



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/228286/>

Version: Accepted Version

Proceedings Paper:

Moglia, V., Smith, L., Cook, G. et al. (2025) Machine Learning Approaches to the Early Detection of Pancreatic Cancer from Time-Series Primary Care Data. In: Artificial Intelligence in Medicine. 23rd International Conference, AIME 2025, 23-26 Jun 2025, Pavia, Italy. Lecture Notes in Computer Science, 15734. Springer, pp. 313-322. ISBN: 978-3-031-95837-3.

https://doi.org/10.1007/978-3-031-95838-0_31

© 2025 The Author(s), under exclusive license to Springer Nature Switzerland AG. This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use (<https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>), but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/978-3-031-95838-0_31.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Machine Learning Approaches to the Early Detection of Pancreatic Cancer from Time-Series Primary Care Data

Victoria Moglia¹[0009-0001-9124-8030], Lesley Smith²[0000-0002-4280-6323],
Gordon Cook^{2,3}[0000-0003-1717-0412], Marc De Kamps¹[0000-0001-7162-4425],
and Owen Johnson¹[0000-0003-3998-541X]

¹ School of Computing, University of Leeds, Leeds, UK

² Leeds Institute for Clinical Trials Research, University of Leeds, Leeds, UK

³ NIHR Leeds Biomedical Research Centre, Leeds Teaching Hospitals Trust, Leeds, UK

Abstract. Pancreatic cancer is notoriously difficult to detect, with diagnosis often relying on symptoms that only develop at advanced stages of the disease. Routine blood tests may signal a developing cancer before these symptoms appear. Limited research has investigated the use of time-varying information from laboratory tests before diagnosis. This study used UK primary care data to compare machine learning approaches for detecting pancreatic cancer at various time-intervals before diagnosis. The machine learning challenge is that such real-world data is irregular and sparse and therefore difficult to use for model creation. In this study, deep learning time-series models (LSTM and GRU-D) were compared to a feature engineering approach. We found that while predictive performance was strongest at diagnosis date (maximum AUROC of 0.85), cases could be detected 18 months before diagnosis, with GRU-D achieving an AUROC of 0.57. Closer to the diagnosis date, where diagnostic signals are stronger, feature engineering approaches outperformed the deep learning models. However, further from diagnosis, the deep learning models, particularly the GRU-D, maintained marginally better performance. Calibration of the models was good at the diagnosis date but was poor across all models at a lead time of greater than 6 months. This study demonstrates that routine blood tests show some predictive capacity for earlier detection of pancreatic cancer. However, this capacity quickly decreases further from diagnosis date, with poor discrimination beyond 6 months. These results should be of interest to researchers interested in using machine learning and electronic health records to support earlier diagnosis of cancer.

Keywords: Machine Learning · Electronic Health Records · Time-Series · Pancreatic Cancer

1 Introduction

Pancreatic cancer (PC) is one of the leading causes of cancer mortality worldwide, accounting for almost 5% of all cancer deaths [2]. A key factor influencing

the mortality of PC is late diagnosis: more than half of PC cases are diagnosed at an advanced stage, resulting in reduced treatment options [10]. As a result, early detection of PC has become an active area of research with the aim of improving patient outcomes by identifying cancers in the early stages [10]. A key barrier to the diagnosis of PC is that symptoms are often silent in the early stages of the disease and are often nonspecific when they do develop [11]. As an alternative, laboratory tests could be investigated in their predictive value for the early detection of PC: routine blood tests have been shown to change in the lead-up to cancer diagnosis, providing a possible opportunity for early detection [12]. Previous works primarily use symptoms and diagnosis data; those that include laboratory tests generally do so in the form of categorical features indicating abnormal tests, or aggregate values (e.g., mean) which may be susceptible to inter-subject variability [8]. Changes in laboratory tests may be more informative, therefore the dynamic nature of measurements should be explored [12]. Previous work by Park et al. investigated the use of time-series laboratory tests for PC detection using neural networks to embed trajectories for each blood test, however, their architecture fails to account for dependencies between time points. There remains a wide range of time-series architectures which may be better suited to the clinical problem.

Time-series data has been utilised in many areas of medicine, including for early detection of PC from symptoms and diagnoses [3]. Previously, we reviewed approaches to the use of time series data for cancer prediction [9]. Long Short-Term Memory networks (LSTMs) are a commonly used model, chosen for the ability to learn longer term dependencies than vanilla recurrent neural networks (RNNs). The GRU-D model (where D denotes decay) was developed to learn from patterns of missingness which may be informative in time-series medical data, making it well-suited to problems involving irregularly sampled laboratory tests [4].

