Pathology Section

Agreement between Ki-67 Proliferative Index in Breast Cancer by Conventional Hotspot Method and International Ki-67 in Breast Cancer Working Group Global Scoring Method: A Cross-sectional Study

NIDHI RADHAKRISHNAN[1], PADMA PRIYA KASUKURTI[2], SULATA M KAMATH[3], CLEMENT WILFRED DEVADASS[4], SR MANGALA GOURI[5]

ABSTRACT

Introduction: The role of the Ki-67 Proliferative Index (PI) in the molecular classification and as a predictive and prognostic biomarker in breast cancer is definitive. To address the inconsistency in Ki-67 interpretation and scoring, the International Ki-67 in Breast Cancer Working Group (IKWG) introduced the visual scoring Android Application (APP) and proposed a standardised scoring method.

Aim: To determine Ki-67 proliferative indices by both the Global Method (GW) with the Ki-67 visual scoring app recommended by IKWG and conventional institutional Hotspot Method (HM), and to analyse the agreement between the indices obtained by these two methods.

Materials and Methods: A cross-sectional study involving 71 Ki-67 immunostained hormone receptor-positive trucut biopsies of breast cancers from January 2022 to January 2024 were retrospectively collected. Two observers independently scored the biopsies using HM and GW by an app in the Department of Pathology of a tertiary care centre. The GW scores of Ki-67 PI obtained were categorised into low, intermediate and high groups based on 2015 St. Gallen guidelines. A two-way random Intraclass Correlation Coefficient (ICC) was used to assess the absolute agreement between two scorers.

Results: The overall intraclass correlation between interobserver values by HM was 0.819 (good), while it was 0.971 (excellent) by GW. The overall interobserver mean difference was five times greater (p-value=0.008, statistically significant) than the overall interobserver global weighted scores (p-value=0.901, not statistically significant).

Conclusion: The study highlighted the significant interobserver variability in HM compared to GW scores by the app. The standardised app scoring method has the potential to broaden the prognostic role of Ki-67 as a companion diagnostic tool.

INTRODUCTION

Ki-67 is a nuclear non histone protein present in all active phases of cell cycle, except the G0 phase [1]. Ki-67 PI is defined as the percentage of positively stained cells within the total number of malignant cells [2]. It serves as both a predictive and prognostic biomarker in breast cancer. The St. Gallen Guidelines recommend the assessment of Ki-67 proliferation for selecting the addition of chemotherapy in hormone receptor-positive breast cancers, and a cut-off was declared to differentiate between luminal A and luminal B breast cancer subtypes [3,4].

The key limitation of Ki-67 as a biomarker is the lack of interlaboratory reproducibility in its measurement and the questionable analytic validity [3]. The difference in interpretation among observers result in consequent diagnostic variability. To address the problem of inconsistency in Ki-67 assessment, as well as for interpretation and scoring, the IKWG introduced the Visual Scoring Android App. The Ki-67 scoring app, accessible from Feb 2023, assists pathologists with scoring Ki-67 using the standardised method proposed by the IKWG. In this background, this study was undertaken to determine Ki-67 proliferative indices by both GW using Ki-67 visual scoring app recommended by IKWG and conventional institutional HM, and to analyse the agreement between the indices by these two methods.

MATERIALS AND METHODS

This was a cross-sectional study conducted in the Department of Pathology at Ramaiah Medical College, Bengaluru, Karnataka, India, from archived cases of January 2022 to January 2024. Institutional Ethics Committee (IEC) clearance was obtained (vide no: MRMC/EC/SP-09/08-2024 dated 28th August 2024).

Inclusion criteria: All cases of treatment-naive breast cancers received as trucut core biopsies with Oestrogen Receptor (ER) positive status, for which immunohistochemistry for Ki-67 marker was performed, were included in the study.

Exclusion criteria: Biopsies with predominantly necrotic tumour cores were excluded from the study.

At the time of the initial assessment, the tissue was collected according to American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines [5-7]. A cold ischaemic time of <30 minutes, fixation in 10% neutral buffered formalin and fixation for 6-72 hours were assured for all specimens. Immunohistochemistry (IHC) was performed using the monoclonal antibody MIB-1, and appropriate staining protocols were used, with reactive lymph node tissue taken as a control. The laboratory is enrolled in the national inter-laboratory quality assessment scheme for IHC by QC Mark and was recognised for par excelence

in run-B17 for MIB-1 immunostaining in February 2023. Whole resection specimens were excluded in order to minimise the associated preanalytic errors including prolonged cold ischaemic time and fixation time. A total of 71 Ki-67 IHC slides meeting the inclusion and exclusion criteria were retrieved. Each slide of each case had multiple cores, thereby ensuring the tumour heterogeneity was not missed.

