Poisson Ordinal Network for Gleason Group Estimation Using Bi-Parametric MRI

Yinsong Xu[1,2], Yipei Wang[1], Ziyi Shen[1], Iani J.M.B. Gayo[1], Natasha Thorley[3], Shonit Punwani[3], Aidong Men[2], Dean Barratt[1], Qingchao Chen[4], and Yipeng Hu[1]

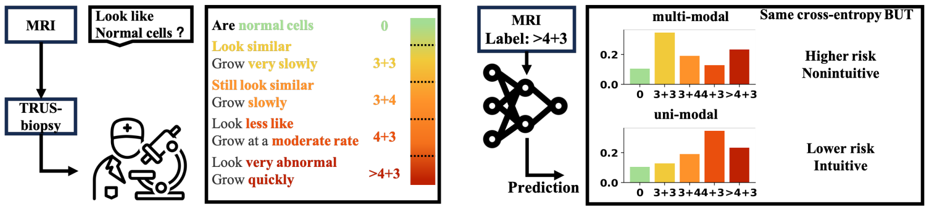
[1] Centre for Medical Image Computing, University College London, London, UK
[2] School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China
[3] Centre for Medical Imaging, University College London, London, UK
[4] National Institute of Health Data Science, Peking University, Beijing, China
xuyinsong@bupt.edu.cn, yipeng.hu@ucl.ac.uk

Abstract. The Gleason groups serve as the primary histological grading system for prostate cancer, providing crucial insights into the cancer's potential for growth and metastasis. In clinical practice, pathologists determine the Gleason groups based on specimens obtained from ultrasound-guided biopsies. In this study, we investigate the feasibility of directly estimating the Gleason groups from MRI scans to reduce otherwise required biopsies. We identify two characteristics of this task, ordinality and the resulting dependent yet unknown variances between Gleason groups. In addition to the inter-/intra-observer variability in a multi-step Gleason scoring process based on the interpretation of Gleason patterns, our MR-based prediction is also subject to specimen sampling variance and, to a lesser degree, varying MR imaging protocols. To address this challenge, we propose a novel Poisson ordinal network (PON). PONs model the prediction using a Poisson distribution and leverages Poisson encoding and Poisson focal loss to capture a learnable dependency between ordinal classes (here, Gleason groups), rather than relying solely on the numerical ground-truth (e.g. Gleason Groups 1-5 or Gleason Scores 6-10). To improve this modelling efficacy, PONs also employ contrastive learning with a memory bank to regularise intra-class variance, decoupling the memory requirement of contrast learning from the batch size. Experimental results based on the images labelled by saturation biopsies from 265 prior-biopsy-blind patients, across two tasks demonstrate the superiority and effectiveness of our proposed method. The source code is available at https://github.com/Yinsongxu/PON.git.

Keywords: Gleason group · Ordinal classification · MRI.

1 Introduction

Prostate cancer (PCa) is the most frequently diagnosed malignancy in 105 countries [3], potentially contributed by both its high prevalence and over-diagnosis



(a) Pathologist grading process (b) Multi-modal v.s. uni-modal prediction

Fig. 1. (a) Pathologists assign Gleason groups based on how the cancer cells look like healthy tissue under the microscope. Although cell growth is a continuous process, pathologists discretize it (dotted lines) into multiple groups. (b) Both predictions are incorrect and yield the same cross-entropy loss. However, a multi-modal distribution tends to produce misclassifications that deviate further from the label, resulting in different treatments. Consequently, it poses a higher clinical risk. Consider an image predicted as `3+3`. The multi-modal distribution appears counter-intuitive, as pathologists will not assign second highest confidence to `>4+3` due to the obvious difference.

using previous blind transrectal ultrasound (TRUS)-guided biopsy. As illustrated in Fig. 1(a), the current diagnostic process has improved the TRUS biopsy, by adopting magnetic resonance (MR) imaging as a triage test before biopsy [1]. However, this MR-targeted biopsy is subject to a significant (albeit much lower than blind biopsies) false-positive rate which leads to considerable unnecessary biopsies in men without cancer. In addition to the inherent specificity of MR imaging, the current false-positive rate may also be explained by the complexity and subjectivity in the biopsy and subsequent histology examination procedures.

