

RESEARCH

Open Access



# The value of artificial intelligence for detection and grading of prostate cancer in human prostatectomy specimens: a validation study

Maíra Suzuka Kudo<sup>1</sup>, Vinicius Meneguette Gomes de Souza<sup>2\*</sup> , Carmen Liane Neubarth Estivallet<sup>2</sup>, Henrique Alves de Amorim<sup>1</sup>, Fernando J. Kim<sup>3</sup>, Katia Ramos Moreira Leite<sup>2</sup> and Matheus Cardoso Moraes<sup>1</sup>

## Abstract

**Background:** The Gleason grading system is an important clinical practice for diagnosing prostate cancer in pathology images. However, this analysis results in significant variability among pathologists, hence creating possible negative clinical impacts. Artificial intelligence methods can be an important support for the pathologist, improving Gleason grade classifications. Consequently, our purpose is to construct and evaluate the potential of a Convolutional Neural Network (CNN) to classify Gleason patterns.

**Methods:** The methodology included 6982 image patches with cancer, extracted from radical prostatectomy specimens previously analyzed by an expert uropathologist. A CNN was constructed to accurately classify the corresponding Gleason. The evaluation was carried out by computing the corresponding 3 classes confusion matrix; thus, calculating the percentage of precision, sensitivity, and specificity, as well as the overall accuracy. Additionally, k-fold three-way cross-validation was performed to enhance evaluation, allowing better interpretation and avoiding possible bias.

**Results:** The overall accuracy reached 98% for the training and validation stage, and 94% for the test phase. Considering the test samples, the true positive ratio between pathologist and computer method was 85%, 93%, and 96% for specific Gleason patterns. Finally, precision, sensitivity, and specificity reached values up to 97%.

**Conclusion:** The CNN model presented and evaluated has shown high accuracy for specifically pattern neighbors and critical Gleason patterns. The outcomes are in line and complement others in the literature. The promising results surpassed current inter-pathologist congruence in classical reports, evidencing the potential of this novel technology in daily clinical aspects.

**Keywords:** Convolutional neural network, Artificial intelligence, Prostate cancer, Pathology, Biopsy image

\*Correspondence: vmeneguette@gmail.com

<sup>2</sup> Laboratory of Medical Investigations Number 55 of the Sao Paulo University Medical School – FMUSP, Avenida Dr Arnaldo, 455, sala 2145 – Cerqueira Cesar, Sao Paulo, SP CEP 01246-903, Brazil  
Full list of author information is available at the end of the article

## Background

Prostate cancer is the fifth deadliest cancer in the world and the second most frequent among men. Globocan 2020 data count 1,414,259 new diagnoses and 375,304 deaths in a single year. It is known that asymptomatic early-stage, well treated disease is associated with up to 98% long-term survival according to guidelines [1–3].



© The Author(s) 2022. **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/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Pathological diagnosis still raises divergence between specialists, even despite using the same classical Gleason grading (GG) studies [4–7]. Tumor grading is a cornerstone to guide cancer therapy, raising concerns among practitioners worldwide due to heterogeneity [8–10] and the need for novel tools. GG, described elsewhere [4], is purely visual analytic, throwing the spotlight on the standardization capability of artificial intelligence, with growing evidence in the deep learning field [11, 12].

Since 1998, Convolutional Neural Network (CNN) has been established and become a popular technique for image classification [13]. Hence, different topologies have been constructed and evaluated for many applications. Some groups presented data on high-performance CNN Gleason grading standards classification [13–18]. However, in every context, specific possible limitations may be found in data organization, computational cost, and scope or focus of evaluation. These possible limitations may compromise more accurate interpretations for specific contexts. Consequently, it is still necessary for medical literature publications in this field to complement, strengthen, and support the CNN potential and its continuous evolution for this application. Therefore, the purpose of this study is to construct and evaluate a deep learning model for graduate the relevant Gleason patterns.

## Method

### Hypothesis

This study's hypothesis is that a CNN system could be efficient to graduate the relevant GG patterns. Accordingly, we managed to merge computing engineering science with high-standard pathological reports to construct, train and evaluate a specific CNN system to classify G3, G4, and G5.

### Design

The study design is divided into two main flow of work, Clinical Actions, and Computational Actions, with corresponding procedures.

### Clinical actions

The laboratory of medical investigations from the Medical School of the University of Sao Paulo (FMUSP) collected 32 previously reported radical prostatectomy specimens. They were colored by hematoxylin and eosin (H&E) method, and scanned by an Aperio<sup>®</sup> microscope, the slide's images were analyzed, and Gleason patterns 3, 4, and 5 were delineated by the corresponding specialists. Additionally, images from "the Prostate Cancer Grade Assessment (PANDA) Challenge" were also added to the dataset. Hence, providing a richer dataset to support to the model to improve performance, alongside robustness and capacity of generalizing. PANDA includes

