



# Time series radiomics for the prediction of prostate cancer progression in patients on active surveillance

Nikita Sushentsev<sup>1</sup> · Leonardo Rundo<sup>1,2</sup> · Luis Abrego<sup>3</sup> · Zonglun Li<sup>4</sup> · Tatiana Nazarenko<sup>3,4</sup> · Anne Y. Warren<sup>5</sup> · Vincent J. Gnanapragasam<sup>6,7</sup> · Evis Sala<sup>1,8</sup> · Alexey Zaikin<sup>3,4</sup> · Tristan Barrett<sup>1</sup> · Oleg Blyuss<sup>9,10</sup>

Received: 11 July 2022 / Revised: 3 January 2023 / Accepted: 9 January 2023  
© The Author(s) 2023

## Abstract

Serial MRI is an essential assessment tool in prostate cancer (PCa) patients enrolled on active surveillance (AS). However, it has only moderate sensitivity for predicting histopathological tumour progression at follow-up, which is in part due to the subjective nature of its clinical reporting and variation among centres and readers. In this study, we used a long short-term memory (LSTM) recurrent neural network (RNN) to develop a time series radiomics (TSR) predictive model that analysed longitudinal changes in tumour-derived radiomic features across 297 scans from 76 AS patients, 28 with histopathological PCa progression and 48 with stable disease. Using leave-one-out cross-validation (LOOCV), we found that an LSTM-based model combining TSR and serial PSA density (AUC 0.86 [95% CI: 0.78–0.94]) significantly outperformed a model combining conventional delta-radiomics and delta-PSA density (0.75 [0.64–0.87];  $p=0.048$ ) and achieved comparable performance to expert-performed serial MRI analysis using the Prostate Cancer Radiologic Estimation of Change in Sequential Evaluation (PRECISE) scoring system (0.84 [0.76–0.93];  $p=0.710$ ). The proposed TSR framework, therefore, offers a feasible quantitative tool for standardising serial MRI assessment in PCa AS. It also presents a novel methodological approach to serial image analysis that can be used to support clinical decision-making in multiple scenarios, from continuous disease monitoring to treatment response evaluation.

## Key Points

- LSTM RNN can be used to predict the outcome of PCa AS using time series changes in tumour-derived radiomic features and PSA density.
- Using all available TSR features and serial PSA density yields a significantly better predictive performance compared to using just two time points within the delta-radiomics framework.
- The concept of TSR can be applied to other clinical scenarios involving serial imaging, setting out a new field in AI-driven radiology research.

**Keywords** Prostatic neoplasms · Magnetic resonance imaging · Artificial intelligence

Tristan Barrett and Oleg Blyuss contributed equally to this study.

✉ Nikita Sushentsev  
ns784@medschl.cam.ac.uk

<sup>1</sup> Department of Radiology, School of Clinical Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge Biomedical Campus, Box 218, Cambridge CB2 0QQ, UK

<sup>2</sup> Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Fisciano, SA, Italy

<sup>3</sup> Department of Women's Cancer, Institute for Women's Health, University College London, London, UK

<sup>4</sup> Department of Mathematics, University College London, London, UK

<sup>5</sup> Department of Pathology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

<sup>6</sup> Department of Urology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK

<sup>7</sup> Cambridge Urology Translational Research and Clinical Trials Office, Addenbrooke's Hospital, Cambridge Biomedical Campus, Cambridge, UK

<sup>8</sup> Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK

<sup>9</sup> Wolfson Institute of Population Health, Queen Mary University of London, London, UK

<sup>10</sup> Center of Photonics, Lobachevsky University, Nizhny Novgorod, Russian Federation

## Abbreviations

ADC	Apparent diffusion coefficient
AS	Active surveillance
AUC	Area under the receiver operating characteristic curve
DR	Delta-radiomics
ISUP	International Society of Urological Pathology
LOOCV	Leave-one-out cross-validation
LSTM	Long short-term memory
MRI	Magnetic resonance imaging
NPV	Negative predictive value
PCa	Prostate cancer
PPV	Positive predictive value
PSA	Prostate-specific antigen
PSAd	Prostate-specific antigen density
PZ	Peripheral zone
RNN	Recurrent neural network
T2WI	T2-weighted imaging
TSR	Time series radiomics
TZ	Transition zone

## Introduction

Prostate cancer (PCa) is the second commonest and the fifth deadliest male cancer globally [1]. Nearly half of newly diagnosed men present with low-risk and favourable intermediate-risk disease [2], for which active surveillance (AS) is the recommended treatment option [3]. However, 27% of AS patients show 5-year histopathological disease progression [4], highlighting the unmet clinical need for improved baseline and follow-up risk stratification. MRI at baseline has been shown to be of benefit in selecting for AS by identifying and excluding patients in whom AS is inappropriate [5, 6]. Simultaneously, the pooled performance of serial MRI for follow-up AS monitoring is only moderate and highly variable [7], in part due to the inherently subjective and reader-dependent nature of its assessment that is particularly evident in non-specialist centres [8]. To address this, the standardised 5-point Prostate Cancer Radiological Estimation of Change in Sequential Evaluation (PRECISE) score [9] was introduced. Despite becoming a gold standard for serial MRI assessment and demonstrating a high negative predictive value ranging between 92 and 100% [10], PRECISE has been exclusively validated in expert centres and still showed only a moderate positive predictive value of 15 to 66% for predicting histopathological PCa progression on AS [10].

