ORIGINAL ARTICLE

TRUSformer: improving prostate cancer detection from micro-ultrasound using attention and self-supervision

Mahdi Gilany[1] · Paul Wilson[1] · Andrea Perera-Ortega[1] · Amoon Jamzad[1] · Minh Nguyen Nhat To[2] ·
Fahimeh Fooladgar[2] · Brian Wodlinger[3] · Purang Abolmaesumi[2] · Parvin Mousavi[1]

Abstract

Purpose A large body of previous machine learning methods for ultrasound-based prostate cancer detection classify small regions of interest (ROIs) of ultrasound signals that lie within a larger needle trace corresponding to a prostate tissue biopsy (called biopsy core). These ROI-scale models suffer from weak labeling as histopathology results available for biopsy cores only approximate the distribution of cancer in the ROIs. ROI-scale models do not take advantage of contextual information that are normally considered by pathologists, i.e., they do not consider information about surrounding tissue and larger-scale trends when identifying cancer. We aim to improve cancer detection by taking a multi-scale, i.e., ROI-scale and biopsy core-scale, approach.

Methods Our multi-scale approach combines (i) an "ROI-scale" model trained using self-supervised learning to extract features from small ROIs and (ii) a "core-scale" transformer model that processes a collection of extracted features from multiple ROIs in the needle trace region to predict the tissue type of the corresponding core. Attention maps, as a by-product, allow us to localize cancer at the ROI scale.

Results We analyze this method using a dataset of micro-ultrasound acquired from 578 patients who underwent prostate biopsy, and compare our model to baseline models and other large-scale studies in the literature. Our model shows consistent and substantial performance improvements compared to ROI-scale-only models. It achieves 80.3% AUROC, a statistically significant improvement over ROI-scale classification. We also compare our method to large studies on prostate cancer detection, using other imaging modalities.

Conclusions Taking a multi-scale approach that leverages contextual information improves prostate cancer detection compared to ROI-scale-only models. The proposed model achieves a statistically significant improvement in performance and outperforms other large-scale studies in the literature. Our code is publicly available at www.github.com/med-i-lab/TRUSFormer.

Keywords Prostate cancer · Micro-ultrasound · Self-attention · Self-supervised learning

Mahdi Gilany and Paul Wilson have contributed equally to this work.

Purang Abolmaesumi and Parvin Mousavi are joint senior authors.

✉ Mahdi Gilany
mahdi.gilany@queensu.ca

[1] School of Computing, Queen's University, Kingston, Canada

[2] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada

[3] Exact Imaging, Markham, Canada

Introduction

Prostate cancer (PCa) is the second most common cancer in men worldwide. Early and accurate diagnosis and staging greatly increases the chances of successful treatment. The standard method for diagnosing and grading PCa is histopathological analysis of tissue samples of the prostate, typically obtained via trans-rectal ultrasound (TRUS) guided core biopsy. Because of the low sensitivity of conventional ultrasound in identifying lesions in the prostate [1], TRUS-guided biopsy is generally *systematic*, meaning that a large number of biopsy cores (typically 10–12) are uniformly sampled from the prostate rather than targeted to suspicious tissue

locations. Systematic biopsy has a high false negative rate [2] and carries considerable risks of biopsy-related adverse effects due to the large number of samples that must be obtained [3].

The emerging state of the art in targeted prostate biopsy pairs ultrasound with multi-parametric MRI (mp-MRI). Compared to ultrasound, mp-MRI has significantly higher sensitivity in detecting cancerous lesions [2], and in combination with ultrasound, it can be used to identify and sample biopsy targets [4]. However, this requires preprocedure MRI imaging, which is generally only available at large, well-funded urban centers and registration of MRI to US images, which can be laborious and error-prone. Therefore, the holy grail for cancer detection is to directly identify biopsy targets via ultrasound.