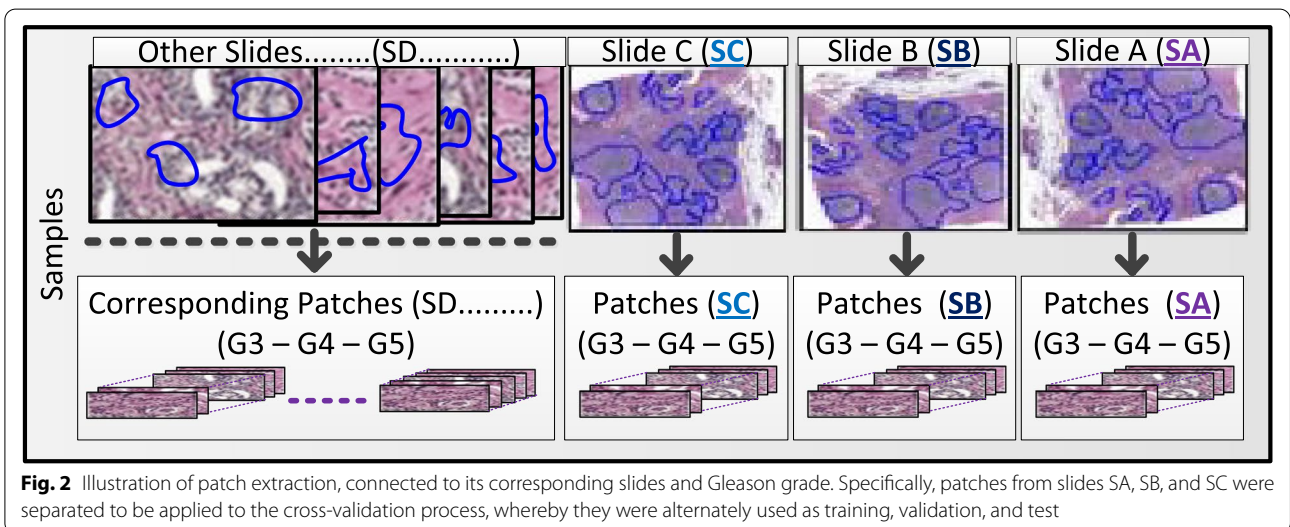
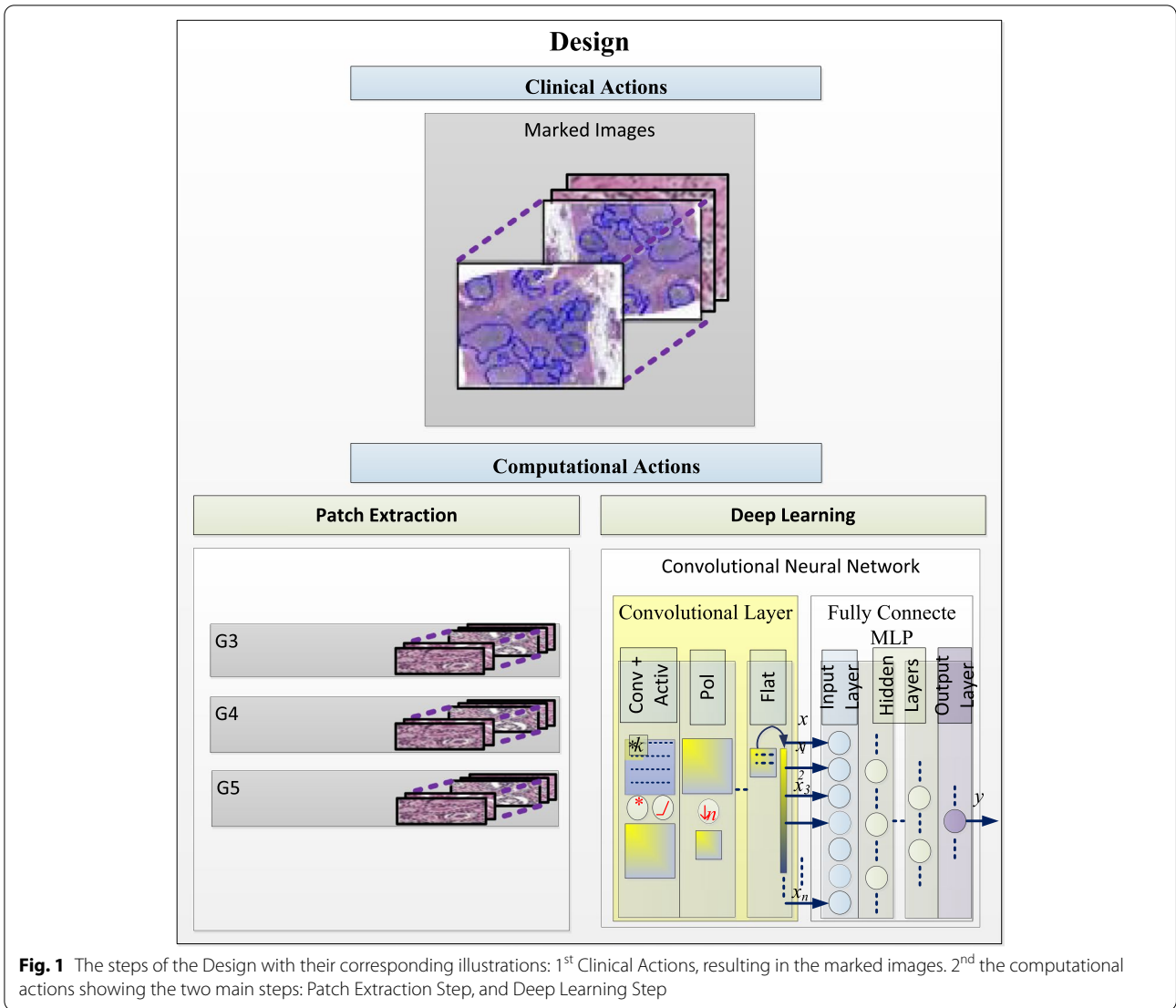
two open-access datasets: Karolinska Institute (images divided into background, benign and cancerous tissue) and Radboud University Medical Center (13 images divided into background, stroma, benign tissue, and Gleason patterns 3, 4, and 5). To ensure methodological similarity, only Radboud images were used to improve the initial training sample. All the samples underwent the same screening process previously presented [19].