The slides were reassessed by global method using IKWG visual scoring Android application by two scorers independently, after the completion of the Ki-67 calibration exercise available on the IKWG website. Another set of values for all the collected cases was obtained using the institutionally standardised conventional HM. The scorers were blinded to the already reported scores and to each other's scores.

Steps in the conventional institutional Hotspot Method (HM): The entire glass slide section was examined under low power (10x). The area with highest nuclear expression of Ki-67 by the tumour cells was identified. 100 tumour cell nuclei in the area were counted, avoiding non invasive cells (normal epithelium, stromal and immune cells) and the positive cell percentage was considered as the Ki-67 PI.

Steps in Global Method Scoring (GW) using Ki-67 Visual scoring app [5]: The entire glass slide section was examined in low power magnification, and the percentages of invasive tumour that exhibited Ki-67 were estimated in four areas- "Negligible," "Low," "Medium," and "High" [Table/Fig-1]. Based on the homogeneity or heterogeneity of staining in the tumour area, the percentages of scoring areas were estimated. Furthermore, one high-powered field was allocated for each category. The positively stained invasive tumour nuclei were counted in a typewriter pattern until 100 invasive tumour nuclei were counted or until all invasive tumour nuclei in the entire scoring field had been counted, whichever was earliest. The relative percentage of invasive tumour nuclei in a particular staining category was entered into the app, and final Ki-67 GW report was obtained upon completion of counting all four staining categories.
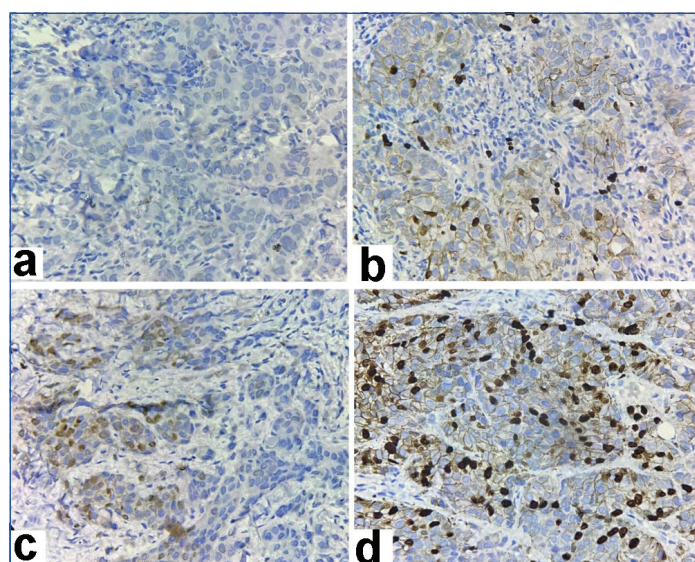
[Table/Fig-1]: Tumour cells exhibiting variable Ki-67 staining in different areas within the same breast core (IHC Stain 400x); Areas with (a) Negligible; (b) Low; (c) Medium; (d) High staining.

Each case now having two values obtained from HM (HM1, HM2) and two values from the global weighting derived from the mobile application (GW1, GW2), reported as a continuous variables. These were divided into three categories: low, intermediate and high. This categorical division is based on the 2015 St. Gallen consensus recommendations, where 10% or more of the median value is taken as high, 10% or less of median value as low and the values in between as intermediate [8]. A two-way random ICC was used to assess the absolute agreement between two scorers.

STATISTICAL ANALYSIS

Data was entered into Microsoft Excel data sheet and was analysed using Statistical Package for the Social Sciences (SPSS) version 22.0 (IBM SPSS Statistics, Somers, NY, USA) version software. Categorical data were represented in the form of frequencies and proportions. The Chi-square test or Fisher's exact test (for 2×2 tables only) was used as a test of significance for qualitative data. Continuous data were represented as mean and standard deviation.