To improve ultrasound-based PCa detection, different ultrasound-based imaging modalities (e.g., shear wave elastography [5], quantitative ultrasound [6], Doppler ultrasound [7], temporal-enhanced ultrasound [8], micro-ultrasound [9], and contrast-enhanced ultrasound [10]) have been used. These methods, together with machine learning, have shown promise in interpreting complex ultrasound/tissue interactions. Classical machine learning models have been successfully used in conjunction with hand-selected features such as quantitative ultrasound [11] or shear wave elastography [12] coefficients. Despite promising results, manual feature selection is inherently limited to a small number and may miss complex properties that correlate with PCa. An alternative to classical machine learning methods are deep networks. Deep learning models have been proposed for PCa detection, for example from temporal-enhanced ultrasound ultrasound [13], micro-ultrasound [14–16], and contrast-enhanced ultrasound [17]. These studies generally formulate the problem of cancer detection as classification, where small regions of interest (ROIs) of an ultrasound image are classified into categories (benign or cancer). A notable exception is [18], which aims to classify *biopsy cores* using temporal-enhanced ultrasound but only achieves 68% AUROC. Classification of PCa at ROI scale is limiting for two reasons: i) ground-truth histopathology annotations (i.e., Gleason Scores) describe tissue properties at the scale of the entire biopsy core. Therefore, these only approximate the spread of cancer at the smaller ROI scale. Such weak labeling makes it challenging to train and evaluate robust ROI-scale models; ii) ROI-scale models only take into account individual ROIs and cannot take advantage of contextual information, such as the properties of surrounding tissue or larger-scale tissue patterns, when identifying cancer. Context is important in identifying PCa: for example, for histopathology labeling, PCa grade is determined from fine-scale features (e.g., degree of cell differentiation), medium-scale features (e.g., gland shape, size and uniformity, and infiltration patterns), and coarse-scale features (e.g., primary vs. secondary tumor types in the tissue sample) [19].

To address the limitations of conventional ROI-scale models, we propose a multi-scale approach (ROI-scale *and* core-scale) to PCa detection. At the ROI scale, following the successful approach of [16], we use a self-supervised convolutional network to extract features from ROIs of raw backscattered ultrasound data. At the core scale, inspired by successful methods in digital pathology (e.g [20]), we treat a core as a bag of ROIs and use multiple instance learning (MIL) to aggregate ROI features and predict tissue type for the core. Specifically, we use an attention-based transformer [21] model to aggregate ROI-level features, similar to attention-based MIL [22]. As a by-product, the attention maps produced by the transformer model allow us to identify the ROIs that contribute most to the prediction of tissue types in a biopsy core, hence localizing the tissue types at the ROI scale.

Using a large multicenter dataset of micro-ultrasound data, we demonstrate that our method significantly outperforms baseline approaches in distinguishing cancerous from benign cores. We demonstrate the additional explainability offered by the model via the attention maps. Furthermore, when visualizing the latent space of our multi-scale model, the data are not only clustered as cancer and benign but also distributed according to the severity of cancer (Gleason score). To the best of our knowledge, this is the first method for micro-ultrasound-based PCa detection that combines ROI-scale *and* core-scale analyses.

Materials

Data Acquisition We use private data from 578 patients, from five centers, who underwent systematic TRUS-guided biopsy as part of a clinical trial (clinicaltrials.gov NCT02079025) and after agreements are provided by patients. Biopsies are performed with an ExactVu micro-ultrasound [11] system (Exact Imaging, Markham, Canada). Prior to firing the biopsy gun, a single radio-frequency (RF) ultrasound image is recorded with 28 *mm* depth and 46 *mm* width in the sagittal plane. Images are paired with the histopathology findings for the corresponding biopsy core. For most patients, 10–12 biopsy cores are obtained. Among the 6607 total cores, 5727 are benign (defined as GS6 or lower), 675 are GS7, 134 are GS9, 60 are GS9, and 11 are GS10.

Data Selection First, the data of each center are divided to three patient-exclusive training, validation, and test sets with proportion 60%-15%-25%, respectively. Then, the patients of all centers are grouped into an overall train, validation, and test cohort. For performing cross-validation, different validations are selected from training set. Additionally following

