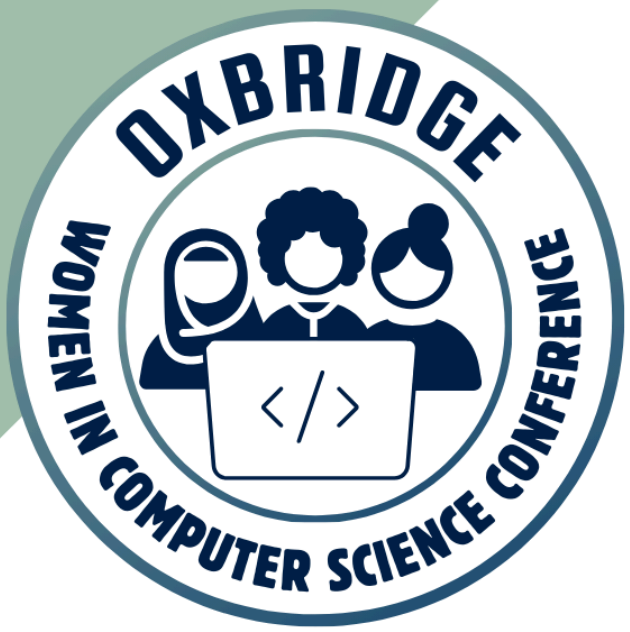
Body



BOOK OF ABSTRACTS 2025

3RD MAY 2025

UNIVERSITY OF CAMBRIDGE

UNIVERSITY OF OXFORD

List of Abstracts

Extracting prognostic markers from radiology reports in chronic liver disease: benchmarking Meta-Llama 3.1 **5**
Hania Paverd, University of Cambridge 5
A Graph-Based Approach to Digital Stressed Brain Modelling **7**
Sonia Koszut, University of Cambridge 7
Synthetic Retinal Imaging with Topology-Preserving VQ-GAN **8**
Zuzanna Skórniewska, University of Oxford 8
Predicting molecular and cellular features from gut histopathological images using deep learning **10**
Kexin Xu, University of Oxford 10
Algorithmic Fairness Beyond Predictions **12**
Silvia Casacuberta, University of Oxford 12
Digital Phenotyping of Emotional Expression in Schizophrenia **13**
Shrankhla Pandey, University of Cambridge 13
AI-Powered Histopathology: Deep Learning for Risk Stratification in Esophageal Cancer in a Large-Scale Clinical Trial **15**
Greta Markert, University of Cambridge 15
Can AI Design the Shape of the Internet? **17**
Akanksha Ahuja, University of Cambridge 17
Computational Detection of 5-Fluorouracil Metabolites in Nanopore-Sequenced RNA: Unraveling Chemotherapy-Induced Transcriptome Damage **19**
Shutong Ye, University of Cambridge 19
Complex-Weighted Diffusion: A New Perspective on Heterophily and Over Smoothing in GNNs **21**
Cristina Lopez Amado, University of Oxford 21
A Non-Intrusive Monitoring Framework for ML Workload: Enabling Bottleneck Optimization and AI Governance across Various Computing Infrastructures **23**
Ziji Chen, University of Oxford 23
Ask Patients with Patience: Enabling LLMs for Human-Centric Medical Dialogue with Grounded Reasoning **25**
Jiayuan Zhu, University of Oxford 25
The Effect of Representational Compression on Flexibility Across Learning **26**
Mia Whitefield, University of Oxford 26
PEARL: Preference Extraction and Reward Learning **27**
Marta Grzeskiewicz, University of Cambridge 27
Colors in Costumes: Computational Analyses of Color in Costumes in 12K Portrait Paintings from 1400 to Today **28**
Christine Li, University of Oxford 28
Contextual IoT Insights through Edge Computing and Image Captioning **30**
Deema Abdal Hafeth, School of Engineering and Physical Sciences, University of Lincoln 30
Automotive SPICE and AI Integration to support software process improvement **32**

Menna Noureldin 32
Subgroups Matter for Robust Bias Mitigation 33
Anissa Alloula, University of Oxford 33
“It’s All Greek to Me”: Benchmarking Semantic Representations Techniques for Idioms with Multimodal Contexts 35
Kirthi Kumar, University of Oxford 35
Accessing Applications Remotely in the Mobility Era: The Role of Zero Trust Network Access 36
Andrea Ibiassi, University of Oxford 36
The impact of England’s CAZ, ULEZ, and COVID-19 lockdowns on cardiovascular and respiratory chronic diseases during the COVID-19 pandemic 37
Millie Zhou, University of Cambridge 37
Full Reference Image Quality Assessment for Medical Images: Pitfalls, Analyses and Generalization 38
Anna Breger, University of Cambridge 38
Enhanced Synthetic MRI Generation from CT Scans Using CycleGAN with Feature Extraction 40
Lachin Naghashyar, Department of Computer Science Sharif University of Technology, Tehran, Iran 40
Solving Large Deterministic POMDPs with Finite State Controllers 41
Alex Schutz, University of Oxford 41
Tau Cell Detection using Object Detection (OD) Models 42
Peiyan (Gracie) Zhou, University of Cambridge 42
Data Integrity Detection in Wearable IoT Devices for Cardiac Monitoring 44
Alexandra Loh, University of Cambridge 44
Sheaf-Based Diffusion for Structured Modality Fusion in Multimodal Graph Learning 45
Mar Gonzàlez i Català, University of Cambridge 45
SPA: Efficient User-Preference Alignment against Uncertainty in Medical Image Segmentation 46
Jiayuan Zhu, University of Oxford 46
Aegis: A Programming Language that enforces Information Flow Security with Types 47
Komal Rathi, University of Cambridge 47
Evaluating Gender Fairness of ML Algorithms for Pain Detection 48
Yuting Shang, University of Cambridge 48
Emission Impossible: 49
Verifiable Carbon Emissions Reporting Without Revealing Private Data 49
Jessica Man, University of Cambridge 49
Instruction Fusion Limit Study for RISC-V 50
Elizabeth Ho, University of Cambridge 50
Profiling neural grammar induction on morphemically tokenised child-directed speech 52
Mila Marcheva, University of Cambridge 52

New Tolerance Factor for Molecular Perovskites **54**
Lauren Wilkes, University of Cambridge 54
Enhancing a Quantum Error Correction Code Simulator **56**
Alison Dauris, University of Cambridge 56
Non categorical models of typed lambda calculi **57**
Jessica Richards, University of Oxford 57
GeoFT: Fine-tuning Foundation Models for Automated OSINT Geolocation **58**
Selena Sun, University of Oxford 58
Harnessing Machine Learning for Heart Failure Management: Opportunities and Limitations **59**
Kristy Poon, University of Cambridge 59
Machine Learning for Building-Level Heat Risk Mapping **60**
Andrea Domiter, University of Cambridge 60
Material discovery - High-temperature superconductor at ambient pressure **61**
Maelie Causse, University of Cambridge 61
The Invisible Handshake: State Pensions and Corporate Political Contributions **62**
Jingshu Wen, University of Oxford 62
Instrumental variable simulation **63**
Ruizi Yan, University of Oxford 63
Non-Automatability: Formal Barriers To Efficient Proof Search **64**
Gaia Carenini, University of Cambridge 64

Extracting prognostic markers from radiology reports in chronic liver disease: benchmarking Meta-Llama 3.1

Hania Paverd, University of Cambridge

Background

Liver cancer almost exclusively develops in patients with chronic liver disease (CLD), but despite this clear at-risk population, most are diagnosed at an advanced stage with poor prognosis. Large language models (LLMs) present an opportunity to extract discrete variables from free-text clinical data in order to identify early prognostic markers of disease outcomes. LLMs have demonstrated efficacy in extracting structured data on tumours, but benchmarking for chronic pre-cancerous diseases like CLD remains unexplored.

Methods

Using an open-source LLM, Meta-Llama 3.1 (8B), we extracted structured data from 62 radiology reports from 5 patients with CLD, including 30 CT/MRI reports of liver imaging. Each report was queried independently. We employed zero-shot prompting, with prompt including the report content, the question and possible answers (Table 1). The ground truth was manually curated by a clinical radiology resident.

Results

The model achieved 100% accuracy in classifying imaging modality and identifying liver imaging studies across all imaging reports. Accuracy of 93% was noted for identifying cirrhosis and 96% for diagnosing complications of CLD such as portal hypertension in liver imaging reports. Performance declined for distinguishing prior cirrhosis diagnoses (63%) and for multi-class questions, such as determining whether other CLD complications, including ascites (66%) and collaterals (61%), were present, absent, or not reported.

Conclusions

Our approach demonstrates high accuracy in extracting specific key clinical variables for CLD using LLMs. It also highlights challenges in separating prior knowledge from current findings, and in distinguishing between absent and unreported features, aligning with the LLM's known tendency towards overconfidence and pattern completion bias.

Question	Possible answers	Assessment type	Accuracy	Precision	Sensitivity	Specificity	F1 score	Rate correct
Does this report describe liver findings?	yes, no	2-class	**1.00**	1.00	1.00	1.00	1.00	
What kind of imaging is described in this report?	CT, MRI, US, fluoroscopy, angio, transjugular biopsy	correct / incorrect						**1.00**
Is cirrhosis explicitly mentioned in this report?	yes, no	2-class	**0.93**	1.00	0.87	1.00	0.93	
Is there current diagnosis of cirrhosis?	yes, no	2-class	**0.84**	0.74	1.00	0.69	0.85	
Was there a prior diagnosis of cirrhosis?	yes, no	2-class	**0.67**	0.48	0.91	0.42	0.63	
Is portal hypertension explicitly diagnosed in this report?	yes, no	2-class	**0.96**	0.75	1.00	0.92	0.86	
Are there shunt vessels, collaterals, or varices present?	yes, no, not reported	3-class	**0.61**	0.49	0.67	0.56		
Is there ascites present?	yes, no, not reported	3-class	**0.66**	0.75	0.70	0.68		
What is the size of spleen?	<size in cm>, not reported	correct / incorrect						**0.90**

Table 1. Results of LLM data extraction. Accuracy = balanced accuracy. For 3-class questions, precision, sensitivity, specificity and F1 score are weighted averages.

A Graph-Based Approach to Digital Stressed Brain Modelling

Sonia Koszut, University of Cambridge

Stress, clinically defined as a state of imbalance when demands exceed an individual's adaptive capacity (McEwan, 1998), is increasingly recognised as a major causal factor in both mental and physical disorders. Traditional diagnostic methods rely on self-reported questionnaires, which can be subjective and may not capture the details required for an accurate assessment. Advances in non-invasive neuroimaging, particularly electroencephalography (EEG), offer a promising means to objectively evaluate stress-induced changes in brain function. Conventional machine learning methods have been successful in distinguishing between stressed and non-stressed states from EEG data (Roy et al., 2023; Malviya & Mal, 2022). However, the potential of geometric machine learning (GML)— capable of processing non-Euclidean data—remains largely unexplored. GML is particularly suited to capturing the complex spatial and temporal patterns in EEG signals (Wagh & Varatharajah, 2020; Jang et al., 2023).

This study proposes a framework that employs geometric machine learning (GML) to construct a digital model of the stressed brain. EEG recordings, acquired via the 10-20 system, are transformed into fully connected graphs where nodes represent individual electrodes and edges connect every pair of nodes (excluding self-loops), serving as input for a Graph Attention Network (GAT). Using a state-of-the-art ensemble classifier's feature selection strategy (Shikha et al., 2023), each trial's features are organised into a 32×13 matrix, with each of the 32 nodes described by 13 features—including statistical measures, Hjorth parameters, and frequency domain metrics—that capture the brain activity under stress.

Preliminary tests on SAM-40 dataset (Ghosh et al., 2022) show that while the state-of-the-art ensemble method (Shikha et al., 2023) achieves 71.53% accuracy on raw EEG data, the initial GAT model reaches 75.69% accuracy after 100 training epochs. Future work will extend this framework to spatio-temporal graph neural networks and additional datasets to advance generalisable mental health diagnostics and real-time stress quantification.

Synthetic Retinal Imaging with Topology-Preserving VQ-GAN

Zuzanna Skórniewska, University of Oxford

The growth of neural network models, driven by advancements in parallel computing, has been constrained in medical imaging by strict privacy regulations, limited data availability, high acquisition costs, and demographic biases. Deep generative models offer a promising solution by generating synthetic data that bypasses privacy concerns and addresses fairness by producing samples for under-represented groups. However, unlike natural images, medical imaging demands validation not only for fidelity (e.g., Frechet Inception Score) but also for morphological and clinical accuracy. This is particularly true for colour fundus retinal imaging, which requires precise replication of the retinal vascular network, including vessel topology, continuity, and thickness.

In this study, we propose a VQ-GAN model trained with a retinal-topology preserving function based on the latent space of RetFound [1] to synthesize high-fidelity colour fundus images that are both morphologically and clinically reliable. Our preliminary results suggest that the generated images can be validated using AutoMorph [2], an automated retinal biomarker extraction tool, to ensure that biomarkers derived from synthetic images follow the same distribution as those from real images. This validation is intended to involve a two-sided t-test to assess the statistical similarity. By leveraging the quantized latent space, the model is designed to generate higher resolution images and class-conditioned samples, facilitating the synthesis of data for rare clinical conditions and under-represented demographic groups. This approach aims to enhance data availability while addressing fairness in medical imaging research and downstream diagnostic tasks.

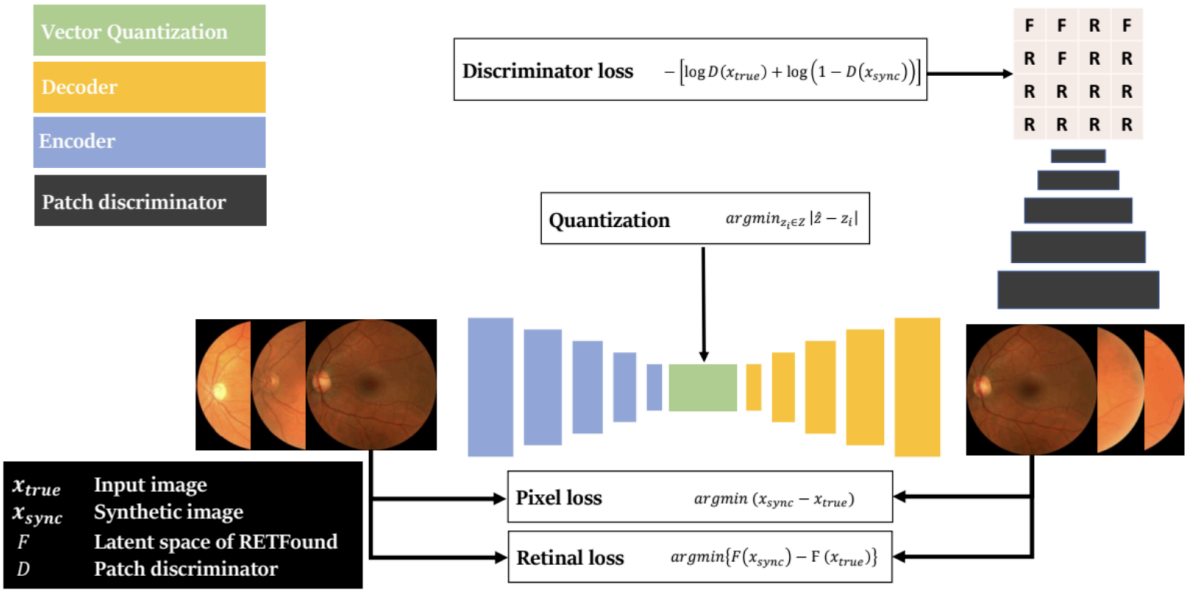


Fig. 1: The model is a standard VQ-GAN with an autoencoder-decoder architecture, incorporating quantization reparametrization in the latent space and patch-based discriminator supervision. The proposed training strategy introduces a loss function defined in the latent space of RETFound, which was pre-trained on 1.6 million retinal images, ensuring that the synthetic images retain clinical and topological accuracy.

References

[1] Y. Zhou et al. A foundation model for generalizable disease detection from retinal images. Nature, 622(7981):156–163, September 2023.
[2] Y. Zhou et al. Automorph: Automated retinal vascular morphology quantification via a deep learning pipeline. Translational Vision Science amp; Technology, 11(7):12, July 2022.

Predicting molecular and cellular features from gut histopathological images using deep learning

Kexin Xu, University of Oxford

Objectives

Deep learning has been applied to digital pathology for computer-aided diagnosis and prognosis. However, most models focused on cancers. Crohn's disease is a type of inflammatory bowel disease. Its diagnosis and prognosis remain challenging due to undefined clinical criteria.

Spatial transcriptomics (ST), named "Method of the Year" by Nature in 2021, maps spatial gene expression and cell localisation. It can identify disease biomarkers for diagnosis and prognosis. However, its high cost and technical complexity limit its clinical application.

To overcome this, I employed deep learning to predict ST and its derived features — such as gene signatures, cell types, and tissue domains — from routinely collected microscopy images of gut biopsies for the stratification of Crohn's disease patients.

Methods

A convolutional neural network was developed to predict molecular and cellular features using ST data of 33 sections of Crohn's disease gut tissues. Strategies to improve prediction were tested, such as gene imputation, model pre-training, image augmentation, and varying model architectures and tissue patch sizes.

