(International Print/Online Journal)

SJIF IMPACT FACTOR: 5.565
PUBMED-National Library of Medicine ID-101739732

ISSN (Print): 2209-2870
ISSN (Online): 2209-2862

IJMSCR

International Journal of Medical Science and Current Research (IJMSCR)
Available online at: www.ijmscr.com
Volume 4, Issue 5, Page No: 1384-1394
September-October 2021

ISUP Grading For Prostate Cancer Pathology Images Using Deep Learning

Dr.Rajasekaran Subramanian[1*], Dr.Rohit Tapadia[2], Dr.R.Devika Rubi[3], Aadithya Pabbisetty[4]

[1*] Professor, [2]Sr. Consultant, Pathologist,[3]Associate Professor,[4]Research Intern,
[1,3]Keshav Memorial Institute of Technology, Hyderabad
[2]Tapadia Diagnostics Ltd, Hyderabad
[4]Neil Gogte Institute of Technology, Hyderabad

***Corresponding Author:**
Dr.Rajasekaran Subramanian
Keshav Memorial Institute of Technology, Hyderabad

Type of Publication: Original Research Paper
Conflicts of Interest: Nil

Abstract

Prostate cancer is the most leading and aggressive cancer among men in the world. Early diagnostics and treatment will reduce disease severity and deathrates. International Society of Urologic Pathologists (ISUP) Grading, grades the growth patterns of the prostate cancer cells by using Gleason scores 1-5 depends on its aggressiveness. Lack of expert Urology pathologists and the inter observer variability among the pathologists delays the treatment of the cancer. AI based deep learning (DL) models helps pathologist for an unbiased prediction and reduces their work load. As DL models learns the most significant cancer features from expert annotations produces fast, accurate and automatic ISUP grade prediction for prostate cancer. This paper proposes EfficientNet and Resnet DL models to predict ISUP grade over Kaggle Prostate Cancer Grade Assessment (PANDA) dataset. These models are trained and tested by 5000 prostate WSI and produced 80% accuracy with Quadratic Weighted Kappa (QWK) score of 0.6898.

Keywords: Prostate pathology grading, Prostate ISUP grades, Prostate Gleason score, Prostate pathology diagnostics, Prostate EfficientNet deep learning, feature extraction, classification

INTRODUCTION

Worldwide prostate cancer is the second most aggressive cancer among men. Every year around 1.4 million people are suffered by Prostate cancer (PC). This cancer affects 1 in 9 men all over the world [1]. Even though there are several methods like MRI are available for initial screening, the gold standard for diagnosis the severity of the PC is Gleason scoring system from prostate biopsy [2]. Diagnostic of prostate cancer reliable on prostate needle biopsy. The prognosis of prostate cancer depends on various features such as cancer type / stage and grading [3]. Currently this is a manual task done by a urological expert pathologist. Developing countries don't have adequate expert pathologists. In some African countries there is one expert pathologist available for one million people [4]. As there is a lot of inter observer variability occurs in this diagnostic process and lack of trained people leads inappropriate treatment disputes and misdiagnosis. The training cost is high due to lack of expert pathologists exist all over the world. [5]

PC is caused by male hormones like testosterone due to the growth of excessive abnormal cells and their accumulation. These DNA mutated cells divide rapidly than the normal cells which are present in and out of glands in cancerous tissue affects the structure and functionalities of prostate glands. High grade cancer fills up epithelial cells for stroma and lumen. Gleason scoring system is the most reliable method for evaluating aggressiveness of prostate cancer. It describes the cancer cells growth, patterns and the severity of the disease. This system grades the disease based on the granular patterns or tumor architectural growth patterns. Its grades or classify tumor varies from 1 (Excellent prognosis) to 5 (Poor prognosis). These histological growth patterns were developed by Dr.Donald Gleason in 1967 and later updated in 2014 [6]. After classifying the Gleason score it is converted to the respective ISUP (International Society of Urologic Pathologists) grade which is 1-5 scale [7]. The Vancouver consensus conference of the International Society of Urological Pathology (ISUP) provided the foundation for much of the 2016 World Health Organization (WHO) renal tumour classification [8].

Gleason grading system focuses not on the occurrence of worst patterns but on tumor architectural features and cytological appearances [9]. This system introduces a heterogeneous grading. The ISUP grade is assigned by both the minority and majority growth patterns of the tissue which is described by chen et al [10] and is listed in the Table 1.

