DOI: https://doi.org/10.4038/gmj.v31i1.8329

The effect of using a morphological image scale on inter-observer variation in Gleason grading of adenocarcinoma of prostate

MKL Manjula[1], KI Jayasinghe[2], AU Kiriwandeniya[3], BV Hasheni[4], PHN Ruwantha[4]

[1]*District General Hospital, Monaragala, Sri Lanka.*
[2]*District General Hospital Negombo, Sri Lanka.*
[3]*NSFaculty of Medical Sciences, University of Sri Jayewardenepura, Sri Lanka.*
[4]*National Hospital Galle, Sri Lanka.*

Correspondence: Dr. M K L Manjula
e-mail: *manjulamkl@hotmail.com*
https://orcid.org/0009-0007-3295-1045
Dr. K I Jayasinghe
https://orcid.org/0000-0002-0949-5367
Submitted on 14.02.2025 and accepted for publication on 18.03.2026

ABSTRACT

Introduction: The Gleason grade (GG) and grade group (Gg) are the most important prognostic indicators in histopathology report of a prostate core biopsy for adenocarcinoma. These indicators influence in determining the treatment options and prognosis. However, due to the presence of complex architectural patterns, inter observer variability (IOV) in assessing GG and Gg of prostate is unavoidable. A morphological image scale (MIS) of varied architectural patterns of GG was used in the current study to aid in reporting. The aim of the study was to assess its influence on IOV.

Methods: The study included 101 Haematoxylin and Eosin (H&E) stained histopathological slides of transrectal ultrasound scan guided prostate core biopsies of adenocarcinoma, reported over a period of two years in a Teaching Hospital in Sri Lanka. Two general histopathologists assessed and scored the slides, according to modified Gleason grading system (MGGs), independently without the use of MIS in the first round. Reassessment and scoring were done in the second round after a period of 3 months, using the MIS. MIS was prepared using the microscopic images published in a reference article.

The inter-rater reliability was assessed using the Kappa statistics (K) to determine consistency between two investigators for Gleason sum score (GSS) and Gg separately, for first and second rounds, to assess the effect of the MIS on IOV.

Results: Exact percentage agreement on GSS achieved in the first and second rounds were 48.5% (K=0.264, $p< 0.001$) and 55.4% (K=0.368, $p<0.001$) respectively. Exact percentage agreement on Gg achieved in first and second rounds were 45.5% (K=0.273, $p<0.001$) and 47.5% (K=0.324, $p< 0.001$) respectively.

The inter-rater reliability for GSS (55.4% vs. 48.5%, p=0.322) and Gg (47.5% vs. 45.5%, p=0.779) were found to be higher with the assistance of MIS. However, the improvement observed after introducing MIS was not statistically significant for both GSS and Gg.

Conclusions: There was a non-significant improvement in percentage inter-observer agreement with MIS assistance. Replication of the study in a larger sample is suggested to further evaluate the usefulness of MIS.

Keywords: *Adenocarcinoma of prostate, Gleason grading, Inter-observer variation, Morphological image scale.*

Introduction

Adenocarcinoma of prostate is one of the leading causes of cancer in males and its incidence varies worldwide (1). Carcinoma of prostate falls in 5th position of the leading causes of cancer in males in Sri Lanka according to National Cancer Incidence Data 2019 (2).

The histopathology report of a prostate core biopsy for adenocarcinoma includes several factors that determine treatment and prognosis. The grade of the tumour is the most important prognostic marker in the histopathology report and the grading of prostate adenocarcinoma is done according to Modified Gleason Grade (MGG), which is the recommended grading system by World Health Organization (WHO) (3, 4).

The Gleason grading system was introduced by Donald F. Gleason in 1966 (5). This has evolved over time with the changes suggested by Gleason himself and by his followers. The currently used grading system is the MGGs published following the consensus meeting of International Society of Urological Pathology (ISUP) held in 2005, 2014 and 2019 (3, 5, 6).