Analysis of Variance (ANOVA) was used as a test of significance to identify the mean difference between more than two quantitative variables. A two-way random ICC was used to assess the absolute agreement between the two scorers. The ICC is a value between 0 and 1, where values below 0.5 indicate poor reliability, between 0.5 and 0.75 indicate moderate reliability, between 0.75 and 0.9 indicate good reliability, and any value above 0.9 indicates excellent reliability. A difference versus mean plot was done through Bland-Altman method. A p-value (probability that the result is true) of <0.05 was considered statistically significant after assuming all the rules of statistical tests.

RESULTS

Seventy-one Ki-67 immunostained slides were available during the study period. All cases had ER-positive status; five cases were PR-negative, and there were four cases each with HER2-positive and equivocal hormonal status. Based on the Ki-67 scores, the cases were categorised as low, intermediate, and high groups within each method [Table/Fig-2,3].
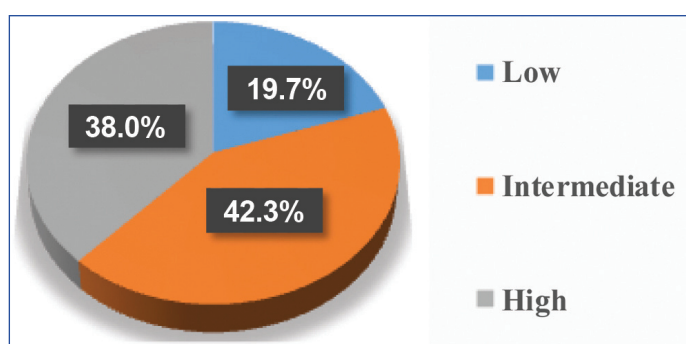
[Table/Fig-2]: Pie chart showing distribution of cases according to classification by hot spot method.
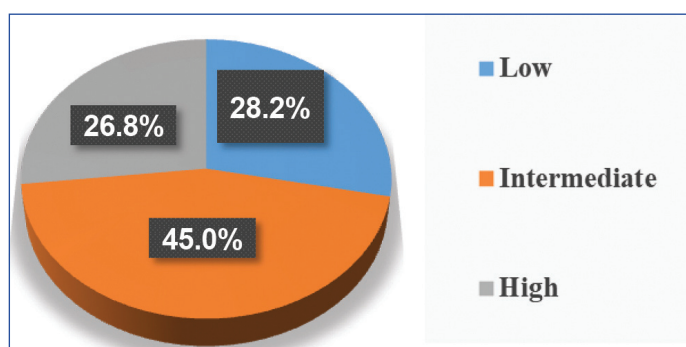
[Table/Fig-3]: Pie chart showing distribution of cases according to classification by global weighted app method.

The degree of reliability between the two scorers for HM was evaluated using a two-way random, absolute agreement, single measures ICC. The ICC values were high (ICC: 0.819, 95% confidence interval 0.725-0.883, p-value <0.001), indicating a good degree of reliability between scorers using the eye-ball method. The ICC values obtained for global weighted scores were even higher (ICC: 0.971, 95% confidence interval 0.954-0.982, p-value <0.001), indicating an excellent degree of reliability between scorers in the global method. However, considering categorical reproducibility, the intermediate category showed an ICC value of 0.77 in the global

method, rendering good reproducibility, whereas the other two categories exhibited excellent reproducibility [Table/Fig-4].

Categories	ICC between HM1 and HM2	ICC between GW1 and GW2
Low	0.136	0.965
	Poor	Excellent
Intermediate	0.59	0.773
	Moderate	Good
High	0.764	0.981
	Good	Excellent
Overall	0.819 95% CI 0.725-0.883	0.971 95% CI 0.954-0.982
p-value	<0.001	<0.001
Overall reliability	Good	Excellent

[Table/Fig-4]: Intraclass Correlation Coefficient (ICC) between two scores by two methods.

Despite good overall reproducibility in HM, a statistically significant difference was found between groups with respect to the mean difference between HM1 and HM2, with a p-value of 0.008. Whereas no such statistically significant differences among the groups between GW1 and GW2 were observed (p-value=0.901) [Table/Fig-5,6].