Many countries have structured primary care systems that serve as the first point of contact for most patients. In these systems, primary care data is crucial for cancer detection as it captures early symptoms and healthcare interactions. Since most patients first present with symptoms in primary care, leveraging this data enables earlier detection of cancer. Time-series models for prediction from laboratory tests often use data with a relatively high sampling rate, e.g. ICU data, where models have more data within a patient’s trajectory to learn temporal patterns. In primary care, measurements are more sporadic and potentially more susceptible to noise. As a result, deep learning networks may not perform as well as they do in settings with higher density data, meaning there is a need to assess these models in primary care settings.

In this research, common approaches to time-series analysis in electronic health record (EHR) data were evaluated for their ability to detect PC early in primary care records in the UK. To the authors’ knowledge, this research is the first to use the GRU-D model in primary care laboratory test data.

2 Methods

2.1 Overview

This work compared the performance of LSTMs, GRU-Ds, and standard tabular ML methods combined with feature engineering when detecting PC in primary care and assessed their performance of the models when detecting PC at various time points before diagnosis. These models were chosen based on prevalence in the literature and performance in preliminary experiments.

2.2 Study Design

This research used data from the SAIL Databank [7] which contains health and social care datasets covering the entire Welsh population. We used data from the Welsh Longitudinal General Practice (WLGP) dataset and the Welsh Cancer Surveillance Unit (WCSU) dataset. The WLGP dataset covers approximately 80% of Welsh General Practices (GPs), or around 2.8 million patients. The WCSU is the Welsh cancer registry, containing data on all cancer cases diagnosed or treated in Wales. Data was used from between 2009 and 2019.

Inclusion and exclusion criteria Patients were included in the study based on the existence of valid registration blocks (see 2.3). Cases were any patient with a diagnosis of PC within the study period, as recorded in the WCSU with an ICD-10 code beginning with C25. Controls were selected from all patients without a recorded PC diagnosis using 1:4 age and sex matching. Patients were excluded if there was insufficient demographic data to perform matching.

Time windows The terminology used in this work to define the windows of interest within a patient’s timeline is provided in Figure 1. We investigated the performance of models at numerous lead times until a visible plateau in performance was reached. Patients were required to have a minimum registration of two years so this served as a minimum observation window, however, all data measured within the study period was used to classify the patients. A follow-up period of 1-year with no PC diagnosis was required for controls.

Candidate Predictor Variables This research used blood test variables from a matched cohort to assess the predictive capacity of routine blood tests. The variables were selected based on prevalence in the PC literature [12, 5], discussions with clinicians, and the frequency with which they are recorded in primary care. This work used 31 blood tests: Haemoglobin, Platelets, White blood cells, Red blood cells, Neutrophils, Mean corpuscular volume, Lymphocytes, Monocytes, Eosinophils, Basophils, Alanine aminotransferase, Total bilirubin, Alkaline phosphatase, Albumin, Total protein, Gamma-glutamyl transferase, Urea, Sodium, Potassium, Calcium, Creatinine, Cholesterol, High-density lipoprotein, Low-density lipoprotein, Triglycerides, C-reactive protein, Erythrocyte sedimentation rate, Amylase, HbA1c, Blood glucose, and Ferritin.

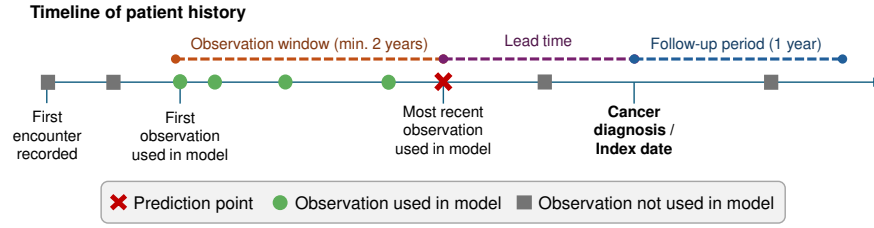


Fig. 1. Diagram showing the terminology used to describe the models.

2.3 Preprocessing

Selection of cohorts using registration blocks The inclusion criteria necessitated sufficient registration at a GP practice. Valid registration was determined using WLGP data, which provided GP practice registration dates. These records were combined to create registration ‘blocks’, indicating periods patients were registered with any GP practice in SAIL. Valid registration blocks were selected for each lead time, ensuring a window within the study period that was at least as long as the observation period plus the lead time. This approach maximized the number of patients eligible for each lead time experiment.