Gleason grading of adenocarcinoma of prostate is based on architectural pattern of the tumour and there are five categories of morphological patterns scored from 1 to 5. For core biopsies the Gleason score is derived from the sum of the most predominant pattern (primary pattern) and the second most predominant pattern (secondary pattern). However, if a higher minor pattern is present, irrespective of its extent, this is used instead of the secondary pattern. Patterns 1 and 2 are not reported any more for core biopsies (3, 5, 6).

Oncological management of patients diagnosed to have adenocarcinoma of prostate mainly depends on histological factors, particularly the GG. Many studies support the predictive and prognostic value of GG (3). However much the modifications are made to the Gleason grading system in different consensus meetings, it continues to have its own limitations. The concept that histological patterns represent a continuum of the morphological patterns of a malignancy, highlights that grey zones are encountered leading to inter-observer and intra-observer variability (4, 7-10). Specialized uropathologists have shown higher concordance rates (7, 10). However, such expertise is currently limited in pathology practice in Sri Lanka.

In 2014, Epstein proposed the concept of Gg, lumping the Gleason score and later ISUP has approved Gg which consist of five prognostic groups and this grading system has been recognised by WHO to be used in conjunction with GG (10). Prognostic groups intended to improve reporting in terms of clinical value, increased interpretability of the tumour grade and reduced discordant rates (10, 11). However, despite efforts to standardize grading of prostate cancer, there is still a considerable variation in grading prostate carcinoma even among experts (12).

Assistance of advanced technology such as artificial intelligence-based tools have recently shown significant improvements in inter-observer agreement in Gleason grading of prostate adenocarcinoma (3, 4).

A study has been done previously on GG of carcinoma of prostate to assess the influence of a set of reference images and shown that using reference images have significantly improved the reproducibility of grading. However, in the above study, microphotographs of carcinoma of prostate have been used as study material (12). In the present study, we have tested the effect of a set of reference images; Morphological Image Scale (MIS), on reproducibility of GG and Gg, using slides as study material.

MIS is a simple and inexpensive tool and, in this study, we have compared IOV of GSS and Gg with and without the assistance of MIS to ascertain its usefulness to improve reproducibility of grading in prostate carcinoma.

Methods

This study included H&E-stained histopathological slides of adenocarcinoma of prostate, diagnosed on transrectal ultrasound scan guided prostate core biopsies over a period of two years starting from January 2013 to December 2014. Slides were retrieved from the archives of Teaching Hospital Karapitiya, Sri Lanka. The ethical approval for the study was taken from the Ethics Review Committee, Faculty of Medicine, University of Ruhuna and administrative approval was obtained

from the Head of the Histopathology Department, National Hospital Galle.

Two general histopathologists with similar experience and working in two different District General Hospitals of Sri Lanka were involved in the study as investigators.

Quality of the slides was initially assessed by an independent histopathologist and slides with poor staining quality were re-stained. A total of 101 H&E slides with reasonable staining quality were selected for the study. All the slides were re-coded to ensure that they could not be identified by the investigating pathologists.

Before commencing the study both investigators read through published articles on ISUP consensus conference 2005, 2014 and 2019 in order to update the knowledge on criteria for Gleason grading.

In the first round the two general histopathologists independently scored all the slides according to MGGs to assign GG and the Gg using mental imagery without using MIS or any other images for assistance.

A MIS was prepared using 15 microscopic images cited in the article based on the consensus reached at the 2005 ISUP consensus conference of Gleason grading of prostate carcinoma, published in *American Journal of Surgical Pathology* in September 2005 (5), (Figure 1). Microscopic images were representative of Gleason morphological patterns 3, 4 and 5. Microscopic images on Gleason morphological patterns 1 and 2 were not taken into MIS as patterns 1 and 2 are not reported any more for core biopsies.

