Use of AI to improve the early detection of upper GI cancers from routine blood tests

Victoria Moglia, Owen Johnson, Dr. Marc de Kamps, Prof. Gordon Cook, Dr. Lesley Smith

UNIVERSITY OF LEEDS

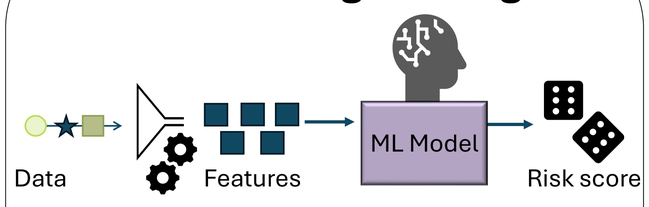
Engineering and Physical Sciences Research Council

Introduction

- Upper gastrointestinal (GI) cancers are characterised by late diagnosis and **poor survival rates**.
- Upper GI cancers are **difficult to diagnose**, largely due to late-presenting, vague symptoms.
- Changes in blood tests over time may show signs of cancer **earlier** than current diagnostic criteria.
- Models using longitudinal data for cancer prediction often fall under two approaches: feature engineering or sequential input. The two approaches are **rarely compared**[1].

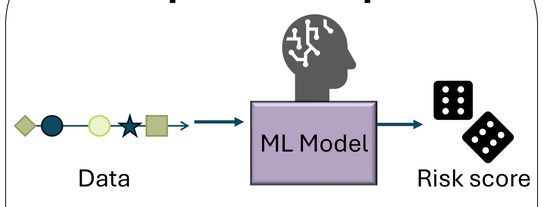
AIM: Develop predictive models to detect upper GI cancers from primary care blood tests, comparing feature engineering and sequential approaches

Feature engineering

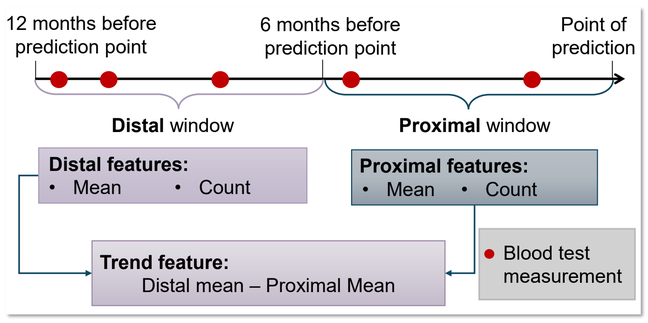


+ Interpretable, can be used alongside simpler classification algorithms
− Requires expert knowledge to craft features, may miss complex patterns

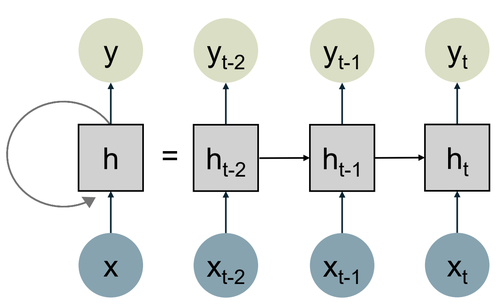
Sequential input



+ Can potentially capture more complex, novel patterns
− Computationally expensive, 'black box' models