	Mean difference between HM1 and HM2		Mean difference between GW1 and GW2	
Categories	Mean difference	SD	Mean difference	SD
Low	-5.714	6.5888	-0.265000	1.597127
Intermediate	-7.000	11.1533	-0.775000	5.613865
High	2.889	14.9134	-0.378947	3.276105
Overall	-2.986	12.8029	-0.525352	4.178848
p-value	0.008		0.901	

[Table/Fig-5]: Mean difference between two scorers by two methods.
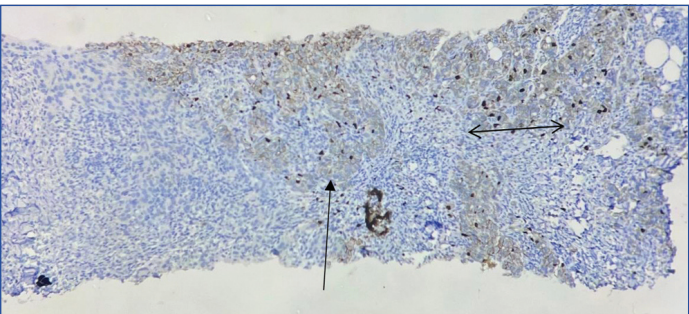
[Table/Fig-6]: Ki-67 scores by observer 1 and 2 in hot spot method for the above core- Score 20 (Horizontal arrow with two arrow heads) and score 15 (Vertical arrow with single arrow head); reflecting high mean difference. Ki-67 scores by observer 1 and 2 in global weighted app method for the above core- score 10.8 and score 11.1; reflecting *low mean difference* (IHC Stain: 40x).
(Note: As shown with arrows, observer 1 and observer 2 might have chosen two different areas of the tumour (it is subjective) as 'HOT SPOT' leading to a significantly high difference in the final score. However, the global method diminishes this subjectivity by taking into account all the heterogeneous areas of the tumour resulting in a lesser mean difference and more agreement)

DISCUSSION

Ki-67 is a promising biomarker in various malignancies, as it shows peak expression in malignancies exhibiting high proliferation and poor differentiation [4]. In the study conducted by Arun I et al., Ki-67 PI in breast cancers and patient survival showed a significant degree of correlation [9]. Limited studies on its role as a predictive marker in candidate selection for further cytotoxic chemotherapy and molecular subclassification into luminal-A and luminal-B groups in hormone-positive breast cancer cases are also available. Moreover, newer studies explore the possibilities of the marker in companion diagnostics and for targeted molecular therapies [4,10,11]. The Food and Drug Administration (FDA) approved Ki-67 IHC MIB-1 pharmDx as a companion diagnostic for adjuvant abemaciclib in high-risk early breast cancer in 2021 [12]. However, questionable analytic validity limits the global adoption of the biomarker to drive patient care. The

recent American Joint Committee on Cancer (AJCC) guidelines also point out the limitations of integrating Ki-67 PI in clinical practice as a reliable factor because of poor reproducibility due to lack of standardised staining techniques and scoring methods [13].

Since 2011, the IKWG in order to determine analytic validity and promote standardisation in Ki-67 PI, has taken up multiphase multi institutional studies and hence could come up with a standard scoring method. The IKWG also introduced a free visual scoring application compatible with Android and iOS. The app provides options for scoring by global or hotspot techniques, following the proposed standardised method. In the global method, two scores are obtained: weighted and unweighted. The weighted score considers the relative percentages of the four areas (negligible, low, intermediate and high) of positively stained invasive tumour cells in the entire glass slide section, whereas the unweighted score considers only the sum of positively stained nuclei in these areas. In this study, the Android application was used to score the Ki-67 immunostained slides and weighted global scores were compared with the scores obtained by conventional Eyeball Estimate (EBE) [14]. Eyeball Estimate usually follows internally standardised laboratory-based assessment protocols, which varies across institutions and hence good intralaboratory reproducibility in these scores are usually attained but it fails in interlaboratory assessments. The phase 1 study by IKWG substantiates the same [15].

After the introduction of uniform scoring method in IKWG phase 2 multi-institutional study, the interobserver variability (ICC=0.92) showed a significant decline compared to that of phase 1 study (ICC=0.71) [15,16]. Similar results with high reproducibility in global weighted and unweighted scores (ICC=0.91) were observed by Arun I et al., without any significant bias in weighted score. The same study also compared app and EBE scores against digital image analysis scores, where the ICC was significantly higher between Digital Image Analysis (DIA) and the app compared to EBE. Several recent studies propose DIA as a faster and more standardised method and has shown to outperformed manual scoring techniques [9,17]. The ICC obtained in present study for global method (ICC=0.971) was also in concordance with these studies where the app has proven to be superior. The study chose global method over IKWG hotspot for evaluation, as a more robust score was obtained with the former method because the differences in individual fields average out and tumour heterogeneity is accounted for. The IKWG phase 3 study focused on understanding the variability from field selection, and both methods were compared. When IKWG HM was associated with higher variability and lower reproducibility (ICC=0.84), the global method met the prespecified criteria of success with better reproducibility (ICC=0.87) [18].