In parallel with working on further iterations of PRECISE, developing quantitative tools to make serial MRI assessment more objective may help further improve its performance and limit variance to achieve consistent expert-level quality [6]. Pilot studies have adopted artificial intelligence (AI) techniques to devise MRI-derived radiomics models for predicting

PCa progression on AS both at baseline [11] and at follow-up [12]. However, in the multiple time point follow-up setting, the only established methodological framework for analysing temporal radiomic patterns is delta-radiomics (DR), which only measures change between two time points [13]. Being more appropriate for treatment response assessment where clinical decisions are often based on a single post-treatment study, it is less applicable to the AS setting where patients undergo multiple scans, each storing quantitative data describing patient-specific dynamics of tumour development.

In this study, we aimed to develop a time series radiomics (TSR) framework for predicting histopathological PCa progression on AS based on longitudinal changes in radiomic features extracted from all MRI scans obtained over the follow-up period. We hypothesised that the TSR predictive model incorporating all imaging and clinical data collected during AS would outperform DR and achieve at least comparable performance to the expert-derived PRECISE scoring as a clinical standard, offering a novel approach towards quantitative serial medical imaging data analysis.

## Materials and methods

### Dataset and study population

This ethically approved (Health Research Authority and Health and Care Research Wales, IRAS Project ID 288,185) retrospective exploratory study included consecutive patients with biopsy-proven PCa enrolled on AS in our institution between July 2013 and October 2019 according to the previously published criteria [14], which are limited to the inclusion of International Society of Urogenital Pathology (ISUP) grade groups 1 and 2 disease with  $\leq 10\%$  Gleason pattern 4. The minimum inclusion criteria for this study were the presence of at least one MR-visible lesion, 2-year follow-up, three 3 T MRI scans (same magnet), and one repeat targeted biopsy within 12 months of the final MRI. The exclusion criteria are described in Supplementary Information. AS progression was defined as biopsy-proven histopathological ISUP grade group progression.

### MRI acquisition

All patients underwent prostate MRI on a 3 T MR750 scanner (GE Healthcare) using a 32-channel receiver coil, with the full protocol described in Supplementary Information. At baseline, all patients underwent multiparametric MRI protocol that included multiplanar high-resolution T<sub>2</sub>-weighted, diffusion-weighted imaging (DWI), and dynamic contrast-enhanced (DCE) MRI. Follow-up scans were acquired using a biparametric MRI protocol, omitting DCE-MRI.

## Ground truth assessment

Since AS eligibility is primarily based on tumour histopathological characteristics, repeat targeted biopsy results were selected as an appropriate ground truth reference standard for confirming the disease progression [15]. Both baseline and follow-up biopsies were targeted at the same lesions delineated on MRI, with biopsy reports including separate sections dedicated to the assessment of target cores. Repeat biopsy procedures were either protocol-driven (12 and 36 months after AS enrolment) or triggered by either three consecutive elevated serum prostate-specific antigen (PSA) levels that were above the pre-defined threshold or radiological progression (PRECISE score  $\geq 4$ ) [9]. Biopsy samples were reviewed by an expert genitourinary pathologist (A.Y.W.) according to ISUP guidelines.

## Image segmentation and analysis

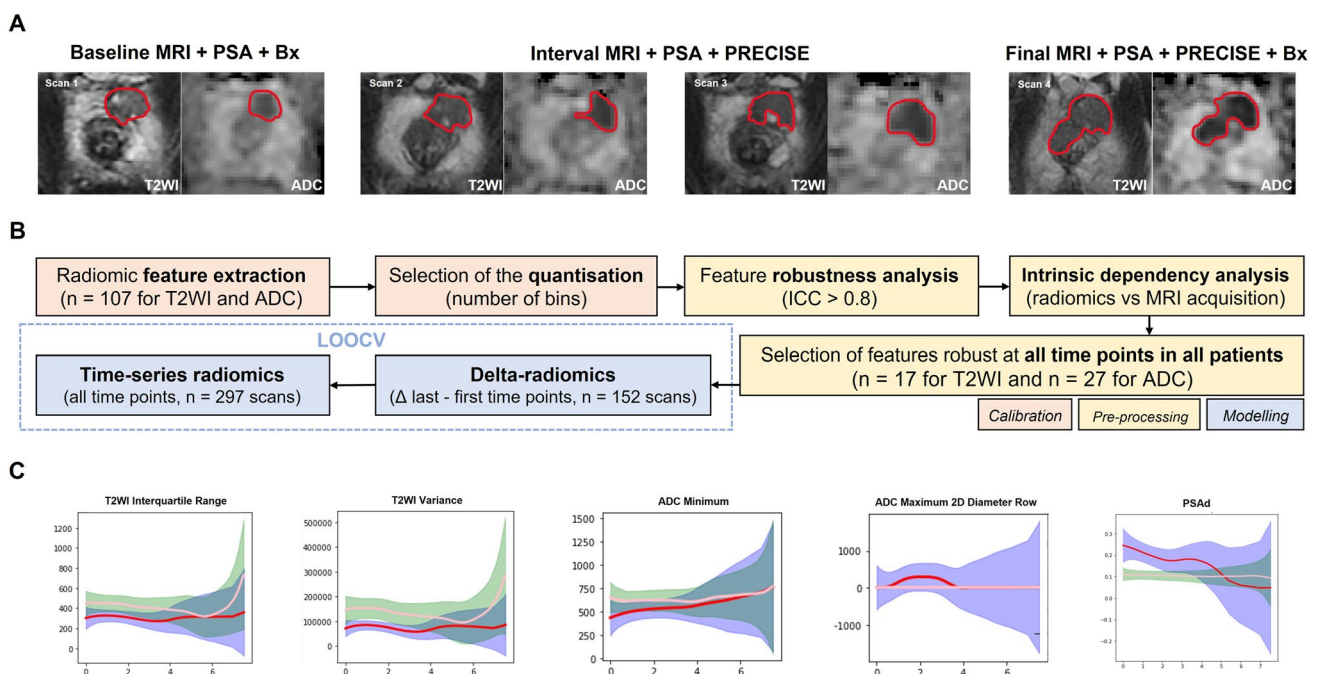
Tumour regions of interest (ROIs) were drawn on de-identified anatomical T2WI and ADC maps (Fig. 1A) by a fellowship-trained urogenital radiologist (T.B.) with a 14-year clinical experience in prostate MRI and a research fellow (N.S.) with a 5-year experience. ROIs were drawn in