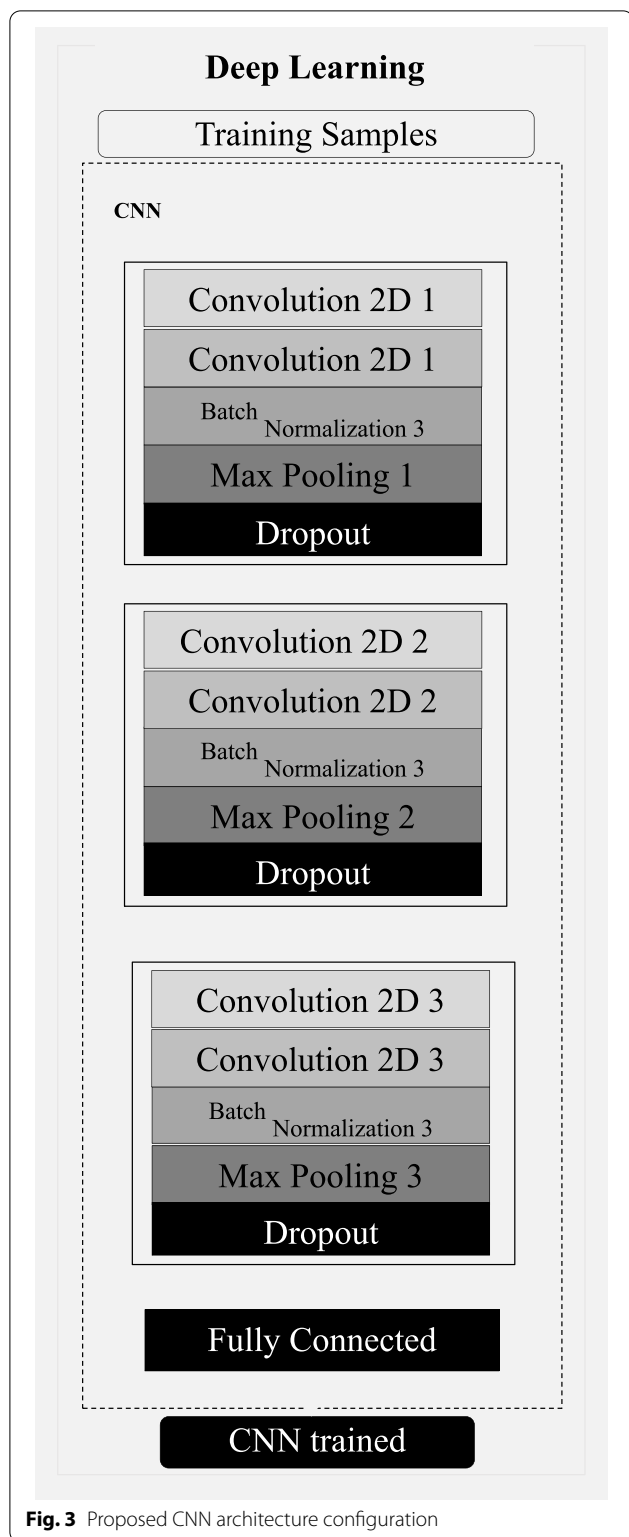
### Computational actions

The computational procedures are composed by *Patch Extraction Step*, *Deep Learning Step* (Fig. 1).

*Patch extraction step* This step consists in building a dataset of extracted patches, small sample images sized  $256 \times 256$  pixels, with a corresponding  $20 \times$  zoom of previously marked regions. The zoom and patch size values were chosen considering they are adequate for individual and combined clinical element identification. Considering this parameter, we obtained a total of 6982 patches (5036 from FMUSP prostatectomy samples and 1946 from the PANDAS Challenge dataset). As a result, patches of Gleason's 3, 4, and 5 were obtained and can be identified by their corresponding Slide (Fig. 2).

*Deep learning step* This step included topology construction considering a combination of multiple blocks. Previous architectures, characteristics, and important elements were considered to establish the structure proposed. Several experiments were performed using features of complex neural nets, combined blocks, and learning methods, resulting in the obtention of a high-performing architecture for this purpose, as shown in Fig. 3. The neural net input starts with two convolutional layers containing 32 filters with  $5 \times 5$  kernel, and 64 filters with  $5 \times 5$  kernel, respectively. The number of filters is related to the feature extraction diversity in the input – the more filters, the more complementary features are extracted and considered to support decision. The batch normalization layer standardizes output values of the corresponding layer, decreasing the chances of value range saturation. Max Pooling decreased the feature matrix dimension, allowing only the best parameters to proceed; the first sequence ends in a dropout layer. The other sequences (Second and Third) work similarly, except for the number of the filter (64 and 128 in the second, 128 and 256 in the third). Lastly, information goes through the Fully Connected layer containing 512 neurons within the hidden layer, batch normalization; the dropout of 0.5. SoftMax was used as an activation function. RMSprop was chosen as an optimizer for training; thus, the neural net can output the images classified as Gleason patterns 3, 4, and 5.





The patch images extracted (Fig. 2) were applied to the architecture described (Fig. 3) for training and evaluation. The separation of training, validation, and test groups was performed using the 80%, 10%, and 10%

ratios. In addition, to obtain the most from our image set, we carried out a 3-time k-fold cross-validation, as shown in Fig. 4.