The 2015 St. Gallen consensus meeting recommended determination of cut-off for categorical classification of Ki-67 PI into low, intermediate and high based on the local laboratory median value, as it takes into account internally standardised preanalytical and analytical practices, scoring methods and population bias [8]. In present study, the median Ki-67 value for the study population was 24, and values 10% more and less from the value were considered as the cut-offs for high and low categories, respectively.

Despite standardisation in the assessment, the concordance among scorers in intermediate category was not as satisfactory as in the high and low categories. Both phase 2 and phase 3 IKWG studies concluded not to recommend clinical decision-making based on Ki-67 values in the intermediate category, for which the expensive commercial multiparameter gene assays have to be relied upon [16,18]. The present study observed that the ICC was lesser in the intermediate category (ICC=0.77) compared to the other two categories (ICC=0.96 in low and 0.98 in high) but was higher than that of institutional HM (ICC=0.59). The mean difference between two observer values was highest in intermediate category in both methods,

but it was not statistically significant in GW scores (p=0.901). It was observed that 21.8% of cases designated as intermediate by HM1 were categorically reclassified into higher and lower categories by HM2; however, this percentage of shift was only 12.5% between GW1 and GW2 app scores. These differences could be attributed to subjective assessments, as there are currently no guidelines regarding the choice of area within the slide for best prognostic information [9].

Multiple scoring methods for Ki-67 PI are available, but their validity has not been proven [19]. A few drawbacks of the IKWG recommended scoring method encountered compared to the other methods include relatively increased median scoring time of nine minutes per case and the requirement of calibration exercises [13]. This level of attention to training and the time required for each case possess a challenge in routine practice. In present study, that median scoring time for EBE was less than a minute, whereas it was seven minutes for app method.

Limitation(s)

The major limitation of our study was that it was a single-institution-based study. Several such studies and inter-institutional collaborations are required to study the precision of the global weighted scoring system and analyse its validity.

CONCLUSION(S)

High-quality tumour biomarker tests with proven analytic validity are critical in clinical decision-making. Utilisation of Ki-67 in determining the residual risk and its applicability as companion diagnostic can be improved by adopting a validated, universally standardised method. The Ki-67 visual scoring Android application is currently available free of cost, both online and offline and is a simple and easily comprehensible method utilising light microscopy. The global weighted scoring system obtained using the app would upgrade the analytic validity of Ki-67 PI, especially in resource-moderate countries like India.

REFERENCES

[1] Nahed AS, Shaimaa MY. Ki-67 as a prognostic marker according to breast cancer molecular subtype. Cancer Biol Med. 2016;13(4):496.

[2] Dowsett M, Nielsen TO, A'Hern R, Bartlett J, Coombes R, Cuzick J, et al. Assessment of Ki67 in breast cancer: Recommendations from the International Ki67 in Breast Cancer working group. J Natl Cancer Inst. 2011;103:1656-64. Doi: 10.1093/jnci/djr393.

[3] Aung TN, Acs B, Warrell J, Bai Y, Gaule P, Martinez-Morilla S, et al. A new tool for technical standardization of the Ki 67 immunohistochemical assay. Mod Pathol. 2021;34(7):1261-70.

[4] Li LT, Jiang G, Chen Q, Zheng JN. Ki67 is a promising molecular target in the diagnosis of cancer (review). Mol Med Rep. 2015;11(3):1566-72.

[5] Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, et al. American Society of Clinical Oncology/College of American pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. Arch Pathol Lab Med. 2010;134(6):907-22.

[6] Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American pathologists clinical practice guideline focused update. J Clin Oncol. 2018;36(20):2105-22.

[7] Arima N, Nishimura R, Osako T, Nishiyama Y, Fujisue M, Okumura Y, et al. The importance of tissue handling of surgically removed breast cancer for an accurate assessment of the Ki-67 index. J Clin Pathol. 2016;69(3):255-59.

[8] Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, et al. Tailoring therapies--improving the management of early breast cancer: StGallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. Ann Oncol. 2015;26(8):1533-46.