Fig. 1 Summary of the proposed method. In Stage 1, batches of RF ultrasound ROIs are randomly selected from different cores (1) and used to train a feature extractor using the VICReg [23] method (2). In Stage 2, ROIs for the cores are concatenated together (3) and features are extracted using the pretrained feature extractor (4). Features are passed to the transformer network that uses self-attention to create a contextualized global representation of the core, from which it predicts the class (cancer or benign) (5)
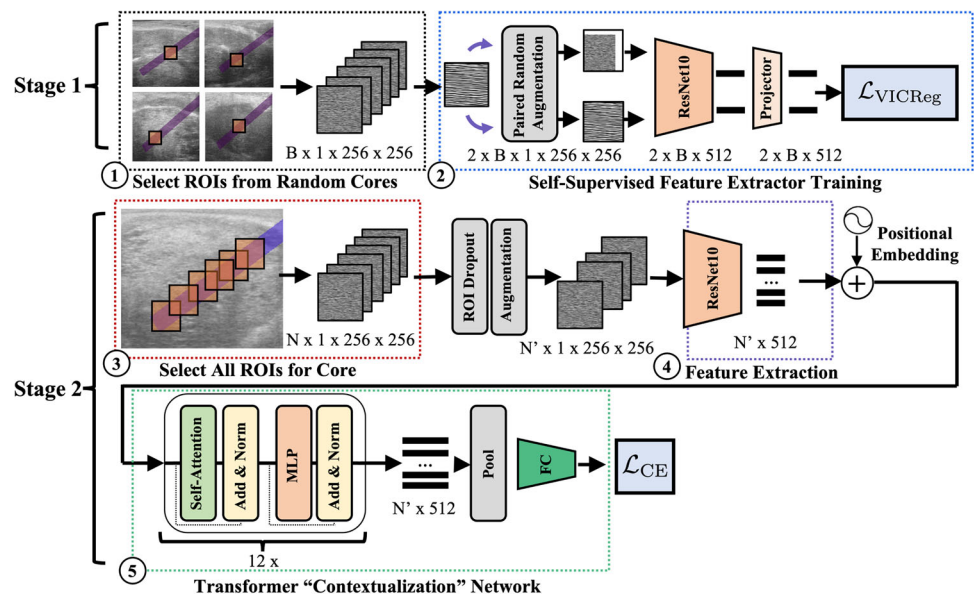
previous works [11, 14], cores with cancer involvement ≤ 40% of the core tissue are excluded in all datasets, and then benign cores are randomly selected to match the number of cancerous cores. Altogether, the training, cross-validation, and test sets include 864, 219, and 315 cores, respectively.

Data Preprocessing For each biopsy core, the needle trace region is identified on the corresponding ultrasound image. From the needle trace region, several overlapping ROIs, corresponding to tissue areas of size 5 *mm* by 5 *mm* are automatically selected, moving in 1 *mm* by 1 *mm* "strides." Following previous work [16], an ROI is considered inside the needle trace region if it overlaps with the region by at least 66 %. This selection process generates a bag of 55 ROIs per core. ROIs are resized from 1780 by 55 to 256 by 256 pixels by upsampling in the lateral and downsampling[1] in the axial directions. ROIs are instance-normalized by truncating outlier pixels ($>$ 4 standard deviations from the mean for that image) and rescaling intensity to the range of (0, 1).

Methods

Our proposed method consists of two main parts (Fig. 1). First, we use self-supervised learning (SSL) to train a convolutional neural network to extract informative feature representations for RF ultrasound ROIs. Second, we utilize MIL: We treat a biopsy core as a bag of ROIs, extract a bag of features from the ROIs using the pretrained network, and use an aggregation network to predict tissue type from the bag of features. The aggregator network uses attention to gather

context from the various input ROIs, allowing it to identify cancer more robustly than looking at individual ROIs in isolation.

Self-supervised feature extraction

Following [16], we use the variance–invariance–covariance regularization (VICReg) [23] SSL method to train a modified ResNet architecture as the backbone network to extract 512 features to represent each ultrasound ROI.

VICReg works as follows: Given an extracted RF ultrasound ROI x_i^{raw}, two augmentations t and t' are sampled from a distribution τ and applied to the ROI to produce two augmented "views," $x_i = t(x_i^{\text{raw}})$ and $x_i' = t'(x_i^{\text{raw}})$. Augmentations for RF ultrasound are studied in [16]. Different from that work, we use a strategy similar to [23], i.e., composition of random crops (scale $0.5-1.0$), and horizontal and vertical flips, which we found to work well. The augmented data are then mapped to representation vectors h_i and h_i' using the backbone network $h_i = f_\theta(x_i)$. Then, the representations are projected to z_i and z_i' using the projection network $z_i = g_\phi(h_i)$.