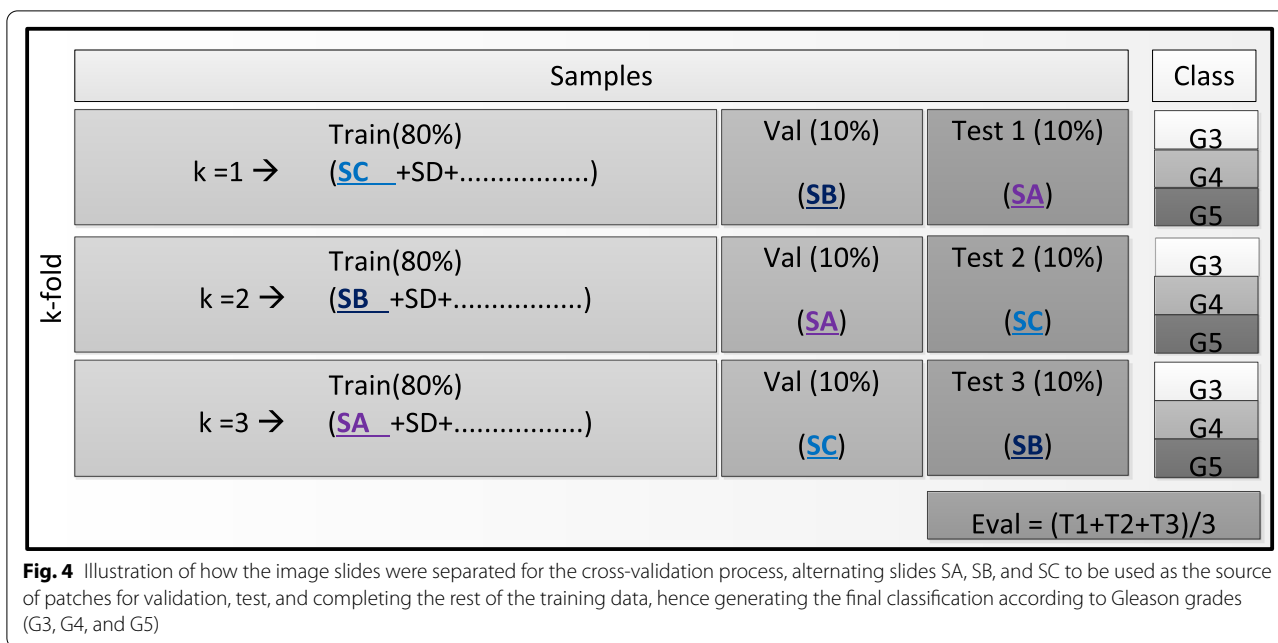
Specifically, this cross-validation took patches from slides SA, SB, and SC (Figs. 2 and 4) to be individually used as validation and test, and the patches from the spare slides complete the training data. This prevented patches of the same patient and slides from being present in the training, validation, and test groups; accordingly, providing a wider context variation, and composed outcome, thus leading to a more reliable and unbiased outcome, supporting better interpretation for corroboration. Slides SA, SB, and SC were chosen because each of them had the most balanced distribution of patches for Gleason 3, 4, and 5. The corresponding distribution and number of patches used for each k of the k-fold can be seen in Table 1. Finally, to improve model accuracy together with robustness, minimizing potential overfitting, data augmentation was performed before being applied to the neural net; specifically, this process includes random rotations, brightness, and zoom.

The evaluation process occurred in two steps for each k during the proposed k-fold cross-validation (Fig. 4). The first was the training and validation step, and the second was the test, computing typical and additional parameters of performance for better interpretation. The training and validation step evaluate the potential of learning the current application; thus, the parameters accuracy and loss were computed to validate this step. Once high rate of accuracy and loss were achieved, the corresponding trained CNN topology were saved and submitted to the test step. The test step corroborates the previously obtained accuracy; as well as, measuring robustness and potential of generalize classification. During the test step, the corresponding image samples were applied to the trained CNN topology to be classified. As a result, generating the regarding confusion matrix; hence, allowing computing precision, sensitivity, and specificity, for interpretation of possible consequences.

**Results**

**Training and validation step**

As can be seen in Fig. 5, among the different groups of the three k-folds, the training and validation curves were convergent in terms of accuracy and loss. For training and validation, the accuracy reached about 98%, and loss variates around 1% to 2.5%. In addition, the differences shown in Fig. 5 demonstrate there is no considerable underfitting or overfitting. Considering the proposed context, using slides from different patients for training and validation, in addition to the two different image sources to train the topology, the learning potential of the network is demonstrated. The total training processing



**Table 1** Dataset separation considering, approximately, 80%, 10%, 10% ratio, for training, validation, and test of each corresponding k, respectively

k-fold	Class	Train(n)~80%	Val(n)~10%	Test(n)~10%
k=1	Slides	SC+SD+...	SB	SA
	Gleason 3	614	73	69
	Gleason 4	2205	271	258
	Gleason 5	2801	348	349
k=2	Slides	SB+SD+...	SA	SC
	Gleason 3	593	69	94
	Gleason 4	2224	259	251
	Gleason 5	2821	352	325
k=3	Slides	SA+SD+...	SC	SB
	Gleason 3	589	94	73
	Gleason 4	2185	281	268
	Gleason 5	2807	339	352

time was estimated at 1200 s using one of our laboratory computers (Intel Core i7 3.50 GHz configurations, NVIDIA GeForce GTX 1060 Graphics, 16 GB RAM, 2 TB Hard Disk).