Sampling of index date The index date for the case cohort was the date of diagnosis. For the control cohort, a pseudo-index was selected by randomly selecting a date within their registration period, while still allowing a minimum of the observation window (2 years) between the block start and the index date, and the follow-up period (1 year) between the index and the block end. This approach reduced temporal bias in control index dates by shifting the distribution of dates closer to uniform to match the distribution of PC diagnosis dates.

Feature cleaning and selection Feature cleaning and selection were performed for the feature engineering models using the development set (see 2.4). Features with a missing rate above a threshold were removed: thresholds of 99% and 90% were used for the count variables and remaining features respectively. Highly correlated features were removed by calculating pairwise Pearson’s correlation coefficients and randomly removing a feature from each pair with a value above 0.7. Further feature selection was achieved by fitting a Logistic Regression model with L1 penalty to the development set and selecting features with a coefficient above $1e-5$. Missing data in the time dimension were not addressed in preprocessing, as the models did not require regular time steps. At time points with present measurements, missing variables were imputed using the mean, except in the GRU-D, which has mechanisms to impute missing values. Blood test values were scaled using z-score normalisation.

2.4 Models

Feature Engineering A gradient boosting classifier (FE+GB) and a random forest (FE+RF) were used to classify patients using engineered features which were generated to describe dynamic characteristics of the data. These were the mean value and number of measurements for each blood test within two windows: less than 6 months before index (the proximal window), and between 6 and 12 months before index (the distal window). A trend feature was generated by subtracting the proximal mean from the distal mean. This gave 5 features for each blood test, resulting in 155 features overall. These were reduced before model training through feature cleaning and feature selection (see 2.3).

Deep Learning Models The LSTM architecture processed the entire patient sequence and used the last hidden state as input to a sigmoid activation layer to predict the probability of cancer. LSTMs do not account for irregular time-gaps between data, which may be an informative feature in primary care records, therefore, the inputs were concatenated with the number of days since the last measurement. The GRU-D model is an adaptation of a standard GRU model designed to learn from patterns of missingness in time-series data. The model, described in detail elsewhere [4], augments the inputs with a mask to indicate missing measurements, and a matrix with time since last measurement. The model learns a decay function for imputing missing values, so values decay to a stable value (the mean) if they have not been measured for a long time, or else they are closer to the most recent measurement.

Training Protocol The models were evaluated using cross-validation. The data set was divided into five folds. Each fold was used as a holdout set in turn, with the remaining four folds used as a development set to train and validate the model. Hyperparameter tuning was performed using Bayesian optimization (implemented using the Optuna framework [1]) for each of the data splits using the development set.

2.5 Evaluation

The discriminative performance of the classifiers was assessed at each of the lead times using area under the receiver operating characteristic curve (AUROC), chosen as a threshold-independent indicator of performance. In addition, balanced accuracy, precision, recall, and F1 score are provided at lead times 0, 90, 180, and 540 days. These metrics were calculated for each fold, and summarised using the mean and estimated upper and lower bounds for the 95% confidence interval. The classification metrics are presented at discrete times using thresholds chosen to maximise balanced accuracy. We also provide calibration plots to show how predicted risks of the models align with the distribution in the dataset.

3 Results

3.1 Cohorts

The number of PC patients available in the cohort ranged from 3764 at lead time 0 and 3040 at lead time 630. This gave control cohorts between 15056 (lead time 0) and 12160 (lead time 630).

3.2 Discriminative performance at intervals before diagnosis

The best AUROC at each lead time is provided in Table 1. AUROC is shown for models trained between 0 and 630 days before diagnosis, where the performance seemed to plateau. Figure 3.2 shows how the AUROC changes with earlier detection for the GRU-D and FE+GB models, also indicating the 95% confidence interval for each time point.

Table 1. Best performing model and corresponding AUROC at each lead time. *Indicates a model where the improvement in mean AUROC was significant at the 95% confidence level.

Lead Time	Best Model	Best AUROC	Lead Time	Best Model	Best AUROC
0	FE + GB*	0.85	180	GRU-D*	0.61
30	FE + GB*	0.72	270	GRU-D*	0.60
60	FE + GB*	0.67	360	GRU-D	0.58
90	GRU-D	0.62	450	GRU-D	0.57
120	GRU-D	0.61	540	GRU-D	0.57

3.3 Performance at discrete times before index date

Table 2 shows the performance of the models at lead times 0, 90, 180, and 540 days. Calibration curves are provided for each model at these times in figure 3.