Following the biopsy, the cancer-severity-indicating Gleason groups (on a scale of 1-5) are assigned to each biopsy-obtained tissue sample, during pathological examination [12]. In this study, we discuss the ordinal nature of this classification system and highlight that the dependency between Gleason groups is highly variable and yet quantified, which may be attributed to a lack of studies for predicting Gleason groups directly in literature. Deep learning approaches have been proposed for detecting and segmenting clinically significant lesions from MR images, based on radiologist labels [15], targeted-biopsy labels [10] and on labels obtained from prostatectomy patients - which is subject to a different set of challenges such as shifting patient cohort and registration [4], among other relevant tasks, e.g. needle placement [5] and registration [11].

Ordinal classification and class dependency. Gleason Groups 1-5 are defined with an inherent order, indicating increasingly aggressive cancer which predicts a poorer prognosis. Fig.1(b) illustrates an example of problems using one-hot encoding with cross-entropy loss, which assumes independence between classes. For instance, given a ground-truth label `>4+3`, the provided two examples, which would be classified as `3+3` and `4+3`, lead to the same loss values (indicating the same risks) if the class-independent encoding is used. However, the former mistake (classified as `3+3`) has a more serious consequence, which should have been indicated by a higher risk.

Ordinal encoding is considered an efficient label representation for this type of task [8], alleviating the above-discussed issue. Furthermore, we also investigate modelling the change between classes with an uni-modal distribution (e.g. Poison described in Sec. 2.2), for incorporating a prior on the dependency between these ordered classes. This uni-modal assumption reflects the fact that, in practice, it is more likely to have class changes (the random variable in the example Poison distribution models) near the "middle" classes (in this case, `3+3`→`3+4` and `3+4`→`4+3`) than at two ends (e.g. `4+3`→`>4+3`). This is also echoed with clinical practice in which more middle-class cancers are considered equivocal, therefore associating a higher possibility of changing their labels.

Unknown inter- and intra-class distributions. Furthermore, the Gleason groups have undergone two discretisation steps, assigning one of the five Gleason Patterns to represent a continuously changing pattern at the cellular scale, before being grouped into the five Gleason groups by ranking majority and secondary Gleason patterns [12]. For example, Gleason groups are assigned based on the appearance characterising Gleason patterns on obtained tissue specimens [12], as depicted in Fig.1(a). Patients may exhibit multiple groups of cells with varying proportions and locations. MRIs within the same group can display diverse appearances, leading to high intra-class variance as well as the dependent inter-class variance. This results in a complex unknown intra-/inter-class covariance, subject to subjectivity and uncertainty in either step for our application, and also to the biopsy sampling and imaging. This may challenge the design of the ordinal encoding with a predefined and/or fixed class distribution.

Motivated by the aforementioned observations, we propose the **P**oisson **O**rdinal **N**etwork (**PON**) for Gleason group estimation. First, we model the prediction distribution using a Poisson distribution. Instead of normalization (such as softmax) on the classifier's output as prediction distribution, we employ the output as the learnable Poisson's distribution parameter. We then introduce Poisson encoding, which assigns values based on the learnable Poisson distribution. For supervision, we minimise the Kullback-Leibler (KL) divergence between the predicted distribution and the encoded labels. It offers two key advantages. First, it encourages the network to learn the inherent order of classes. Second, it prompts the model to adapt the label distribution rather than the numerical groundtruth. Finally, we further propose a memory-bank-based contrastive learning to enhance representation by regulating the inter-/intra-class covariance.

Our contributions are threefold. 1) We introduce a novel Poisson encoding approach to model ordinal Gleason groups. 2) We propose the PONs, a novel ordinal classification network and its training strategy for the challenging Gleason group estimation. 3) We evaluate our method across two tasks, on a densely-labelled clinical bi-parametric MR data set, demonstrating its superiority over existing state-of-the-art approaches.