consensus using open-source segmentation software ITK-SNAP [16], matching the original ROIs used for targeted biopsy of suspicious lesions with a PI-RADS score of  $\geq 3$ .

Follow-up MRI scans were prospectively scored on a 5-point PRECISE scale [9] by four expert urogenital radiologists with 6–17-year experience in prostate MRI reporting and considered experts based on the number of MRIs reported [17, 18]. For predictive modelling, the PRECISE scores were dichotomised at a cut-off value of 4, with patients scoring  $\geq 4$  at any point deemed as those exhibiting radiological PCa progression. At the time of clinical reporting of the PRECISE scores, the readers were aware of clinical information that included PSA and PSA density dynamics, which reflects the real-life clinical use of this scoring system.

## Image calibration and pre-processing

Figure 1B summarises the radiomics pipeline used to develop and validate DR and TSR predictive models for PCa progression on AS. Tumour-derived radiomic features were extracted using PyRadiomics version 2.0 and Python version 3.7.5 [19]; full features are presented in Supplementary Information. Only features deemed robust at all



**Fig. 1** Example T2WI and ADC images (A) obtained from a patient with a PZ lesion (red ROI) that showed progression from ISUP grade group 1 at baseline to ISUP grade group 3 at repeat biopsy triggered by radiological progression noted at scan 4. Flow diagram (B) summarising the key stages of radiomic analysis used in this study, from radiomic feature extraction to leave-one-out cross-validation (LOOCV) of baseline

radiomics, delta-radiomics, and time series radiomics predictive models. Locally weighted scatterplot smoothing (LOWESS) curves (C) demonstrating time series changes in representative T2WI- and ADC-derived radiomic features and PSA density in progressors (red lines) and non-progressors (yellow lines), with x-axes denoting years to progression or end of follow-up and y-axes denoting corresponding values

time points for all patients, separately for T2WI ( $n = 17$ ) and ADC ( $n = 27$ ), were included for predictive modelling (Supplementary Information), with representative examples shown in Fig. 1C.

### Prostate-specific antigen predictive modelling

The addition of non-radiomic clinical parameters to texture features is expected to result in more holistic predictive models [20]. We retrieved all serial prostate-specific antigen (PSA) measurements and derived PSA density (PSAd; Fig. 1C) as a known independent baseline predictor used in clinical practice [14, 15], which was prioritised for predictive modelling following the rationale described in Supplementary Information.

### Delta-radiomics predictive modelling

DR features were computed as the difference between the final and baseline features as:

$$\Delta f = f_{\text{final}} - f_{\text{base}}$$

where  $f_{\text{base}}$  and  $f_{\text{final}}$  correspond to features derived from baseline and final scans, respectively. Predictive modelling was employed using parenchitic networks [21] as described previously [12].

### Time series radiomics predictive modelling

Our data considered the longitudinal patient history of different features as shown in Fig. 2. For the  $i$ -th patient, the  $k$  input is expressed by the sequence  $(x_{i1k}, x_{i2k}, \dots, x_{ijk}, \dots, x_{iT_k})$ , where  $k = 1, 2, \dots, K$ . These measurements are collected at the time points  $t_j, j = 1, 2, \dots, T_i$  for every input  $k$ .

Here, we evaluated multivariate longitudinal observations, separated into three different inputs corresponding to a specific feature type (namely, T2WI- and ADC-derived radiomics, and PSA density). The index  $k$  described above labels each group of features T2WI, ADC, and PSAd. In turn, each of these groups represents a multivariate input composed of  $d_k$  feature columns, for each  $k$ .

In this paper, the probability of prostate cancer is estimated from a parallel long short-term memory (LSTM) architecture [22–25]. The architecture is described by the following set of equations:

$$h_{ijk} = o_{ijk} \odot \tanh(c_{ijk})$$

where

$$\begin{aligned} c_{ijk} &= f_{ijk} \odot c_{ij-1k} + i_{ijk} \odot \widetilde{c}_{ijk} \\ \widetilde{c}_{ijk} &= \tanh(h_{ij-1k} U_c + x_{ijk} W_c + b_c) \\ i_{ijk} &= \sigma(h_{ij-1k} U_f + x_{ijk} W_i + b_i) \\ f_{ijk} &= \sigma(h_{ij-1k} U_f + x_{ijk} W_f + b_f) \\ o_{ijk} &= \sigma(h_{ij-1k} U_o + x_{ijk} W_o + b_o) \end{aligned}$$

Here,  $i_{ijk}$ ,  $f_{ijk}$  and  $o_{ijk}$  refer to the input, forget, and output gates, respectively.  $\widetilde{c}_{ijk}$  is the candidate cell state,  $c_{ijk}$  is the current cell state, and  $h_{ijk}$  is the hidden state. The weight matrices  $W$  and  $U$  were adjusted during learning along with the biases  $b$ . Both  $W$  and  $U$  had dimensions  $d_k \times H_k$  and  $H_k \times H_k$ , respectively. Here  $d_k$  indicates the number of feature columns of each multivariate input  $k$ , and  $H_k$  is the number of hidden units; biases have dimension  $1 \times H_k$ ;  $\odot$  denotes the point-wise multiplication.