4 Discussion

We show that laboratory tests can be used to predict PC status, without information on symptoms, diagnoses, or demographics. The models show strong predictive ability at lead time of 0, however, this diminishes as the lead time is increased. While the decrease in performance appears to be steep, we show that the models are still able to discriminate between PC patients and controls up to a year before diagnosis. Some of the models maintain a small degree of discriminative ability past this point, however, the performance seems to plateau at around 0.55 and does not drop to 0.5 for three of the models, indicating potential bias in the dataset.

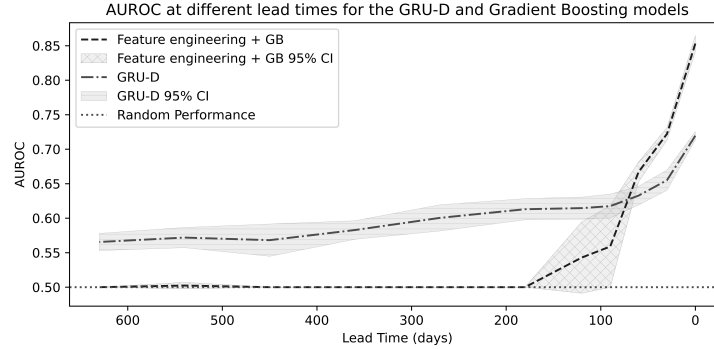


Fig. 2. AUROC of the FE+GB model and the GRU-D against lead time. These FE+GB model performed best up to 3 months before diagnosis. Beyond 3 months, the GRU-D performed best. 95% confidence interval estimated from the cross-validated AUROC scores are shown.

Between 0 and 90 days before the index, the FE+GB model achieves significantly higher discrimination than the other models. It is possible that within this window there are highly predictive features that are explicitly included in the feature engineering models, for example, a count of a blood test used in the diagnostic process. These may be missed by the more complex models. At lead times earlier than 3 months, the GRU-D model achieves the best performance. At the 95% confidence level, this improvement in performance is significant at 180 and 270 months before diagnosis, maintaining an AUROC of 0.6 in this window. The confidence intervals of all models widen as the lead time increases, indicating more variable performance and increased uncertainty over longer prediction horizons. We report an AUROC of 0.58 at 12 months before diagnosis; similar to results achieved by Park et. al [6] in a US dataset (AUROC of 0.6 at 12 months pre-index), however, comparisons are limited due to the differing health systems and apparent data leakage in their methodology.

The FE+GB model shows the best calibration at the index date, however past 90 days the performance seems to collapse. The LSTM and GRU-D models both show worse calibration at greater lead times, but the calibration of the deep learning models is much better than that of the feature engineering models.

Beyond performance, there are additional considerations that make a model appropriate for a clinical setting. While GRU-D outperforms feature engineering models, it does not provide any interpretability or model reasoning. If feature engineering models can be improved, they may be more acceptable to clinicians and patient’s than black-box models as the features used are interpretable. In addition, GRU-D doesn’t extend well to categorical features such as symptoms/comorbidities. Including these features may improve performance significantly, and while they can be incorporated using fusion techniques, it is a much

Table 2. Classifications metrics for the models at four lead times, presented as mean $\pm$ standard error. Bold indicates the highest performing model at each lead time.