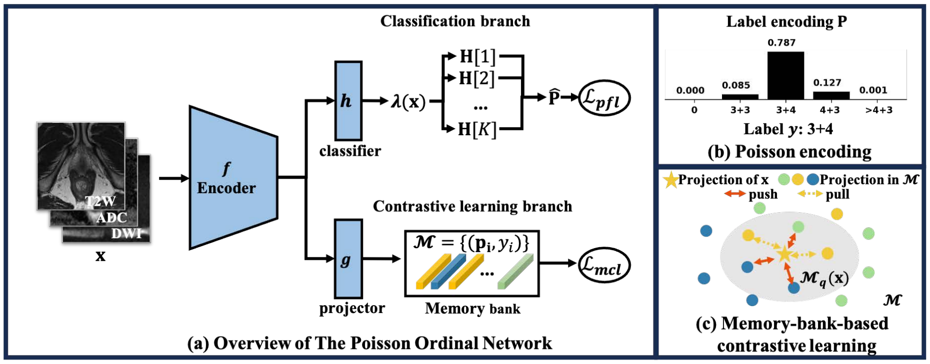


Fig. 2. (a) Our method represents the prediction probability using a Poisson distribution with the parameter $\lambda(\mathbf{x})$ output from the classifier, and stores the output from projector in a memory bank. (b) For ordinal classification, we introduce the inherent class ordering through Poisson encoding. (c) To regulate the inter-/intra-class covariance, we conduct contrastive learning between the sample and the memory bank, which pulls together features with the same label and pushes away different labeled features

2 Methods

2.1 Overview

The Poisson Ordinal Network (PON), depicted in Fig.2(a), consists of an encoder $f(\cdot)$ followed by two parallel branches: a classification branch and a contrastive learning branch. In the classification brunch, we model prediction distribution as a Poisson distribution with parameters derived from the output of classifier $h(\cdot)$. Then, we propose Poisson encoding to introduce the inherent class ordering into supervision, as shown in Fig.2(b), and the Poisson focal loss, $\mathcal{L}_{pfl}$ to adjust the loss contributions for different samples. In the contrastive learning branch, the projection output from the projector $g(\cdot)$ is stored in the memory bank $\mathcal{M}$. We conduct contrastive learning between the projection of input and those stored in the memory bank to improve the representation ability of the encoder as shown in Fig.2(c). The details are described as follows.

2.2 Poisson-based Prediction

We enforce uni-modal constraints by modelling the predicted probability using a Poisson distribution [2]. The Poisson distribution describes the probability of a specific number of events occurring within a fixed time interval. For a discrete random variable v, the probability mass function with parameter $\lambda > 0$ is:

$$\Pr(v = k) = \frac{\lambda^k \exp(-\lambda)}{k!}, \tag{1}$$

where k is the number of occurrences ($k = 0, 1, 2, \dots$).

For Gleason group estimation, we consider an advancing by 1 group as the "event", e.g. transitioning from 3+4 to 4+3. Consequently, the number of events corresponds to the class index (0-4). To model this, we replace the λ with the scalar output of classifier $\lambda(\mathbf{x}) = h \circ f(\mathbf{x})$, where $\circ$ denotes function composition ($\lambda(\mathbf{x}) \in \mathbb{R}^+$ is enforced by softplus function). Considering the K-class classification, we obtain the predicted distribution $\hat{\mathbf{P}} \in \mathbb{R}^K$ after normalization:

$$\hat{\mathbf{P}}[k] = \frac{\lambda(\mathbf{x})^k \exp(-\lambda(\mathbf{x}))/k!}{\sum_{k=1}^{K} \lambda(\mathbf{x})^k \exp(-\lambda(\mathbf{x}))/k!}. \tag{2}$$

For ease of implementation, we introduce the auxiliary variable $\mathbf{H} \in \mathbb{R}^K$:

$$\mathbf{H}[k] = \log[\lambda(\mathbf{x})^k \exp(-\lambda(\mathbf{x}))/k!] = k\log(\lambda(\mathbf{x})) - \lambda(\mathbf{x}) - \log(k!). \tag{3}$$

In this way, $\hat{\mathbf{P}}$ can be derived using the softmax function applied to $\mathbf{H}$.