For the patient  $i$ , time-step  $j$ , and input  $k$ , we calculated the sequence of hidden states  $(h_{i1k}, h_{i2k}, \dots, h_{ijk}, \dots, h_{iT_k})$ . The last element in the sequence  $h_{iT_k}$  was extracted and then concatenated with other hidden states from the remaining  $K - 1$  LSTMs to have a single vector  $h_i$ , accounting for the whole multivariate time series. That is,

$$h_i = [h_{iT_1}, h_{iT_2}, \dots, h_{iT_K}]$$

Finally, a dropout was added during training followed by a regular densely connected neural network layer of 1 unit with sigmoid activation,

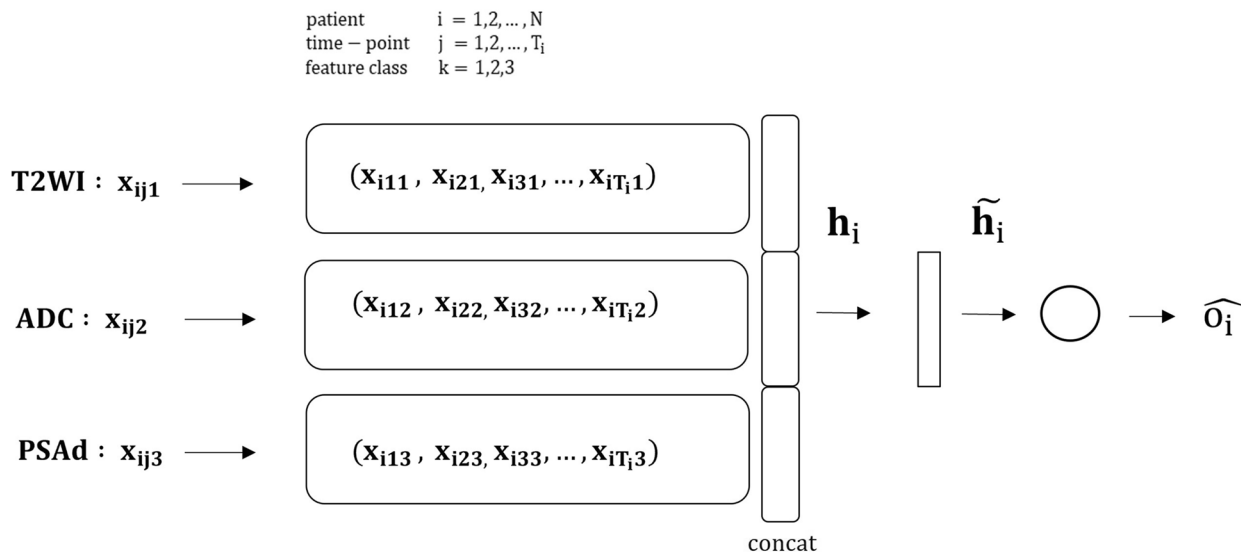
$$\hat{o}_i = \sigma(W_l \widetilde{h}_i + b_l)$$

where  $\sigma(x) = 1/(1 + \exp(-x))$ ,  $\widetilde{h}_i$  denotes the last hidden state after dropout with a set rate,  $W_l$  is a weight vector of size  $(H_1 + H_2 + \dots + H_K) \times 1$ , and  $b_l$  is a scalar bias.  $\hat{o}_i$  is the estimated probability to be compared with the target outcome (0 for controls and 1 for cases) using binary cross entropy.

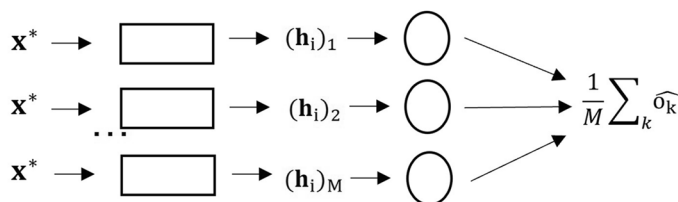
In this study, we had  $K = 3$  associated with T2WI-derived radiomic features, ADC-derived radiomic features, and PSA density. Each of these had  $d_k$  column features 17, 27, and 1, respectively. Each input was passed to its associated LSTM cell, from where we estimated the probability of cancer.

For the training stage, we used batch gradient descent with dynamic learning rates updated through Adam Optimiser [22]. Since neural networks are sensitive to feature scaling, each feature in the data was standardised appropriately during cross-validation by centring each feature value around its mean and then dividing the difference by its standard deviation. This ensured that training converged faster [26]. This scaling transformation was then kept and applied to subsequent validation samples to estimate the probability of PCa progression.