Gleason Score (Minority + Majority)	ISUP Grade
(3+3)=6	1
(3+4)=7	2
(4+3)=7	3
(4+4),(5+3),(3+5)=8	4
(5+4),(4+5),(5+5)>=9	5

Table 1 The Gleason Score and their respective ISUP Grades

The Gleason score varies from 6 to 10 depending on the structural growth patterns of cancer cells in the tissue. The challenge of pathologists is to identify anomalies in H&E stained tissue which requires adequate training and immense concentration. There is a lot of inter observer variability between the diagnosis of the pathologist. Due to the bulk amount of diagnosis, pathologist feels heavy workload which increases the chances for misdiagnosis.

Computer Aided Diagnosis systems help pathologists for accurate and unbiased diagnosis. Early detection can be increased by the CAD system. The advent of massively parallel GPU system increases the speed and the accuracy of deep learning in the Gleason scoring system. Deep learning automatically extracts highly similar features in WSI which are learned from the annotations (ground truth) of the skilled/expert pathologist [11].

This paper is organised as follows: Section 2 describes the related work, Section 3 explained the proposed model, Section 4 discusses the results and evaluation parameters of the proposed model and Section 5 concludes the work.

RELATED WORK

Karimi D et al. [12] discusses the Gleason scoring system and the importance of CAD system in the diagnosis and prognosis of prostate cancer. It also discusses the vital role of deep learning in the analysis of H&E stained Whole Slide Images (WSI). Data pre-processing, post-processing techniques and limitations of deep learning (DL) in CAD are discussed. Santiago Toledo-Cortés et al [13] aims to develop a DL model based on the probabilistic regression and quantum measurement regression (QMR) to predict the Gleason score (GS) of prostate WSI. They use the public dataset TCGA-PRAD from the cancer Genome Atlas [14]. Xception CNN model extracts the features in patch level which is inputted to QMR. QMR predicts the disease grade of a single WSI. The model used 189 WSI for training and validation and 46 for testing. This model produces 0.5 mean and standard deviation accuracy and a macro F1 score of 0.293. The Multi-channel and Multi-spatial (MCMS) Attention model [15] were designed to enhance the feature extraction of CNN. The channel and spatial informatics are enhanced in this model. It uses Kaggle PANDA (16) dataset for prostate cancer grading. The model uses 1000 to 2000 images to train each ISUP grade and around 200 for to test each ISUP grade.

Volume 4, Issue 5; September-October 2021; Page No 1384-1394

Three different models such as Resnet, CBAM, and MCMS are used for training/testing and the highest accuracy among these models is 0.7179 with the kappa value of 0.85. Yet another automated Gleason grading system (YAAGGS) [17] deep learning model generates regional level automatic annotations. Deep learning models extracted common patterns from the huge expert annotated data and alsovthey designed a two-stage architecture. The first stage extracts Gleason patterns 3,4 and 5 which are then inputted to another model to predict Gleason scores. This model first extracts the WSI features at a 10x magnification level of size 360*360 and produces 1024 features to a two-dimensional feature map. The second CNN classifies it into six ISUP gradings by using the feature map. The inter-observer Gleason scoring model is done by Hangyang University Medical Center (HUMC) and Korea University Guro Hospital with a kappa value between 0.56-0.70 [18]. The Graph Neural Network (GNN) model [19] was designed to classify the Gleason score 7. This model considers WSI as graphs so that patch relationship and topological information about the WSI can be obtained. GNN classifies 7 (ie 3+4 and 4+3) Gleason score patterns. The performance of the model is based on a graphical conventional network (GCN) which gives global and local dependencies among graph nodes [20]. The model uses 406 WSI at 40x from TCGA [21] for training and testing which produces a classification accuracy of 79.5. Woulter bullin et al [22] discussed the assistance of AI-based deep learning models to pathologists. They investigated a panel of 14 pathologists graded 161 WSI with and without the Deep learning model which was developed by them [23]. The model was trained and tested by 5759 H&E 20x magnification WSI from Radboud university. The investigation concluded that the system with AI assistance improved the kappa value from 0.799 to 0.872 over the stand-alone AI model.