During self-supervised training, batches Z and Z' are collected from z_i and z_i', respectively (i.e., $Z = |concat|(\{z_i : i \in |batch_indices|\}))$. The weights θ and ϕ are adjusted to minimize the loss of VICReg network, $\ell(Z, Z')$. This loss is the weighted sum of three regularization terms, i.e., invariance $s(Z, Z')$, variance $v(Z)$, and covariance $c(Z)$ losses defined as follows:

$$\ell(Z, Z') = \lambda s(Z, Z') + \mu[v(Z) + v(Z')] \\ + \nu[c(Z) + c(Z')], \tag{1}$$

---

[1] The RF is bandlimited in the axial direction such that no loss of information occurs by downsampling.

where the weights λ, μ, and ν are hyperparameters (following [23]; we used 25, 25, 1, which worked well). The invariance loss $s(Z, Z')$ is the mean square error loss, maximizing the agreement between projection vectors and ensuring that only related information is captured. The variance loss $v(Z)$ acts to maintain the variance of features across batches, avoiding the collapse of the representation. The covariance loss $c(Z)$ minimizes feature redundancy by minimizing cross-correlations projected features.

Context-aware core classification

The second stage of our method is to train a biopsy core classification network. The input to this network is the bag of ROIs from the needle region of a core, stacked along the first axis, denoted by X ($X \in \mathbb{R}^{n \times 256 \times 256}$, where n is the number of ROIs within the core). The output of the network is a two-dimensional class score $\hat{y} \in \mathbb{R}^2$. The network first applies the pretrained feature extractor f_θ from the previous stage to each ROI, obtaining the bag of features $Z \in \mathbb{R}^{n \times 512}$. This is passed to the feature aggregator model g_ϕ to obtain class scores. During training, a dropout layer randomly drops some ROIs from the core to avoid seeing the same bag for the core each time.

The feature aggregator g_ϕ is a self-attention model based on a Transformer architecture [21] with 12 attention blocks, 8 heads, inner dimension 256 and MLP dimension 512.[2] The model uses learnable positional embeddings, which are added to the input features to make the model aware of the relative positions of ROIs within the core. The model also includes a linear layer which projects the 512 input dimensions to a smaller dimension d (we use $d = 256$). The main body is a transformer encoder consisting of alternating multi-head self-attention layers and multilayer perceptrons. A single self-attention head with input $Y \in \mathbb{R}^{n \times d}$ and output $Y' \in \mathbb{R}^{n \times d}$ uses the rule:

$$Y' = AV; \quad A = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right); \tag{2}$$

$$Q = W^Q Y; \quad K = W^K Y; \quad V = W^V Y, \tag{3}$$

for the learnable weights W^Q, W^K, and W^V (A is the so-called attention matrix). Self-attention blocks combine information about each ROI with relevant information about other ROIs in bag at each layer, gathering context into the output state which is pooled and classified using a linear layer.

---

[2] We empirically found this to be the best configuration after a manual search through: layers 1, 8, 12, 16, inner dimensions 64, 128, 256, 768, and heads 4, 8.

Attention mapping

Our model takes as input a sequence of ultrasound image ROIs and outputs a class prediction. We use the technique of Chefer et al. [24] to determine the extent to which each ROI contributes to either the "benign" or "cancerous" class of a core. Given an input sequence of length n, the relevancy map R is initialized as an identity matrix I_n, and propagated through the attention layers via the update rule $R = R + \bar{A}R$. Here, $\bar{A}$ can be thought of as a modified attention matrix that measures only contributions to a desired class. It is given by:

$$\bar{A} = E_h[(A \odot \nabla A)^+], \tag{4}$$

where ∇ denotes the gradient with respect to a class score (cancer or benign), $\odot$ denotes the Hadamard product, and E_h denotes the mean of different attention heads in a layer. At the last layer, the relevancy map is pooled to an $n \times 1$ vector whose i'th entry is the relative contribution of the i'th ROI for this class.

Experiments

Stage 1: We train the feature extractor for 200 epochs using a batch size of 64 and the Adam optimizer, warming up the learning rate for 10 epochs to $1e - 4$ and cosine annealing (Schedule A). The model is assessed, as it trains, by "online evaluation," periodically fitting a linear classifier to predict a class label (cancer or benign) for each ROI based on the features generated by the model. The best feature extractor is chosen based on online evaluation performance on the validation set. *Stage 2*: the core classification and feature extractor models are trained end-to-end for 70 epochs using the Adam optimizer. The best model is also chosen based on validation performance tested after each epoch. The learning rate (LR) for the transformer network is warmed up over 5 epochs to 1e-4 and cosine annealed. The LR for the feature extractor is warmed up over 10 epochs to 3e-5 and cosine annealed.