2.3 Poisson Encoding and Poisson Focal Loss

The commonly-used combination of one-hot encoding and cross-entropy does not account for the inherent order of classes or explicitly encourage models to follow uni-modal distributions, as discussed in Sec. 1. To address this limitation, we propose a novel approach called Poisson encoding. Formally, we transform the label y to a Poisson distribution $\mathbf{P} \in \mathbb{R}^K$:

$$\mathbf{P}[k] = \frac{(y^k \exp(-y)/k!)^t}{\sum_{k=1}^{K} (y^k \exp(-y)/k!)^t}, \tag{4}$$

where t is the temperature hyperparameter to control the smoothness of distribution. As a supervision, we introduce Poisson focal loss, $\mathcal{L}_{pfl}$, which adapts focal loss [7] to Poisson encoding for ordinal classification:

$$\mathcal{L}_{pfl} = -(\mathbf{P}[y] - \hat{\mathbf{P}}[y])^\gamma \mathrm{KL}(\mathbf{P}||\hat{\mathbf{P}}), \tag{5}$$

where $\gamma \geq 0$ serves as the focusing parameter, and $\mathrm{KL}(\cdot||\cdot)$ is Kullback-Leibler divergence. When a sample is accurately classified, $(\mathbf{P}[y] - \hat{\mathbf{P}}[y])^\gamma$ approaches 0, contributing minimally to the overall loss. Consequently, the training emphasizes challenging samples. To summarise, the Poisson focal loss offers two key advantages over the with vanilla focal loss [7] and cross entropy: (1) it introduces the inherent order of classes through label encoding. (2) it encourages the modal to consider not only the expected numerical ground-truth class but also the distribution of the predictions.

2.4 Memory-bank-based Contrastive Learning

Taking into consideration of the diverse appearances of MRIs within the same group, discussed in Sec. 1, we propose contrastive learning to improve the representation learning. However, existing methods often heavily depend on batch size to enhance the range of contrast [14,16]. When dealing with 3D MRI data, the memory constraints associated with small batch sizes may introduce bias towards samples in the mini-batch. To decouple the contrastive learning from the batch-size, we introduce memory-bank-based contrastive learning. Formally, we propose the memory bank denoted as $\mathcal{M}$, which stores the projections of all samples $\mathbf{x}_i$, $\mathbf{p}_i = g \circ f(\mathbf{x}_i)$, along with their labels y_i, i.e. $\mathcal{M} = \{(\mathbf{p}_i, y_i)\}_{i=1}^{N}$, where N is dataset size. We update $\mathcal{M}$ at each forward iteration. Given an image $(\mathbf{x}, y)$, we select the q nearest elements to $g \circ f(\mathbf{x})$ in $\mathcal{M}$ denoted as $\mathcal{M}_q(\mathbf{x}) \subseteq \mathcal{M}$. We propose $\mathcal{L}_{mcl}$ to facilitate contrastive learning:

$$\mathcal{L}_{mcl} = -\frac{\sum_{(\mathbf{p}_i, y_i) \in \mathcal{M}_q(\mathbf{x})} \exp sim(\mathbf{p}_i, g \circ f(\mathbf{x})) \mathbb{1}\{y_i = y\}}{\sum_{(\mathbf{p}_i, y_i) \in \mathcal{M}_q(\mathbf{x})} \exp sim(\mathbf{p}_i, g \circ f(\mathbf{x}))}, \tag{6}$$

where $sim(\cdot, \cdot)$ is cosine similarity. With this configuration, features with the same label are pulled together and different labeled features are pushed away.

PONs could be trained in an end-to-end manner with the overall objective:$\mathcal{L} = \mathcal{L}_{pfl} + \mathcal{L}_{mcl}$.

3 Experiments

3.1 Experiment Setup

Dataset. Our method is evaluated on the PROMIS dataset [1] consisting of 265 publicly available MRIs from 262 patients. The Gleason Scores, obtained through template prostate mapping biopsy[13], were determined by expert uropathologists based on core biopsies taken at 5 mm intervals. Among all patients, 67 patients have no cancer, 49, 89, 31, and 26 patients were reported with Gleason group `3+3`, `3+4`, `4+3`, and `>4+3`, respectively. We utilize 3 image volumes from bi-parametric sequences, including T2-weighted (T2W), diffusion-weighted with high b-value (DWI), and apparent diffusion coefficient (ADC). All sequences are resampled to voxels of $0.5 \times 0.5 \times 2$ mm^3 with trilinear interpolation. All sequences are aligned with T2W images as the reference, and cropped into a $128 \times 128 \times 32$ region of interest (ROI) centred at the prostate's centroid.