Oscar Jimenez–del–Toroa et al. [24], proposed a fully automatic approach that detects prostatectomy WSIs with high– grade Gleason score. This binary classification model classifies the WSI as low or high Gleason grade. The model used a binary tissue mask generated by using the Blue Ratio image (BR) processing technique in the 5× resolution tile. This preprocessing method can remove more than 80% of empty content of the image. The model used 47000 tiles of size 227*227 which were extracted from 235 H&E stained WSIs and achieved a classification accuracy of 73.52%. Hongming Xu et al. [25] proposed a model which uses statistical local binary pattern (CSLBP) descriptors for feature extraction. Along with this preprocessing technique Otsu threshold were applied on the patches of size 128*128 to distinguish the nucleus regions. Finally, multi-class SVM classifier produces a classification accuracy of 79% over 312 WSI from TCGA. Ali Tabesh et al [26] designed a fully automatic system that extracts color, texture, and morphometric features from the prostate tissue. The system enhanced the features by a sequential forward search (SFS) algorithm and support vector machine (SVM) classifiers is used for classification. Total of 367 tiff images of size 1600*1200 are used as dataset and achieved an accuracy of 81% with a 5-fold CV classification.

PROPOSED WORK

This section discusses the description of dataset, preprocessing techniques and proposes two different deep learning (DL) models to predict ISUP grade along with Gleason score classification for prostate WSI.

a.Dataset Description:

The proposed DL model uses Kaggle Prostate Cancer Grade Assessment (PANDA) challenge dataset provided by Radboud University Medical Center and Karolinska Institute.[27].This dataset consists of 5000 prostate WSI and its respective 5000 mask along with the semantic annotations of most relevant Gleason score patterns. These Gleason patterns help to predict ISUP grade. The description of semantic annotations are shown in Table 2.

Volume 4, Issue 5; September-October 2021; Page No 1384-1394

Radboud	Karolinska
0: background (non -issue) or unknown 1: stroma (connective tissue, non-epithelium tissue) 2: healthy (benign) epithelium 3: cancerous epithelium (Gleason 3) 4: cancerous epithelium (Gleason 4) 5: cancerous epithelium (Gleason 5)	0: background (non-tissue) or unknown 1: benign tissue (stroma and epithelium combined) 2: cancerous tissue (stroma and epithelium combined)

Table 2 Semantic annotation descriptions of the mask

Radboud Dataset contains a .CSV file which consists of img_id,data provider information, ISUP grade and Majority and minority Gleason scores.

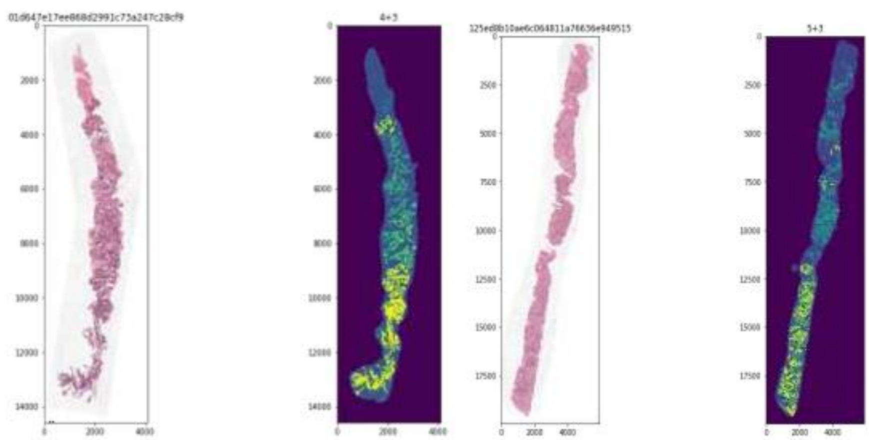


Figure 1 Sample WSI and its respective semantic mask

Figure 1 shows sample prostate WSI images with image id and their corresponding semantic mask which represents in different colors for Majority and minority Gleason scores.

a.Proposed Models

DEEP LEARNING MODEL-1:

Pre-processing technique:

Prostate WSI consist, a lot of empty pixels which need to be removed to enhance the performance of DL. For this pre- processing, the model uses tileing method [28] and concatenate tile pooling method [29]. According to these methods the entire WSI is divided into the tiles of size 256*256