Lead Time	Model	Balanced Accuracy	F1 Score	Precision	Recall
0	<i>FE+GB</i>	0.75 $\pm$ 0.01	0.66 $\pm$ 0.01	0.87 $\pm$ 0.01	0.53 $\pm$ 0.02
	<i>FE+RF</i>	0.71 $\pm$ 0.00	0.55 $\pm$ 0.00	0.46 $\pm$ 0.01	0.67 $\pm$ 0.00
	<i>GRU-D</i>	0.66 $\pm$ 0.00	0.50 $\pm$ 0.01	0.45 $\pm$ 0.03	0.56 $\pm$ 0.02
	<i>LSTM</i>	0.74 $\pm$ 0.00	0.62 $\pm$ 0.01	0.73 $\pm$ 0.01	0.53 $\pm$ 0.01
90	<i>FE+GB</i>	0.51 $\pm$ 0.01	0.04 $\pm$ 0.03	0.29 $\pm$ 0.18	0.02 $\pm$ 0.02
	<i>FE+RF</i>	0.53 $\pm$ 0.01	0.35 $\pm$ 0.00	0.22 $\pm$ 0.0	0.87 $\pm$ 0.05
	<i>GRU-D</i>	0.56 $\pm$ 0.02	0.38 $\pm$ 0.01	0.24 $\pm$ 0.01	0.83 $\pm$ 0.03
	<i>LSTM</i>	0.57 $\pm$ 0.01	0.36 $\pm$ 0.004	0.24 $\pm$ 0.005	0.721 $\pm$ 0.018
180	<i>FE+GB</i>	0.50 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	<i>FE+RF</i>	0.53 $\pm$ 0.01	0.35 $\pm$ 0.00	0.22 $\pm$ 0.00	0.84 $\pm$ 0.03
	<i>GRU-D</i>	0.58 $\pm$ 0.00	0.36 $\pm$ 0.00	0.27 $\pm$ 0.01	0.57 $\pm$ 0.04
	<i>LSTM</i>	0.55 $\pm$ 0.01	0.34 $\pm$ 0.01	0.23 $\pm$ 0.01	0.69 $\pm$ 0.04
540	<i>FE+GB</i>	0.50 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00	0.00 $\pm$ 0.00
	<i>FE+RF</i>	0.52 $\pm$ 0.00	0.34 $\pm$ 0.00	0.21 $\pm$ 0.00	0.87 $\pm$ 0.03
	<i>GRU-D</i>	0.54 $\pm$ 0.00	0.34 $\pm$ 0.01	0.21 $\pm$ 0.01	0.83 $\pm$ 0.04
	<i>LSTM</i>	0.52 $\pm$ 0.00	0.34 $\pm$ 0.00	0.21 $\pm$ 0.00	0.91 $\pm$ 0.02

simpler process to add the features into a technique such as gradient boosting or random forest.

A key strength of this work is the use of a large, population-based, multi-centre dataset. We demonstrate how predictive capacity of routine blood tests changes in the months before diagnosis, providing potential windows where future research should focus. We also provide calibration of the models, which is often unreported in clinical prediction model research. This work provides the first evaluation of GRU-D in a primary care setting, showing that the model achieves improved performance in detecting weak signals, but that it may miss simple, stronger features that can be summarised by simple feature engineering, as seen close to diagnosis.

We acknowledge limitations to our work. Firstly, the engineering of the trend features was relatively simple, meaning more complex patterns may have been missed. We do not perform external validation due to the limited availability of open-access datasets, however, we aim to mitigate this by leveraging cross-validation to reduce optimism in the performance. While this research demonstrates the potential for earlier detection of PC in primary care, we do not explore how such models could be integrated into clinical workflows, such as forming decision support tools for GPs. The performance in this research is not comparable to real-world performance due to the matched cohort. We also do not discuss the potential implications of earlier detection, such as increased false

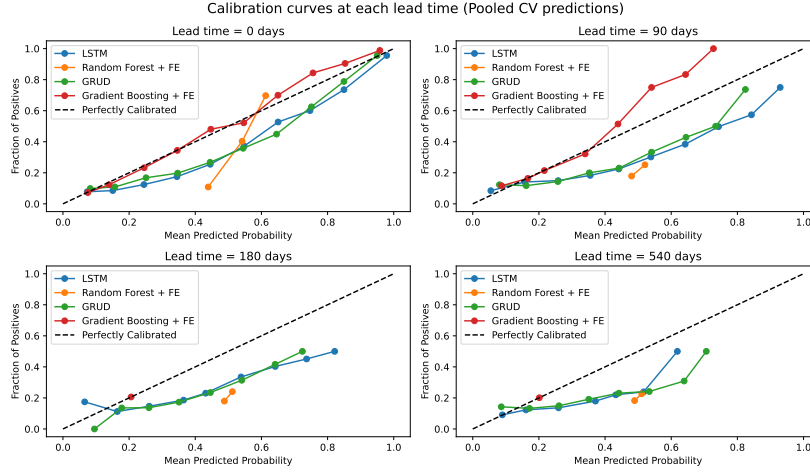


Fig. 3. Calibration curves for each model at four time points. The curves show the relationship between predicted risk and observed event rates.

positives, unnecessary investigations, and increased anxiety for patients, which are important considerations when evaluating clinical utility.