Evaluation Metrics. We evaluate two tasks: Gleason group prediction (five classes), and detection of clinically significant cancer (primary and secondary definitions). For all tasks, we utilize five-fold cross-validation. Each experiment is repeated five times with different random seeds, and we report the mean and standard deviation. We adopt the following metrics for evaluation: accuracy (Acc), macro area under curve (AUC), quadratic weighted kappa (QWK), F1-score, sensitivity at specificity(Sen@Spec), and specificity at sensitivity(Spec@Sen).

Implementation Details. We employ the 3D ResNet18 architecture as the encoder. The classifier consists of a fully connected layer, and the projector is a two-layer multilayer perceptron. For optimization, we adopt the Adam optimizer with an initial learning rate of 1e-4, training for 60 epochs. All experiments are conducted using PyTorch on RTX6000 GPUs. The hyperparameters t, γ, and q are set to 0.1, 2 and 20, respectively. To keep a balanced number of samples for each class, we utilise a weighted sampler during training.

Table 1. Results of Gleason group prediction. (mean±standard)

Method	Acc↑	AUC↑	QWK↑	F1↑
CE	28.08±2.33	53.91±0.61	0.03±0.01	20.62±1.49
FocalLoss [7]	30.70±1.81	55.02±1.54	0.06±0.03	22.82±1.89
EMD [6]	**31.47**±2.10	57.84±2.58	**0.08**±0.04	25.03±2.18
ExpLoss [9]	31.15±0.62	56.78±1.89	**0.08**±0.01	24.67±1.45
OrdinalEncoding [8]	29.49±1.60	-	0.06±0.02	22.91±2.38
Softlabel[8]	25.13±3.66	56.19±3.40	0.05±0.03	22.39±3.36
Ours	29.35±1.82	**60.06**±1.22	0.07±0.02	**25.58**±2.25

Table 2. Results of detection of clinically significant cancer (mean±standard).

Method	Spec@Sen80%	Sen@Spec80%	Spec@Sen90%	Sen@Spec90%
Primary definition: Gleason score $\geq$ 4+3				
CE	32.46±5.31	31.76±3.55	15.74±5.35	16.77±5.47
FocalLoss[7]	34.02±5.20	38.53±7.64	19.34±5.64	21.47±6.61
EMD [6]	38.77±7.02	39.12±3.56	23.44±7.27	24.12±2.56
ExpLoss [9]	35.08±7.62	36.47±5.77	21.97±8.23	21.77±3.00
Softlabel[8]	40.82±3.68	31.17±1.71	24.26±8.32	15.59@9.33
Ours	**43.60**±6.06	**44.12**±6.77	**27.62**±8.80	**26.18**±5.91
Secondary definition: Gleason score $\geq$ 3+4				
CE	41.51±3.93	36.00±3.77	26.64±5.15	22.45±2.57
FocalLoss[7]	44.70±4.98	36.33±6.64	28.33±4.20	23.11±6.97
EMD [6]	**45.91**±1.03	36.44±7.84	**30.91**±3.08	20.44±4.99
ExpLoss [9]	44.39±3.44	38.78±5.29	26.06±4.20	22.67±2.66
Softlabel[8]	40.00±2.22	37.00±5.64	25.46±3.31	17.24±8.66
Ours	43.79±8.76	**39.00**±7.32	26.51±7.22	**25.00**±5.60

3.2 Results

Comparison Results. To evaluate the performance of PON, we compare our method with ordinal classification methods, including CrossEntropy (CE), re-weighting methods (FocalLoss [7]), ordinal classification loss (Earth mover's distance, EMD[6], and ExpLoss [9]), and encoding methods (OrdinalEncoding [8], Softlabel[8]). To ensure fairness, we reproduce all methods on our datasets with the same experimental settings. The results of Gleason group prediction and detection of clinically significant cancer are shown in Tab.1 and 2, respectively. It can be seen that PON shows a significant advantage with the highest performance in most metrics. Noticeably, PON outperforms other methods by great gains in the detection of primary definition significant cancer e.g., outperforms EMD 4.83% in Spec@Sen80%, and 5.00% in Sen@Spec 80% (p-value=1.58e-3 with the Mann–Whitney U test).