**A**



**B**



**Fig. 2** Multistream classification using longitudinal data. The input (A) consists of multiple parallel synchronised input sequences  $x_{ijk}$  grouped into T2WI-derived radiomic features, ADC-derived radiomic features, and PSA density. Features were first standardised and fed into their respective LSTM cell. Then, we concatenated the last hidden state  $h_i$ , which was associated with the LSTM output from each clinical feature. We computed  $\tilde{h}_i$  by adding a dropout layer during training. The final output  $\hat{o}_i$  was estimated from a regular densely

connected neural network layer of 1 unit with sigmoid activation. Subscript  $i$  indicates the patient label,  $j$  corresponds to the time point, and  $k$  denotes the feature class. Input data for the LSTM was masked and padded to handle variable sequence lengths. **B** The network was trained  $M$  times using longitudinal data from  $N-1$  patients (following a LOOCV procedure), and with a fixed set of hyperparameters. Once the network was trained, we estimated the average probability of risk of PCa progression on AS for the remaining patient

The kernel weights matrix for the recurrent state was initialised by a random orthogonal matrix, while for the inputs we used a kernel weights matrix initialised using Glorot’s scheme. Bias vectors were initialised at zero [27]. Finally, to handle variable sequence lengths, masking and post-padding were applied to the inputs of the network [28–30].

As a preliminary stage, we performed hyperparameter tuning using random sampling with Leave-One-Out Cross-Validation (LOOCV). LOOCV is a special case of cross-validation where the number of folds equals the number of instances in the dataset. Thus, the learning algorithm is applied once for each instance, using all other instances as a training set and using the selected instance as a single-item test set, which is particularly suitable for small datasets where avoiding overfitting is critical [31]. The performance was determined by the mean of binary cross

entropy obtained from the cross-validation procedure. The empirical loss was estimated in each case using different combinations of the number of hidden neurons, dropout rate, number of epochs, and learning rates. From this set of simulations, we selected an optimal set of hyperparameters, which we used to report our results: 130 epochs, learning rate of 0.002, dropout 0.2, 16 units for each separate group of features (PSA density, T2WI radiomics, ADC radiomics).

To estimate the repeatability of the estimations provided by our models, the network was trained multiple times using this set of hyperparameters (100 iterations for each of the patients in the cohort following the LOOCV procedure) from which we obtained a distribution of outputs. Then, we computed the mean probability and variance associated with the estimation. LSTM architectures were implemented



in Python 3.8.8 using TensorFlow version 2.7.0 and Keras version 2.7.0.

## Statistical and data analysis

Data normality was assessed using the D'Agostino-Pearson test (threshold  $p \geq 0.05$ ). Intergroup comparison of patient clinicodemographic characteristics was performed using the Mann–Whitney  $U$  test and Fisher's exact test as appropriate. DR, TSR, and PRECISE performance against targeted biopsy results (ground truth) were assessed using standard measures of discrimination, with radiomic models validated using a LOOCV scheme to avoid overfitting given the limited sample size compared to the number of radiomic features. In this study, a true positive result was recorded when any of the above models made a correct prediction of tumour progression that was confirmed histopathologically; a false positive result was noted when a model predicted progression in a patient with histopathologically stable disease. In turn, a true negative result was recorded when a model correctly identified stable disease in a patient with no histopathological tumour progression; a false negative result was recorded when a model falsely predicted stable disease in a patient demonstrating histopathological disease progression. Statistical analysis was performed in R version 3.5.1 (R Foundation for Statistical Computing) using the “pROC” and “reportROC” packages as detailed in Supplementary Information.

## Results

### Study population

Of 364 PCa patients enrolled on AS in our centre, 76 were included in this study, with the study flowchart presented in Fig. 3A. Baseline clinicopathological characteristics of the study cohort are listed in Table 1, the final cohort including 28 progressors and 48 non-progressors. Each patient

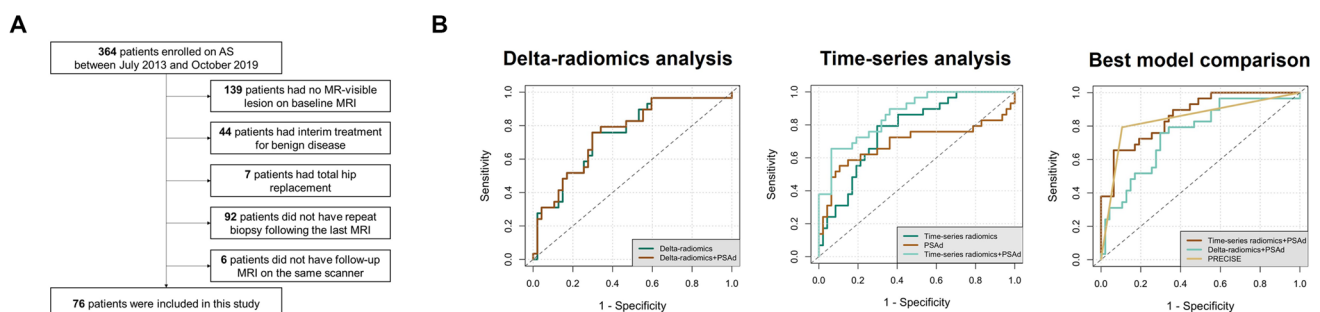
included in this study had only one MR-visible lesion at baseline, and no patient developed new lesions in the follow-up. While baseline gland volume was significantly higher in non-progressors ( $p = 0.004$ ; Table 1), baseline PSA<sub>d</sub> was significantly increased in progressors ( $p = 0.003$ ; Table 1), further supporting its inclusion in the predictive modelling.