Results

Prediction correlation was good for genes that are spatially heterogeneous or markers for tissue domains. Predicting seven gene principal components recovered 3,000 constituent genes, offering a computationally efficient way to predict gene expression. Cell types and clusters could be predicted well when abundant in training data. Gene signature for a cell was more predictable than the cell type. Predicting a single cell type was easier than multi-target prediction. Performance was improved by using bigger patches, smaller models, gene imputation, model pre-training and fine tuning, and image augmentation. Models showed cross-technology transferability.

Conclusions

I demonstrated the predictability of various molecular and cellular features from gut histopathological images and identified strategies to improve predictions. With limited gut ST data, imaging-derived predictions might identify biomarkers with diagnostic, prognostic, or therapeutic values for Crohn's disease and other gut pathologies.

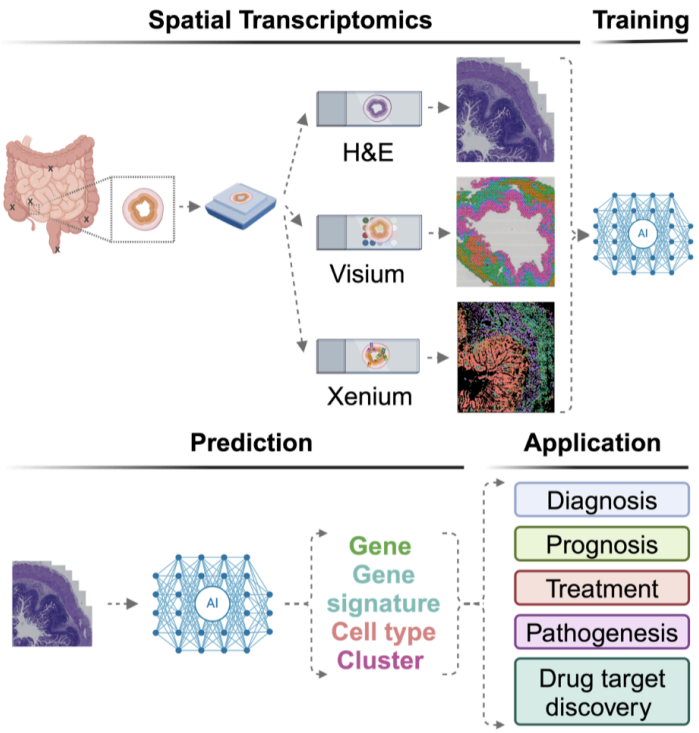


Figure 1. Workflow for training a neural network using gut spatial transcriptomics data to predict molecular and cellular features from H&E images for biological and clinical applications. Spatial transcriptomics was measured by Visium and Xenium technologies.

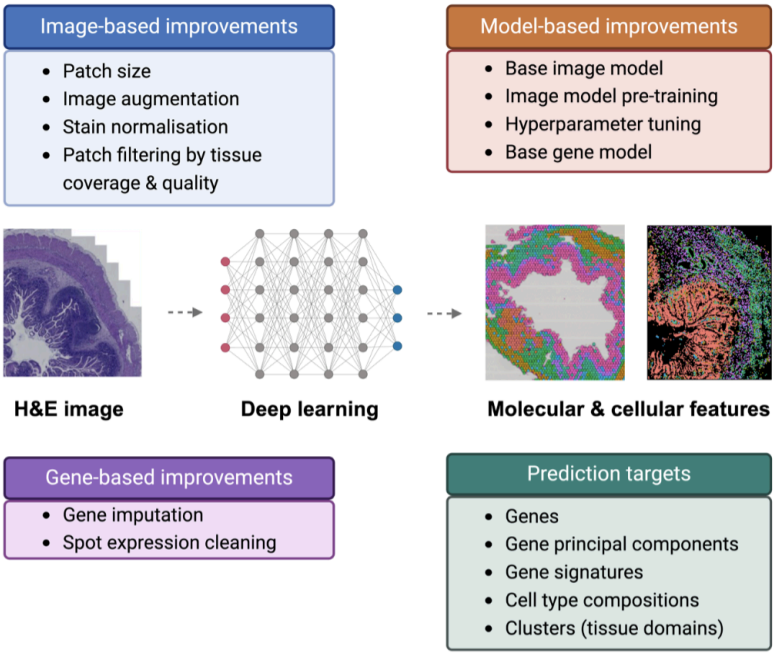


Figure 2. Various prediction targets and strategies to improve predictions were explored.

Algorithmic Fairness Beyond Predictions

Silvia Casacuberta, University of Oxford

In recent years, algorithms have impacted all sectors of society, such as healthcare, teaching, and finance. In many such settings, algorithms are used in decision-making processes to obtain individual predictions: should someone receive a loan? Will someone develop a certain illness? Along with the rise of algorithmic decision-making, many studies have demonstrated how algorithmic predictions can be biased against certain groups, and other ways in which individual predictions can be faulty. In turn, this has led to the development of the field of algorithmic fairness, which studies and tries to resolve instances of potential algorithmic harms. However, much of the field has focused on studying the potential biases embedded within these individual predictions, for example by proposing various individual and group fairness metrics. In this talk, we present research that aims to widen this narrow traditional focus on the prediction values themselves in three different new directions.

First, we focus on the problem of model multiplicity. What occurs when we have multiple competing classifiers that disagree on certain individual predictions? We present a comprehensive analysis of a recently-introduced method for reconciling two disagreeing models and demonstrate its effectiveness on many different fairness-related datasets. We also show how to extend this algorithm theoretically and practically to the setting of causal inference. Second, we discuss the problem of prediction uncertainty: why should we focus so heavily on the potential biases present in individual predictions if these turn out to be highly arbitrary and unreliable? We propose a new algorithm for learning with abstentions: that is, we want to build a classifier that returns either a numerical value or refuses to abstain on particular inputs. We provide theoretical guarantees in the rigorous framework of agnostic learning. Finally, we ask: in many settings, do we actually need to produce individual predictions in the first place? We focus on the setting of resource allocation, and provide a new algorithm for performing an optimal allocation without requiring the usual framework of allocation-through-prediction. Together, our work demonstrates various ways in which we can expand our understanding of the pitfalls of algorithmic prediction.

This research stems from three different joint works with 1) Varun Kanade (University of Oxford), 2) Moritz Hardt (Max Planck Institute for Intelligent Systems), & 3) Tina Behzad (Stony Brook University), Emily Ruth Diana (Carnegie Mellon University), and Alexander Williams Tolbert (Emory University).

Digital Phenotyping of Emotional Expression in Schizophrenia

Shrankhla Pandey, University of Cambridge
Sandra A. Just Charité, University Medicine Berlin, Germany

Background

Schizophrenia diagnosis and treatment currently rely heavily on subjective clinical assessments, creating challenges due to the disorder's heterogeneous nature.Recent advances in computational methods, particularly natural language processing (NLP) and acoustic analysis, have demonstrated potential to detect subtle patterns in speech and language that predict symptom development. This highlights the necessity of such research.

Aim

This research aims to develop objective, scalable computational metrics for assessing emotional expression in speech and language of individuals with schizophrenia or schizoaffective disorder. These novel metrics will complement traditional clinical scales (SAPS, PANSS, SANS, mini-ICF, CDSS) and reveal emotional patterns which are disrupted in schizophrenia.

Methods

The study will analyse responses from a longitudinal study [1] to the Narrative Emotions Task (NET) [2], which prompts patients (n=95) to (a) define emotions, (b) recall emotional experiences, and (c) explain causal relationships between events and emotions. Clinical data related to Methodological approaches include: 1) Predicting symptoms & emotion category using NLP features and eGeMAPS-88 acoustic features – also revealing how patients shift between conceptual, episodic, and causal aspects of emotion 2) measure emotional dissonance using cross-modal synchrony between linguistic content and vocal expression; 3) Comparing sentiment trajectories across narrative aspects and emotions between patients and healthy controls; 4) Analysing stability and flexibility of emotional expressions using state space grids; and 5) Examining emotional inertia differences across narrative aspects.

Expected Results

We anticipate identifying distinctive patterns in schizophrenia patients,including: deficits in emotional prosody processing, diminished emotional coherence throughout narratives, abrupt transitions between emotional states, difficulty shifting between emotions, and altered emotional inertia compared to healthy controls. These patterns may correlate with clinical symptom severity and provide quantifiable markers of emotional dissonance.

Conclusion

This research can complement existing clinical assessments. Quantifying emotional expression patterns and deriving meaningful metrics can assist clinicians in early prediction,

diagnosis, and more precise evaluation of outcomes. It can also offer tools for symptom monitoring for the patients living with schizophrenia.

References

[1] Haebler et al, 2015, Modified psychodynamic psychotherapy for patients with schizophrenia - a randomized controlled efficacy study (MPP-S study) https://www.ipu-berlin.de/en/mpp-s-study/
[2] Buck et al, 2014, The Narrative of Emotions Task (NET) - The use of narrative sampling in the assessment of social cognition https://doi.org/10.1016/j.psychres.2014.03.014
[3] Eyben et al, 2015, The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing | IEEE Journals & Magazine | IEEE Xplore 10.1109/TAFFC.2015.2457417

AI-Powered Histopathology: Deep Learning for Risk Stratification in Esophageal Cancer in a Large-Scale Clinical Trial

Greta Markert, University of Cambridge

Histopathological assessment is the gold standard for diagnosing oesophageal cancer, but it is time-intensive, labour-intensive, and prone to human error. In our large-scale clinical trial involving 120,000 patients, pathologists are required to analyse vast numbers of slides, increasing both workload and the risk of missed high-risk cases. To address these challenges, I developed a fully automated AI pipeline to assist in tissue segmentation, gland detection, multi-stain registration, and colour-based filtering for risk stratification.

Our AI model processes whole slide images stained with H&E, TFF3, and p53, which provide crucial insights into Barrett's oesophagus and tumour progression. A novel segmentation and registration algorithm enables the precise alignment of multi-stain images at the glandular level, improving the identification of high-risk cellular patterns. Additionally, we incorporate patient metadata (e.g., age, BMI, Barrett's segment length) to compute personalised risk scores, allowing for a more refined assessment of disease progression.

This pipeline serves two key purposes:
1. Massively reducing pathologists' workload by automating tedious and repetitive tasks.
2. Enhancing diagnostic accuracy by ensuring that no high-risk cases are missed.

Already, our system has detected cases that were initially overlooked by pathologists, demonstrating its potential to significantly improve early cancer detection and patient outcomes. By integrating AI into histopathology workflows, we aim to revolutionise clinical practice, making cancer diagnostics both more efficient and more reliable.

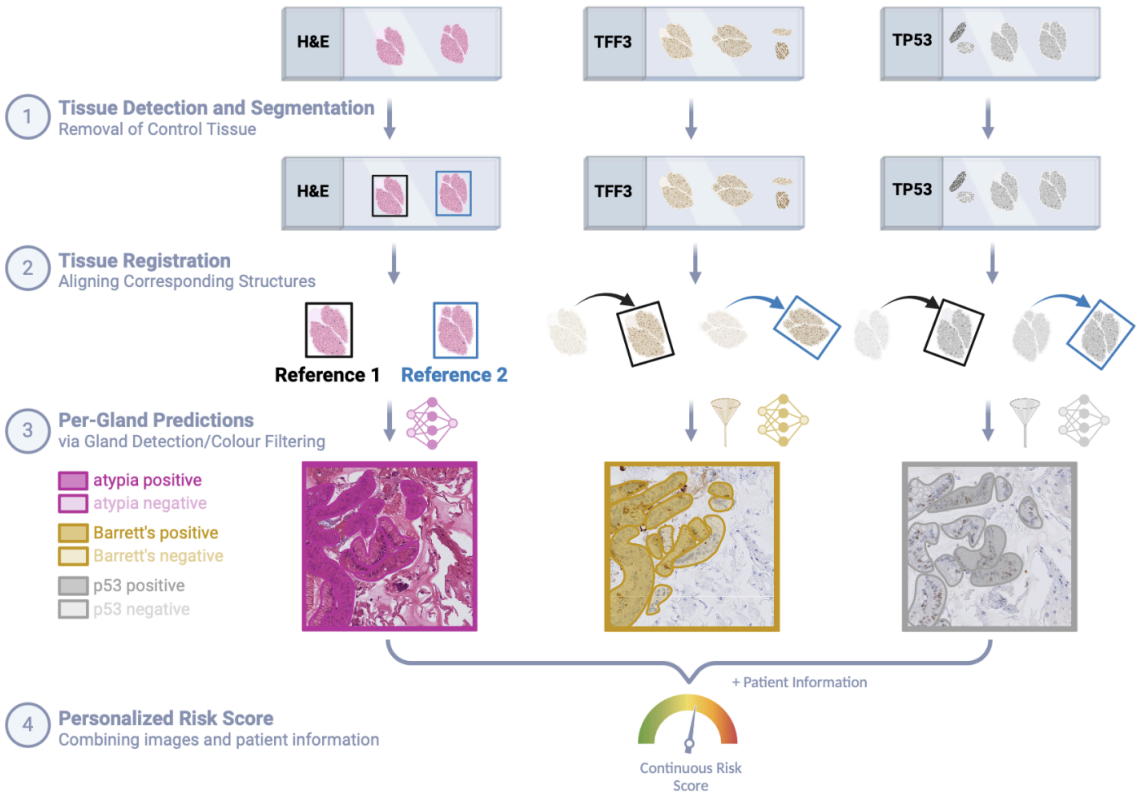


Fig 1: Pipeline from whole slide images to personalised risk score.

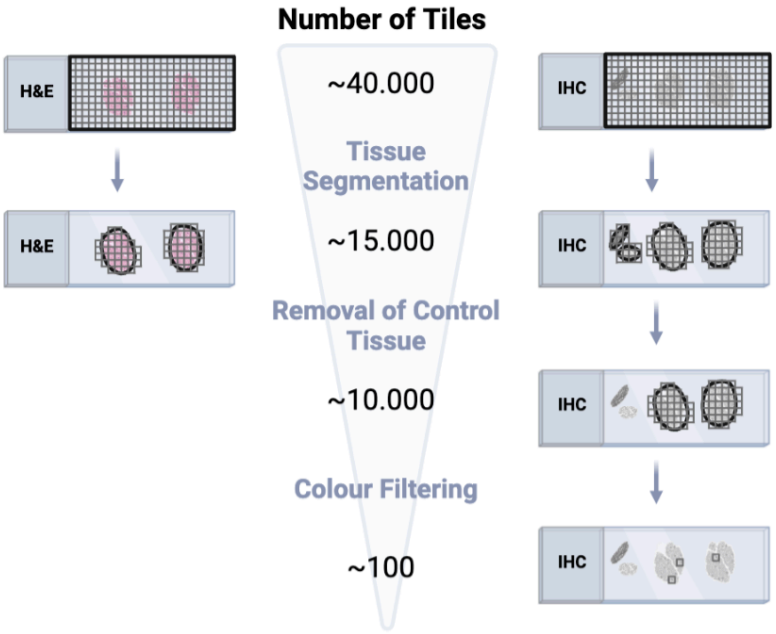


Fig 2: In order to reduce the number of input tiles, we developed different techniques in-house, such as IHC colour filtering and removing control tissue.

Can AI Design the Shape of the Internet?

Akanksha Ahuja, University of Cambridge

Optical network performance depends on its physical topology, which supports critical infrastructure for national security and defense and connects millions of users nationwide. Designing an optimal topology is inherently complex due to scalability limitations and computational constraints. Conventional graph generators fail to scale beyond 40 nodes. Similarly, manual or optimization based design processes are computationally intractable for large scale networks, requiring months of effort. Consequently, automating the design of optical networks remains an open challenge, further compounded by limited access to real-world data.

Can graph machine learning models infer network design principles from past topologies to generate scalable topologies for the future?

To address the dataset gap, we introduce Topology Bench, a collection of 105 real-world optical networks spanning diverse regions and scales. We introduce Topology Architect, the first application of graph machine learning for unsupervised topology generation. We evaluate this model using graph similarity metrics and show that it achieves 95% spectral and 84% structural similarity to real networks, as measured by Wasserstein distances. It generates topologies by sampling edge probabilities from the latent space based on user-specified parameters, such as node count and locations, in less than a second. To assess latent space quality, we compare clustering in Topology Bench using graph-theoretic properties and graph embeddings from Topology Architect, finding the latter 20 times more effective.

This research makes three key contributions: (i) provides an open-access dataset of real-world network topologies, (ii) captures the graph-theoretic properties underlying topology design, and (iii) automates topology generation. This further enables the synthesis of realistic topologies in regions with sparse or proprietary data. While developed for optical networks, this framework is transferable to large-scale infrastructure systems, including transportation, urban planning, and quantum communication.

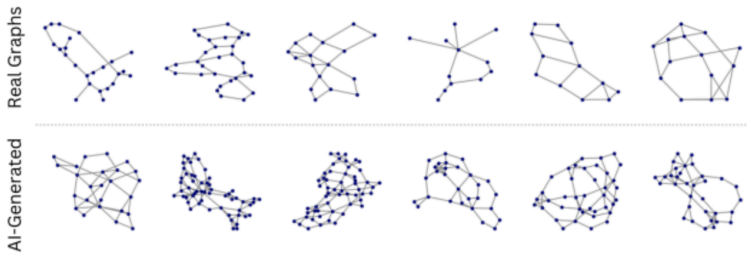


Figure 1: Real (training dataset) and generated (sampled) graphs with different graph sizes and nodal locations, scalable up to 100 nodes.