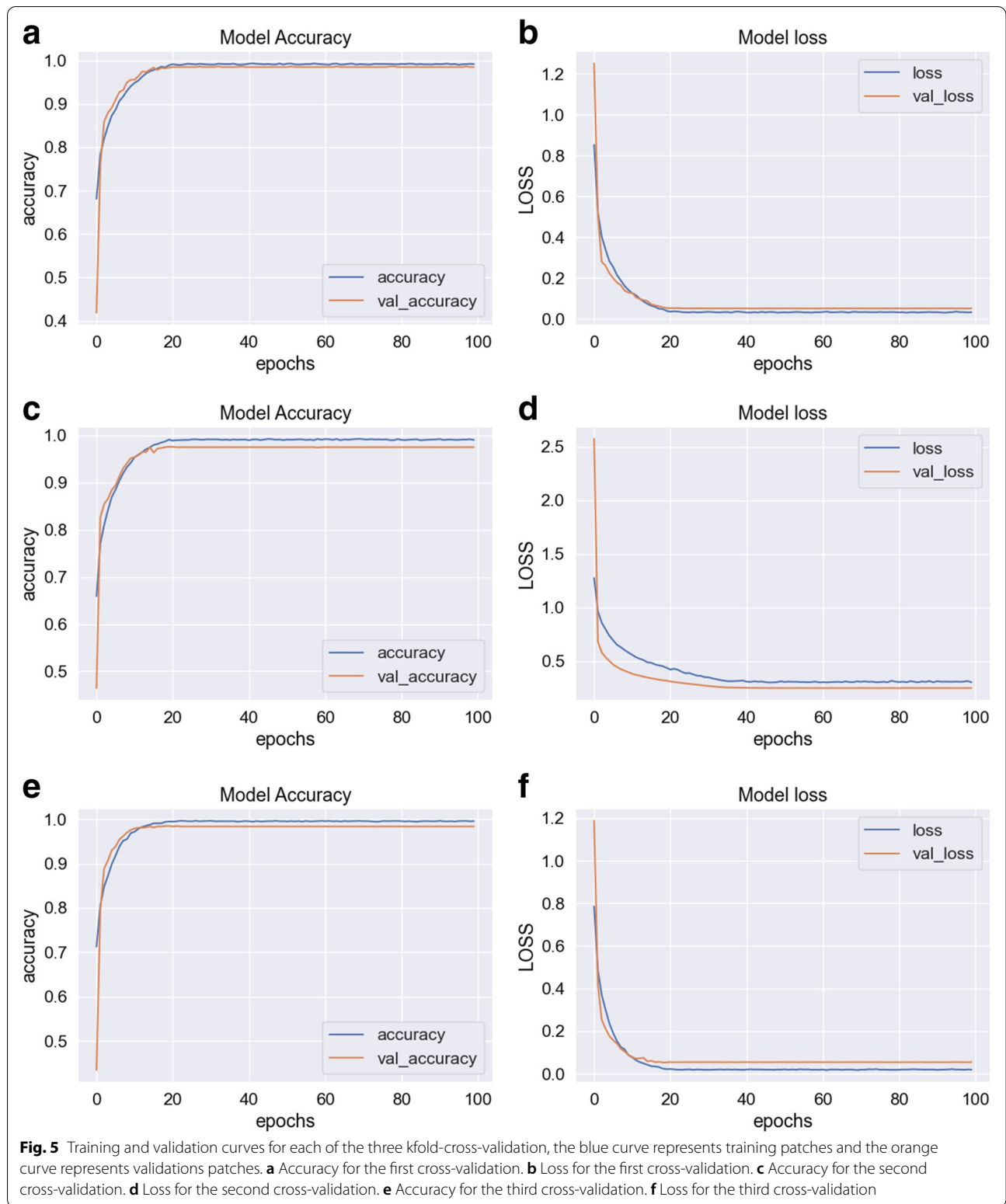
**Test step**

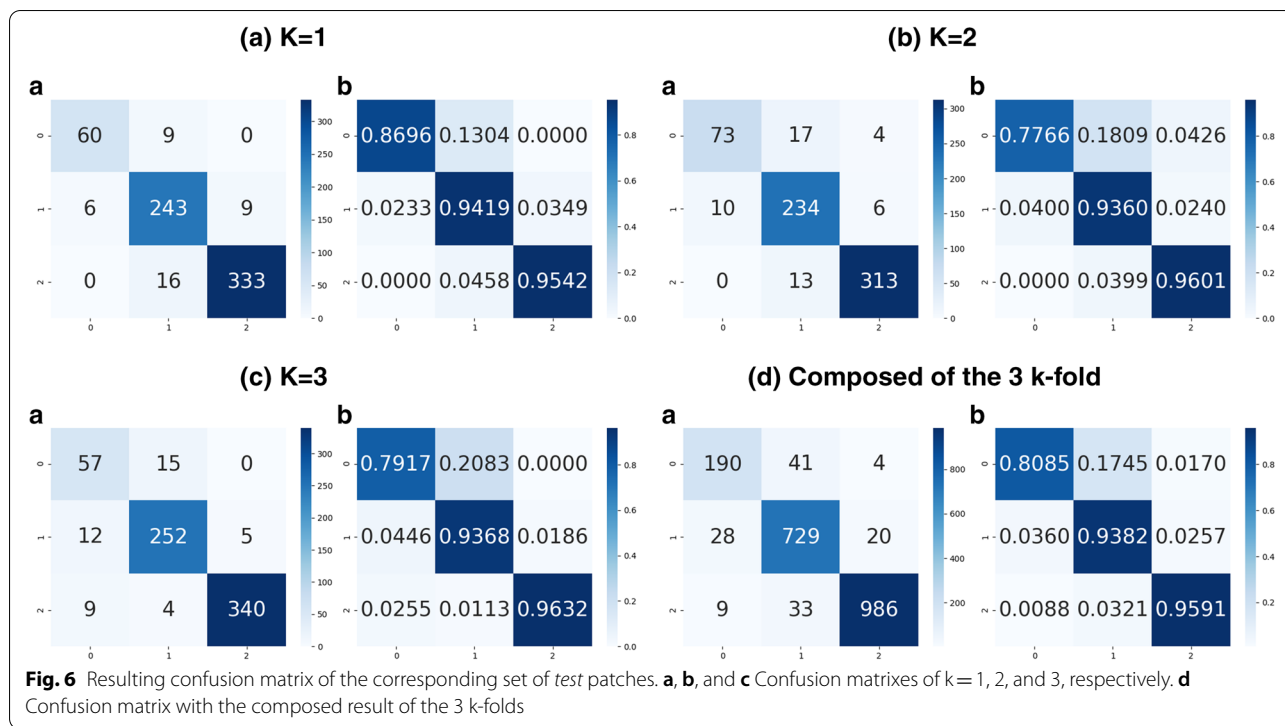
After the training and validation of each k, the corresponding trained model was subjected to the test step (Fig. 4). The test step was carried out using the corresponding test data of each cross-validation (Table 1). For each k, the corresponding test data had never been seen by the model; hence, the network was blinded for