### Predictive modelling: delta-radiomics, time series radiomics, and PRECISE

Following LOOCV, standalone DR and TSR predictive models achieved AUCs of 0.75 [0.64–0.86] and 0.77 [0.66–0.87], respectively ( $p = 0.714$ ; Fig. 3B), which was similar to the AUC of PRECISE at 0.84 [0.76–0.93] ( $p = 0.108$  and 0.199, respectively; Supplementary Fig. 2). The addition of delta-PSA<sub>d</sub> (0.61 [0.47–0.75]; Supplementary Fig. 3) to DR did not result in a better performance of the combined model (0.75 [0.64–0.87]) compared to the standalone DR model ( $p = 0.503$ ; Fig. 3B). Conversely, the addition of time series PSA<sub>d</sub> (0.69 [0.55–0.83]; Supplementary Fig. 3) to TSR significantly increased its performance to 0.86 [0.78–0.94] compared to the standalone TSR model ( $p = 0.015$ ; Fig. 3B). As a result, the AUC of the combined TSR + PSA<sub>d</sub> model was significantly higher compared to that of the DR + PSA<sub>d</sub> model ( $p = 0.048$ ; Fig. 3B). In addition, the AUCs of both DR + PSA<sub>d</sub> and TSR + PSA<sub>d</sub> models were also similar to that of PRECISE ( $p = 0.121$  and 0.710, respectively; Fig. 3B). Numerically, TSR + PSA<sub>d</sub> had the highest specificity, PPV, and AUC of 0.94, 0.86, and 0.86, while PRECISE had the highest NPV at 0.88, with 95% CI for each metric reported in Table 2.

## Discussion

We developed an RNN-based framework for time series analysis of MRI-derived radiomic features and PSA<sub>d</sub> for the prediction of PCa histopathological progression on AS. The standalone TSR model showed similar performance



**Fig. 3** Study flowchart (A). ROC curves for standalone and combined radiomic models (B) developed as a result of delta-radiomics and time series analyses, along with a summary graph presenting ROC curves for the best-performing models

**Table 1** Baseline clinicopathological characteristics of the study cohort. The data for age, PSA, gland volume, and PSAd are presented as median (interquartile range). AS active surveillance, ISUP International Society of Urological Pathology, PSA prostate-specific antigen, PSAd prostate-specific antigen density, PZ peripheral zone, TZ transition zone

Variable	Total cohort (n=76)	Progressors (n=28)	Non-progressors (n=48)	p, progressors vs non-progressors
Age, years	66 (61–69)	66 (59–69)	66 (61–69)	0.79
PSA, ng/mL	5.0 (3.6–7.5)	5.7 (4.0–8.2)	4.5 (3.3–7.3)	0.17
Gland volume, mL	44.8 (36.0–72.0)	39.0 (26.0–47.8)	55.0 (39.3–80.8)	0.004
PSAd	0.10 (0.07–0.17)	0.13 (0.08–0.28)	0.09 (0.06–0.12)	0.003
AS follow-up, mo	42 (31–60)	39 (32–47)	44 (30–68)	0.09
Biopsy ISUP grade 1, n	58 (77%)	21 (28%)	37 (49%)	> 0.99
Biopsy ISUP grade 2, n	18 (23%)	7 (9%)	11 (14%)	
Target lesion in PZ, n	59 (78%)	21 (28%)	38 (50%)	0.80
Target lesion in TZ, n	17 (22%)	7 (9%)	10 (13%)	

to both expert-derived PRECISE and conventional DR, which is expected in a dataset where the final time point was linked to the ground truth assessment. This may have artificially increased the delta in progressors, leading to an overoptimistic performance of DR. However, in the real-life AS scenarios, the “final time point” used in DR cannot be known in advance, particularly if a decision to biopsy is based on visible tumour radiological progression that may or may not occur at a given time point. Hence, TSR provides a more viable and unbiased framework capable of tracking real-time personalised longitudinal changes in tumour imaging features at all time points throughout AS. Importantly, the improved performance of TSR with the addition of time series PSAd, which was not seen when delta-PSAd was added to DR, further demonstrates the added value of utilising all available serial clinical data. As a result, the AUC of the most holistic model combining TSR and time series PSAd was the closest to that of an expert-derived

PRECISE assessment, and it was also considerably higher than previously reported serial MRI performance pooled across different centres [7]. Finally, some radiomic features that passed the robustness analysis including *Maximum 2D Diameter Row* and *Surface Area* directly correspond to previously identified tumour measurements that are known to differ between progressors and non-progressors [32, 33]. These findings suggest that TSR with or without the addition of time series PSAd may offer a quantitative and interpretable solution to standardising serial MRI assessment in AS, with potential to increase radiologists’ confidence and improve performance across centres and healthcare systems by achieving the expert-derived PRECISE standard.