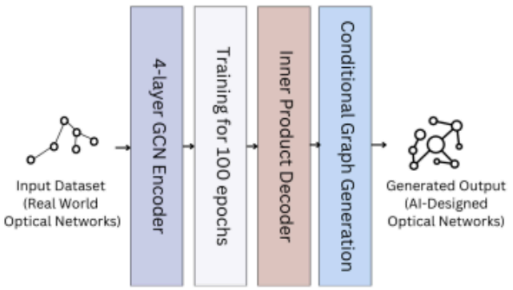


Figure 2: Topology Architect’s Pipeline using Graph Variational Autoencoder.

Computational Detection of 5-Fluorouracil Metabolites in Nanopore-Sequenced RNA: Unraveling Chemotherapy-Induced Transcriptome Damage

Shutong Ye, University of Cambridge
Dr. Nicole Simms, Division of Cancer Sciences, School of Medical Sciences, University of Manchester
Dr. John Knight, Division of Cancer Sciences, School of Medical Sciences, University of Manchester
Dr. Michael Beomo, Department of Genetics, Department of Pathology, University of Cambridge

Attaining insight into RNA modifications and base analogs is increasingly important in molecular biology and genetics. As a widely used chemotherapeutic drug for the treatment of solid cancer (colorectal, breast cancer), the efficacy of 5-Fluorouracil (5-FU) is partial and often associated with drug tolerance [5, 9]. Although 5-FU has been applied clinically for more than 60 years, nearly half of the patients are unable to respond to this therapy positively [2, 8]. Methods for identifying and quantifying 5-FU have often relied on time-consuming and labor-intensive experimental techniques [1, 2, 6, 7] but computational methods for detecting 5-FU are lacking. There is an urgent requirement for new tools to finely map the location of 5-FU within transcriptome to improve our understanding its role in cytotoxicity .

Recent advances in direct RNA sequencing using Oxford Nanopore R10 platform provide a promising approach to discriminate and identify different RNA modifications in native RNA sequence. This next-generation tool introduces development challenges—such as version conflicts and analytical pipeline refinements—but enables deeper exploration of emerging biological questions [3, 4]. We developed RNAscent, which is, to our knowledge, the first algorithm capable of pinpointing 5-FU positions in individual RNA molecules using the latest ONT R10 platform. RNAscent combines robust data processing pipelines with a deep neural network to achieve high-sensitivity localization of 5-FU. Trained on in vitro samples with complete 5-FU incorporation, RNAscent achieved a sensitivity of 0.8858 and specificity of 0.8810, demonstrating robust performance even at lower incorporation rates. Overall, RNAscent offers a streamlined, user-friendly solution for researchers seeking to detect RNA modifications, including base analogs, at the single-molecule level. Notably, this is the first reported instance of a chemotherapy metabolite being successfully detected in nanopore-sequenced RNA. RNAscent represents a substantial leap forward in quantifying 5-FU at single-base resolution, promising new insights into RNA damage pathways, drug tolerance, and personalized therapeutic strategies.

References

[1] Blondy, S., David, V., Verdier, M., Mathonnet, M., Perraud, A., Christou, N.: 5-fluorouracil resistance mechanisms in colorectal cancer: From classical path- ways to promising

processes. Cancer Science 111(9), 3142–3154 (aug 2020). https://doi.org/10.1111/cas.14532, https://doi.org/10.1111%2Fcas.14532

[2] Chalabi-Dchar, M., Fenouil, T., Machon, C., Vincent, A., Catez, F., Marcel, V., Mertani, H.C., Saurin, J.C., Bouvet, P., Guitton, J., Venezia, N.D., Diaz, J.J.: A novel view on an old drug, 5-fluorouracil: an unexpected RNA modifier with intriguing impact on cancer cell fate.

NAR Cancer 3(3) (jul 2021). https://doi.org/10.1093/narcan/zcab032

[3] Furlan, M., Delgado-Tejedor, A., Mulroney, L., Pelizzola, M., Novoa, E.M., Leonardi, T.: Computational methods for RNA modification detection from nanopore direct RNA sequencing data. RNA Biology 18(sup1), 31–40 (sep 2021). https://doi.org/10.1080/15476286.2021.1978215

[4] Leger, A., Amaral, P.P., Pandolfini, L., Capitanchik, C., Capraro, F., Miano, V., Migliori, V., Toolan-Kerr, P., Sideri, T., Enright, A.J., Tzelepis, K., van Werven, F.J., Luscombe, N.M., Barbieri, I., Ule, J., Fitzgerald, T., Birney, E., Leonardi, T., Kouzarides, T.: RNA modifications detection by comparative nanopore direct RNA sequencing. Nature Communications 12(1) (dec 2021). https://doi.org/10.1038/s41467-021-27393-3

[5] Longley, D.B., Harkin, D.P., Johnston, P.G.: 5-fluorouracil: mechanisms of action and clinical strategies. Nature Reviews Cancer 3(5), 330–338 (5 2003). https://doi.org/10.1038/nrc1074

[6] Pawłowski, P.H., Szczęsny, P., Rempoła, B., Poznańska, A., Poznański, J.: Combined iin silico/i and 19f NMR analysis of 5-fluorouracil metabolism in yeast at low ATP conditions. Bioscience Reports 39(12) (dec 2019). https://doi.org/10.1042/bsr20192847

[7] Procházková, A., Liu, S., Friess, H., Aebi, S., Thormann, W.: Determination of 5-fluorouracil and 5-fluoro-2′-deoxyuridine-5′-monophosphate in pancreatic cancer cell line and other biological materials using capillary electrophoresis. Journal of Chromatography A 916(1-2), 215–224 (may 2001). https://doi.org/10.1016/s0021-9673(00)01171-7

[8] Shahi, S., Ang, C.S., Mathivanan, S.: A high-resolution mass spectrometry-based quantitative metabolomic workflow highlights defects in 5-fluorouracil metabolism in cancer cells with acquired chemoresistance. Biology 9(5), 96 (may 2020). https://doi.org/10.3390/biology9050096

[9] Therizols, G., Bash-Imam, Z., Panthu, B., Machon, C., Vincent, A., Ripoll, J., Nait-Slimane, S., Chalabi-Dchar, M., Gaucherot, A., Garcia, M., Laforêts, F., Marcel, V., Boubaker-Vitre, J., Monet, M.A., Bouclier, C., Vanbelle, C., Souahlia, G., Berthel, E., Albaret, M.A., Mertani, H.C., Prudhomme, M., Bertrand, M., David, A., Saurin, J.C., Bouvet, P., Rivals, E., Ohlmann, T., Guitton, J., Venezia, N.D., Pannequin, J., Catez, F., Diaz, J.J.: Alteration of ribosome function upon 5-fluorouracil treatment favors cancer cell drug-tolerance. Nature Communications 13(1) (jan 2022). https://doi.org/10.1038/s41467-021-27847-8

Complex-Weighted Diffusion: A New Perspective on Heterophily and Over Smoothing in GNNs

Cristina Lopez Amado, University of Oxford

Graph Neural Networks (GNNs) have gained significant attention due to their applications ranging from social sciences to drug discovery. Most GNN architectures rely on the message-passing paradigm, which, despite its success, faces two major challenges: (1) poor performance on heterophilic graphs and (2) oversmoothing, where node representations become indistinguishable after multiple layers. Graph Convolutional Networks (GCNs) can be interpreted as an augmented heat diffusion process, but [1] shows that heat diffusion fails to linearly separate classes in many node-classification tasks. Inspired by [1], we address these limitations by introducing a novel geometric perspective on graph diffusion. Specifically, we propose equipping the graph with a complex-weighted structure, where each edge is assigned a complex number. This enables a diffusion process that extends random walks to graphs with complex weights. Our main theoretical result shows that, with an appropriate complex structure, every node-classification task can be separated in the time limit of a complex random walk. Building on this insight, we introduce the Complex-Weighted Convolutional Network (CWCN), which augments the complex random walk diffusion and surpasses the expressiveness of GCNs. Finally, we discuss the advantages of our approach over [1] and expect competitive performance on standard node-classification benchmarks. Our findings suggest that incorporating complex-weighted diffusion provides a powerful approach to designing more expressive GNNs. To the best of our knowledge, this is the first work to leverage complex weights to enhance GNN expressiveness, and we hope it inspires further exploration of their potential in graph-based learning tasks.

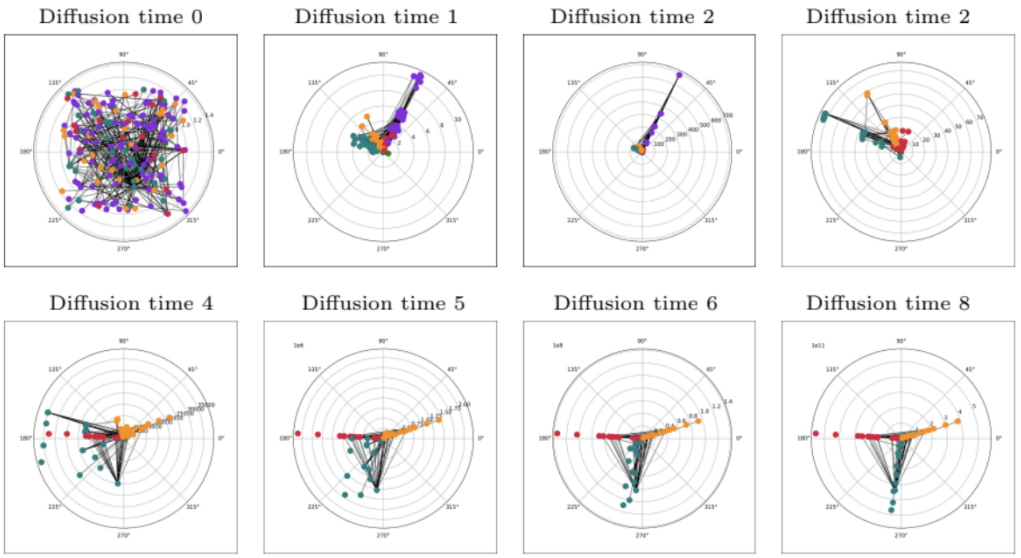


Figure 1: Complex random walk diffusion progressively separates the classes of a complex-weighted graph, where nodes in the same class converge to complex numbers with the same argument. From the third figure onward, the purple class is omitted for better visualization.

References

[1] C. Bodnar, F. Di Giovanni, B. Chamberlain, P. Lio, and M. Bronstein, “Neural sheaf diffusion: A topological perspective on heterophily and over-smoothing in gnns,” Advances in Neural Information Processing Systems, 2022.
https://doi.org/10.48550/arXiv.2202.04579

A Non-Intrusive Monitoring Framework for ML Workload: Enabling Bottleneck Optimization and AI Governance across Various Computing Infrastructures

Ziji Chen, University of Oxford
Noa Zilberman, University of Oxford

The rapid adoption of machine learning (ML) across scientific and industrial domains has significantly accelerated their deployment in diverse computing infrastructures. Technologies such as virtual machines and containers abstract physical resources, enabling efficient resource sharing while preserving isolated user environments. However, this isolation introduces a critical challenge: computing service providers and Artificial Intelligence (AI) regulators often lack sufficient visibility into user ML activities within these virtualized settings. Such limited transparency raises substantial concerns regarding the potential misuse of ML technologies.

To address this issue, we propose a lightweight, privacy-preserving, and non-intrusive monitoring framework designed to systematically collect and analyse host-level system performance metrics. These metrics include CPU and GPU utilization, memory usage, cache behaviour, storage I/O operations, and network communication across distributed ML training nodes. Figure 1 presents the architecture and conceptual overview of the proposed monitoring framework. By capturing system performance data, the framework provides insights into ML workload behaviours without compromising user privacy.

Beyond supporting effective technical AI governance, this framework enhances ML efficiency. Improved speed and performance in ML applications directly lead to faster data processing, more efficient simulations, and improved prediction accuracy, thereby advancing both scientific discoveries and industrial innovations. Given the increasing ubiquity of ML applications, systematic approaches to enhance training and inference performance are essential. Furthermore, this study identifies performance bottlenecks by analysing resource consumption patterns of ML workloads. Pinpointing these constraints allows developers and service providers to optimise resource configurations, thereby improving operational efficiency, reducing energy consumption, and lowering carbon emissions. As illustrated in Figure 2, systematic analysis reveals significant variability among ML models in resource utilization patterns, validating the core hypothesis of this research.

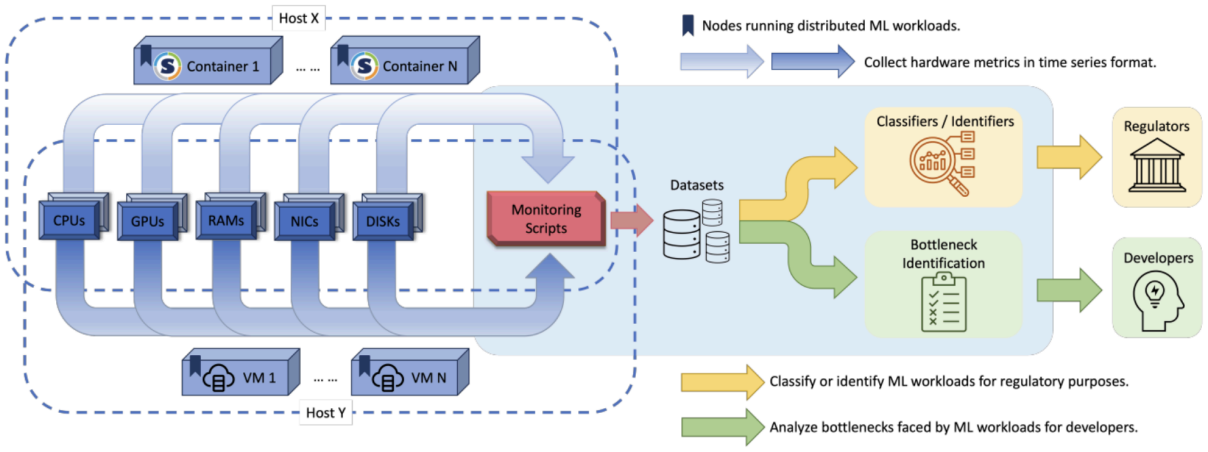
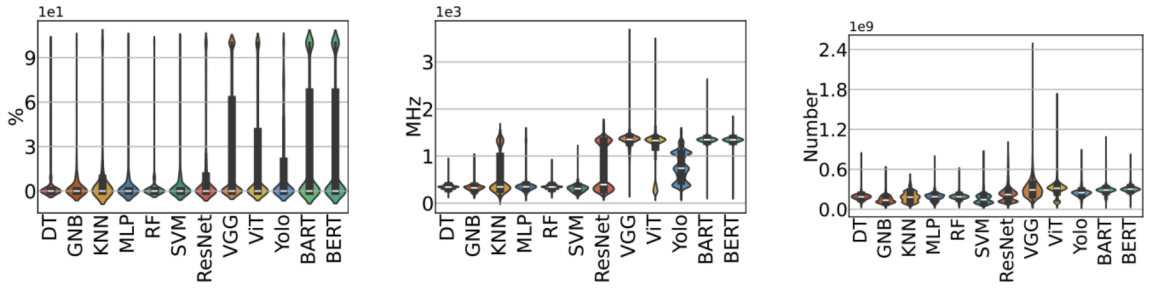


Figure 1: Description of non-intrusive monitoring framework.



(a) CPU Utilisation (b) CPU Frequency (c) Branches

Figure 2: Violin plots of CPU-related metrics when different ML workloads.

Ask Patients with Patience: Enabling LLMs for Human-Centric Medical Dialogue with Grounded Reasoning

Jiayuan Zhu, University of Oxford

The severe shortage of medical doctors limits access to timely and reliable healthcare, leaving millions underserved. Large language models (LLMs) offer a potential solution but struggle in real-world clinical interactions. Their language is often rigid and mechanical, lacking the human-like qualities essential for patient trust. To address these challenges, we propose Ask Patients with Patience (APP), a multi-turn LLM-based medical assistant designed for grounded reasoning and human-centric interaction. APP enhances communication by eliciting user symptoms through empathetic dialogue, significantly improving accessibility and user engagement. It also incorporates Bayesian active learning to ensure reliable and transparent diagnoses. The framework is built on verified medical guidelines from the MSD Manual, ensuring grounded and evidence-based reasoning. To evaluate its performance, we develop a new benchmark that simulates a realistic clinical consultation environment using real-world interview cases. We compare APP against SOTA one-shot and multi-turn LLM baselines. Results show that APP improves diagnostic accuracy, reduces uncertainty, and enhances user experience. By integrating medical expertise with transparent, human-like interaction, APP bridges the gap between AI-driven medical assistance and real-world clinical practice.