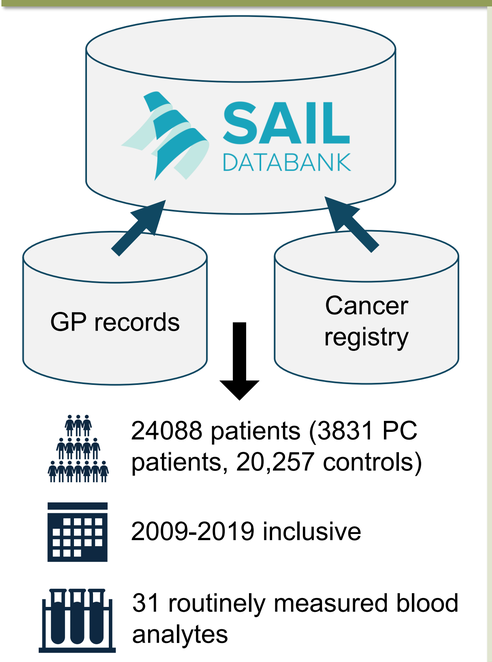


Feature engineering approach



Unfolded representation of an RNN model

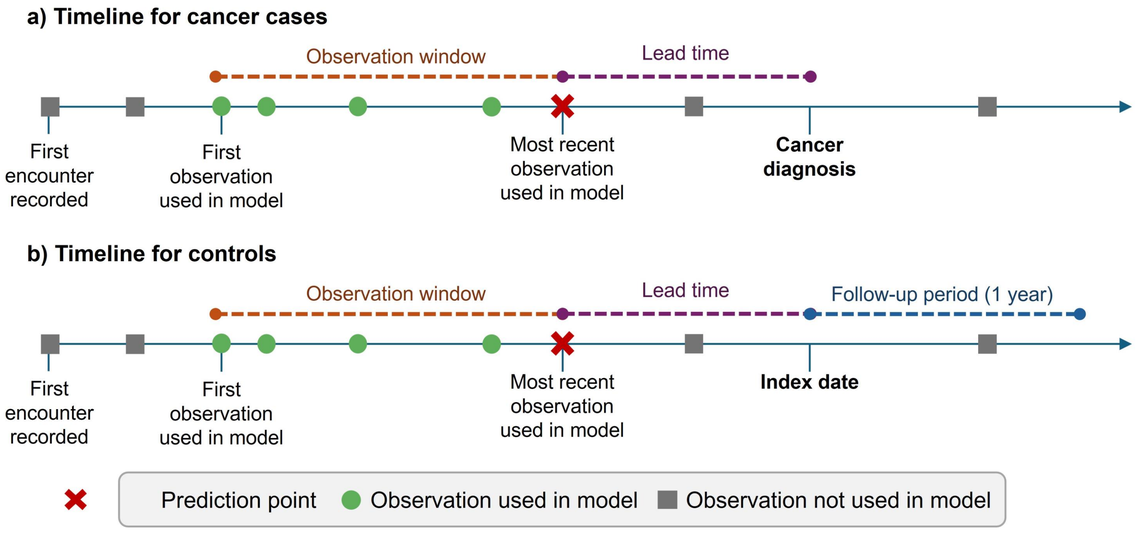
Data



- **S**ecure **A**nonymised **I**nformation **L**inkage (SAIL) databank[2]
  - Linked GP records and cancer registry data for Wales, covering 2.8 million patients
- **Case-control** dataset matched 1:4 on age and sex at each lead time. Index date was date of diagnosis for PC patients; for controls, a random index date was chosen within their trajectories.

Methodology

- Models were trained and evaluated at different **lead times** before diagnosis
- 5-fold cross-validation
- Feature engineering + Logistic Regression (best performing in prior work)
- Sequential input models were LSTM, GRU, & GRU-D[3]



Acknowledgments
This work uses data provided by patients and collected by the NHS as part of their care and support and uses anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. We would like to acknowledge all the data providers who make anonymised data available for research

References
[1] V. Moglia *et al., BMC Med Res Methodol*, vol. 25, no. 1, p. 24, 2025.
[2] R. A. Lyons *et al.*, *BMC Medical Informatics and Decision Making*, vol. 9, no. 1, p. 3, Jan. 2009
[3] Z. Che *et al.,* Sci Rep, vol. 8, no. 1, p. 6085, 2018

Findings

- **Trend features improve performance** compared to traditional machine learning models using a single timepoint:

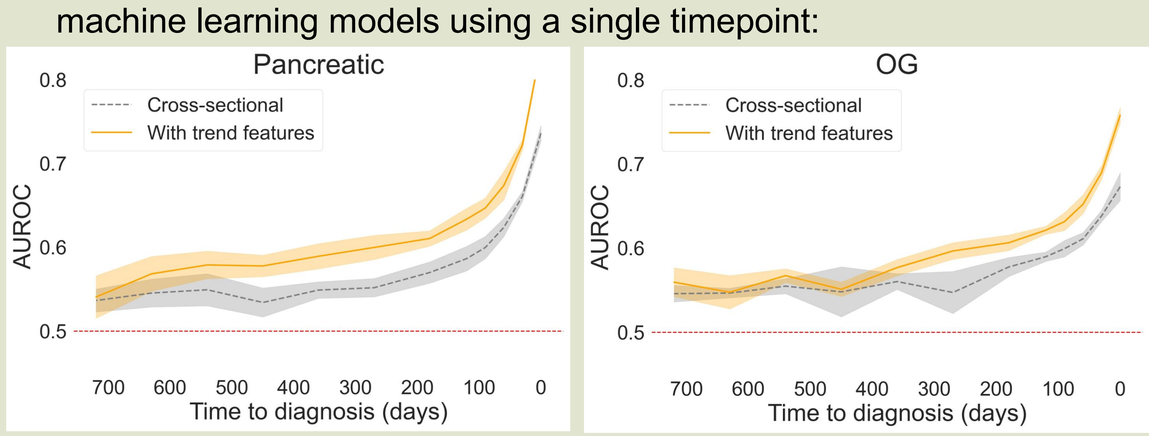


Figure 1: Performance of logistic regression models with and without trend features when predicting pancreatic and oesophageal cancer. The red dotted line is random classification.

- **Deep learning models** did not improve upon logistic regression with engineered features

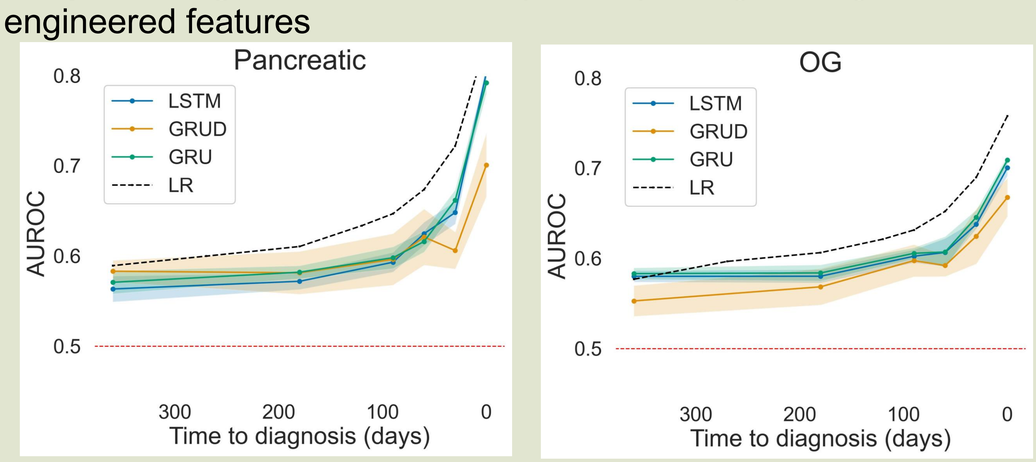


Figure 2: Performance of deep learning models with logistic regression (LR) for comparison, shown with a black dotted line.

Note: higher AUROC means a **model is better** at differentiating between a cancer patient and a non-cancer patient

In real numbers…

At **2 months before diagnosis**, the pancreatic cancer model has an AUROC of **0.67**. At the optimal threshold, if we applied the model to 100 patients, of whom 20 have cancer:

- ⊕ Correctly identify 12 cancer patients.
- ⊕ 30 false positives.
- ⊖ Miss 8 cancers.
- ⊖ 50 true negatives.

Takeaways

- Pre-cancer signals in blood tests are **difficult to detect**, but some differences **can be detected months** before diagnosis
- Blood test **trends** could be used to **improve** prediction models
- **More complex models don't necessarily mean better performance**