We compared TRUSFormer to several other methods which we implemented, trained and tested on our dataset. First, ROI-scale approaches: (a) end-to-end supervised learning using the same ResNet architecture as TRUSformer's feature extractor, (b) the method of [15] which enhanced supervised learning with methods to handle weak labels and uncertainty quantification, and (c) the method of [16] which used SSL pretraining followed by supervised ROI-scale fine-tuning. As in these studies, ROI predictions were converted to biopsy core predictions by taking the average of ROI predictions across the set of ROIs inside the core. Second, we compared our method to attention-based MIL [22] which is another example of a core-scale approach, and similar to our approach, aggregates ROI feature embeddings using an attention mechanism, but does not use the SSL pretraining step.

Table 1 Comparison of the proposed method (TRUSformer) with other proposed models for PCa detection

Method	AUROC	Avg-Prec	Sens	Spec
ROI-scale classifier				
Supervised	76.3 ± 1.7	73.7 ± 2.2	63.6 ± 12.9	**71.1** ± 7.7
EDL + Coteaching [15]	75.9 ± 0.5	72.6 ± 0.1	63.3 ± 7.5	70.8 ± 5.9
SSL + Linear Eval [16]	77.9 ± 1.1	74.8 ± 1.0	75.6 ± 4.3	66.8 ± 3.9
SSL + Fine-tuning [16]	78.2 ± 3.8	75.2 ± 4.4	**90.9** ± 7.0	41.3 ± 18.8
Core-scale classifier				
Attention MIL [22]	63.0 ± 3.2	60.4 ± 2.5	60.0 ± 3.6	59.9 ± 3.4
Gated Attention MIL [22]	65.0 ± 2.5	62.2 ± 2.9	56.7 ± 1.9	67.9 ± 4.1
TRUSformer (ours)	**80.3** ± 2.0	**78.7** ± 1.9	88.0 ± 11.6	51.2 ± 14.6

Bold values indicate the best performance
Metrics are based on performance in classifying benign versus cancerous (Gleason score 7 and above) biopsy cores. For ROI-scale classifiers, the core class prediction is the average of ROI for the core

Metrics are reported for the test data as average $+/-$ standard deviation across 16 runs.

Results

Table 1 summarizes the core classification performance of TRUSFormer compared to other methods implemented on our dataset.[3] It has the best performance in AUROC and Average Precision, with a considerable improvement over fully supervised ROI-scale methods (+4% AUROC, +5% Average Precision), large improvement over fully supervised core-scale methods (+15% AUROC, +16% Average Precision), and small but statistically significant[4] (+2% AUROC, +2.5% Average Precision) improvement over SSL methods. We conclude that (1) improvements in performance are due both to the introduction of SSL feature extraction from ROIs *and* the core-scale MIL, and (2) SSL seems necessary for core-scale methods to work, likely because good ROI features cannot be learned from scratch using core-scale MIL. Although TRUSformer has worse specificity than some methods, this could be improved by raising the detection threshold (at the cost of lowering sensitivity).

Figure 2 depicts visualizations of the model output. For 4 cores, we compare TRUSformer's prediction (bottom row) for the core to the ground truth (second row from bottom). We show attention maps produced by the model to see which ROIs in the core contribute to the "benign" (second row from top) vs. "cancer" (third row from top) class predictions. Cores 1–3 belong to one patient (A) and core 4 belongs to another patient (B). The model was consistently accurate for cores of patient A. For cores 1 and 2, the a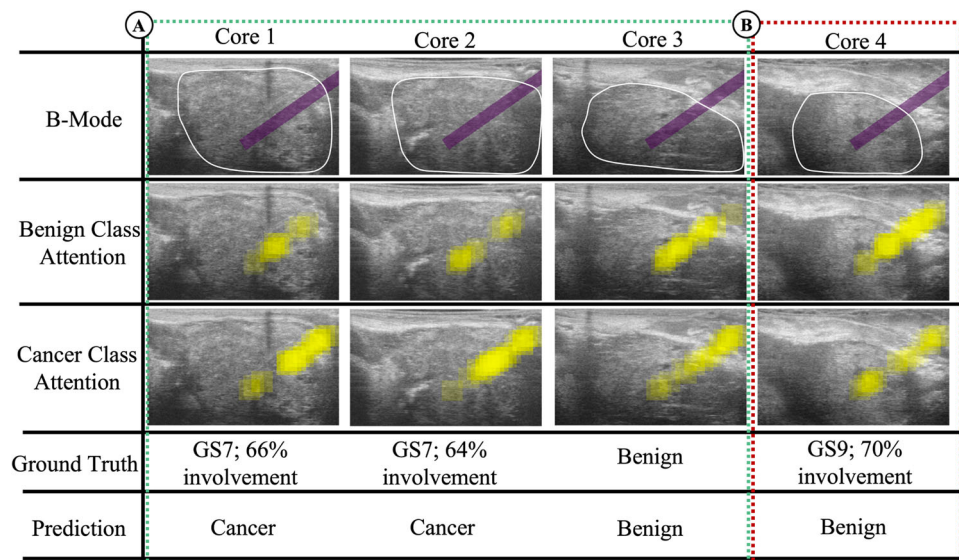ttention values for cancer are dominant compared to the attention values for benign, and the ratio of benign to cancer attention areas approximately correspond to the involvement of cancer in the cores. For core 3, benign attention was higher in the prostate region, whereas cancer attention appeared higher outside the prostate. The final core from patient B is a failure case: This core was incorrectly predicted as benign, and the model appears primarily attentive to the top right corner (outside the prostate region). Since non-prostatic tissue can neither identify nor exclude cancer, excessive attention to this region may explain the model's failure here. Because the model sees fewer non-prostatic ROIs during training, features of these patches may appear as outliers and draw model attention. This could be improved with future work by adding more non-prostatic ROIs in training.