The Effect of Representational Compression on Flexibility Across Learning

Mia Whitefield, University of Oxford

The representational geometry framework formalises how information is structured in intelligent systems and suggests that abstraction involves a trade-off between generalisation and flexibility. However, how the geometry of task representations evolves across learning and how it corresponds to performance remains unclear. Here, we tested the hypothesis that task representations become compressed throughout learning, sacrificing flexibility for task efficiency. Using an extra-dimensional shifting task, we manipulated the pretraining length to control the degree of compression. In both humans and artificial neural networks, longer pre-training was associated with decreased flexibility. Analysis of network dynamics suggested that greater compression incurred a higher representational reorganisation cost, restricting flexibility. However, the introduction of an auxiliary reconstruction loss maintained higher dimensionality, mitigating the flexibility impairment. Our findings point towards a representational geometry-based explanation for flexibility dynamics across learning.

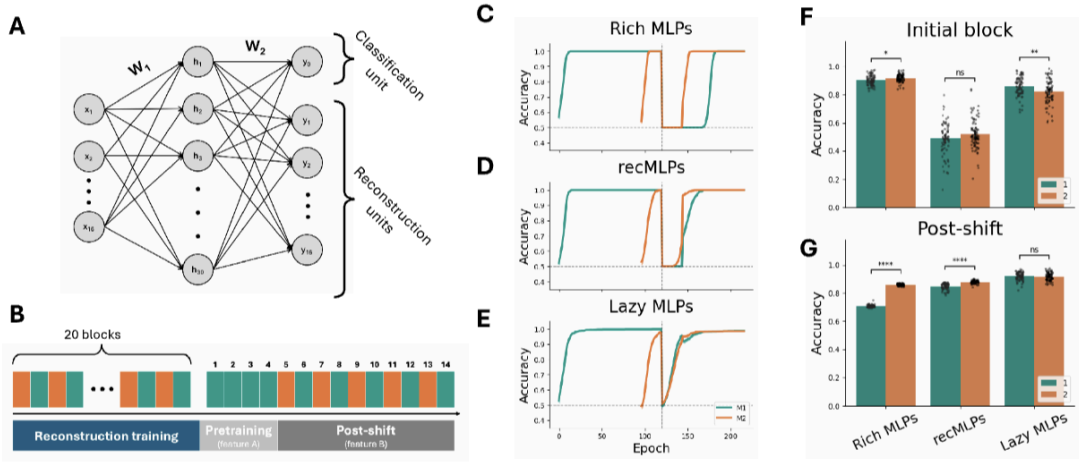


Figure 1. ANN setup and behaviour

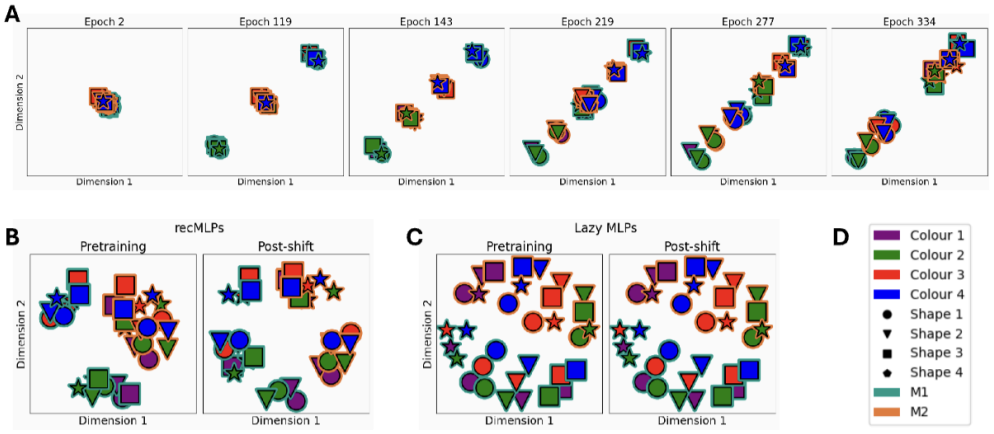


Figure 2. Representational geometry. MDS plots

PEARL: Preference Extraction and Reward Learning

Marta Grzeskiewicz, University of Cambridge

Preferences and utility are foundational concepts in economics and are central to determining consumer behaviour through the utility-maximising consumer decision-making process. However, preferences and utilities are not observable and may not even be known to the individual making the choice; only the outcome is observed in the form of demand. Without the ability to observe the decision-making mechanism, demand estimation becomes a challenging task. Even so, machine learning-based regression models are widely but erroneously used in an attempt to solve this problem. The alternative is using econometric methods, which rely on specifying instrumental variables (IVs) for each product and combination of products, which rapidly faces scalability issues. To address the shortcomings of existing methods, we introduce our algorithm, Preference Extraction and Reward Learning (PEARL), where we treat the consumer as an agent acting in an economic environment, maximising utility by choosing what to consume while constrained by how much they can afford. PEARL is an algorithmic approach which combines revealed preference (RP) theory and optimisation methods in an algorithmic structure often used in inverse reinforcement learning. The effects of unobserved demand shocks are decoupled by ensuring that observations are consistent under RP theory. We present a functional form for utility, the Input-Concave Neural Network, as a neural network with concave activation functions: concave-tanh, concave-sigmoid and concave-log. Experimental results show that on noise-free synthetic purchasing data, PEARL predicts with near-zero error when the functional form of the utility function is known and when the ICNN is used. PEARL outperforms machine learning regression benchmarks on both noise-free and noisy synthetic purchasing data. PEARL is also shown to successfully compute elasticities, and to out-perform the benchmark on real-world supermarket data.

Colors in Costumes: Computational Analyses of Color in Costumes in 12K Portrait Paintings from 1400 to Today

Christine Li, University of Oxford

Clothing is a lens through which a society expresses its culture and history. Its stylized portrayal in painting adds an immensely rich layer of cultural self introspection—how artists see themselves and their contemporaries, expressed through art. Particularly of interest in this study is color: how has color in costumes in portraiture painting changed over time, across art styles, and for different genders? In this study, we apply computational methods drawn from computer vision, machine learning, economics, and statistics to a large open-sourced corpora of over 12k portrait paintings to analyze trends in color in Western art over the past 600 years [1]. For each painting, we obtained clothing segmentation masks using a fine-tuned SegFormer model [2], performed gender classification using CLIP (Contrastive Language-Image Pre-Training), extracted dominant colors via clustering analysis, and computed the Color Contrast Index (CI) and the Diversity Index (DI) [3]. We highlight particularly noteworthy observations. For example, Pop Art and Fauvist paintings rank highest in CI and DI values, respectively, while Symbolist and Mannerist paintings rank the lowest (Fig 1). With the exception of the latter half of the 20th century, the DI of female portraits is always higher than that of male portraits, and this DI gender gap largely increases from the 15th to 18th centuries but shrinks in the past 300 years (Fig 2). This study is, to our knowledge, the most comprehensive, large-scale analysis of colors of clothing in paintings. Through sharing this work, we hope to make more widely accessible the methodology for applying state-of-the-art computer vision and other computational techniques to art historic analysis, to aid scholars studying the history and development of style in fine art paintings—with the hopes of fostering more cross disciplinary collaboration between computer science and art history.

References:

[1] W. Kan, “Painter by numbers,” 2016. https://kaggle.com/competitions/painter-by-numbers.

[2] A. C. Gallagher and T. Chen, “Using context to recognize people in consumer images,” IPSJ Transactions on Computer Vision and Applications 1, pp. 115–126, 2009.

[3] A. Cifuentes and V. Charlin, The Worth of Art: Financial Tools for the Art Markets, Columbia Business School Publishing, New York, NY, 2023.

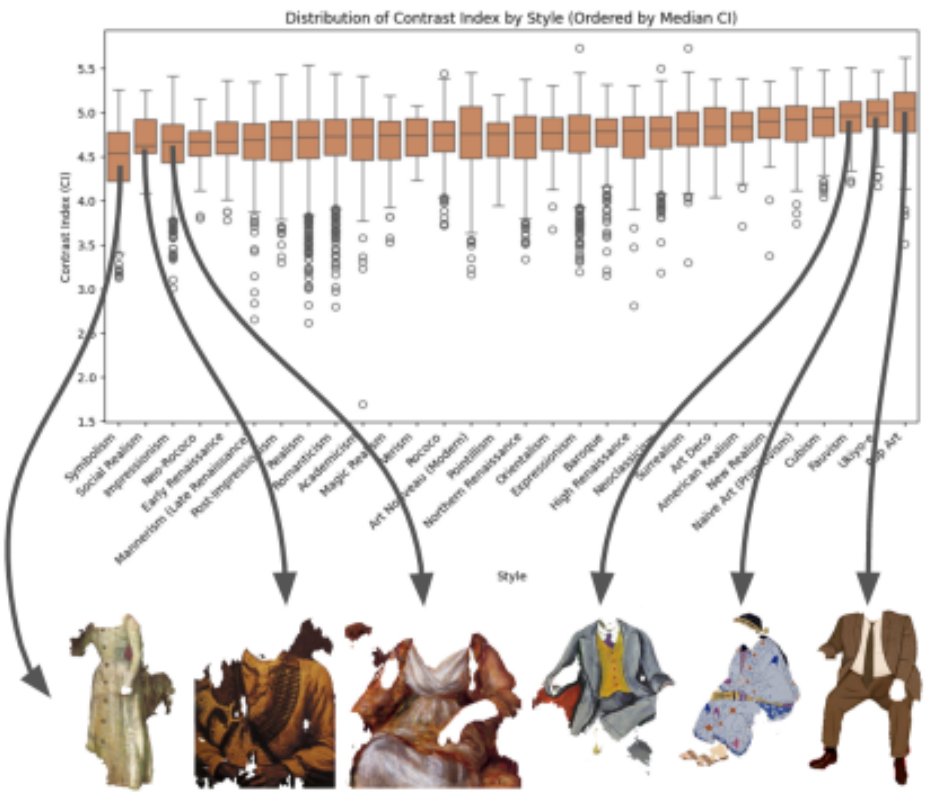


Fig 1.

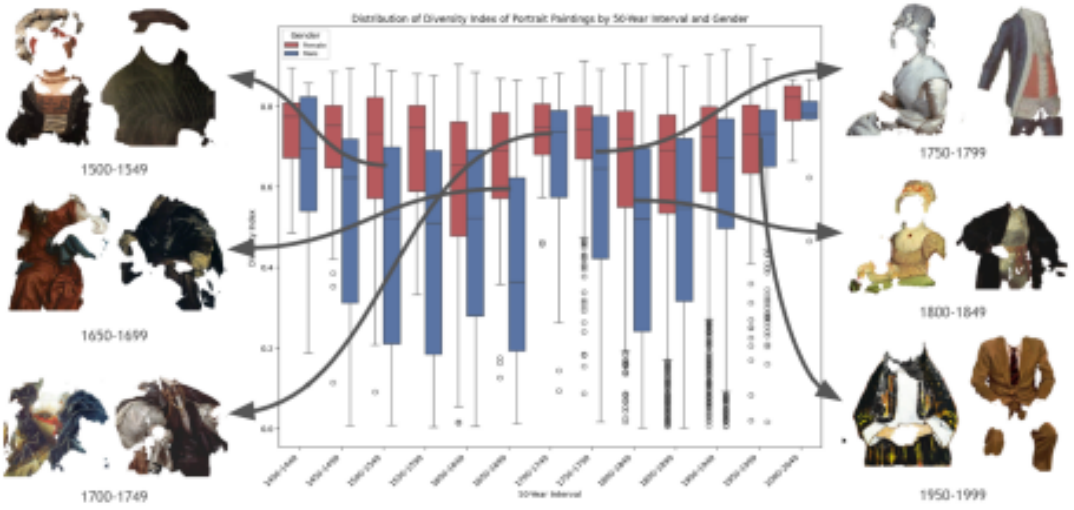


Fig 2: Boxplots of DI for art styles by gender

Fig 2.

Contextual IoT Insights through Edge Computing and Image Captioning

Deema Abdal Hafeth, School of Engineering and Physical Sciences, University of Lincoln
Mohammed Al-khafajiy, University of Lincoln
Stefanos Kollias, University of Lincoln

The rise of Edge Computing has brought processing capabilities closer to the data sources of the Internet of Things, providing solutions to latency and bandwidth limitations in various applications. This allies with Cloud Computing, particularly in managing real-time data processing, decreasing the need for constant cloud communi cation, reducing network congestion and associated costs. The cooperation between image processing and natural language processing, particularly image captioning, significantly enhances the effectiveness of smart monitoring systems. Although early image captioning approaches performed well with encoder-decoder frameworks and attentional mechanisms, they frequently overlooked the importance of semantic representations, which are crucial for a deeper understanding of images. This research aims to utilize Edge Computing for image analysis and seman tic feature extraction near the data sources to enhance the generation of more informative and enriched image captions, ultimately improving the accuracy of captioning. The proposed model architecture is designed to i) learn distinct feature representations of important image regions, and ii) integrate visual attention and semantic attributes into a shared space for feature fusion. This enhances the models' ability to interpret and react to data in a mean ingful manner, which is especially important for IoT applications that rely on a comprehensive understanding of the semantics of diverse and continuously changing data for optimal performance. We conducted extensive experi ments on the MS-COCO dataset to showcase the superiority of the proposed model in terms of both quantitative and qualitative performance. The model achieved the highest CIDEr score of 120.9, marking significant progress in IoT-driven image understanding. In future, we will develop lightweight models for efficient edge-device performance.

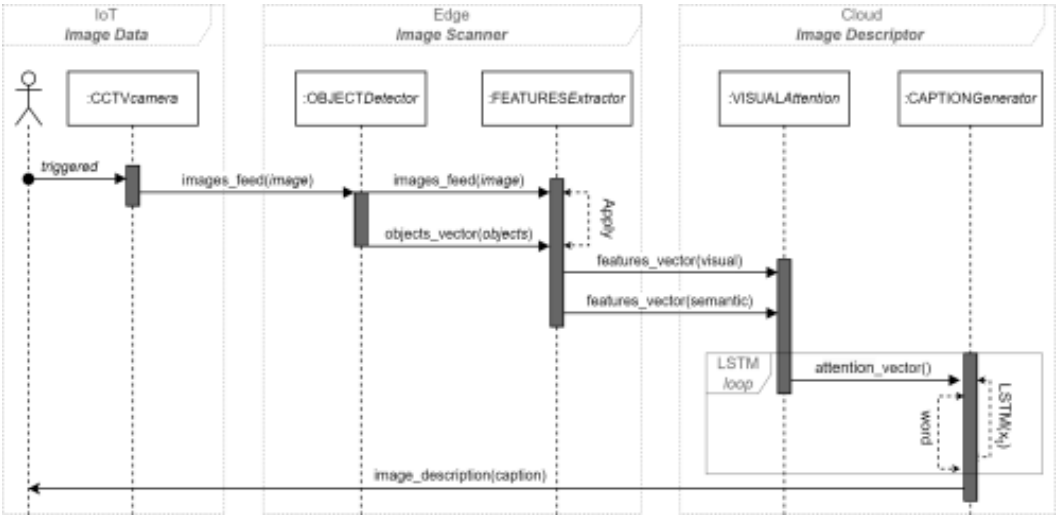


Figure 1: Sequence Diagram for Model Execution.

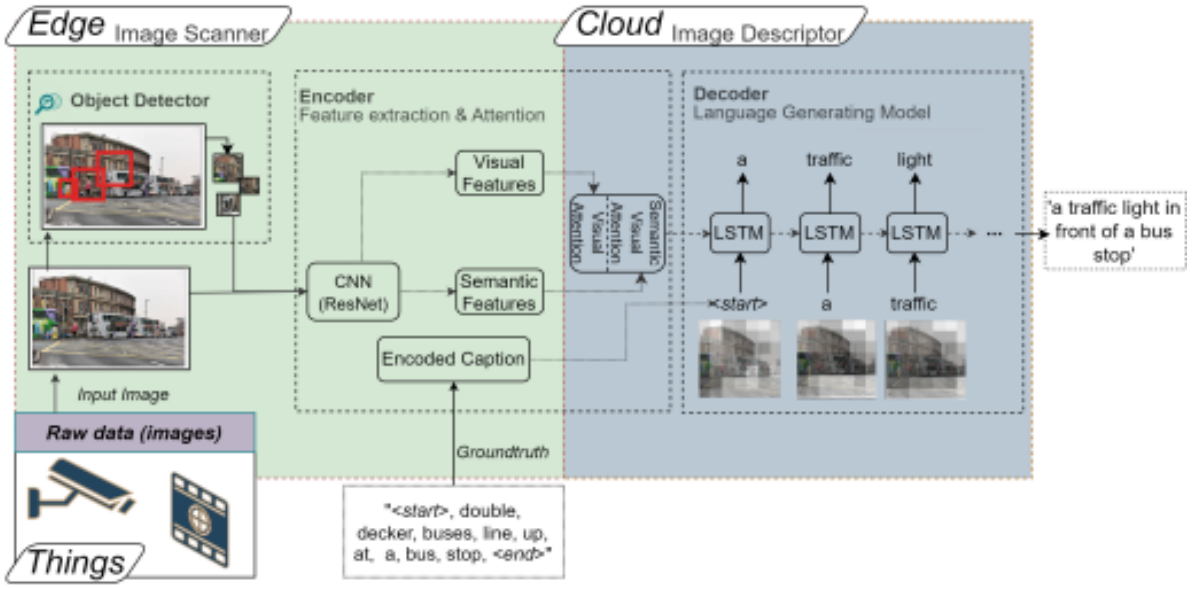


Figure 2: Model Architecture for Caption Generation.