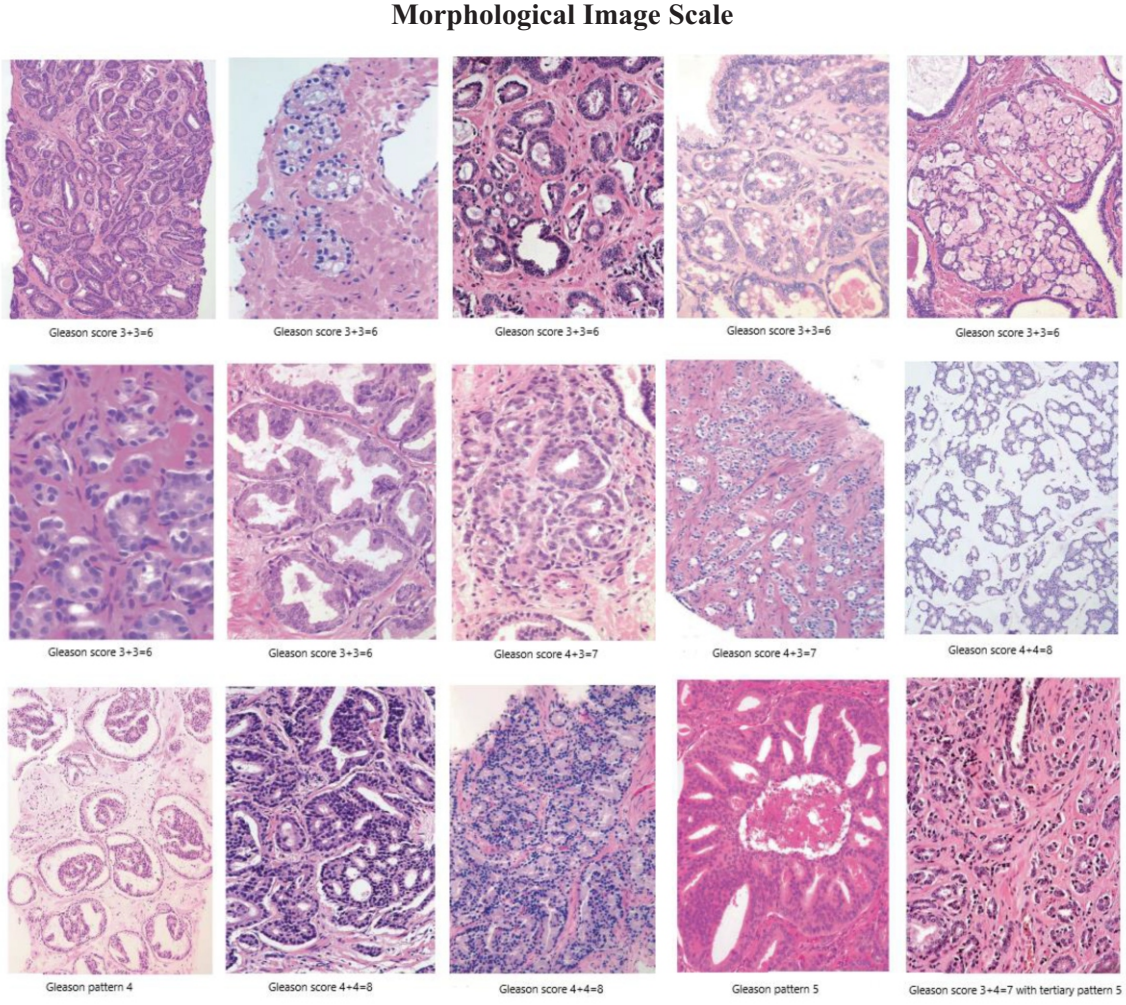

Figure 1: Morphological Image Scale (MIS) for grading prostatic carcinoma (5)

After a period of three months the set of slides were re-coded by the independent histopathologist to change the identification code assigned in the first round. Second round of scoring was done by the same two investigating histopathologists according to MGGs using the MIS for assistance. No other reference materials were used for assistance. The two histopathologists used the MIS on computer screen to assist grading process and consistently matched with the images before assigning the final score of GG and Gg. During the two rounds of scoring two pathologists entered their respective score in separate excel data sheets.