Future work should aim to focus on improving prediction between 3-12 months, where there is evidence of weak diagnostic signals that may aid early detection. More extensive validation of models in primary care data is required before tools can be assessed in a prospective setting, which is a critical step for clinical translation of decision support tools. Further investigation of appropriate feature engineering techniques could improve the feature engineering models while maintaining lower computation requirements, for example, by using analysis of population level trends to inform engineering. In addition, differences in performance between different patient subgroups should be investigated to assess bias. The GRU-D shows promise in primary care, with implications for broader research in primary care, however, the increase in performances comes with increases in computation and training times.

Acknowledgments. This research was financially supported by the UKRI EPSRC (grant number EP/S024336/1/). This work uses data provided by patients and collected by the NHS as part of their care and support.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M.: Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th

- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)
2. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R.L., Soerjomataram, I., Jemal, A.: Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* **74**(3), 229–263 (2024). <https://doi.org/10.3322/caac.21834>
 3. Carrasco-Ribelles, L.A., Llanes-Jurado, J., Gallego-Moll, C., Cabrera-Bean, M., Monteagudo-Zaragoza, M., Violán, C., Zabaleta-del-Olmo, E.: Prediction models using artificial intelligence and longitudinal data from electronic health records: A systematic methodological review. *Journal of the American Medical Informatics Association* **30**(12), 2072–2082 (2023). <https://doi.org/10.1093/jamia/ocad168>
 4. Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y.: Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports* **8**(1), 6085 (2018). <https://doi.org/10.1038/s41598-018-24271-9>
 5. Cranfield, B.M., Koo, M.M., Abel, G.A., Swann, R., McPhail, S., Rubin, G.P., Lyratzopoulos, G.: Primary care blood tests before cancer diagnosis: National Cancer Diagnosis Audit data. *British Journal of General Practice* **73**(727), e95–e103. <https://doi.org/10.3399/BJGP.2022.0265>
 6. Jia, K., Kundrot, S., Palchuk, M.B., Warnick, J., Haapala, K., Kaplan, I.D., Rinnard, M., Appelbaum, L.: A pancreatic cancer risk prediction model (Prism) developed and validated on large-scale US clinical data. *eBioMedicine* **98** (2023). <https://doi.org/10.1016/j.ebiom.2023.104888>
 7. Lyons, R.A., Jones, K.H., John, G., Brooks, C.J., Verplancke, J.P., Ford, D.V., Brown, G., Leake, K.: The SAIL databank: Linking multiple health and social care datasets. *BMC Medical Informatics and Decision Making* **9**(1), 3 (2009). <https://doi.org/10.1186/1472-6947-9-3>
 8. Mishra, A.K., Chong, B., Arunachalam, S.P., Oberg, A.L., Majumder, S.: Machine Learning Models for Pancreatic Cancer Risk Prediction Using Electronic Health Record Data—A Systematic Review and Assessment. *Official journal of the American College of Gastroenterology* **119**(8), 1466 (2024). <https://doi.org/10.14309/ajg.0000000000002870>
 9. Moglia, V., Smith, O., Cook, G., Smith, L.: Artificial intelligence methods applied to longitudinal data from electronic health records for prediction of cancer: A scoping review. *BMC Medical Research Methodology* **25**(1), 24 (2025). <https://doi.org/10.1186/s12874-025-02473-w>
 10. Pereira, S.P., Oldfield, L., Ney, A., Hart, P.A., Keane, M.G., Pandol, S.J., Li, D., Greenhalf, W., Jeon, C.Y., Koay, E.J., Almario, C.V., Halloran, C., Lennon, A.M., Costello, E.: Early detection of pancreatic cancer. *The Lancet Gastroenterology & Hepatology* **5**(7), 698–710 (2020). [https://doi.org/10.1016/S2468-1253\(19\)30416-9](https://doi.org/10.1016/S2468-1253(19)30416-9), publisher: Elsevier
 11. Rawla, P., Sunkara, T., Gaduputi, V.: Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors. *World Journal of Oncology* **10**(1), 10–27 (2019). <https://doi.org/10.14740/wjon1166>
 12. Virdee, P.S., Collins, K.K., Friedemann Smith, C., Yang, X., Zhu, S., Roberts, S.E., Roberts, N., Oke, J.L., Bankhead, C., Perera, R., Hobbs, F.R., Nicholson, B.D.: The Association between Blood Test Trends and Undiagnosed Cancer: A Systematic Review and Critical Appraisal. *Cancers* **16**(9), 1692 (2024). <https://doi.org/10.3390/cancers16091692>