Automotive SPICE and AI Integration to support software process improvement

Menna Noureldin

Recently Automotive software complexity has been raised which led to the adoption of Automotive SPICE “Automotive Software Process Improvement and Capability determination” as a Standard framework that evaluates OEMs and Automotive suppliers process maturity to ensure Quality, safety and efficiency of Automotive Software Development, Concurrently AI is becoming a key tool to automate code generation, Test cases and traceability.

This Research analyzes to what extent AI could be integrated in ASPICE based projects and ASPICE Assessment to support different process areas in multiple aspects e.g.: Software Development, Testing and Quality Assurance through process automation, predictive analytics, and continuous improvement.

Subgroups Matter for Robust Bias Mitigation

Anissa Alloula, University of Oxford

A significant barrier to the deployment of machine learning (ML) models is their tendency to fail when tested on distributions that differ from their training data. One major concern is performance degradation for population subgroups, often caused by bias in training data such as spurious correlations or under-representation [1]. Bias mitigation methods aim to address these issues by adapting model training to avoid learning these biases and improve disadvantaged subgroup performance.

Despite the number of mitigation methods which have been proposed, benchmarks increasingly report their inconsistent performance when tested in new settings, often failing to surpass the empirical risk minimisation baseline (no mitigation) [2], [3]. This raises a fundamental question: when and why do bias mitigation techniques fail? We hypothesise that a key reason may be linked to an often-overlooked but crucial step shared by many methods: the definition of subgroups. Indeed, most mitigation methods rely on some form of grouping to identify disadvantaged subgroups and then to implement group-based mitigation strategies. However, no research has explicitly addressed how to actually define these subgroups, and most papers default to the same, coarse subgroups (e.g. Male/Female, or White/non-White).

We aim to understand whether we can optimise this crucial step of subgroup definition, thereby improving mitigation. We conduct a comprehensive evaluation of state-of-the-art mitigation methods across multiple semi-synthetic image classification tasks, systematically varying subgroup definitions, including coarse, fine-grained, intersectional, and noisy subgroups (Fig. 1). Our findings reveal that subgroup choice significantly impacts performance, with certain groupings paradoxically leading to worse outcomes than no mitigation at all (Fig. 2). Through theoretical analysis, we explain these discrepancies and uncover a counter-intuitive insight that, in some cases, improving fairness with respect to a particular set of subgroups is best achieved by using a different set of subgroups for mitigation. Our work highlights the necessity of carefully considering subgroup definition before applying bias mitigation and offers new perspectives for improving the robustness and fairness of ML models.

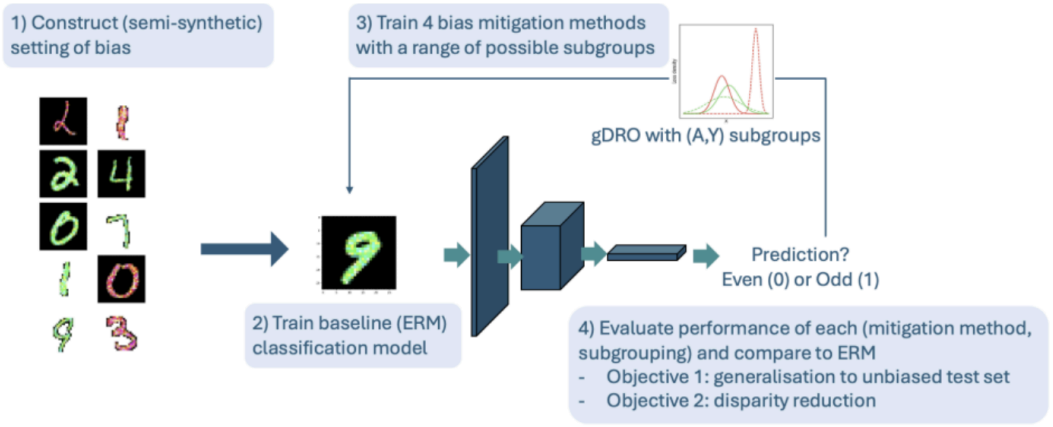


Fig. 1. High level overview of our method.

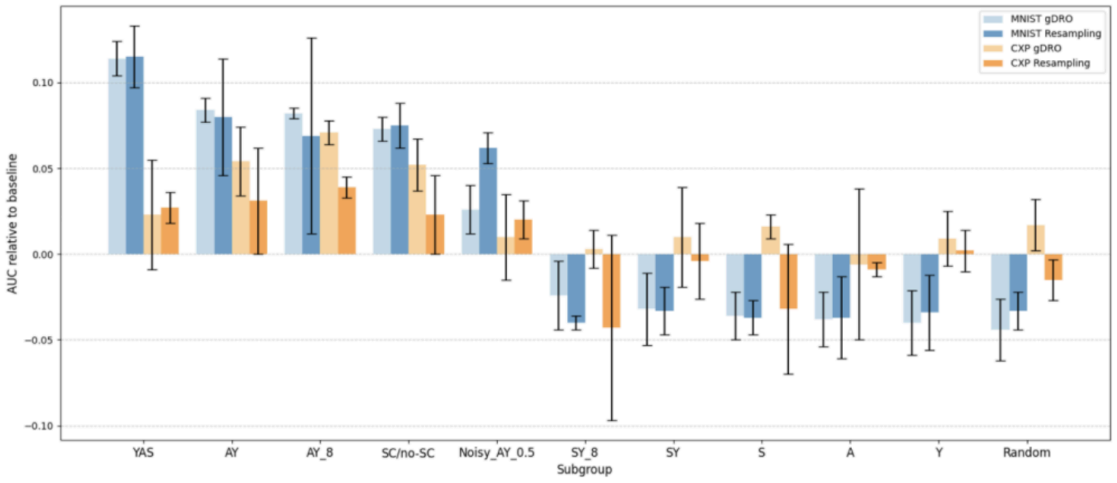


Fig. 2. Performance on the unbiased test set in gDRO and resampling is highly dependent on the subgroups used, and follows similar patterns across mitigation methods and datasets. Bars represent overall change in AUC relative to the ERM baseline, with error bars indicating the standard deviation across 3 seeds.

References

[1] C. Jones, D. C. Castro, F. De Sousa Ribeiro, O. Oktay, M. McCradden, and B. Glocker, “A causal perspective on dataset bias in machine learning for medical imaging,” Nature Machine Intelligence, vol. 6, no. 2, pp. 138–146, Feb. 2024.

[2] Y. Zong, Y. Yang, and T. Hospedales, “Medfair: Benchmarking fairness for medical imaging,” in International Conference on Learning Representations (ICLR), 2023.

[3] R. Shrestha, K. Kafle, and C. Kanan, “An Investigation of Critical Issues in Bias Mitigation Techniques,” in 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, Jan. 2022, pp. 2512–2523.

“It’s All Greek to Me”: Benchmarking Semantic Representations Techniques for Idioms with Multimodal Contexts

Kirthi Kumar, University of Oxford

One of the crucial areas in which language models are challenged today are idioms. Phrases such as “it’s all Greek to me” have multiple meanings, and the non-literal meaning can be difficult for models to interpret. The non-literal meaning can evolve over time and be relevant more in certain cultural contexts than others.

My work approaches this problem using the SemEval 2025 dataset containing noun contexts with two subtasks: A) rank the best contexts out of 5 labels and B) after a sequence of 3 labels, predict the next best context given some options. The input could be images or the input could be text labels, and each of these inputs provide different aspects of idioms. Building better representations of idioms and their contexts will have increased significance with the rise of automated translation.

My project benchmarks various techniques for this task: testing different embedding models and different architectures for this problem. I apply this directly to the text captions, and then work on incorporating the image for additional context. I also explore if any topics or subset of idioms are most highly correlated with misinterpretation. These techniques can also be helpful in other relevant contexts, such as verifying model hallucinations and fact-checking.

Dataset: https://semeval2025-task1.github.io/

Accessing Applications Remotely in the Mobility Era: The Role of Zero Trust Network Access

Andrea Ibiassi, University of Oxford

The shift to remote work has transformed how employees access corporate applications. Organisations are no longer confined to hosting applications in centralized data centres. Instead, they increasingly rely on cloud platforms like AWS and Azure. As a result, the internet has effectively become the new corporate network, challenging traditional security models.

Virtual Private Networks (VPNs) were originally designed to extend corporate networks securely, but they struggle to meet modern security demands. A compromised device with VPN access can jeopardize the en6re corporate network, making VPNs an inadequate solution for today's cybersecurity challenges. Notably, 28.6% of ransomware aSacks worldwide leveraged VPNs as the ini6al access vector especially during the surge in remote work during COVID-19.

Unlike the traditional "castle-and-moat" security model, where everyone inside the moat is implicitly trusted, Zero Trust Network Access (ZTNA) shi's from network-based access to iden6ty- and context-aware application access. ZTNA enforces strict authen6ca6on methods such as Single Sign-On (SSO) and continuous user and device verification. By minimizing implicit trust and restricting access on a per-session basis, this approach significantly reduces the risk of lateral movement and data breaches.

This presentation will explore the evolution of remote access security, highlighting the shortcomings of traditional VPNs and demonstrating how ZTNA provides a more secure, scalable, and flexible approach to accessing corporate applications in an era of mobility and cloud-first strategies.

The impact of England's CAZ, ULEZ, and COVID-19 lockdowns on cardiovascular and respiratory chronic diseases during the COVID-19 pandemic

Millie Zhou, University of Cambridge
Elena Raffetti, Alexia Sampri, Angela Wood, Michael Inouye

Air pollution is the world's leading environmental health risk and trafficked air pollution is associated with adverse health outcomes across the life span. Chronic diseases are positively associated with infectious diseases, such as coronavirus disease (COVID-19), and a COVID-19 infection may also exacerbate pre-existing chronic diseases. Studies have hypothesised that air pollutants may contribute to the severity of COVID-19. Various studies have reported a direct relationship between the spread and contagion of viruses with the mobility of air pollutants and atmospheric levels. This study exploits a quasi-experiment and leverages the fact that traffic policies and COVID-19 lockdowns were not introduced in regions across England simultaneously.

The primary objective aims to assess the effects of air quality interventions (traffic-related policies and COVID-19 lockdowns) on the rates of hospitalisation due to chronic diseases across regions in England. The secondary objective will investigate potential effect modifications due to COVID-19, demographic attributes, or pre-existing conditions. An exploratory analysis will examine potential clustering patterns of chronic diseases.

This study will utilise the CVD-COVID-19 dataset, encompassing multiple NHS electronic health records sources and representing approximately 57 million adults in England. The spatial and temporal variation in the introduction of interventions on hospitalisation rates due to chronic diseases will be modelled using a controlled interrupted time series design. The impact of time-varying air quality interventions on hospitalisation rates due to chronic diseases will be estimated using a segmented regression model; stratified analyses will be conducted to assess effect modifiers. Potential clustering patterns of multiple long-term conditions will be investigated using latent class analysis and graph algorithms.

This study contributes original evidence by quantifying the health benefits of reduced traffic emissions on public health. It specifically highlights how the decline in air pollutants across England, attributed to traffic-related and COVID-19 policies, has affected hospitalisation rates for chronic diseases.

Full Reference Image Quality Assessment for Medical Images: Pitfalls, Analyses and Generalization

Anna Breger, University of Cambridge

Image quality assessment (IQA) is standard practice in the development stage of novel machine learning algorithms that operate on images to evaluate the quality of the outputs. Commonly used IQA measures have been primarily developed and optimised for natural images. Therefore, reported inconsistencies arising when applied for medical images are not surprising as they have different properties than natural images. Here, I will first systematically discuss diverse pitfalls observed by medical experts when applying standard full-reference IQA measures to medical images. Next, the applicability of common IQA measures for medical image data is tested by comparing their assessment to novel gathered and manually rated chest X-ray (5 experts) and photoacoustic image data (2 experts). In order to foster the lack of available data in this domain, the data sets have been made available. In the experiments, the IQA measure HaarPSI (based on Haar wavelets) shows exceptional behaviour in terms of generalisability for natural and medical images. Due to the promising results, we studied the measure more closely, specifically the parameters of the framework that have originally been aligned for natural images. We observe that in the medical seRng the specific parameter choices in the IQA measure have a bigger impact on the performance than for the natural images. When optimising the two parameters in HaarPSI towards the two medical image data sets the performance, measured with Spearman/Kendall rank correlation, increases significantly for the optimised as well as independent medical data. Those results suggest that adapting HaarPSI for medical images can provide a valuable, generalizable addition to the employment of more specific task-based measures.

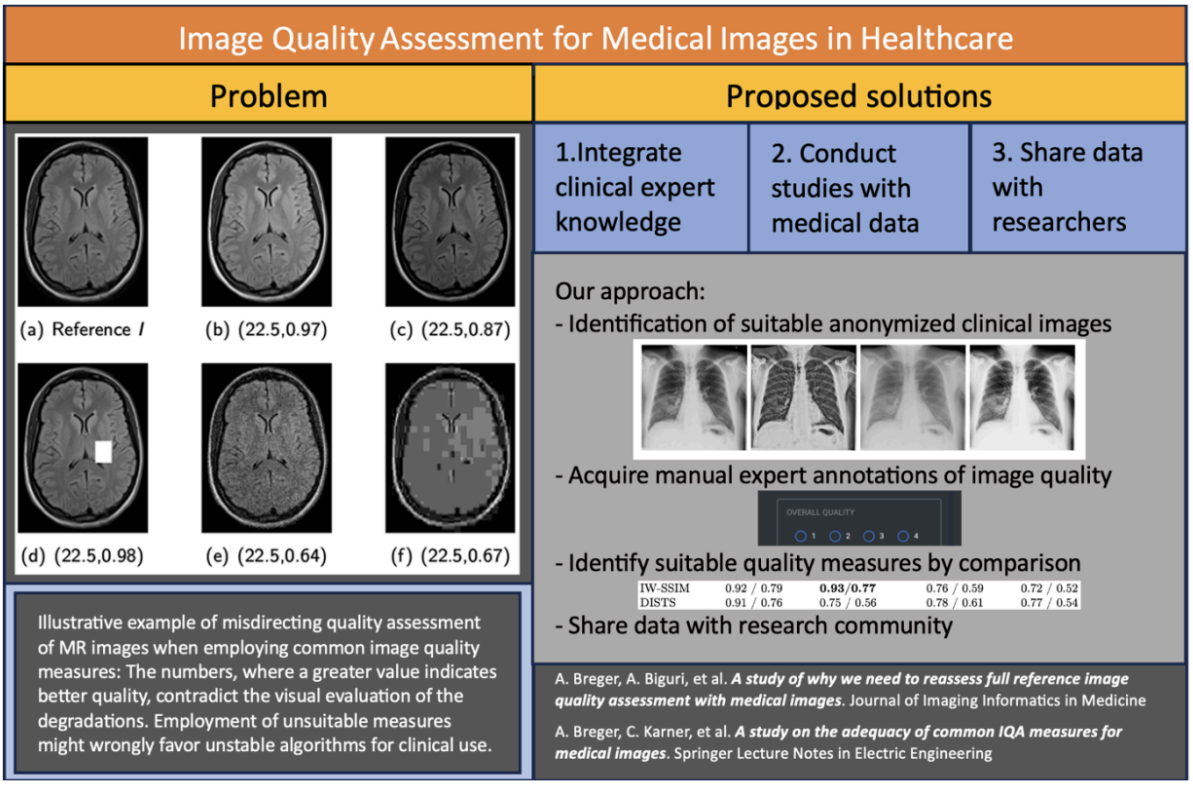


Image Quality Assessment for Medical Images in Healthcare
Problem
Proposed solutions
(a) Reference I
(b) (22.5,0.97)
(c) (22.5,0.87)
(d) (22.5,0.98)
(e) (22.5,0.64)
(f) (22.5,0.67)
Illustrative example of misdirecting quality assessment of MR images when employing common image quality measures: The numbers, where a greater value indicates better quality, contradict the visual evaluation of the degradations. Employment of unsuitable measures might wrongly favor unstable algorithms for clinical use.
1.Integrate clinical expert knowledge
2. Conduct studies with medical data
3. Share data with researchers
Our approach:
- Identification of suitable anonymized clinical images
- Acquire manual expert annotations of image quality
OVERALL QUALITY
1 2 3 4
- Identify suitable quality measures by comparison
IW-SSIM 0.92 / 0.79 0.93/0.77 0.76 / 0.59 0.72 / 0.52
DISTS 0.91 / 0.76 0.75 / 0.56 0.78 / 0.61 0.77 / 0.54
- Share data with research community
A. Breger, A. Biguri, et al. A study of why we need to reassess full reference image quality assessment with medical images. Journal of Imaging Informatics in Medicine
A. Breger, C. Karner, et al. A study on the adequacy of common IQA measures for medical images. Springer Lecture Notes in Electric Engineering

Figure 1.