Ablation Study. To verify the impact of each component in the proposed PON, we conducted an ablation study, as detailed in Tab.3. We propose cross-entropy as the baseline. Notably, the inclusion of Poisson-based predictions and Poisson encoding significantly enhances the AUC and QWK, e.g. a considerable 5.13%

Table 3. Ablation study. PP: Poisson-based Prediction, PE:Poisson encoding.

PP	PE	$\mathcal{L}_{pfl}$	$\mathcal{L}_{mcl}$	Acc↑	AUC↑	QWK↑	F1↑
				28.08±2.33	53.91±0.61	0.03±0.01	20.62±1.49
✓				27.37±1.86	59.04±0.95	0.04±0.02	21.17±2.55
✓	✓			29.62±2.13	60.43±1.49	**0.07**±0.02	22.92±1.90
✓	✓	✓		28.85±1.93	60.41±2.49	**0.07**±0.02	24.14±1.09
			✓	**30.51**±2.89	55.36±0.62	0.06±0.03	23.18±2.15
✓	✓	✓	✓	29.35±1.82	**60.06**±1.22	**0.07**±0.02	**25.58**±2.25

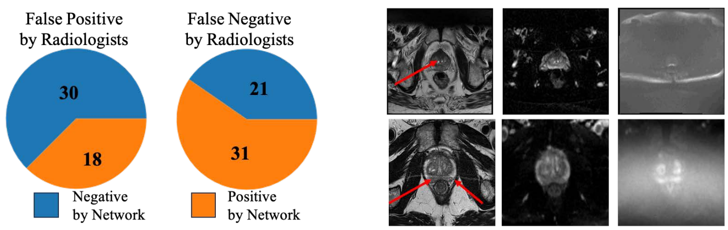


Fig. 3. Left: our method helps distinguish 30 negatives in 48 false positive samples under the same sensitivity, and 31 positive samples in 52 false negative samples under the same specificity by radiologists. **Right**: Two examples of the network predicting correctly and radiologists predicting wrong. Arrows indicate lesions.

improvement in AUC. It demonstrates the effectiveness of uni-modal constraints on prediction distributions, enhancing the model's ability to learn from ordinal data. Furthermore, when combined with $\mathcal{L}_{pfl}$, F1-score further improves. Subsequently, we introduced $\mathcal{L}_{mcl}$ to the baseline. The improvement suggests that this strategy can significantly enhance the network's data representation ability by regulating inter-/intra-class covariance. In summary, PON outperforms other models across most evaluation metrics, highlighting its suitability for ordinal classification tasks characterized by large intra-class variance.

Comparison with Radiologists on clinically significant cancer detection. MRI scans were reported at each centre by experienced urologic radiologists prior (and blind) to obtaining densely sampled histopathology labels [1]. This provides a unique opportunity to compare the machine-predicted with radiologists' ability in grading prostate cancer, with histopathology as ground-truth. Radiologists achieved sensitivity 75.82%/63.63%, and specificity 75.00%/71.11%, for primary and secondary definitions respectively. We compare our method with radiologists by adjusting the threshold to control sensitivity and specificity. Although the network did not outperform expert radiologists, it can assist radiologists' decision-making. As shown in Fig.3-left, the network can distinguish 30 negative samples in a total of radiological 48 false-positives with the same sensitivity, and 31 positive samples in 52 radiological false-negatives with the same specificity, quantitatively confirming its added clinical value. Two example cases in which the machine predictions can be used to co-pilot radiologists, either as a first- or second reader, are provided in Fig.3-right.

4 Conclusion

In this work, we present a Poisson ordinal network (PON) for Gleason group estimation. We identify two characteristics of this task, ordinality and a dependent and unknown variance between Gleason groups. To address this challenge, PONs models the prediction using a Poisson distribution and leverages Poisson encoding and Poisson focal loss to capture a learnable dependency between ordinal classes. Furthermore, to improve this modelling efficacy, PONs also employ contrastive learning with a memory bank. Experimental results based on the images labelled by saturation biopsies across two tasks demonstrate the superiority and effectiveness of our proposed method.