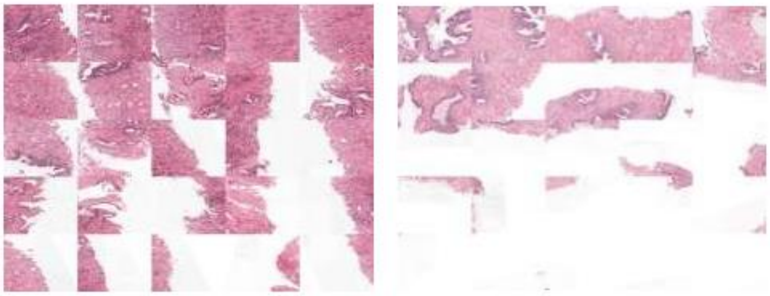
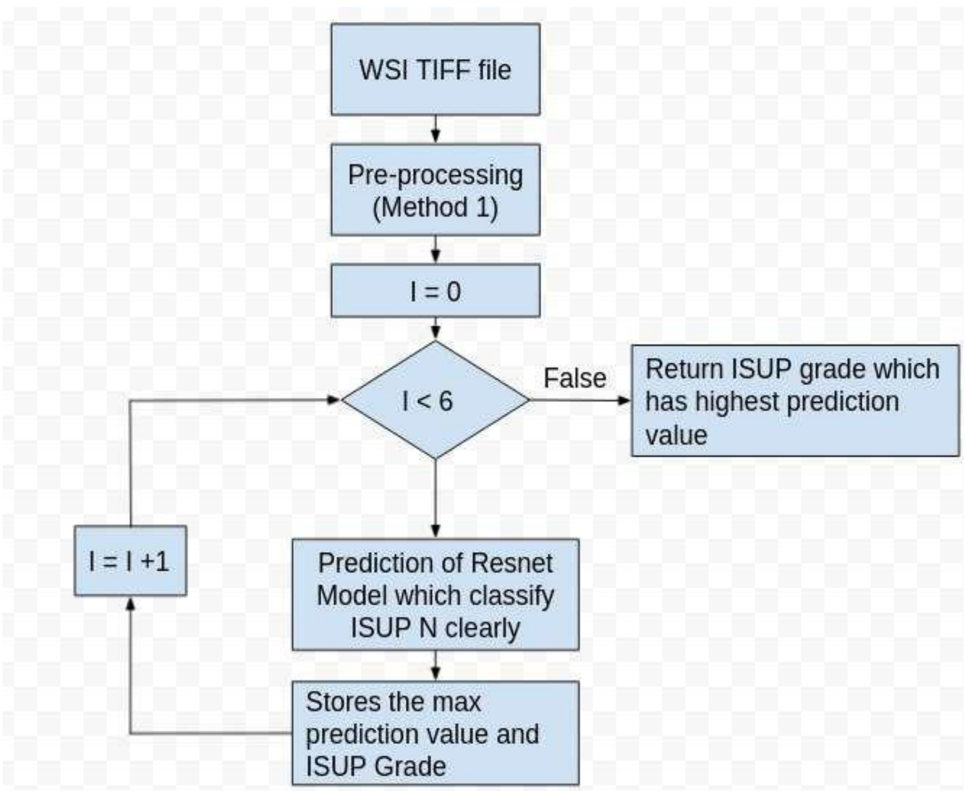


Figure 2 DL Model 1 preprocessed WSI

The top 25 tiles which are having more than 95% percentage of tissue are filtered out and concatenate into a single WSI of size 1280*1280 with the same ISUP label given in the annotation. Figure 2 shows sample pre-processed WSI by using tile concatenate pooling method.

Proposed Model-1

Resnet-101 [30] is used as a multi-classification prediction model for ISUP on prostate WSI. The pre-processed images having label ISUP from 0 to 5(0 for normal and 1-5 cancerous) are trained and tested. Around 4000 images are used for training and 1000 images are used for testing. As the model produces a poor 47% of accuracy, the model is redesigned as having individual Resnet for each ISUP grade and the predicted ISUP



grade with highest accuracy has been considered as an output of the model. The workflow is defined in Figure 3.

Figure 3 Workflow for ResNet prediction

The proposed model produced Accuracy around 68% after 8 epochs for each ISUP classification model with a quadratic weighted kappa of 0.674. The limitation of this model is each prediction needs to run 6 Resnet models to find the respective ISUP between 0 to 5.