The statistical analysis was performed using Statistical Package for Social Sciences *(SPSS)* software (Version 22.0). Based on scores of first and second rounds, an inter-rater reliability analysis using the Kappa statistics was performed to determine consistency between two investigators for GSS and Gg separately. Intra-rater reliability for Gleason scoring without using morphological image scale (first round) and with using morphological image scale (second round) were compared to assess the effect of the MIS on IOV. Percentage agreement (concordance rate) was calculated for both rounds using Kappa coefficient of agreement and compared using the z-test for difference between two proportions. The level of probability was set at 0.05.

Results

GSS and the Gg were separately assessed for first and second rounds of scoring to determine the level of agreement between two investigators without MIS and with MIS respectively. The level of agreement with regards to GSS for each round of scoring is presented in Tables 1- 4.

Tables 5 - 8 show the level of agreement with regards to Gg for each round of scoring by the two investigators.

Table 9 summarizes the level of agreement between the two investigators for GSS and Gg with and without the assistance of the MIS.

Table 1: Inter-observer agreement of the Gleason Sum Score (GSS) without the assistance of MIS

		Investigator 2 - GSS (without MIS)				**Total**
	GSS	**6**	**7**	**8**	**9**	
Investigator 1 - GSS (without MIS)	**6**	0 (0.0%)	2 (100.0%)	0 (0.0%)	0 (0.0%)	2 (100.0%)
	7	6 (12.8%)	16 (34.0%)	18 (38.3%)	7 (14.9%)	47 (100.0%)
	8	0 (0.0%)	5 (15.2%)	22 (66.7%)	6 (18.2%)	33 (100.0%)
	9	0 (0.0%)	1 (5.9%)	5 (29.4%)	11 (64.7%)	17 (100.0%)
	10	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (100.0%)	2 (100.0%)
	Total	6 (5.9%)	24 (23.8%)	45 (44.6%)	26 (25.7%)	101(100.0%)

Table 2: Level of agreement of Gleason Sum Score (GSS) between the two histopathologists without assistance of MIS (N=101)

GSS	Agreement for GSS No. of slides (%)
6	0 (0.0%)
7	16 (15.84%)
8	22 (21.78%)
9	11 (10.89%)
10	0 (0.0%)
Level of agreement	**48.51% (K=0.264, *p*<0.001)**

Percentage agreement on GSS achieved in the first round (without assistance of MIS) was 48.5% (K=0.264, *p*<0.001), 95% CI (0.135-0.393).

Table 4: Level of agreement of Gleason Sum Score (GSS) between the two histopathologists with assistance of MIS (N=101)

GSS	Agreement for GSS No. of slides (%)
6	0 (0.0%)
7	25 (24.75%)
8	17 (16.83%)
9	14 (13.86%)
10	0 (0.0%)
Level of agreement	**55.44% (K=0.368, *p*<0.001)**

Percentage agreement on GSS achieved in the second round (with assistance of MIS) was 55.44% (?=0.368, p<0.001), 95% CI (0.239-0.497).