every set of the test image to prevent bias. The confusion matrix with the results of each k-fold analysis group is presented in Figs. 6a, b, and c; additionally, the composition of all k-fold results is presented in Fig. 6d. From the resulting confusion matrixes, we obtained the pertinent metrics of efficacy, in addition to the general accuracy; precision, sensitivity, and specificity were also computed considering each class as a target object, therefore measuring the potential performance of the model to separate each class. Table 2 summarizes the findings with accuracy, precision, sensitivity, and specificity. Accuracy of around 95% was achieved in the final evaluation data from tests results (Table 2). Additionally, values above 80% and close to 98% were achieved for precision, sensitivity, and specificity for the different classes of Gleasons (Table 2). Considering the blind test applied, the classification potential of the network was evidenced.

Specifically, the lowest performance values for Gleason 3, compared to patterns 4 and 5 (approximately 81% and 87%, for precision and sensitivity), can be explained by the lower number of samples for this class. Considering the way the patches were obtained, Gleason 3 is lower grade, more similar to noncancer tissue, and has lower volume in prostate specimens resulting in less material for analytical purposes. The consequences of this difference in the number of samples can also be seen with the lower specificity of the other two classes compared to Gleason 3, see Table 2. Considering the clinical relevance of patterns 4 and 5 and eventual non-relevant pattern 3 findings, this disadvantage may represent minimal clinical relevance.







**Table 2** Precision, Sensitivity, Specificity, and Accuracy for each of the corresponding  $k$

<i>k</i> -fold	classe	Precision(%)	Sensitivity (%)	Especificity (%)	Accuracy(%)
<i>k</i> = 1	Gleason 3	90.91	86.95	99.01	94
	Gleason 4	90.67	94.18	94.02	
	Gleason 5	97.36	95.41	97.25	
<i>k</i> = 2	Gleason 3	87.95	77.65	98.26	95.02
	Gleason 4	88.63	93.6	92.85	
	Gleason 5	96.9	96.01	97.09	
<i>k</i> = 3	Gleason 3	73.07	79.16	96.62	95.67
	Gleason 4	92.98	93.68	95.52	
	Gleason 5	98.55	96.31	98.53	
média geral	<b>Gleason 3</b>	<b>83.7</b>	<b>80.85</b>	<b>97.95</b>	<b>94.89</b>
	<b>Gleason 4</b>	<b>90.78</b>	<b>93.82</b>	<b>94.14</b>	
	<b>Gleason 5</b>	<b>97.62</b>	<b>95.91</b>	<b>97.62</b>	

**Discussion**

Inter-pathologist grading discordance is known to be a relevant issue in prostate cancer treatment with numerous clinical consequences. Classical studies show inter-pathologists concordance varies between 51–78%, with a greater effect on patterns 3 and 4 differentiation [8–10]. Artificial intelligence (AI) is growing in importance among novel technologies in diagnostic procedures, mainly involving pattern recognition, being widely used in pathology and radiology [20–22].

Multiple recent literature reports on prostate cancer pathology and AI usage show the relevance of this theme [13, 14, 23, 24]. In a recent paper from the Karolinska Institute, Ström et al. [23], the authors obtained 96–99% accuracy in terms of benign-malignant differentiation and Gleason Grade concordance kappa of 0.62. Bulten et al. [25], used a dataset composed of 5759 prostate biopsies and reached a kappa agreement of 91.8% with pathologist reports. Patches were used with sufficient zoom to find fundamental structures and to

classify samples between cancer and non-cancer as well as to provide GG differentiation. Tolkach et al. [17] have separated cancer and non-cancer (stroma, glandular, and non-glandular benign tissue) patches obtaining more than 1.67 million patches.