Acknowledgments. This work was supported by the International Alliance for Cancer Early Detection, a partnership between Cancer Research UK [C28070/A30912; C73666/A31378], Canary Center at Stanford University, the University of Cambridge, OHSU Knight Cancer Institute, University College London and the University of Manchester. This work was also supported by the China Scholarship Council.

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

References

1. Ahmed, H.U., Bosaily, A.E.S., Brown, L.C., Gabe, R., Kaplan, R., Parmar, M.K., Collaco-Moraes, Y., Ward, K., Hindley, R.G., Freeman, A., et al.: Diagnostic accuracy of multi-parametric mri and trus biopsy in prostate cancer (promis): a paired validating confirmatory study. The Lancet **389**(10071), 815–822 (2017)
2. Beckham, C., Pal, C.: Unimodal probability distributions for deep ordinal classification. In: International Conference on Machine Learning. pp. 411–419. PMLR (2017)
3. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **68**(6), 394–424 (2018)
4. Cao, R., Bajgiran, A.M., Mirak, S.A., Shakeri, S., Zhong, X., Enzmann, D., Raman, S., Sung, K.: Joint prostate cancer detection and gleason score prediction in mp-mri via focalnet. IEEE transactions on medical imaging **38**(11), 2496–2506 (2019)
5. Gayo, I.J., Saeed, S.U., Barratt, D.C., Clarkson, M.J., Hu, Y.: Strategising template-guided needle placement for mr-targeted prostate biopsy. In: MICCAI Workshop on Cancer Prevention through Early Detection. pp. 149–158. Springer (2022)
6. Hou, L., Yu, C.P., Samaras, D.: Squared earth mover's distance-based loss for training deep neural networks. arXiv preprint arXiv:1611.05916 (2016)
7. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
8. Potdar, K., Pardawala, T.S., Pai, C.D.: A comparative study of categorical variable encoding techniques for neural network classifiers. International journal of computer applications **175**(4), 7–9 (2017)

9. Qin, Z., Zhang, P., Li, X.: Ultra fast deep lane detection with hybrid anchor driven ordinal classification. IEEE transactions on pattern analysis and machine intelligence (2022)
10. Saha, A., Bosma, J., Twilt, J., van Ginneken, B., Yakar, D., Elschot, M., Veltman, J., Fütterer, J., de Rooij, M., et al.: Artificial intelligence and radiologists at prostate cancer detection in mri—the pi-cai challenge. In: Medical Imaging with Deep Learning, short paper track (2023)
11. Shen, Z., Yang, Q., Shen, Y., Giganti, F., Stavrinides, V., Fan, R., Moore, C., Rusu, M., Sonn, G., Torr, P., et al.: Collaborative quantization embeddings for intra-subject prostate mr image registration. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 237–247. Springer (2022)
12. Stark, J.R., Perner, S., Stampfer, M.J., Sinnott, J.A., Finn, S., Eisenstein, A.S., Ma, J., Fiorentino, M., Kurth, T., Loda, M., et al.: Gleason score and lethal prostate cancer: does 3+ 4= 4+ 3? Journal of Clinical Oncology **27**(21), 3459 (2009)
13. Valerio, M., Anele, C., Charman, S.C., van der Meulen, J., Freeman, A., Jameson, C., Singh, P.B., Emberton, M., Ahmed, H.U.: Transperineal template prostate-mapping biopsies: an evaluation of different protocols in the detection of clinically significant prostate cancer. BJU international **118**(3), 384–390 (2016)
14. Wang, P., Han, K., Wei, X.S., Zhang, L., Wang, L.: Contrastive learning based hybrid networks for long-tailed image classification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 943–952 (2021)
15. Yan, W., Chiu, B., Shen, Z., Yang, Q., Syer, T., Min, Z., Punwani, S., Emberton, M., Atkinson, D., Barratt, D.C., Hu, Y.: Combiner and hypercombiner networks: Rules to combine multimodality mr images for prostate cancer localisation. Medical Image Analysis **91**, 103030 (2024)
16. Zhang, Y., Chen, J., Wang, K., Xie, F.: Ecl: Class-enhancement contrastive learning for long-tailed skin lesion classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 244–254. Springer (2023)