Table 3: Inter-observer agreement of the Gleason Sum Score (GSS) with the assistance of MIS

		Investigator 2-GSS (with MIS)					Total
Investigator 1 - GSS (with MIS)	**GSS**	**6**	**7**	**8**	**9**	**10**	
	7	10 (21.7%)	25 (54.3%)	6 (13.0%)	5 (10.9%)	0 (0.0%)	46 (100.0%)
	8	0 (0.0%)	8 (30.8%)	17 (65.4%)	1 (3.8%)	0 (0.0%)	26 (100.0%)
	9	0 (0.0%)	2 (7.1%)	8 (28.6%)	14 (50.0%)	4 (14.3%)	28 (100.0%)
	10	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (100%)	0 (0.0%)	1 (100%)
	Total	10 (9.9%)	35 (34.7%)	31(30.7%)	21 (20.8%)	4 (4.0%)	101 (100.0%)

Table 5: Inter-observer agreement of the Grade group (Gg) without the assistance of MIS

		Investigator 2 - Gg (without MIS)					Total
Investigator 1 - Gg (without MIS)	**Gg**	**1**	**2**	**3**	**4**	**5**	
	1	0 (0.0%)	1 (50.0%)	1 (50.0%)	0 (0.0%)	0 (0.0%)	2 (100.0%)
	2	1 (4.3%)	5 (21.7%)	3 (13.0%)	10 (43.5%)	4 (17.4%)	23 (100.0%)
	3	5 (20.8%)	2 (8.3%)	6 (25.0%)	8 (33.3%)	3 (12.5%)	24 (100.0%)
	4	0 (0.0%)	1 (3.0%)	4 (12.1%)	22 (66.7%)	6 (18.2%)	33 (100.0%)
	5	0 (0.0%)	0 (0.0%)	1 (5.3%)	5 (26.3%)	13 (68.4%)	19 (100.0%)
	Total	6 (5.9%)	9 (8.9%)	15 (14.9%)	45 (44.6%)	26 (25.7%)	101 (100.0%)

Table 6: Level of agreement of Grade group (Gg) between both pathologists without assistance of MIS (N=101)

Gg	Agreement for Gg No. of slides (%)
1	0 (0.0%)
2	5 (4.95%)
3	6 (5.94%)
4	22 (21.78%)
5	13 (12.87%)
Level of agreement	**45.54% (K=0.273, p<0.001)**

Percentage agreement on Grade group achieved in the first round (without assistance of MIS) was 45.54% (?=0.273, p<0.001), 95% CI (0.154-0.392).

Table 8: Level of agreement of Grade group (Gg) between the two histopathologists with assistance of MIS (N=101)

Gg	Agreement for Gg No. of slides (%)
1	0 (0.0%)
2	5 (4.95%)
3	7 (6.93%)
4	17 (16.83%)
5	19 (18.81%)
Level of agreement	**47.52% (K=0.324, p<0.001)**

Percentage agreement on Gg achieved in the second round (with assistance of MIS) was 47.52% (?=0.324, p<0.001), 95% CI (0.209-0.439).

Table 7: Inter-observer agreement of the Grade group (Gg) with the assistance of MIS

		Investigator 2-Gg (with MIS)					Total
Investigator 1 - Gg (with MIS)	**Gg**	**1**	**2**	**3**	**4**	**5**	
	2	7 (43.8%)	5 (31.2%)	1 (6.2%)	2 (12.5%)	1 (6.2%)	16 (100.0%)
	3	3 (10.0%)	12 (40.0%)	7 (23.3%)	4 (13.3%)	4 (13.3%)	30 (100.0%)
	4	0 (0.0%)	4 (15.4%)	4 (15.4%)	17 (65.4%)	1 (3.8%)	26 (100.0%)
	5	0 (0.0%)	0 (0.0%)	2 (6.9%)	8 (27.6%)	19 (65.5%)	29 (100.0%)
	Total	10 (9.9%)	21 (20.8%)	14 (13.9%)	31 (30.7%)	25 (24.8%)	101 (100.0%)

Table 9: Level of agreement of Gleason Sum Score (GSS) and Grade group (Gg) with and without MIS.