DEEP LEARNING MODEL-2:

Pre-processing Technique:

The mask image of the dataset contains annotations for different Gleason scores. Deep Learning model-1 considers entire WSI and its respective mask for training and testing whereas Deep Learning model-2 divides the WSI and its respective semantic mask into four Gleason subgroups which is shown in Fig 4.

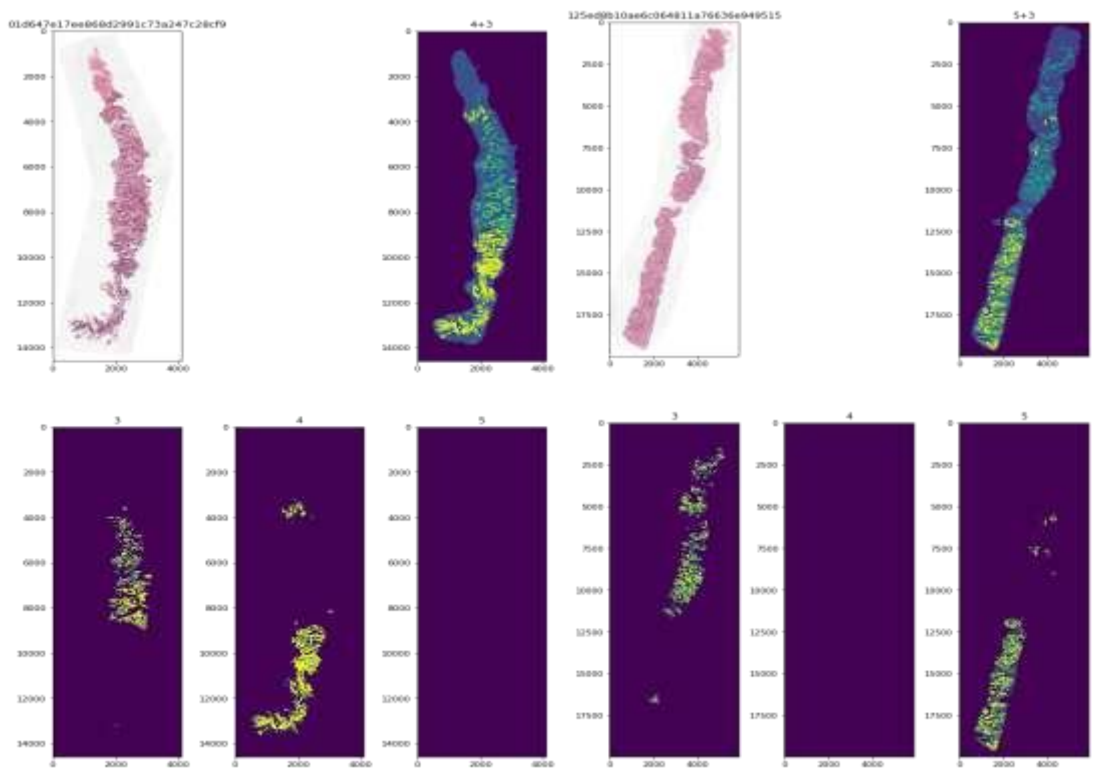


Figure 4: Sample WSI, Mask WSI and the respective Gleason Patterns

Deep Learning Model 2, divides entire WSI and its respective mask by 224*224.The pre-processing method removes all empty tiles i.e., the tile without tissue and then the tiles are categorised into four groups as