Enhanced Synthetic MRI Generation from CT Scans Using CycleGAN with Feature Extraction

Lachin Naghashyar, Department of Computer Science Sharif University of Technology, Tehran, Iran
Saba Nikbakhsh, Department of Electrical Engineering, Urmia University of Technology, Urmia, Iran

In the field of radiotherapy, accurate imaging and image registration are of utmost importance for precise treatment planning. Magnetic Resonance Imaging (MRI) offers detailed imaging without being invasive and excels in soft-tissue contrast, making it a preferred modality for radiotherapy planning. However, the high cost of MRI, longer acquisition time, and certain health considerations for patients pose challenges. Conversely, Computed Tomography (CT) scans offer a quicker and less expensive imaging solution. To bridge these modalities and address multimodal alignment challenges, we introduce an approach for enhanced monomodal registration using synthetic MRI images. Utilizing unpaired data, this paper proposes a novel method to produce these synthetic MRI images from CT scans, leveraging CycleGANs and feature extractors. By building upon the foundational work on Cycle-Consistent Adversarial Networks and incorporating advancements from related literature, our methodology shows promising results, outperforming several state-of-the-art methods. The efficacy of our approach is validated by multiple comparison metrics.

Solving Large Deterministic POMDPs with Finite State Controllers

Alex Schutz, University of Oxford

How can robots make plans in a world full of uncertainty? Deterministic Partially Observable Markov Decision Processes (DetPOMDPs) are used to model decision-making problems in which the agent is able to act and observe deterministically, but the true state of the environment is not known. We use the motivating problem of robot navigation in a forest, where a path may not be discovered to be impassable until the robot attempts to traverse the route (Figure 1). Each uncertain path in the environment doubles the number of possible environments that the robot must plan for. Like the robot navigation example, many situated AI problems involve independent uncertainties which create exponential growth in the size of the state space. Existing offline solution approaches for DetPOMDPs typically involve generating tree-based policies which branch with each new action-observation history, but this approach does not scale to problem sizes encountered in real-world planning situations.

Our key innovation is adapting a continuous-state POMDP solver for the discrete-state DetPOMDP planning domain to solve problems with very large state spaces. We present DetMCVI, an adaptation of the Monte Carlo Value Iteration (MCVI) algorithm designed specifically for goal-oriented DetPOMDPs. DetMCVI constructs policies in the form of finite-state controllers (FSCs), enabling effective decision-making in large-scale scenarios. Our results demonstrate that DetMCVI achieves high success rates and significantly outperforms existing baselines. We validate the algorithm in a real-world mobile robot forest mapping task. This work advances planning under uncertainty and opens new possibilities for deploying DetPOMDP solvers in robotics and situated AI domains.

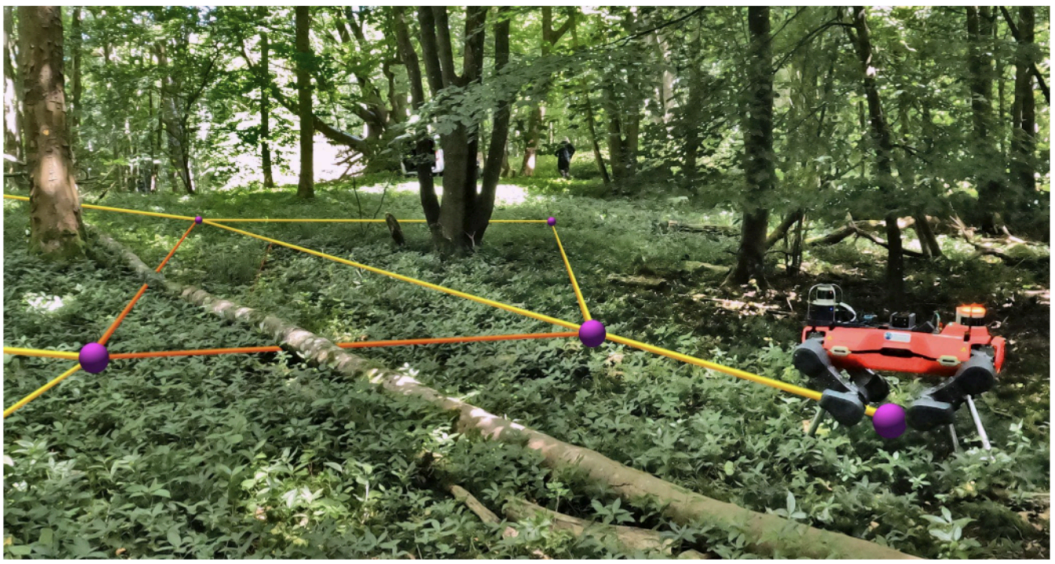


Figure 1: A topological map used for navigation in a forest where possibly obscured terrain leads to uncertain traversability.

Tau Cell Detection using Object Detection (OD) Models

Peiyan (Gracie) Zhou, University of Cambridge

Abnormal accumulation of tau proteins in the brain is a hallmark of neurodegenerative diseases such as Alzheimer's and Progressive Supranuclear Palsy (PSP). Current post-mortem assessments rely on manual evaluation, where pathologists visually identify tau lesion types to understand the disease progression. This process is labour-intensive and prone to variability. Existing machine learning models in this field commonly involve using a general segmentation model to select potential objects, followed by separate models (typically random forests) for classification. Deep learning-based object detection models offer an end-to-end solution by combining detection and classification into a single stage.

This project involves tuning and modifying open-source YOLO and Faster R-CNN to automate tau lesion identification in a dataset of 16 postmortem brain slides containing 24,642 tau objects from PSP patients. The OD models generate bounding boxes for tau objects and classify them into one of four lesion types, achieving a best mAP@50 score of 0.702 on the held-out test set.

To improve reliability, we evaluate detection uncertainty using Probabilistic Detection Quality (PDQ), which quantifies both label and spatial uncertainty. The uncertainty measures provide pathologists with a more comprehensive assessment of model performance when interpreted alongside traditional evaluation metrics like mAP. Label uncertainty reflects classification confidence, while spatial uncertainty concerns the precision of detected object location. We incorporate spatial uncertainty by integrating dropout layers and applying Monte Carlo Dropout at inference time, treating detected coordinates as probabilistic distributions. YOLO with a dropout rate of 0.5 achieved a PDQ of 3.4%, outperforming the vanilla YOLO model's PDQ of 1.9%.

Our custom OD models are integrated into QuPath, a widely used pathology imaging tool, as a plugin that supports tiling, detection, and merging detections across tiles. Developed using the Deep Java Library, the plugin enables seamless integration of automated tau detection into pathologists' workflows.

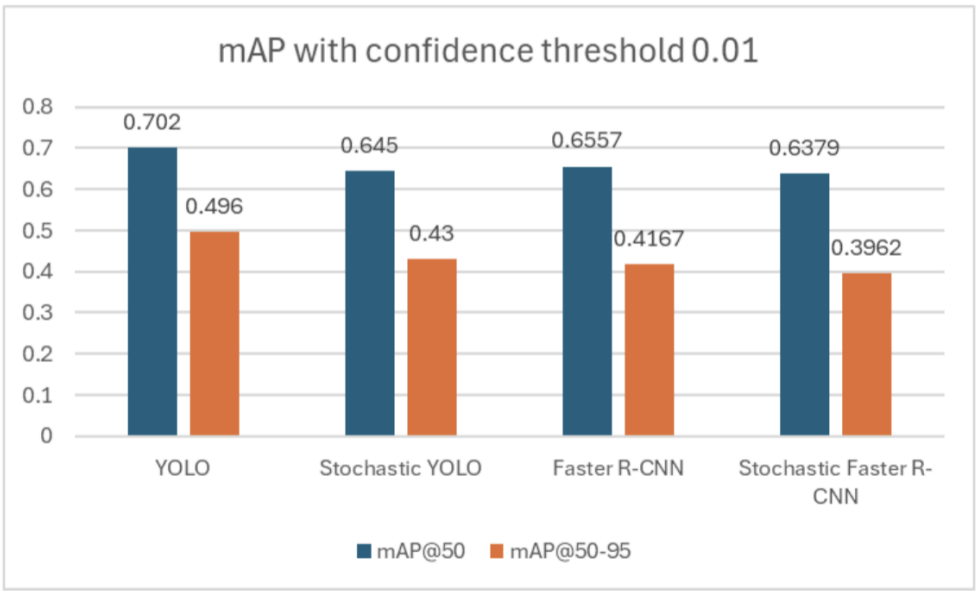


Figure 1: mAP evaluation of different models with and without dropout layers, using confidence threshold of 0.01 for Non-Maximum Suppression.

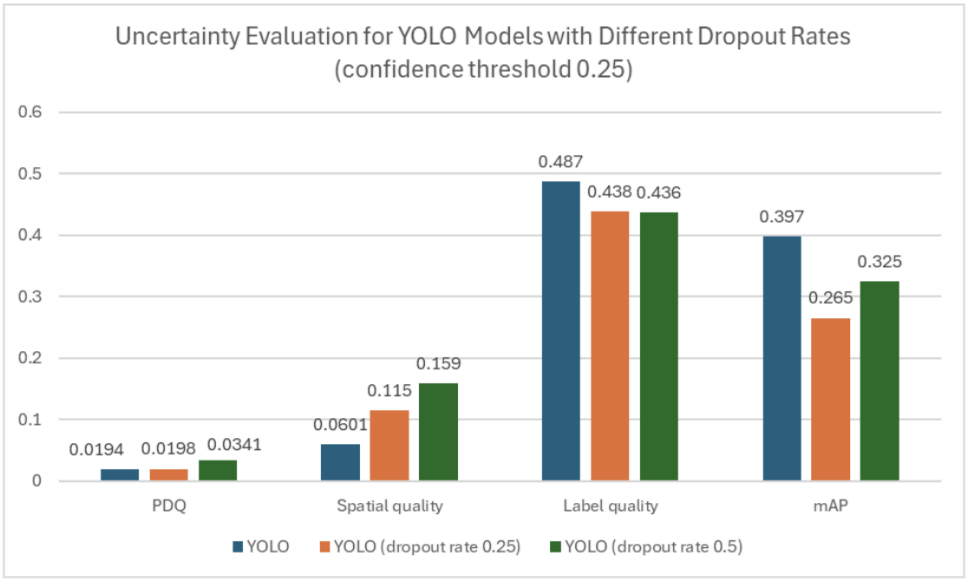


Figure 2: PDQ, spatial and label quality evaluation for YOLO models with different dropout rates.

Data Integrity Detection in Wearable IoT Devices for Cardiac Monitoring

Alexandra Loh, University of Cambridge

This abstract describes a joint project I carried out with a partner and a supervisor. Cardiovascular diseases (CVDs) are the leading cause of death globally. In 2021, CVDs accounted for almost a third of deaths worldwide. Thus, the constant monitoring of cardiac health has become more important in today's world. Fortunately, the recent advent of IoT devices has made it more practical and aU'ordable to monitor cardiac health in real-time. These devices typically use the electrocardiogram (ECG), a non-invasive method to measure the heart's electrical activity and detect abnormalities. However, for analysis and storage of this ECG data, it must be transmitted from the user's IoT to the cloud. This process is energy-consuming, which is at a premium for battery-operated devices.

In this project, we design and implement an energy-eU'icient monitoring system that determines the usability of incoming ECG data, filtering out data corrupted by motion artefacts and other noises prior to transmission. This reduces the size of transmitted data, thereby minimising energy wastage. Based on the training dataset obtained from Physionet/CinC 2017 challenge, our method achieved a 90% accuracy in classifying ECG signals into usable or corrupted data and is expected to filter out 95% of corrupted data when employed on ECG wearable devices. Our monitoring system is largely successful in detecting and removing noise generated during real-time ECG monitoring, optimising the energy and storage on IoT wearable devices.

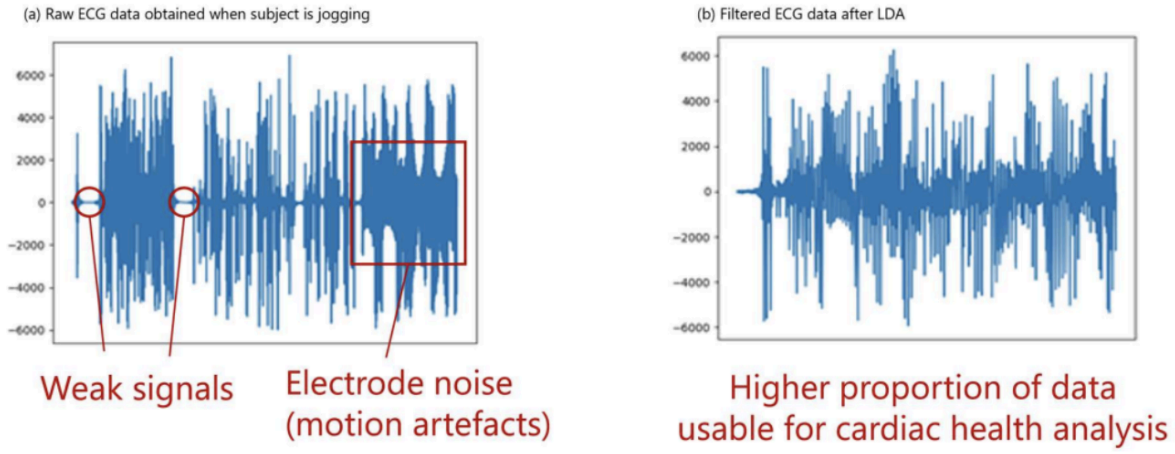


Figure (a) shows raw ECG data obtained when the subject is jogging and (b) shows that same data after being filtered by our method.

Sheaf-Based Diffusion for Structured Modality Fusion in Multimodal Graph Learning

Mar Gonzàlez i Català, University of Cambridge

Multimodal graph learning requires integrating diverse modalities while preserving their structural relationships. Existing methods typically perform modality-wise processing followed by late-stage fusion, failing to exploit cross-modal dependencies during learning. Inspired by recent advances in cellular sheaf theory, we propose sheaf-based diffusion as a structured approach to node-specific modality fusion in multimodal graphs, leveraging its ability to model structured dependencies.

We investigate whether modeling multimodal data as cellular sheaves—assigning vector spaces to nodes and edges with structured restriction maps—improves node classification by ensuring adaptive, modality-aware information propagation.

Standard graph neural networks store node information as a single feature vector, which presents a challenge: we either have to process each modality with a separate neural network—losing cross-modality interactions—or merge all modalities into one feature vector, sacrificing their individuality. In contrast, sheaf neural networks assign each node a vector space, known as a stalk, where different modalities can be stored in separate dimensions. Communication between these vector spaces happens through restriction maps, which facilitate information exchange between stalks. This approach preserves the distinct characteristics of each modality while still enabling interaction between them.

Our approach is evaluated against graph convolutional networks (GCNs), graph attention networks (GATs), and classical graph-based baselines on synthetic multimodal datasets where classification requires non-trivial information fusion across modalities. Our model shows improved performance in scenarios where each modality provides only partial information, and meaningful patterns emerge only through structured modality interactions. Sheaf-based diffusion provides a flexible, expressive framework for multimodal graph learning, offering a viable alternative to traditional graph-based approaches. Future work will explore its applications to real-world multimodal datasets.

SPA: Efficient User-Preference Alignment against Uncertainty in Medical Image Segmentation

Jiayuan Zhu, University of Oxford

Medical image segmentation data inherently contain uncertainty, often stemming from both imperfect image quality and variability in labeling preferences on ambiguous pixels, which depend on annotators' expertise and the clinical context of the annotations. For instance, a boundary pixel might be labeled as tumor in diagnosis to avoid under-assessment of severity, but as normal tissue in radiotherapy to prevent damage to sensitive structures. As segmentation preferences vary across downstream applications, it is often desirable for an image segmentation model to ofer user-adaptable predictions rather than a fixed output. While prior uncertainty-aware and interactive methods ofer adaptability, they are inefficient at test time: uncertainty-aware models require users to choose from numerous similar outputs, while interactive models demand significant user input through click or box prompts to refine segmentation. To address these challenges, we propose SPA, a segmentation framework that efficiently adapts to diverse test time preferences with minimal human interaction. By presenting users a select few, distinct segmentation candidates that best capture uncertainties, it reduces clinician workload in reaching the preferred segmentation. To accommodate user preference, we introduce a probabilistic mechanism that leverages user feedback to adapt model's segmentation preference. The proposed framework is evaluated on a diverse range of medical image segmentation tasks: color fundus images, CT, and MRI. It demonstrates 1) a significant reduction in clinician time and effort compared with existing interactive segmentation approaches, 2) strong adaptability based on human feedback, and 3) state-of-the-art image segmentation performance across diverse modalities and semantic labels.

The code is released at: https://github.com/SuperMedIntel/SPA

Aegis: A Programming Language that enforces Information Flow Security with Types

Komal Rathi, University of Cambridge

Data confidentiality and preventing unintended information leakage are critical challenges in software systems. Inspired by Francois Pottier and Vincent Simonet's Information Flow Inference in ML, this project develops Aegis, a new programming language with a functional style, which integrates built-in information flow security directly into the type system, with a focus on both static enforcement and manual annotations.

Aegis aims to enforce confidentiality by ensuring that high-security data can not influence low-security outputs. The language provides static enforcement of information flow using a type system that tracks security levels, preventing data leaks and ensuring that information flow complies with confidentiality requirements.