Limitations of this study include its small sample size, which was dictated by the stringent inclusion criteria related to MRI acquisition, ground truth assessment based on repeat targeted biopsy results, and presence of MRI-visible disease. In future work, we will aim to further develop the TSR

**Table 2** Summary performance characteristics of the best-performing MRI-derived radiomic models and PRECISE scores for predicting prostate cancer progression on AS. Individual performance characteristics are presented along with their 95% confidence intervals. Values with the best performance across the three approaches

are highlighted in bold. AUC area under the receiver operating characteristic curve, DR delta-radiomics, NPV negative predictive value, PPV positive predictive value, PRECISE Prostate Cancer Radiological Estimation of Change in Sequential Evaluation, TSR time series radiomics

Model	Sensitivity	Specificity	PPV	NPV	AUC
DR + PSAd	0.76 (0.60–0.91)	0.70 (0.57–0.83)	0.61 (0.45–0.77)	0.83 (0.71–0.94)	0.75 (0.64–0.87)
TSR + PSAd	0.66 (0.48–0.83)	<b>0.94 (0.87–1)</b>	<b>0.86 (0.72–1)</b>	0.82 (0.71–0.92)	<b>0.86 (0.78–0.94)</b>
PRECISE	<b>0.79 (0.65–0.94)</b>	0.89 (0.81–0.98)	0.82 (0.68–0.96)	<b>0.88 (0.78–0.97)</b>	0.84 (0.76–0.93)

framework by including studies performed on different vendors using different study protocols, as well as performing external validation alongside whole-gland radiomic feature extraction to include patients with no visible MRI lesions. Importantly, whole-gland segmentation will also be inclusive of patients developing new lesions over the course of AS, thereby overcoming an important limitation of the present approach. These steps will require even more rigorous pre-processing, calibration, feature selection, and validation steps to ensure that the resulting predictive models are generalisable and robust to factors unrelated to the true tumour progression.

In conclusion, we have developed the first predictive model analysing the combination of radiomic and clinical features in a time series fashion, which outperformed the conventional DR approach and showed comparable performance to expert radiologists. The proposed TSR concept could be applied to any clinical scenario involving serial imaging, ranging from continuous monitoring to evaluating treatment response, thus establishing a novel field in AI-driven radiology research.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00330-023-09438-x>.

**Funding** This research was supported by the NIHR Cambridge Biomedical Research Centre (NIHR203312). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. The authors also acknowledge support from Cancer Research UK [Cambridge Imaging Centre grant number C197/A16465; ACED Pilot Award A095792], Wellcome Trust Innovator Award [RG98755], CRUK National Cancer Imaging Translational Accelerator (NCITA) [C42780/A27066], and the Engineering and Physical Sciences Research Council Imaging Centre in Cambridge and Manchester. NS acknowledges support from the Gates Cambridge Trust. OB acknowledges support from Cancer Research UK and EPSRC joint award EDDCPJT/100022, as well as «Center of Photonics» supported by the Ministry of Science and Higher Education of the Russian Federation, contract No 075-15-2022-293.

**Data Availability** Access to the de-identified data can be provided upon a reasonable request to the corresponding author and is subject to appropriate regulatory approvals and data transfer agreements.

## Declarations

**Guarantor** The scientific guarantor of this publication is Prof Tristan Barrett.

**Conflict of interest** The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

**Statistics and biometry** Six of the authors have significant statistical expertise (LR, LA, ZL, TN, AZ, OB).

**Informed consent** Written informed consent was waived by the Institutional Review Board (NRES Committee East of England, UK).

**Ethical approval** Institutional Review Board approval was obtained.

**Study subjects or cohorts overlap** Some study subjects or cohorts have been previously reported in:

1. Sushentsev N, Rundo L, Blyuss O, et al (2021) Comparative performance of MRI-derived PRECISE scores and delta-radiomics models for the prediction of prostate cancer progression in patients on active surveillance. *Eur Radiol* 2021 1–10. <https://doi.org/10.1007/S00330-021-08151-X>
2. Sushentsev N, Rundo L, Blyuss O, et al (2021) MRI-derived radiomics model for baseline prediction of prostate cancer progression on active surveillance. *Sci Rep* 11:12917. <https://doi.org/10.1038/s41598-021-92341-6>

## Methodology

- Retrospective
- Observational
- Performed at one institution

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. <https://doi.org/10.3322/caac.21660>
2. Negoita S, Feuer EJ, Mariotto A et al (2018) Annual Report to the Nation on the Status of Cancer, part II: recent changes in prostate cancer trends and disease characteristics. *Cancer* 124:2801–2814. <https://doi.org/10.1002/cncr.31549>
3. Kinsella N, Helleman J, Bruinsma S et al (2018) Active surveillance for prostate cancer: a systematic review of contemporary worldwide practices. *Transl Androl Urol* 7:83–97
4. Van Hemelrijck M, Ji X, Helleman J et al (2019) Reasons for discontinuing active surveillance: assessment of 21 centres in 12 countries in the Movember GAP3 Consortium. *Eur Urol* 75:523–531. <https://doi.org/10.1016/j.eururo.2018.10.025>
5. Barrett T, Haider MA (2017) The emerging role of MRI in prostate cancer active surveillance and ongoing challenges. *AJR Am J Roentgenol* 208:131–139. <https://doi.org/10.2214/AJR.16.16355>
6. Ploussard G, Rouvière O, Rouprêt M et al (2022) The current role of MRI for guiding active surveillance in prostate cancer. *Nat Rev Urol* 2022:1–9. <https://doi.org/10.1038/s41585-022-00587-0>
7. Rajwa P, Pradere B, Quhal F et al (2021) Reliability of serial prostate magnetic resonance imaging to detect prostate cancer progression during active surveillance: a systematic review and meta-analysis. *Eur Urol*. <https://doi.org/10.1016/j.eururo.2021.05.001>
8. Ploussard G, Renard-Penna R (2021) MRI-guided active surveillance in prostate cancer: not yet ready for practice. *Nat Rev Urol* 18:77–78. <https://doi.org/10.1038/s41585-020-00416-2>
9. Moore CM, Giganti F, Albertsen P et al (2017) Reporting magnetic resonance imaging in men on active surveillance for prostate