Label 0 for non-cancerous

Label 1 for Gleason score 3

Label 2 for Gleason score 4

Label 3 for Gleason score 5

Volume 4, Issue 5; September-October 2021; Page No 1384-1394

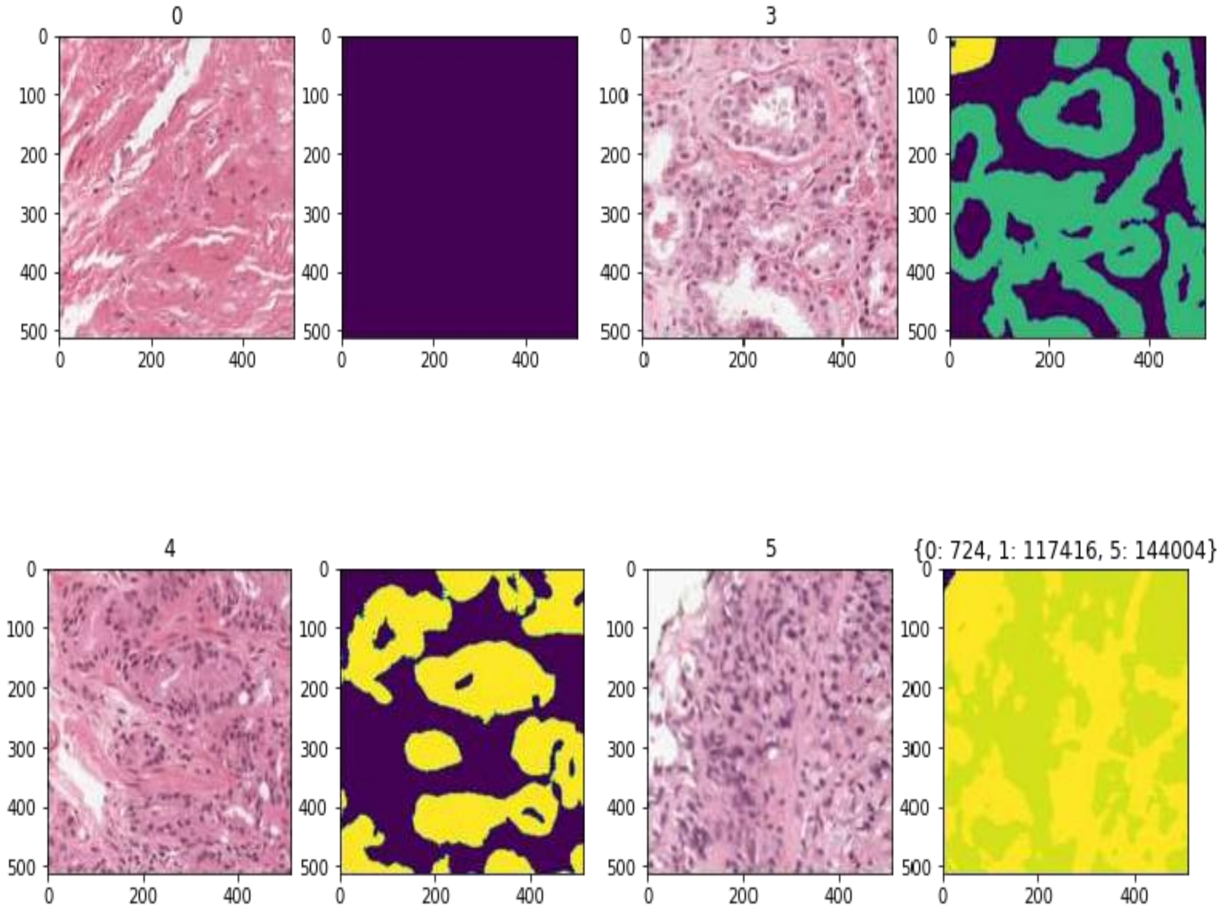


Figure 5 Sample input and Mask after pre-processing of WSI

These groups are classified with the help of semantic labelling in mask WSI. The pre-processing method first finds the respective tiles for each category and then the corresponding WSI tiles are separated. Figure 5 shows the pre-processed sample tiles for Deep Learning model-2 training.

The training dataset after pre-processing contains four categories of tiles of size 224*224 with Gleason score. The dataset consists of 5000 WSI which are divided into 92783 tiles with different Gleason score. EfficientNet [31][32] model is used as a deep learning model for training and to predict the Gleason score. After predicting the Gleason score of each tile, they were sorted out and the most occurred top two Gleason scores are calculated which decides the ISUP grade of the WSI. The model achieved accuracy of 75.8 % for Gleason score classification and got an ISUP grade accuracy of 80%.

RESULTS AND DISCUSSION

Large number of white space tiles (empty tiles) which have no tissue are found in Prostate WSI. As Deep learning models learn the features from tiles, these white tiles need to be removed to improve the performance of AI. Deep Learning model-1 used a complete WSI which is large in size for DL training, makes the model to learn lesser features and produce lesser accuracy around 68%. To increase the accuracy model 2 uses tiles for Deep Learning training/testing and produced an accuracy around 80%. If the Dataset size is increased then the accuracy will also further improve. The sample output of the model to predict ISUP for prostate WSI and the performance table of the models are shown in Figure 6 and Table 4