	Level of agreement		
	Without MIS	**With MIS**	**Significance**
GSS	48.51%	55.44%	Z = -0.9858 p = 0.322
Gg	45.54%	47.52%	Z = -0.9858 p = 0.779

In summary, the inter-rater reliability for the GSS was found to be higher with the assistance of MIS than without assistance (55.44% vs. 48.51% respectively, p=0.322), however, this difference was not statistically significant. A similar non-significant improvement in inter-rater reliability was observed for the Gg with the use of MIS (47.52% vs. 45.54%, p=0.779).

Discussion

Grading of cancer is aimed at stratifying a biological continuum into a limited number of discrete categories. The borders between these categories remain subjective (9, 12, 13). Prostate cancer is a heterogeneous disease that is manifested in a diverse range of histological patterns. Therefore, grading of prostate cancer is associated with inter-observer variation among pathologists leading to under- or over- treatment of the patients (9, 13).

Numerous studies have been done in the past to assess the reproducibility of Gleason grading of prostate carcinoma and the results range from poor to substantial agreement and specialist genitourinary pathologists have shown better agreement compared to general pathologists (7, 9, 10, 12, 14, 15).

We have compared the Gleason grading by two general histopathologists, with and without assistance of a MIS to assess its effect on the level of agreement.

Our results have shown some improvement of inter observer agreement after introducing the MIS as indicated by a change of kappa value from 0.264 to 0.368 for GSS and from 0.273 to 0.324 for Gg. However, in both assessments kappa value remains in the range of fair agreement (0.21-0.40).

A study has been done previously in Sweden, on carcinoma of prostate to assess the influence of a set of reference images on GG. Seventy-three specialists in genitourinary pathology and 12 residents have participated in the study and a set of 40 reference images have been used to grade a set of 20 test microphotographs. The study has shown that using reference images have significantly improved the reproducibility of grading from average kappa value of 0.58 to 0.78 (12).

Participants of this study had experienced difficulty in assigning grade in small areas of glandular fusion vs. tangential sectioning of glands to decide on grade 4 vs. 3 and poorly formed acinar structures vs. solid cord and nests to decide on grade 4 vs. 5. Similar observations have been made in some other published studies (9). Difficulties have been encountered in tumours with complex morphological heterogeneity, particularly due to presence of minute foci of minor higher-grade patterns. What appeared to be a small focus of single cell infiltrate was found to pose difficulty in assigning minor higher grade into Gleason grading.

When considering limitations of the study, finding validated microphotographs for MIS was a challenge and microphotographs to formulate MIS were taken from the special article published following ISUP consensus conference on Gleason grading of prostatic carcinoma. The microphotographs were considered as validated for the score assigned for each of them. However, one of the pictures on cribriform pattern 3 was not taken for MIS as current criteria assign all cribriform glands as Gleason pattern 4, irrespective of their size (3). Picture labelled as "3+4 = 7 with tertiary pattern 5" was considered as 3+5 = 8 according to current criteria. There were sufficient numbers of pictures to represent the morphological spectrum of morphological patterns 3 and 4; however, number of microphotographs to represent pattern 5 were not sufficient in our opinion. Total number of reference images included in the MIS was 15.

Furthermore, the results of this study highlight the necessity for improvement of the inter-observer agreement of grading prostate carcinoma among general pathologists of Sri Lanka. Studies have shown that further training programmes for pathologists have improved inter-observer reproducibility of grading prostate carcinoma (15).

Conclusions

There was a non-significant improvement in percentage inter-observer agreement in GSS and Gg with MIS assistance.

Replication of the study in a larger sample is suggested with improved MIS including more validated reference images and with wider participation of histopathologists to further evaluate the usefulness of MIS.

