia, focusing on M-estimators, which generalise maximum likelihood and with appropriate robust loss functions (e.g., Huber), reduce sensitivity to outliers and modest misspecification. A core part was leveraging dual numbers – numeric types that extend real numbers with an infinitesimal ϵ where $\epsilon^2 = 0$ – to enable forward-mode automatic differentiation, which lets us get derivatives accurate up to floating-point precision (forward-mode AD) and Jacobians of large objective functions efficiently instead of using finite-difference numerical derivatives, improving optimisation speed and stability. I extended an M-estimation package in Julia with profile likelihood: fixing a parameter of interest, re-optimising nuisance parameters, and tracing that profile to get inference and identifiability diagnostics.

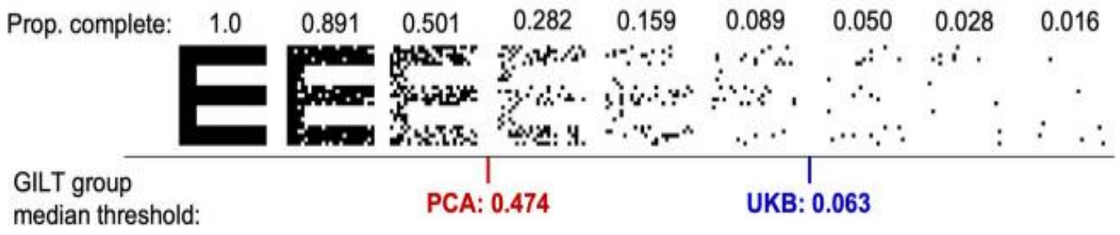
Deep Learning Approaches to Misinformation Detection

Built and evaluated end-to-end NLP systems for social-media misinformation: data collection, feature engineering, and text vectorisation. Compared traditional ML (logistic regression, SVMs, random forests, gradient boosting), deep learning (CNNs, LSTMs), and LLM-based transformers. Assessed both ordinal (graded truthfulness) and binary (true/false) labels to test how models capture factual nuance. Findings: retaining ordinal labels for training improved accuracy by ~10%, and strong feature engineering added a further ~8% gain. LLM-driven augmentation (richer sentiment/subjectivity signals) further lifted a binary baseline from ~80% to ~85%, indicating better nuance capture. The trade-off was increased compute and latency, challenging real-time use without optimisation.

PhD Research - Visual system vulnerability in dementia: from detection to determinants

Supervised by Dr Keir Yong (UCL Dementia Research Centre) & Prof Andre Altmann (UCL Hawkes Institute)

Cortical visual impairments ("brainsight", not eyesight) are disabling yet under-recognised consequences of dementia. They are common in Alzheimer's disease (AD) and especially posterior cortical atrophy (PCA) (visual-led dementia), where visual symptoms can precede memory, language and insight loss. Many people first present to eye care and are misdiagnosed with ocular or psychological, leading to unnecessary interventions and years-long delays. Clinic-ready tests of cortical visual function are scarce, and there is limited understanding about why some individuals are more susceptible.



Projects

1. Cortical visual test (GILT): Validate the Graded Incomplete Letter Test across UK Biobank repeat-imaging and clinic cohorts; relate scores to diagnosis/MRI; optimise sensitivity to visual-cortex damage.
2. Latent factor analysis: Combine visual tests + MRI to model a cortical visual factor; test associations with AD risk and functional measures.
3. Discovery genetics: Perform GWAS on the latent factor; compare SNP effects with visual-led AD/PCA.

My Broader Research Interests

- Cortical visual processing and neurodegeneration
- Bayesian and statistical modelling
- Genomics and multi-omics
- Multimodal and representation learning
- Interpretability and Explainable AI (XAI)
- Biomedical imaging and computational biology
- Methodological applications to antimicrobial resistance (AMR)

Contact QR Code

LinkedIn shaunyap0122

Engineering and Physical Sciences Research Council