- cancer: the PRECISE recommendations—a report of a European School of Oncology Task Force. *Eur Urol* 71:648–655. <https://doi.org/10.1016/j.eururo.2016.06.011>
10. Giganti F, Kasivisvanathan V, Allen C, Moore CM (2021) The importance of being PRECISE in prostate magnetic resonance imaging and active surveillance. *Eur Urol* 79:560–563. <https://doi.org/10.1016/j.eururo.2021.01.016>
  11. Sushentsev N, Rundo L, Blyuss O et al (2021) MRI-derived radiomics model for baseline prediction of prostate cancer progression on active surveillance. *Sci Rep* 11:12917. <https://doi.org/10.1038/s41598-021-92341-6>
  12. Sushentsev N, Rundo L, Blyuss O et al (2021) Comparative performance of MRI-derived PRECISE scores and delta-radiomics models for the prediction of prostate cancer progression in patients on active surveillance. *Eur Radiol* 2021:1–10. <https://doi.org/10.1007/S00330-021-08151-X>
  13. Nardone V, Reginelli A, Grassi R et al (2021) Delta radiomics: a systematic review. *Radiol Med* 126:1571–1583. <https://doi.org/10.1007/S11547-021-01436-7>
  14. Caglic I, Sushentsev N, Gnanaprasasam V et al (2020) MRI-derived PRECISE scores for predicting pathologically-confirmed radiological progression in prostate cancer patients on active surveillance. *Eur Radiol* 2020:1–10. <https://doi.org/10.1007/s00330-020-07336-0>
  15. Gnanaprasasam VJ, Barrett T, Thankapannair V et al (2019) Using prognosis to guide inclusion criteria, define standardised endpoints and stratify follow-up in active surveillance for prostate cancer. *BJU Int* 124:758–767. <https://doi.org/10.1111/bju.14800>
  16. Yushkevich PA, Piven J, Hazlett HC et al (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* 31:1116–1128. <https://doi.org/10.1016/j.neuroimage.2006.01.015>
  17. de Rooij M, Israël B, Tummers M et al (2020) ESUR/ESUI consensus statements on multi-parametric MRI for the detection of clinically significant prostate cancer: quality requirements for image acquisition, interpretation and radiologists' training. *Eur Radiol*. <https://doi.org/10.1007/s00330-020-06929-z>
  18. Barrett T, Padhani AR, Patel A et al (2020) Certification in reporting multiparametric magnetic resonance imaging of the prostate: recommendations of a UK consensus meeting. *BJU Int*. <https://doi.org/10.1111/bju.15285>
  19. van Griethuysen JJM, Fedorov A, Parmar C et al (2017) Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 77:e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
  20. Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14(14):749–762. <https://doi.org/10.1038/nrclinonc.2017.141>
  21. Nazarenko T, Whitwell HJ, Blyuss O, Zaikin A (2021) Parenchymal and synolytic networks revisited. *Front Genet* 12:1999. <https://doi.org/10.3389/FGENE.2021.733783/BIBTEX>
  22. Kim KG (2016) Book review: deep learning. MIT Press
  23. Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9:1735–1780. <https://doi.org/10.1162/NECO.1997.9.8.1735>
  24. Bouaziz M, Morchid M, Dufour R, et al (2017) Parallel long short-term memory for multi-stream classification. 2016 IEEE Work Spok Lang Technol SLT 2016 - Proc 218–223. <https://doi.org/10.48550/arxiv.1702.03402>
  25. Wu X, Wang HY, Shi P et al (2022) Long short-term memory model – a deep learning approach for medical data with irregularity in cancer prediction with tumor markers. *Comput Biol Med* 144:105362. <https://doi.org/10.1016/J.COMPBIOMED.2022.105362>
  26. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift. 32nd Int Conf Mach Learn ICML 2015 1:448–456. <https://doi.org/10.48550/arxiv.1502.03167>
  27. Saxe AM, McClelland JL, Ganguli S (2013) Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. 2nd Int Conf Learn Represent ICLR 2014 - Conf Track Proc. <https://doi.org/10.48550/arxiv.1312.6120>
  28. Kaji DA, Zech JR, Kim JS, et al (2019) An attention based deep learning model of clinical events in the intensive care unit. *PLoS One* 14:. <https://doi.org/10.1371/JOURNAL.PONE.0211057>
  29. N A, B H, H L, M L G, (2019) Breast lesion classification based on dynamic contrast-enhanced magnetic resonance images sequences with long short-term memory networks. *J Med Imaging (Bellingham)* 6:1. <https://doi.org/10.1117/1.JMI.6.1.011002>
  30. Sanyal J, Tariq A, Kurian AW et al (2021) Weakly supervised temporal model for prediction of breast cancer distant recurrence. *Sci Reports* 11(11):1–11. <https://doi.org/10.1038/s41598-021-89033-6>
  31. Sammut C, Webb GI (2010) Leave-one-out cross-validation. *Encyclopedia of Machine Learning*. Springer, US, Boston, MA, pp 600–601
  32. Sushentsev N, Caglic I, Rundo L et al (2022) Serial changes in tumour measurements and apparent diffusion coefficients in prostate cancer patients on active surveillance with and without histopathological progression. *Br J Radiol* 95:20210842. <https://doi.org/10.1259/BJR.20210842>
  33. Giganti F, Stavrinides V, Stabile A et al (2020) Prostate cancer measurements on serial MRI during active surveillance: it's time to be PRECISE. *Br J Radiol* 93:20200819. <https://doi.org/10.1259/bjr.20200819>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.