The studies above have greatly contributed to the current knowledge of AI application to this field. Nonetheless, investigations and alternative approaches with new topologies must be continuously carried out and constructed to complement the current knowledge. We have thus developed a topology with tuned parameters regarding the number of filters, kernel sizes, layer sequence, number of hidden layers, activation function, and optimizer. Our proposed investigation and implementation complement and support the studies carried out in different contexts. The promising results, showing the performance of grading G3, G4, and G5, are in line with the literature, hence reinforcing the high potential of AI methods for this classification. Additionally, alternatives are offered to be used and evolved, contributing to the growing knowledge and evidence in this field.

As a limitation, the limited number of samples is noted for Gleason 3 pattern. However, Fig. 6 demonstrates that misclassification between classes in terms of numbers and percentages still statistically motivating. Furthermore, most misinterpretations are between neighboring Grades 3 and 4 (only 1 patch of Gleason 3 was interpreted as Gleason 5) with a minimal potential of clinical repercussion. Accordingly, considering that patches represent small portions of a large area, these few misinterpretations have minimal significance for the classification of the whole pathological area of the slide, minimizing possible interpretation effects.

Future work will focus on gathering additional collaborators and performing investigations, parallelly evaluating different promising topologies with the same dataset. With a dataset with wider variances, we will obtain the differences among topologies performance. Finally, we will include the construction of a mosaic from classified patches, creating heat map images, and provide a classification of the whole digital slides.

## Conclusion

Pathology is a cornerstone to support intervention discussion between practitioner and patient in actual customized prostate cancer care involving novel therapies (active surveillance, focal) [26, 27] and classic radical ones (radiation therapy and radical prostatectomy) [28]. Artificial intelligence has demonstrated its great potential in helping pathology pattern recognition with high accuracy. Our proposed CNN model added evidence to supports this potential and provides

a new alternative to be used and evolved, following the trend towards clinical usage in medical daily practices, consequently increasing the standards on pattern recognition to optimize clinical decisions, enabling best therapeutical results.

## Abbreviations

AI: Artificial Intelligence; CNN: Convolutional Neural Network; GG: Gleason grading scale; H&E: Hematoxylin and eosin; KRML: Initials of the uropathology specialist's name; PANDA: The Prostate Cancer Grade Assessment; CAPES: Coordination for the Improvement of Higher Education Personnel; FMUSP: University of São Paulo Scholl of Medicine; UNIFESP: Federal University of São Paulo.

## Acknowledgements

Institute of Science and Technology of the Federal University of São Paulo (ICT-UNIFESP), LAPIS – Laboratory of Image and Signal Processing of ICT-UNIFESP. University of Sao Paulo Medical School, Medical Investigation Laboratory number 55 (LIM-55), Urology Discipline of the Surgery Department, the University Ethics Committee that approved this study (approval number: 3.004.858). CAPES—Coordination for the Improvement of Higher Education Personnel, and CNPq—National Council of Scientific and Technological Development.

## Authors' contributions

Study concept and design: Souza, Leite, Moraes. Acquisition of data: Kudo, Souza, Estivallet, Amorim, Leite. Analysis and interpretation: Kudo, Souza, Kim, Moraes. Drafting of the manuscript: Kudo, Estivallet, Amorim. Critical revision of the manuscript for important intellectual content: Souza, Moraes. Statistical analysis: Kudo, Estivallet, Amorim, Moraes. Administrative, technical, and material support: Kudo, Estivallet, Amorim, Moraes. Supervision: Kim, Moraes. Other (pathology assessment): Leite. The author(s) read and approved the final manuscript.

## Funding

CNPq: National Council of Scientific and Technological Development, Brazil. CAPES: Coordination for the Improvement of Higher Education Personnel.

## Declarations

### Ethics approval and consent to participate

FMUSP: 99896918.3.0000.0065  
UNIFESP: 8657261119

### Competing interests

The authors declare that they have no conflict of interest.

### Author details

<sup>1</sup>Laboratory of Image and Signal Processing of the Institute of Science and Technology of Federal University of São Paulo (Universidade Federal de São Paulo – UNIFESP), Rua Talim 330 Sala 108 - Jardim Aeroporto, São José Dos Campos, SP CEP 12231-280, Brazil. <sup>2</sup>Laboratory of Medical Investigations Number 55 of the Sao Paulo University Medical School – FMUSP, Avenida Dr Arnaldo, 455, sala 2145 – Cerqueira Cesar, Sao Paulo, SP CEP 01246-903, Brazil. <sup>3</sup>Denver Health Medical Center, University of Colorado Anschutz Medical Center, Aurora, CO, USA.

Received: 14 October 2022 Accepted: 23 October 2022

Published online: 23 November 2022

## References

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209–49. <https://doi.org/10.3322/caac.21660>.