Figure 3 visualizes the feature space of TRUSformer's second last layer (the pooling layer just before the classification layer) by projecting the features for all of the cores in the test set to two dimensions using UMAP [25]. Although the model was trained for binary classification (cancer vs. benign), we see apparent clustering of the cores according to their Gleason score. As Gleason scores partly depend on the distribution of tissue types throughout the core (not just on isolated ROIs), we hypothesize that the ability to see clusters with different Gleason scores is due to the core-scale nature of our model.

Table 2 summarizes our method to other imaging methods (note the limitations of this comparison due to differences in data, cross-validation methods, and performance reporting). Our method compares favorably with other micro-ultrasound methods, achieving similar sensitivity but much better specificity than PRIMUS [9] and comparable performance to QUS [11], despite the latter method requiring calibration to obtain the transducer point spread function using a phantom. Our model has the same sensitivity, but higher specificity than mp-MRI performance reported by the PROMIS clinical trial [2]. Since MRI fusion biopsy is the current state of the art for

---

[3] Performance is reported across all data centers. No statistically significant differences in performance between different data centers were observed.

[4] Two-tailed p-value < 0.05

	Core 1	Core 2	Core 3	Core 4
B-Mode				
Benign Class Attention				
Cancer Class Attention				
Ground Truth	GS7; 66% involvement	GS7; 64% involvement	Benign	GS9; 70% involvement
Prediction	Cancer	Cancer	Benign	Benign

Fig. 2 Qualitative analysis of model predictions and attention maps. The purple box on B-mode images in the first row show the biopsy core locations, and the white line shows the prostate outlines. For four cores, we compare TRUSformer's prediction (bottom row) for the core to ground truth (second row from bottom). We show attention maps produced by the model to see which ROIs in the core contribute to the "benign" (second row from top) vs. "cancer" (third row from top) class predictions. The left three cores (A) are from a patient for whom the model was consistently accurate, whereas the final core (B) is a failure case
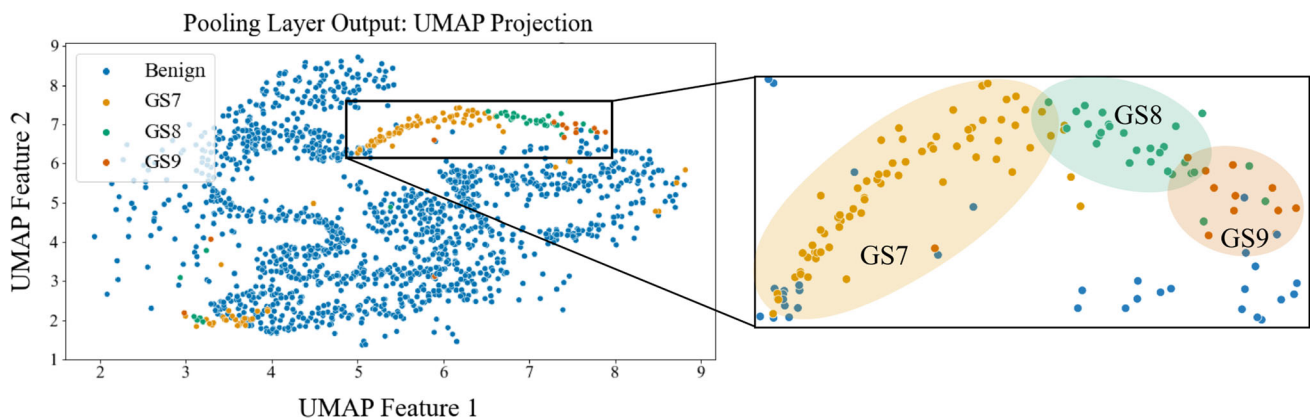
Fig. 3 TRUSformer pooling layer features are visualized for each core in test set via UMAP. Despite not using the Gleason score labels in training, the UMAP representations show clusters of different Gleason scores