Volume 4, Issue 5; September-October 2021; Page No 1384-1394

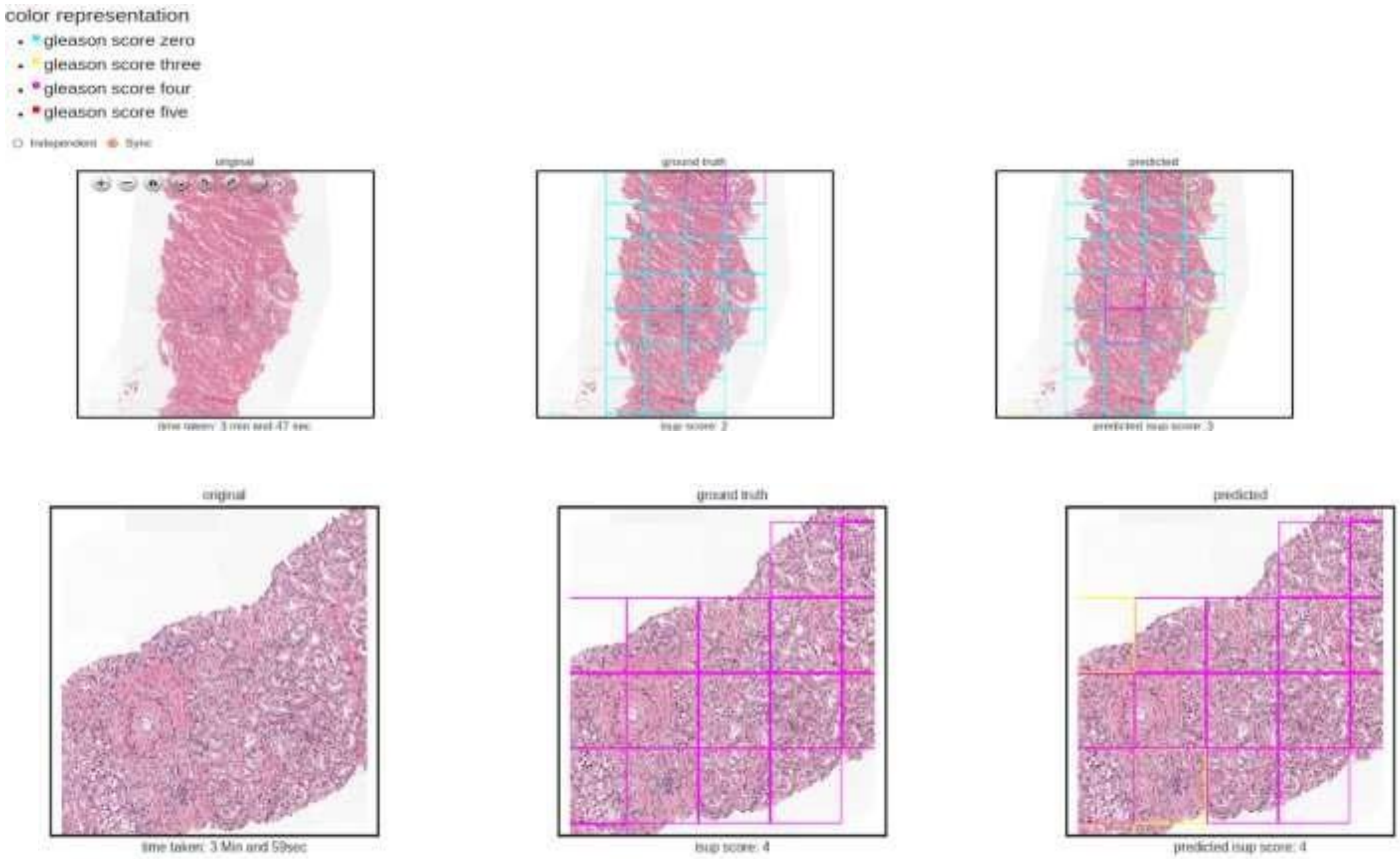


Figure 6 Sample prediction of ISUP for Prostate WSI

Model	Acc	Qwk
ResNet (Submitted to Kaggle)	0.6819	0.6746
EfficientNet [Gleason score]	0.7536	0.8019
EfficientNet [ISUP]	0.8016	0.6898

Table 4 Performance of Each Deep Learning Model

The results of the deep learning model conclude that the model trained after pre-processing will improve accuracy and more WSI with ground truth will increase the performance of AI. The Efficient Net model produces an accuracy of 75% and 80% with Quadradic kappa value .80 and .68 for Gleason score classification and prediction of ISUP grade for prostate WSI respectively