2. Sanda MG, Cadeddu JA, Kirkby E, et al. Clinically Localized Prostate Cancer: AUA/ASTRO/SUO Guideline. Part I: Risk Stratification, Shared Decision Making, and Care Options. *J Urol*. 2018;199(3):683–90. <https://doi.org/10.1016/j.juro.2017.11.095>.
3. Sanda MG, Cadeddu JA, Kirkby E, et al. Clinically Localized Prostate Cancer: AUA/ASTRO/SUO Guideline. Part II: Recommended Approaches and Details of Specific Care Options. *J Urol*. 2018;199(4):990–7. <https://doi.org/10.1016/j.juro.2018.01.002>.
4. Epstein JI, Egevad L, Amin MB, et al. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol*. 2016;40(2):244–52. <https://doi.org/10.1097/PAS.0000000000000530>.
5. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol*. 1974;111(1):58–64.
6. Gleason DF, Mellinger GT, Group VACUR. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. 1974. *J Urol*. 2002;167(2 Pt 2):953–8 (discussion 959).
7. Gleason DF, Mellinger GT, Group VACUR. Prediction of Prognosis for Prostatic Adenocarcinoma by Combined Histological Grading and Clinical Staging. *J Urol*. 2017;197(2S):S134–9. <https://doi.org/10.1016/j.juro.2016.10.099>.
8. Ozkan TA, Eruyar AT, Cebeci OO, Memik O, Ozcan L, Kuskonmaz I. Interobserver variability in Gleason histological grading of prostate cancer. *Scand J Urol*. 2016;50(6):420–4. <https://doi.org/10.1080/21681805.2016.1206619>.
9. Meliti A, Sadimin E, Diolombi M, Khani F, Epstein JI. Accuracy of Grading Gleason Score 7 Prostatic Adenocarcinoma on Needle Biopsy: Influence of Percent Pattern 4 and Other Histological Factors. *Prostate*. 2017;77(6):681–5. <https://doi.org/10.1002/pros.23314>.
10. Sadimin ET, Khani F, Diolombi M, Meliti A, Epstein JI. Interobserver Reproducibility of Percent Gleason Pattern 4 in Prostatic Adenocarcinoma on Prostate Biopsies. *Am J Surg Pathol*. 2016;40(12):1686–92. <https://doi.org/10.1097/PAS.0000000000000714>.
11. Monica M, Vadladi VK, Karuna G, Sowmya P. Comprehensive study of pathology image analysis using a deep learning algorithm. *Mater Today Proc*. 2020. <https://doi.org/10.1016/j.matpr.2020.09.515>.
12. Regnier-Coudert O, McCall J, Lothian R, Lam T, McClinton S, N'Dow J. Machine learning for the improved pathological staging of prostate cancer: A performance comparison on a range of classifiers. *Artif Intell Med*. 2012;55(1):25–35. <https://doi.org/10.1016/j.artmed.2011.11.003>.
13. Li Y, Huang M, Zhang Y, et al. Automated Gleason Grading and Gleason Pattern Region Segmentation Based on Deep Learning for Pathological Images of Prostate Cancer. *IEEE Access*. 2020;8:117714–25. <https://doi.org/10.1109/ACCESS.2020.3005180>.
14. Nagpal K, Foote D, Liu Y, et al. Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *NPJ Digit Med*. 2019;2(1):48. <https://doi.org/10.1038/s41746-019-0112-2>.
15. Pantanowitz L, Quiroga-Garza GM, Bien L, et al. An artificial intelligence algorithm for prostate cancer diagnosis in whole slide images of core needle biopsies: a blinded clinical validation and deployment study. *Lancet Digit Health*. 2020;2(8):e407–16. [https://doi.org/10.1016/S2589-7500\(20\)30159-X](https://doi.org/10.1016/S2589-7500(20)30159-X).
16. Mun Y, Paik I, Shin S-J, Kwak T-Y, Chang H. Yet Another Automated Gleason Grading System (YAAGGS) by weakly supervised deep learning. *NPJ Digit Med*. 2021;4(1):99. <https://doi.org/10.1038/s41746-021-00469-6>.
17. Tolkach Y, Dohmgörge T, Toma M, Kristiansen G. High-accuracy prostate cancer pathology using deep learning. *Nat Mach Intell*. 2020;2(7):411–8. <https://doi.org/10.1038/s42256-020-0200-7>.
18. Hayashi Y. New unified insights on deep learning in radiological and pathological images: Beyond quantitative performances to qualitative interpretation. *Inf in Med Unlocked*. 2020;19:100329. <https://doi.org/10.1016/j.imu.2020.100329>.
19. Kudo MS, de Souza VMG, de Souza AG, et al. The potential of convolutional neural network diagnosing prostate cancer. *Res Biomed Eng*. 2021;37(1):25–31. <https://doi.org/10.1007/s42600-020-00095-3>.
20. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Adv Neural Inf Process Syst*. 2012;25:1090–8.
21. Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng*. 2018;2:158–64.
22. Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis detection in breast cancer histology images with deep neural networks. *Med Image Comput Comput Assist Interv*. 2013;16(Pt 2):411–8.
23. Strom P, Kartasalo K, Olsson H. Artificial intelligence for diagnosis and grading of prostate cancer in biopsies: a population-based, diagnostic study. *Lancet Oncol*. 2020;21(2):E70–E70.
24. Arvaniti E, Fricker KS, Moret M, et al. Automated Gleason grading of prostate cancer tissue microarrays via deep learning. *Sci Rep*. 2018;8(1):12054. <https://doi.org/10.1038/s41598-018-30535-1>.
25. Bulten W, Pinckaers H, van Boven H, et al. Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study. *Lancet Oncol*. 2020;21(2):233–41. [https://doi.org/10.1016/S1470-2045\(19\)30739-9](https://doi.org/10.1016/S1470-2045(19)30739-9).
26. Carlsson S, Benfante N, Alvim R, et al. Long-Term Outcomes of Active Surveillance for Prostate Cancer: The Memorial Sloan Kettering Cancer Center Experience. *J Urol*. 2020;203(6):1122–7. <https://doi.org/10.1097/ju.0000000000000713>.
27. Sivaraman A, Barret E. Focal Therapy for Prostate Cancer: An "À la Carte" Approach. *Eur Urol*. 2016;69(6):973–5.
28. Wilt TJ, Ullman KE, Linskens EJ, et al. Therapies for clinically localized prostate cancer: a comparative effectiveness review. *J Urol*. 2021;205(4):967–76.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