Table 2 PCa detection performance reported in the literature for various imaging modalities, compared to our approach. Performances are reported for cancer with Gleason score≥7

Method	No. Patients	AUROC	Sens	Spec
PROMIS: TRUS biopsy [2]	576	–	48	99
PROMIS: PI-RADS + mp-MRI [2]	576	–	88	45
PRIMUS + micro-ultrasound [9]	93	60	80	37
QUS + micro-ultrasound [11]	75	79	85	56
TRUSformer + micro-US (ours)	125 (test set)	80	88	51

PCa detection, matching the performance of mp-MRI means that this model has the potential for allowing targeted biopsy with ultrasound alone. However, the method still has limited specificity and cannot confirm a cancer case without the need for pathology (from biopsy).

Conclusion

We proposed TRUSformer, a multi-scale and context-aware method for PCa detection from trans-rectal ultrasound. TRUSformer improves PCa detection by eliminating the weak labeling issue of ROI-scale models, and allowing the model to analyze multiple ROIs together in context to more robustly identify PCa. Although it was not trained explicitly to do so, TRUSformer appears to identify biopsy core-scale features that correlate well with distinct Gleason scores. Future work should further test TRUSFormer's generalization performance for cancer detection and grading via prospective studies.

Declarations

Conflict of interest Brian Wodlinger is the Vice President Engineering and Clinical at Exact Imaging, Canada.

Ethical approval Data used in this study was part of a clinical trial, approved by the ethical boards, and followed the rules of patient recruiting and informed consent of the trial (clinicaltrials.gov NCT02079025).