[9] Arun I, Venkatesh S, Ahmed R, Agrawal SK, Leung SCY. Reliability of Ki-67 visual scoring app compared to eyeball estimate and digital image analysis and its prognostic significance in hormone receptor-positive breast cancer. APMIS. 2021;129(8):489-502.

[10] Raab R, Ismaila N, Andre F, Stearns V, Kalinsky K. Biomarkers for adjuvant endocrine and chemotherapy in early-stage breast cancer: ASCO guideline update Q and A. JCO Oncol Pract. 2022;18(9):646-48.

[11] Johnston SRD, Toi M, O'Shaughnessy J, Rastogi P, Campone M, Neven P, et al. Abemaciclib plus endocrine therapy for hormone receptor-positive, HER2-negative, node-positive, high-risk early breast cancer: Results from a preplanned interim analysis of a randomised, open-label, phase 3 trial. Lancet Oncol. 2023;24(1):77-90.

[12] Center for Drug Evaluation, Research. FDA approves abemaciclib with endocrine therapy for early breast cancer [Internet]. U.S. Food and Drug Administration. FDA; 2022. [Internet]. [cited 2024 Apr 20]. Available from: https://www.fda.gov/drugs/resources-informationapproved-drugs/fda-approves-abemaciclib-endocrine-therapy-early-breastcancer.

[13] Nielsen TO, Leung SCY, Rimm DL, Dodson A, Acs B, Badve S, et al. Assessment of Ki67 in breast cancer: Updated recommendations from the International Ki67 in Breast Cancer Working Group. J Natl Cancer Inst. 2021;113(7):808-19.

[14] International Ki67 in Breast Cancer Working Group. [Internet]. [cited 2024 Feb 9]. Available from: https://www.Ki 67 in breast cancerwg.org/wp-content/uploads/2018/12/Ki 67-Phase-3b-WS-protocol-v1.pdf.

[15] Polley MY, Leung SC, McShane LM, Gao D, Hugh JC, Mastropasqua MG, et al. International Ki67 in Breast Cancer Working Group of the Breast International Group and North American Breast Cancer Group. An International Ki67 reproducibility study. J Natl Cancer Inst. 2013;105(24):1897-906.

[16] Polley MY, Leung SC, Gao D, Mastropasqua MG, Zabaglo LA, Bartlett JM, et al. An international study to increase concordance in Ki67 scoring. Mod Pathol. 2015;28(6):778-86.

[17] Zhong F, Bi R, Yu B, Yang F, Yang W, Shui R. A comparison of visual assessment and automated digital image analysis of Ki 67 labeling index in breast cancer. PLoS One. 2016;11(2):e0150505.

[18] Leung SC, Nielsen TO, Zabaglo L, Arun I, Badve SS, Bane AL, et al. Analytical validation of a standardized scoring protocol for Ki67: Phase 3 of an International multicenter collaboration. NPJ Breast Cancer. 2016;2(1):16014.

[19] Kadivar M, Aram F. Assessment of Ki67 in breast cancer: A comparison between the eye-10 method, stepwise counting strategy, and International system of Ki67 evaluation. Iran J Pathol. 2020;15(1):13-18.

PARTICULARS OF CONTRIBUTORS:
1. Resident, Department of Pathology, Ramaiah Medical College, RUAS, Bengaluru, Karnataka, India.
2. Assistant Professor, Department of Pathology, Ramaiah Medical College, RUAS, Bengaluru, Karnataka, India.
3. Professor, Department of Pathology, Ramaiah Medical College, RUAS, Bengaluru, Karnataka, India.
4. Professor, Department of Pathology, Ramaiah Medical College, RUAS, Bengaluru, Karnataka, India.
5. Professor, Department of Pathology, Ramaiah Medical College, RUAS, Bengaluru, Karnataka, India.

NAME, ADDRESS, E-MAIL ID OF THE CORRESPONDING AUTHOR:
Dr. Padma Priya Kasukurti,
Assistant Professor, Department of Pathology, Ramaiah Medical College, RUAS, Gokula, Bengaluru-560054, Karnataka, India.
E-mail: padmakasukurti@gmail.com

AUTHOR DECLARATION:
• Financial or Other Competing Interests: None
• Was Ethics Committee Approval obtained for this study? Yes
• Was informed consent obtained from the subjects involved in the study? No
• For any images presented appropriate consent has been obtained from the subjects. NA