References

1. Wang L, Lu B, He M, *et al.* Prostate Cancer Incidence and Mortality: Global Status and Temporal Trends in 89 Countries From 2000 to 2019. Front Public Health. 2022 Feb 16; **10**: 811044. DOI: 10.3389/fpubh.2022.811044.
2. National cancer Incidence and Mortality data 2019, 21st publication, National Cancer control programme, Ministry of health Sri Lanka. Available from: https://www.nccp.health.gov.lk/storage/post/pdfs/Cancer Incidence Data Book-2019_compressed.pdf.
3. Van Leenders GJLH, van der Kwast TH, Grignon DJ, *et al.* ISUP Grading Workshop Panel Members. The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. *Am J Surg Pathol.* 2020 Aug; **44**(8): e87-e99. DOI: 10.1097/PAS.0000000000001497.
4. Bulten W, Balkenhol M, Belinga JA, *et al.* ISUP Pathology Imagebase Expert Panel; van Boven H, Vink R, van der Laak J, Hulsbergen-van der Kaa C, Litjens G. Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists. *Mod Pathol.* 2021 Mar; **34**(3): 660-671. DOI: 10.1038/s41379-020-0640-y.
5. Epstein JI, Allsbrook WC Jr, Amin MB et al. ISUP Grading Committee. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol.* 2005 Sep; **29**(9): 1228-42. DOI: 10.1097/01.pas.0000173646.99337.b1.
6. Epstein, JI & Egevad, L. & Amin, *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. *American Journal of Surgical Pathology.* 40. 244-52.
7. Al-Maghrabi JA, Bakshi NA, Farsi HM. Gleason grading of prostate cancer in needle core biopsies: a comparison of general and urologic pathologists. *Ann Saudi Med.* 2013 Jan-Feb; **33**(1): 40-4. DOI: 10.5144/0256-4947.2013.40.
8. Qureshi A, Lakhtakia R, AL Bahri M, *et al.* Gleason's Grading of Prostatic Adenocarcinoma: Inter-Observer Variation Among Seven Pathologists at a Tertiary Care Center in Oman. *Asian Pac J Cancer Prev.* 2016 Nov 1; **17**(11): 4867-4868. DOI: 10.22034/APJCP.2016.17.11.4867.
9. Singh RV, Agashe SR, Gosavi AV, *et al.* Interobserver reproducibility of Gleason grading of prostatic adenocarcinoma among general pathologists. *Indian J Cancer.* 2011 Oct-Dec; **48**(4): 488-95. DOI: 10.4103/0019-509X.92277.
10. Ahsan F, Khan A, Asif M, *et al.* Inter-observer variability in Gleason scoring system for histological grading of adenocarcinoma prostate. *PAFMJ*. 2022; **72**(1): 327. [cited 15Jan.2023]
11. De Nunzio C, Pastore AL, Lombardo R, *et al.* The new Epstein Gleason score classification significantly reduces upgrading in prostate cancer patients. *Eur J Surg Oncol*. 2018 Jun; **44**(6): 835-839. DOI: 10.1016/j.ejso.2017.12.003.
12. Egevad L. Reproducibility of Gleason grading of prostate cancer can be improved by the use of reference images. *Urology.* 2001 Feb; **57**(2): 291-5. DOI: 10.1016/s0090-4295(00)00922-5.
13. Nir G, Hor S, Karimi D, *et al.* Automatic grading of prostate cancer in digitized histopathology images: Learning from multiple experts. *Med Image Anal.* 2018; **50**: 167-180. DOI: 10.1016/j.media.2018.09.005.
14. Oyama T, Allsbrook WC Jr, Kurokawa K, *et al.* A comparison of interobserver reproducibility of Gleason grading of prostatic carcinoma in Japan and the United States. *Arch Pathol Lab Med.* 2005; **129**(8): 1004-10. DOI: 10.5858/2005-129-1004-ACOIRO. Erratum in: *Arch Pathol Lab Med.* 2005; **129**(11): 1368.
15. Abdollahi A, Sheikhbahaei S, Meysamie A, *et al.* Inter-observer reproducibility before and after web-based education in the Gleason grading of the prostate adenocarcinoma among the Iranian pathologists. *Acta Med Iran.* 2014; **52**(5): 370-374.