References

1. Smeenge M, de la Rosette JJ, Wijkstra H (2012) Current status of transrectal ultrasound techniques in prostate cancer. Curr Opin Urol 22(4):297–302
2. Ahmed HU, Bosaily AE-S, Brown LC, Gabe R, Kaplan R, Parmar MK, Collaco-Moraes Y, Ward K, Hindley RG, Freeman A, Kirkham A, Oldroyd R, Parker C, Emberton M (2017) Diagnostic accuracy of multi-parametric mri and trus biopsy in prostate cancer (promis): a paired validating confirmatory study. Lancet 389(10071):815–822
3. Madej A, Wilkosz J, Różański W, Lipiński M (2012) Complication rates after prostate biopsy according to the number of sampled cores. Cent Eur J Urol 65(3):116
4. Siddiqui MM, Rais-Bahrami S, Truong H, Stamatakis L, Vourganti S, Nix J, Hoang AN, Walton-Diaz A, Shuch B, Weintraub M, Kruecker J, Amalou H, Turkbey B, Merino M, Choyke P, Wood B, Pinto P (2013) Magnetic resonance imaging/ultrasound-fusion biopsy significantly upgrades prostate cancer versus systematic 12-core transrectal ultrasound biopsy. Eur Urol 64(5):713–719
5. Salomon G, Köllerman J, Thederan I, Chun FK, Budäus L, Schlomm T, Isbarn H, Heinzer H, Huland H, Graefen M (2008) Evaluation of prostate cancer detection with ultrasound real-time elastography: a comparison with step section pathological analysis after radical prostatectomy. Eur Urol 54(6):1354–1362
6. Oelze ML, Mamou J. Review of quantitative ultrasound: envelope statistics and backscatter coefficient imaging and contributions to diagnostic ultrasound. IEEE TUFFC
7. Kelly I, Lees W, Rickards D (1993) Prostate cancer and the role of color doppler us. Radiology 189(1):153–156
8. Moradi M, Abolmaesumi P, Siemens DR, Sauerbrei EE, Boag AH, Mousavi P (2008) Augmenting detection of prostate cancer in transrectal ultrasound images using svm and rf time series. IEEE Trans Biomed Eng 56(9):2214–2224
9. Ghai S, Eure G, Fradet V, Hyndman ME, McGrath T, Wodlinger B, Pavlovich CP (2016) Assessing cancer risk on novel 29 mhz micro-us images of the prostate: creation of the micro-us protocol for prostate risk identification. J Urol 196(2):562–569
10. Mannaerts CK, Engelbrecht MR, Postema AW, van Kollenburg RA, Hoeks CM, Savci-Heijink CD, Van Sloun RJ, Wildeboer RR, De Reijke TM, Mischi M, Wijkstra H (2020) Detection of clinically significant prostate cancer in biopsy-naïve men: direct comparison of systematic biopsy, multiparametric mri-and contrast-ultrasound-dispersion imaging-targeted biopsy. BJU Int 126(4):481–493
11. Rohrbach D, Wodlinger B, Wen J, Mamou J, Feleppa E (2018) High-frequency quantitative ultrasound for imaging pca using a novel micro-us scanner. Ultrasound Med Biol 44(7):1341–1354
12. Secasan CC, Onchis D, Bardan R, Cumpanas A, Novacescu D, Botoca C, Dema A, Sporea I (2022) Artificial intelligence system for predicting prostate cancer lesions from shear wave elastography measurements. Curr Oncol 29(6):4212–4223
13. Fooladgar F, To MNN, Javadi G, Samadi S, Bayat S, Sojoudi S, Eshumani W, Hurtado A, Chang S, Black P, Mousavi P, Abolmaesumi P (2022) Uncertainty-aware deep ensemble model for targeted ultrasound-guided prostate biopsy. In: 2022 IEEE 19th International symposium on biomedical imaging (ISBI), pp 1–5
14. Shao Y, Wang J, Wodlinger B, Salcudean SE (2020) Improving prostate cancer (pca) classification performance by using three-player minimax game to reduce data source heterogeneity. IEEE Trans Med Imaging 39(10):3148–3158
15. Gilany M, Wilson P, Jamzad A, Fooladgar F, To MNN, Wodlinger B, Abolmaesumi P, Mousavi P (2022) Towards confident detection of pca using high resolution micro-ultrasound. In: International conference on medical image computing and computer-assisted intervention, pp 411–420
16. Wilson PF, Gilany M, Jamzad A, Fooladgar F, To MNN, Wodlinger B, Abolmaesumi P, Mousavi P (2022) Self-supervised learning with limited labeled data for prostate cancer detection in high frequency ultrasound. arXiv preprint arXiv:2211.00527
17. Feng Y, Yang F, Zhou X, Guo Y, Tang F, Ren F, Guo J, Ji S (2018) A deep learning approach for targeted contrast-enhanced ultrasound based prostate cancer detection. IEEE/ACM Trans Comput Biol Bioinform 16(6):1794–1801
18. Javadi G, Samadi S, Bayat S, Pesteie M, Jafari MH, Sojoudi S, Kesch C, Hurtado A, Chang S, Mousavi P, Black P, Abolmaesumi P (2020) Multiple instance learning combined with label invariant synthetic data for guiding systematic prostate biopsy: a feasibility study. Int J Comput Assist Radiol Surg 15(6):1023–1031
19. Gordetsky J, Epstein J (2016) Grading of prostatic adenocarcinoma: current state and prognostic implications. Diagn Pathol 11(1):1–8
20. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ (2019) Clinical-grade computational pathology using

weakly supervised deep learning on whole slide images. Nat Med 25(8):1301–1309

21. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. Adv Neural Inf Process syst 30

22. Ilse M, Tomczak J, Welling M (2018) Attention-based deep multiple instance learning. In: International conference on machine learning. PMLR, pp 2127–2136

23. Bardes A, Ponce J, LeCun Y (2022) Vicreg: variance-invariance-covariance regularization for self-supervised learning. In: International conference on learning representations

24. Chefer H, Gur S, Wolf L. Generic attention-model explainability for interpreting bi-modal and encoder-decoder transformers. In: ICCV

25. McInnes L, Healy J, Melville J (2018) Umap: uniform manifold approximation and projection for dimension reduction. J Open Source Softw 3(29)