Implemented in OCaml, Aegis uses a static type system with a Program Context (pc) to track execution security levels. Aegis introduces a lattice structure for security levels, ensuring data can only flow upward and never downward. Formal typing rules, such as a no write-down policy, high watermark propagation, and explicit transitions (classify/declassify), enforce security constraints, while manual annotations allow developers to specify security levels.

Testing involved writing example programs and running unit tests to confirm the type system prevents unauthorised data leakage. Alcotest and Expect tests validated the Abstract Syntax Tree (AST), while example programs ensured correct compiler functionality and error handling for both valid and invalid cases.

Aegis provides a secure and practical solution for information flow security by integrating flow control directly into the type system, eliminating the need for external verification tools or heavy runtime enforcement. With support for manual annotations, Aegis enables developers to create more secure applications. Its functional programming style sets it apart from Java-based approaches like Jif, offering a fresh perspective on secure software development. Future work will focus on adding support for exceptions and monitoring their effects to further prevent information leakage.

Evaluating Gender Fairness of ML Algorithms for Pain Detection

Yuting Shang, University of Cambridge

Automated pain detection through machine learning (ML) and deep learning (DL) algorithms holds significant potential in healthcare, particularly for patients unable to self-report pain levels. However, the accuracy and fairness of these algorithms across different demographic groups (e.g., gender) remain under-researched. This paper investigates the gender fairness of ML and DL models trained on the UNBC-McMaster Shoulder Pain Expression Archive Database, evaluating the performance of various models in detecting pain based solely on the visual modality of participants' facial expressions.

We compare traditional ML algorithms, Linear SVM (L SVM) and Radial Basis Function SVM (RBF SVM), with DL methods, Convolutional Neural Network (CNN) and Vision Transformer (ViT). Data augmentation, hyperparameter tuning and training were performed, and evaluation was done using both performance and fairness metrics. While ViT achieved the highest accuracy and a selection of fairness metrics, all models exhibited gender-based biases. These findings highlight the persistent trade-off between accuracy and fairness, emphasising the need for fairness-aware techniques to mitigate biases in automated healthcare systems.

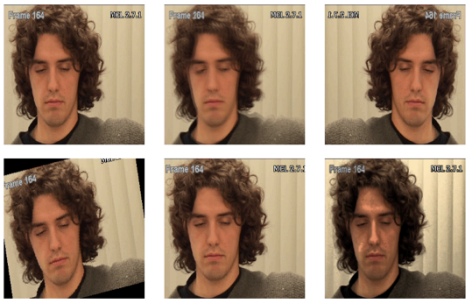


Figure 1: From left to right, the images represent: (1) Original, (2) SMOTE, (3) Flipped, (4) Rotated, (5) Cropped, and (6) Histogram Equalisation versions. All images are resized to squares for consistency in this diagram.

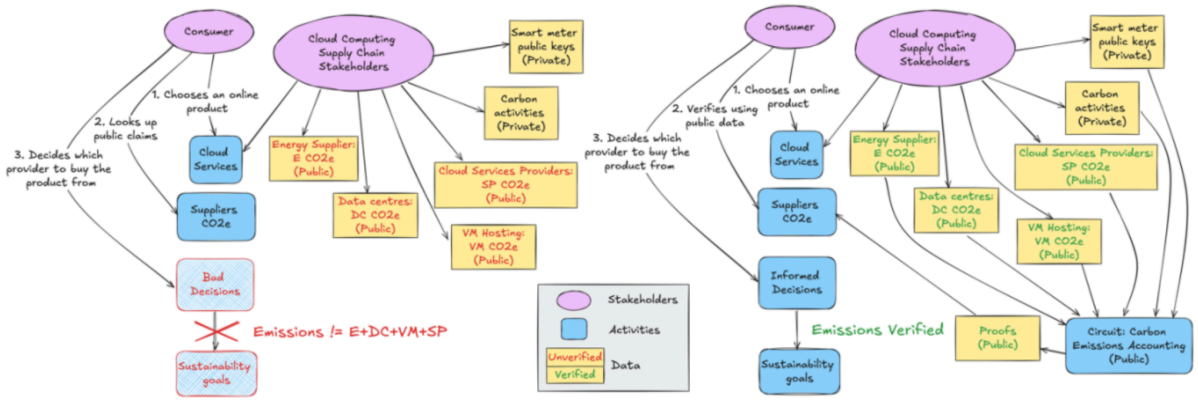
Metric	CNN	ViT	L SVM	RBF SVM
Overall Acc ↑	0.9637	**0.9806**	0.9509	0.9712
F1 Score ↑	0.9639	**0.9808**	0.9514	0.9707
ROC AUC ↑	0.9902	**0.9969**	0.9234	0.9331
Equal Acc	0.0098	0.0039	0.0051	**0.0027**
Equal Opp	0.0801	**0.0268**	0.0763	0.0908
Equal Odds	0.0488	**0.0169**	0.0423	0.0472
Dis Impact ↑	0.4650	0.4987	**0.5099**	0.4572
Demo Parity	0.1300	0.1208	**0.1175**	0.1180
Tr Equality	0.5564	0.1406	**0.0030**	0.7825
Test Fairness	**0.0118**	0.0161	0.0512	0.0142

Note: This was not a solo project, I worked on it with 1 other person for the Affective AI course in Part II

Emission Impossible:
Verifiable Carbon Emissions Reporting Without Revealing Private Data

Jessica Man, University of Cambridge

To achieve sustainability goals, it is important that organisations can track and share trustworthy carbon emissions data with their customers, investors and the authorities. However, businesses have strong incentives to make their numbers look good, while at the same time less so to publish their accounting methods along with all the input data, due to the risk of revealing sensitive information. As a result, carbon emissions reporting in supply chains can only rely on unverified and therefore untrusted data. For my research I am proposing a methodology that applies cryptography and zero-knowledge proof for carbon emissions claims that can be subsequently verified without the knowledge of the private input data. As a practical use case, I am testing the concept on cloud computing, with an electricity supplier, a data centre operator and a cloud services customer in a chain to produce trustworthy carbon emission data for the customer. In this presentation I will explore the conflicting incentives around carbon emissions reporting and why existing systems are not adequate, and how applied cryptography can be used to allow for more accurate claims whilst protecting business sensitive data.



Instruction Fusion Limit Study for RISC-V

Elizabeth Ho, University of Cambridge
Jonathan Woodruff

Instruction Fusion

Instruction fusion, otherwise known as macro-op fusion, is common practice in modern x86 and Arm processors. They look for opportunities to merge multiple assembly instructions into a single fused instruction within the instruction pipeline. Instruction fusion is a good way to allow gains in performance while maintaining the simplicity and flexibility of the ISA. Microarchitects can choose to fuse certain combinations of instructions into one for their specific application, giving the performance benefits of treating the work as a single instruction, while not bloating the ISA with unnecessary instructions for other applications [1].

RISC-V is in a unique position to exploit instruction fusion. Its core ISA is much simpler than alternatives like ARM and x86, but its wide range of applications call for a method to merge complex functionality into a single instruction. To show this, we perform a limit study of instruction fusion opportunities on the SPECInt2006 benchmark suite.

Fusion Results

As seen in Figure 1 below, a fusion rate of over 50% can be achieved in the limit, meaning that the resulting effective instruction count is less than half of the original number of instructions. Fusing pairs alone gives close to 30% fusion rate, suggesting that instruction fusion might be able to provide a big benefit without adding substantial amounts of complexity.

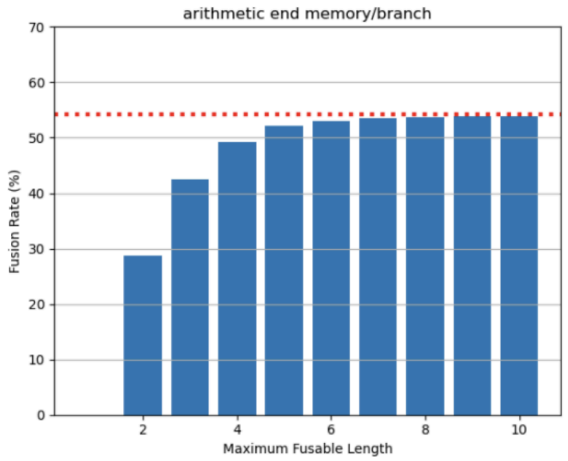


Figure 1: Fusion rate versus maximum fusable length

Conclusion

Instruction fusion is a promising technique that allows microarchitects to capture the improved performance of more complex instructions without bloating the ISA, and should be further utilised in RISC-V microarchitectures.

References

[1] Christopher Celio, Palmer Dabbelt, David Patterson, and Krste Asanovic. The Renewed Case for the Reduced Instruction Set Computer: Avoiding ISA Bloat with Macro-Op Fusion for RISC-V, 2016.

Profiling neural grammar induction on morphemically tokenised child-directed speech

Mila Marcheva, University of Cambridge

Virtually all humans acquire language, but first language acquisition (FLA) remains a central yet unresolved challenge in linguistics. Computational modelling offers a powerful approach to studying FLA, allowing to test precise hypotheses in a simulation environment without the real-world constraints. The computational task of grammar induction (GI) provides a lower bound on the types of grammatical structures that can be inferred from raw text alone. Re cent advances in GI (Kim, Dyer, & Rush, 2019) necessitate a reevaluation of this lower bound in the context of FLA and this project begins to bridge the gap between the State of the Art (SotA) in GI and FLA. Functional morphemes are a key focus of SotA generative research due to their role in shaping the overall structure of language (Dye, Kedar, & Lust, 2018; Biberauer, 2019). Thus, to provide a more cognitively realistic setup I propose a modification to the GI input: I morphemically tokenize (see Figure 1) the CHILDES treebank (Pearl & Sprouse, 2013).

Figure 1: (R) Example of morphemic tokenization and the necessary modifications to the original phrase-structure tree (L).

I train three types of SotA neural GI models—C-PCFG (Kim et al., 2019), N PCFG (Zhu, Bisk, & Neubig, 2020), and TN-PCFG (Yang, Zhao, & Tu, 2021)— on regularly and morphemically tokenized child-directed speech. I evaluate them using F1 measures. Additionally, I introduce depth-of-morpheme and sibling-of morpheme metrics to assess the attachment of functional morphemes. I find that high F1 scores do not always correspond to linguistically meaningful structures for functional morpheme attachment (see Figure 2), highlighting a key challenge for cognitively realistic GI. Future research will explore augmenting neural GI with cognitive biases (Culbertson, Smolensky, & Wilson, 2013) and applying the above framework on morphologically rich languages.

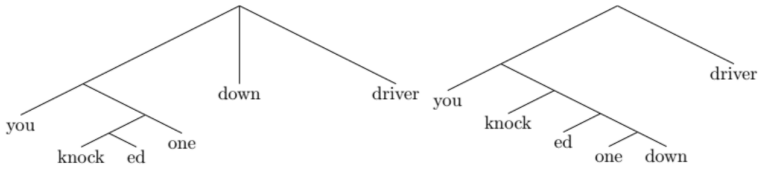


Figure 2: Predicted trees for the sentence “You knocked one down, driver.”; TN-PCGF (L) and N-PCFG (R). TN-PCFG (F1 = 73.81) attaches the functional morpheme -ed to knock which is linguistically plausible, whereas N-PCFG (F1 = 78.56) combines -ed with the unlikely constituent one down instead of with the verb.

References

[1] Biberauer, T. (2019, October). Children always go beyond the input: The maximise minimal means perspective. Theoretical Linguistics, 45 (3–4), 211–224. Retrieved from http://dx.doi.org/10.1515/tl-2019-0013

[2] Culbertson, J., Smolensky, P., & Wilson, C. (2013, May). Cognitive biases, linguistic universals, and constraint-based grammar learning. Topics in Cognitive Science, 5 (3), 392–424. Retrieved from https://doi.org/10.1111/tops.12027

[3] Dye, C., Kedar, Y., & Lust, B. (2018, December). From lexical to functional categories: New foundations for the study of language development. First Language, 39 (1), 9–32. Retrieved from http://dx.doi.org/10.1177/0142723718809175

[4] Kim, Y., Dyer, C., & Rush, A. (2019, July). Compound probabilistic context free grammars for grammar induction. In A. Korhonen, D. Traum, & L. Marquez (Eds.), Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 2369–2385). Florence, Italy: Association for Computational Linguistics. Retrieved from https://aclanthology.org/P19-1228

[5] Pearl, L. S., & Sprouse, J. (2013). Syntactic islands and learning biases: Combining experimental syntax and computational modeling to investigate the language acquisition problem. Language Acquisition. Retrieved from https://ling.auf.net/lingbuzz/001493

[6] Yang, S., Zhao, Y., & Tu, K. (2021, August). Neural bi-lexicalized PCFG induction. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers) (pp. 2688–2699). Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2021.acl-long.209

[7] Zhu, H., Bisk, Y., & Neubig, G. (2020). The return of lexical dependencies: Neural lexicalized PCFGs. Transactions of the Association for Computational Linguistics, 8 , 647–661. Retrieved from https://aclanthology.org/2020.tacl-1.42

New Tolerance Factor for Molecular Perovskites

Lauren Wilkes, University of Cambridge

Perovskites are a class of crystalline materials that have shown immense potential for solar energy applications, including efficient photovoltaic cells. Perovskite-based photovoltaic cells are advantageous due to their low cost, simple manufacturing process, and increased power conversion efficiency compared to lead-based cells. However, perovskites are prone to faster degradation, motivating scientists to design new perovskites that mitigate these difficulties.

Designing new perovskites is challenging because the conventional experimental approaches are time and resource-intensive. Perovskites are combinatorial materials formed from 3 molecular components (A,B, and x-sites), and there are many chemical composition possibilities, resulting in a massive chemical search space (> 1060). One well-established metric to predict the formation of a perovskite given a chemical composition is the tolerance factor; this is a geometrically-derived metric relying on discrete size measurements of the different molecular components of the perovskite. While powerful, this metric fails to generalize to more complex compositions. We recently improved the performance of this metric by reformulating it to incorporate a shape analysis method.

In this work, we continue to refine the tolerance factor. To incorporate information on newly available electrostatic properties, we explore the impact of two additional variables: the HOMO-lumo gap and dipole moment.

We develop a new tolerance factor formulation that incorporates a weighted additive dipole moment parameter while maintaining physical intuition. We find that as the A-site cation complexity increases, the dipole moment becomes more important and results in a better model fit.

This work shows that newly available electrostatic information shows promise in improving perovskite prediction in more complex perovskite structures. In future work, we plan to leverage the potential of electrostatic information to test different tolerance factor formulations to develop more reliable perovskite prediction methods and facilitate faster perovskite discovery.

	Original Weighting	Additive Dipole Weighting
Silva's structures	1541.85	**1443.54**
Sebastien's structures	**486.63**	486.64
Chem. Sci. structures	4.05	4.05
J. Solid State Chem. structures	**1083.49**	1083.50

Table 1: AIC Results for Tolerance Factor Formulations

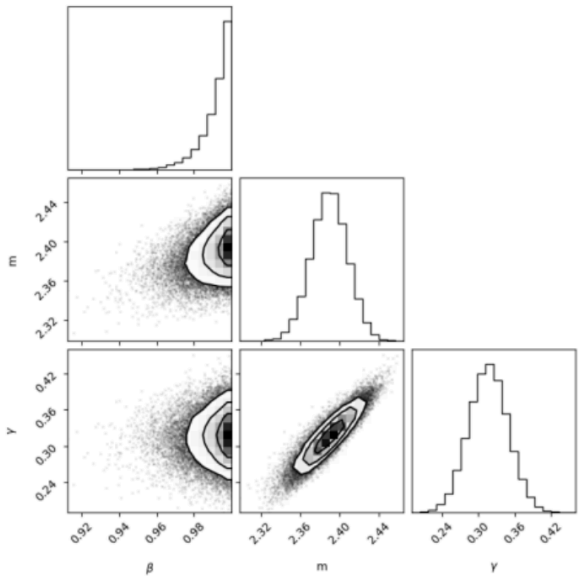


Figure 1: MCMC corner plot for Silva’s Structures

Enhancing a Quantum Error Correction Code Simulator

Alison Dauris, University of Cambridge

Following recent advances in physical quantum computers, quantum error correction (QEC) codes are essential for managing noise. QEC involves encoding a single logical qubit across multiple physical qubits, enabling fault-tolerant quantum computation. Using stabilisers, errors can be detected and then corrected without collapsing the quantum state. Fast stabiliser simulators are crucial for analysing different QEC codes.

This work presents an application for visualising topological QEC codes. After defining a noise model, the QEC representation is converted into the input of a high-speed stabiliser simulator. Multiple simulation runs generate stabiliser outputs, which are processed by a decoder which determines which corrections are required. By comparing the corrections to the state of the data qubits simulated it can be determined if the errors would be successfully corrected.

The system features a JavaScript frontend and a Python backend, evaluated through a user study. A faster but less accurate Union-Find decoder was also implemented, in addition to using a library's Minimum Weight Perfect Matching decoder.

Results indicate that the interface simplifies the use of stabiliser simulators and that it could be used for surface codes designed to deal with chip irregularities. Furthermore, the interface provides a universal representation of surface codes, making key details of QEC codes more accessible. This tool increases understanding of QEC and enables exploration of different decoders which are based on different error models.

Non categorical models of typed lambda calculi

Jessica Richards, University of Oxford

The aim of this work is to give denotational semantics to certain programs in an understandable way. So, given some program calculus, we want a representation of each program as a function-like construct in some mathematical object, typically a category. When this translation is sound and complete it enables us to reason about programs and objects in terms of each other. For example, given any category with products and exponentials, the familiar equation AZ × BZ ~= (A × B)Z holds up to natural isomorphism. This corresponds to the observation that the programs \z -> (fst x z, snd x z) and (\z -> fst (x z),\ z -> snd (x z)) are mutually inverse in the presence of extensionality.

However translations into categories can be intimidating for non mathematicians. For example, in order to interpret the concept of a variable environment requires the use of categorical products and can mean terms are not modelled as themselves. Our approach aims to mitigate both these issues via the use of abstract clones. These naturally represent the variable and substitution features of simple type theories and enable translations to be largely syntactic. It has been shown by Saville [2] that the simply typed lambda calculus both with and without products can be translated into clones in this way.

We look firstly at how sum types fit into this picture. We define structure on a clone analogous to a categorical coproduct and show there is a direct sound and complete translation from the sum calculus into clones featuring this structure. Furthermore, when distributive products are added, this translation aligns with the standard categorical one. The current work is on the translation of effectful programs in the style of Moggi [1]. These can be written into a clone with corresponding structure, and in the case that the language has products it again aligns with the categorical interpretation, but it is less clear what structure this represents in isolation.

References

[1] Eugenio Moggi. Notions of computation and monads. Information and Computation, 93(1):55–92, 1991. Selections from 1989 IEEE Symposium on Logic in Computer Science.

[2] Philip Saville. Clones, closed categories, and combinatory logic. In Naoki Kobayashi and James Worrell, editors, Foundations of Software Science and Computation Structures, pages 160–181, Cham, 2024. Springer Nature Switzerland.

GeoFT: Fine-tuning Foundation Models for Automated OSINT Geolocation

Selena Sun, University of Oxford

Open source intelligence (OSINT) investigators face the challenge of verifying the location of media shared online. Traditional geolocation requires manual effort and cannot scale with the ever-growing volume of images and videos shared on social media. We present GeoFT, a fine tuned version of GeoCLIP specifically optimized for geolocation in Russia and Ukraine. By focusing on street-level imagery and leveraging community-validated datasets, our model achieves significantly improved accuracy compared to existing solutions. On our test set, GeoFT reduces the average error from 3,520km to 2,150km while maintaining interpretable confidence scores. We demonstrate the model's potential for aiding OSINT investigations and discuss pathways for deployment in real-world applications.

Harnessing Machine Learning for Heart Failure Management: Opportunities and Limitations

Kristy Poon, University of Cambridge

Heart Failure is a significant global medical concern affecting over 64 million people worldwide. Its prevalence is increasing due to the aging population and improved survival rates from other cardiovascular conditions. The economic burden of Heart Failure is very large, and is one of the costliest conditions to manage in high-income countries. Globally, HF incurs an estimated annual expenditure surpassing $346 billion, stemming from diagnostic tests, therapies, hospitalizations, and invasive procedures.

Among the emerging approaches for addressing the complexities of Heart Failure management, Machine Learning posits the possibility of disease classification, risk stratification and prognosis detection, offering potential advantages in identifying subtle patterns in patient data, improving diagnostic accuracy, and personalizing treatment in areas that traditional analytical methods were challenged by. AI-based clinical decision support systems have demonstrated high diagnostic accuracy, improving Heart Failure diagnosis in areas with limited access to specialists. One such AI-based is Markov Models which allow us to explore modelling disease progression, offering a probabilistic framework to estimate long-term outcomes and inform decision-making in heart failure care. However, Machine Learning inevitably comes with its limitations in clinical care. Issues such as data bias, interpretability, and integration into real-world practice pose challenges to its widespread adoption.

This review discusses the role of Machine Learning in heart failure management, exploring the feasibility of using different Machine Learning techniques to provide insight into improving diagnostic accuracy, intervention outcomes,and underscoring the importance of combining advanced analytics with traditional clinical expertise to tackle the growing HF burden effectively.

Machine Learning for Building-Level Heat Risk Mapping

Andrea Domiter, University of Cambridge

Climate change is intensifying the frequency and severity of heat waves, increasing risks to public health and energy systems worldwide. However, many existing heat vulnerability assessments focus primarily on outdoor temperatures, overlooking indoor conditions that directly a>ect occupants. Although building modelling can reveal the types of buildings whose occupants are most at risk, they do not pinpoint the locations of these vulnerable structures. In this paper, we present a data-driven workflow to identify where high-risk buildings are located. Specifically, we explore two labeling approaches - rule-based heuristics and multimodal large language models - to classify real-world buildings from satellite imagery and features and visualize heat risks using an interactive heat map. Our preliminary results illustrate how evaluating building data at the stock level can reveal critical heat vulnerabilities and emphasize the importance of automated building archetype identification for grid reliability and thermal equity.

Material discovery - High-temperature superconductor at ambient pressure

Maelie Causse, University of Cambridge

Computational methods have revolutionised material discovery, driving interest in designing new advanced materials, e.g. superconductors, which exhibit zero resistivity below a critical temperature (Tc).

Reaching pressures in the 100 GPa range enables the synthesis of superhydrides, a new class of hydrogen-rich materials that exhibit remarkable properties such as superconductivity, hydrogen diffusion, and hydrogen storage. A striking example is LaH_{10}, which holds the record for the highest superconducting transition temperature at Tc = 250 K. However, stabilising such compounds at ambient or low pressure remains a major challenge for practical applications. This has driven the search beyond binary hydrides toward ternary super hydrides, aiming to discover new superconductors with high critical temperatures that can persist near ambient pressure.

In this work, we explore the Y-Fe-H system using a combination of experimental synthesis and *ab initio* calculations. By compressing the Laves-phase compound YFe_2 under high hydrogen pressure in a diamond anvil cell, we synthesised a new ternary hydride, which, remarkably, was recovered in a metastable state at ambient pressure. Its unusual structure provides a promising model for the discovery of ternary superconducting hydrides stable at ambient conditions. This is the first instance of a high-pressure superhydride being successfully retained in a metastable state at ambient pressure.

The discovery of new superconducting hydrides traditionally relies on *ab initio* structure prediction and experimental synthesis. However, *ab initio* calculations become computationally prohibitive for complex systems such as ternary superhydrides. To overcome this limitation, our group has successfully developed machine learning-based high-throughput calculations, enabling a great acceleration in structure predictions.

Building on this approach, the next step is to apply machine learning-driven materials discovery to our newly synthesised metastable prototype. This will pave the way for identifying the first high temperature superconducting ternary hydride stable at ambient pressure, marking a crucial step toward practical applications of hydrogen-based superconductors.

The Invisible Handshake: State Pensions and Corporate Political Contributions

Jingshu Wen, University of Oxford
Jun Tu, Shanghai Jiao Tong University, Singapore Management University
Fan Zhang, Bentley University
Haofei Zhang, Ningbo University of Technology

We investigate whether U.S. politicians exchange favors with corporations by influencing state pension investments for campaign donations. Unlike revolving doors through which politicians obtain private sector jobs, politicians use their power over public resources to further their political careers in the quid pro quo. We find that portfolio firms of state pension funds donate more to state officials, who may influence the investment decisions of state pension funds. Exploiting state-level variation on soft money restrictions, our difference-in differences tests show that campaign finance freedom increases such political donations. Using textual analysis of campaign speeches and news articles, we find that politicians with lower level of integrity are more likely to solicit campaign contributions and to influence public pension funds. Despite politicians' alleged concern about public pension funding shortfalls, political influence hurts state pension performance. These effects are stronger for states with higher corruption levels, for firms with worse corporate governance, for donations to super PACs, and after the Supreme Court loosened soft money restrictions in 2010. Our results suggest that the recent campaign finance deregulation reduces social welfare by increasing corruption.

Instrumental variable simulation

Ruizi Yan, University of Oxford

We proposed a controllable data generating process for instrumental variable (IV) models that can directly target causal estimands (e.g. average treatment effect (ATE) and average treatment effect on the treated (ATT)). This provides a unified simulation framework for evaluating new IV-based estimators. One motivation is that most parametric IV simulation approaches rely heavily on assumptions of linearity and Gaussian distributions which restrict the ability to capture complex dependencies that is often present in real-world data. The proposed strategy under frugal framework (Evans and Didelez, 2024) overcomes these limitations by using pair copulas to link the ‘past’ and the causal distribution, allowing for more flexible dependency structures and supporting general parametric settings beyond traditional linear-Gaussian frameworks. Another advantage is that we have control on the causal margins, meaning that we know the actual causal estimands. A key add-on compared to the original frugal parametrization work (Evans and Didelez, 2024) is the ATT simulation. We show that the original idea of simulating the treatment through conditional distributions is not directly applicable to the ATT case. Instead, we specify a marginal distribution and a sequence of pair-copulas for the treatment. The proposed method allows simulated data to have any specific ATE or ATT, which offers an efficient way for estimator evaluation.

References

[1] R. J. Evans and V. Didelez. Parameterizing and simulating from causal models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 86(3):535–568, 2024.

Non-Automatability: Formal Barriers To Efficient Proof Search

Gaia Carenini, University of Cambridge

Automatability, in general, tackles the question: How difficult is it to find mathematical proofs? More formally, a proof system P is said to be automatable in time f (n) if there exists an algorithm that given as input an unsatisfiable formula F outputs a refutation of F in the proof system P in time f (n), where n is the size of the smallest P-refutation of F (plus the size of F). Automatability (or lack thereof) for well-studied proof systems is a central question for automated theorem proving and SAT solving.

Numerous findings in the late 1990s and early 2000s showed that several proof systems are not automatable under generally accepted complexity and cryptographic assumptions [BP96; Iwa97; KP98; BPR00; ABMP01; BDG+04; AB04; AR08; GL10]. While cryptographic assumptions are still the only ones under which we can prove the non-automatability of Frege systems (above AC0 -Frege) [ACG24], research in non-automatability of weaker proof systems recently gained new momentum [MPW19; AM20; GKMP20a; Gar20; Bel20; dRGN+21; dRez21; IR22; Pap24] mainly due to the breakthrough result of Atserias and Müller [AM20] who showed by means of an ingenious, and in hindsight natural, reduction that automating resolution is NP-hard. Despite significant advances in the area, several fundamental questions remain unresolved.

One of the frontiers of automatability lies in understanding tree-like proof systems [GKMP20b; dRez21] that embrace several natural algorithms that navigate the solution space through different branching heuristics. Automating tree-like proof systems requires an algorithm able to efficiently “guess” the suitable branching choices at each step. Such non-trivial automating algorithms are known to exist for tree-like resolution [BP96] and tree-like k-DNF resolution [AB02]. Since these proof systems can be automated in a non-trivial way, the results related to their non-automatability differ from those of weaker systems. Specifically, classical techniques for proving non-automatability often rely on achieving upper bounds that most tree-like proof systems cannot attain.

In this presentation, we will give a general overview of automatability, with a focus on tree-like k-DNF resolution, for which we essentially resolved the automatability issue under rETH in a joint work with Susanna F. de Rezende [CdR25].

References

[AB02] A. Atserias and M. L. Bonet, “On the automatizability of resolution and related propositional proof systems,” Electronic Colloquium on Computational Complexity (ECCC), Technical Report TR02-010, 2002. [Online]. Available: https://eccc.weizmann.ac.il/report/2002/010/

[AB04] A. Atserias and M. L. Bonet, “On the automatizability of resolution and related propositional proof systems,” Information and Computation, vol. 189, no. 2, 182–201, Mar. 2004, Preliminary version in CSL ’02. doi: 10.1016/j.ic.2003.10.004.

[ABMP01] M. Alekhnovich, S. Buss, S. Moran, and T. Pitassi, “Minimum propositional proof length is NP-hard to linearly approximate,” Journal of Symbolic Logic, vol. 66, no. 1, 171–191, 2001. doi: 10.2307/2694916.

[ACG24] N. Arteche, G. Carenini, and M. Gray, “Quantum automating TC0-Frege is LWE-hard,” in 39th Computational Complexity Conference (CCC 2024), ser. Leibniz International Proceedings in Informatics (LIPIcs), 2024, isbn: 978-3-95977-331-7. doi: 10.4230/LIPIcs.CCC.2024.15.

[AM20] A. Atserias and M. Müller, “Automating resolution is NP-hard,” Journal of the ACM, vol. 67, no. 5, 2020, Preliminary version in FOCS ’19. doi: 10.1145/3409472.

[AR08] M. Alekhnovich and A. A. Razborov, “Resolution is not automatizable unless W[P] is tractable,” SIAM Journal on Computing, vol. 38, no. 4, 1347–1363, 2008, Preliminary version in FOCS ’01. doi: 10.1137/06066850X.

[BDG+04] M. L. Bonet, C. Domingo, R. Gavaldà, A. Maciel, and T. Pitassi, “Non-automatizability of bounded-depth Frege proofs,” Computational Complexity, vol. 13, no. 1-2, 47–68, Dec. 2004, Preliminary version in CCC ’99. doi: 10.1007/s00037-004-0183-5.

[Bel20] Z. Bell, “Automating regular or ordered resolution is NP-hard,” Electronic Colloquium on Computational Complexity (ECCC), Tech. Rep. TR20-105, 2020. [Online]. Available: https://eccc.weizmann.ac.il/report/2020/105/.

[BP96] P. Beame and T. Pitassi, “Simplified and improved resolution lower bounds,” in Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science (FOCS ’96), Oct. 1996, 274–282. doi: 10.1109/SFCS.1996.548486.

[BPR00] M. L. Bonet, T. Pitassi, and R. Raz, “On interpolation and automatization for Frege systems,” SIAM Journal on Computing, vol. 29, no. 6, 1939–1967, 2000, Preliminary version in FOCS ’97. doi: 10.1137/S0097539798353230.

[CdR25] G. Carenini and S. F. de Rezende, “On the automatability of tree-like k-DNF Resolution,” Submitted, 2025.

[dRez21] S. F. de Rezende, “Automating tree-like resolution in time no (logn) is ETH-hard,” Electronic Colloquium on Computational Complexity (ECCC), Technical Report TR21-033, 2021. [Online]. Available: https://eccc.weizmann.ac.il/report/2021/033/.

[dRGN+21] S. F. de Rezende, M. Göös, J. Nordström, T. Pitassi, R. Robere, and D. Sokolov, “Automating algebraic proof systems is NP-hard,” in Proceedings of the 53rd Annual ACM

Symposium on Theory of Computing (STOC ’21), To appear, Jun. 2021. [Online]. Available: https://eccc.weizmann.ac.il/report/2020/064/

[Gar20] M. Garlík, “Failure of feasible disjunction property for k-DNF resolution and NP-hardness of automating it,” Electronic Colloquium on Computational Complexity (ECCC), Tech. Rep., 2020. [Online]. Available: https://eccc.weizmann.ac.il/report/2020/037/

[GKMP20a] M. Göös, S. Koroth, I. Mertz, and T. Pitassi, “Automating cutting planes is NP-hard,” in Proceedings of the 52nd Annual ACM Symposium on Theory of Computing (STOC ’20), Jun. 2020, 68–77. doi: 10.1145/3357713.3384248.

[GKMP20b] M. Göös, S. Koroth, I. Mertz, and T. Pitassi, “Automating cutting planes is NP-hard,” in Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing. 2020, pp. 68–77, isbn: 9781450369794. [Online]. Available: https://doi.org/10.1145/3357713.3384248.

[GL10] N. Galesi and M. Lauria, “On the automatizability of polynomial calculus,” Theory of Computing Systems, vol. 47, no. 2, 491–506, 2010. doi: 10.1007/s00224-009-9195-5.

[IR22] D. Itsykson and A. Riazanov, “Automating OBDD proofs is NP-hard,” in 47th International Symposium on Mathematical Foundations of Computer Science (MFCS 2022), S. Szeider, R. Ganian, and A. Silva, Eds., ser. Leibniz International Proceedings in Informatics (LIPIcs), vol. 241, Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022, 59:1–59:15, isbn: 978-3-95977-256-3. doi: 10.4230/LIPIcs.MFCS.2022.59.

[Iwa97] K. Iwama, “Complexity of finding short resolution proofs,” in Proceedings of the 22nd International Symposium on Mathematical Foundations of Computer Science (MFCS ’97), 1997, 309–318. doi: 10.1007/BFb0029974.

[KP98] J. Krajícek and P. Pudlák, “Some consequences of cryptographical conjectures for S12 and EF,” Information and Computation, vol. 140, no. 1, 82–94, 1998. doi: 10.1006/inco.1997.2674.

[MPW19] I. Mertz, T. Pitassi, and Y. Wei, “Short proofs are hard to find,” in Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP ’19), 2019, 84:1–84:16. doi: 10.4230/LIPIcs.ICALP.2019.84.

[Pap24] T. Papamakarios, “Depth-d frege systems are not automatable unless p = np,” in 39th Computational Complexity Conference (CCC 2024), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2024.