CONCLUSION

The World's leading and most common cancer for men's death is prostate cancer which can be diagnosed by Urology pathologist and its severity is measured with the patterns defined by ISUP grades. The unbiased and early detection reduce the severity of the cancer and its death rate. Lack of experienced pathologist occurred all over the world. Deep

Volume 4, Issue 5; September-October 2021; Page No 1384-1394

learning models provide solution to this problem and producing an unbiased automatic prediction of ISUP grade for prostate cancer. This paper implemented ResNet and EfficientNet deep learning AI models over PANDA data set which consists of around 5000 prostate WSI with semantic annotations for Gleason scores. The model produced 80% accuracy with QWK 0.68 which is almost better than other existing models. The limitation of model is lack of data and the increment in the size of the dataset and their respective annotations will improve the accuracy of the DL models.

ACKNOWLEDGEMENTS

Prof. Neil Gogte, Director, Keshav Memorial Institute of Technology for the Project Guidance, Finance and Material support.

REFERENCES

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R., Torre, L., Jemal, A. “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries (vol 68, pg 394, 2018)”, CA Cancer J. Clin. 2020, 70, 313 mine is latest [1]
2. Faraj, S.F., Bezerra, S.M., Yousefi, K., Fedor, H., Glavaris, S., Han, M.,Partin, A.W., Humphreys, E., Tosoian, J., Johnson, M.H., Davicioni, E.,Trock, B.J., Schaeffer, E.M., Ross, A.E., Netto, G.J. “Clinical validation of the 2005 isup gleason grading system in a cohort of intermediate and highrisk men undergoing radical prostatectomy”, PLoS ONE 11(1), 1–13 (2016). https://doi.org/10.1371/journal.pone.0146189
3. Buhmeida, A., Pyrhönen, S., Laato, M. & Collan. Y, “Prognostic factors in prostatecancer”, Diagn. Pathol. 1, 4, https://doi.org/10.1186/1746-1596-1-4 (2006)
4. Egevad, L., Delahunt, B., Samaratunga, H., Leite, K.R., Efremov, G., Furusato, B., Han, M., Jufe, L., Tsuzuki, T., Wang, Z., et al, “The International Society of Urological Pathology Education web—A web-based system for training and testing of pathologists”, Virchows Arch. 2019, 474, 577–58
5. Li, Y., Huang, M., Zhang, Y., Chen, J., Xu, H., Wang, G., Feng, W., “Automated Gleason Grading and Gleason Pattern Region Segmentation Based on Deep Learning for Pathological Images of Prostate Cancer”, IEEE Access 8, 117714–117725(2020). https://doi.org/10.1109/ACCESS.2020.3005180
6. Gleason Donald F, “Histologic grading of prostate cancer: a perspective”, HumPathol 1992,23(3):273–9. https://doi.org/10.1016/0046-8177(92) 90108-F.ISSN 0046–8177. [4]
7. 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.
8. Moch, H., Cubilla, A. L., Humphrey, P. A., Reuter, V. E.,and Ulbright, T. M., “The 2016 who classification of tumours of the urinary system and male genital organs—part a: Renal, penile, and testicular tumours”, European Urology, 70(1):93–105
9. Short Emma, Warren Anne Y, Murali Varma, “Gleason grading of prostate cancer:a pragmatic approach”, Diagn Histopathol 2019,25(10):371–8.
10. Chen Ni, Qiao Zhou, “The evolving Gleason grading system”, Chin J Canc Res 2016,28(1):58. [5]
11. Xu, H., Park, S., Hwang, T.H, “Computerized classification of prostate cancer gleason scores from whole slide images”, IEEE/ACM Trans. Comput. Biol. Bioinform. 2019, 17, 1871–1882
12. Karimi D, Nir G, Fazli L, Black PC, Goldenberg L, Salcudean SE, “Deep Learning-Based Gleason Grading of Prostate Cancer From Histopathology Images-Role of Multiscale Decision Aggregation and Data Augmentation”, IEEE J Biomed Health Inform. 2020 May,24(5):1413-1426. doi: 10.1109/JBHI.2019.2944643. Epub 2019 Sep 30. PMID: 31567104.

13. Santiago Toledo-Cortés, Diego H. Useche, Fabio A. González, “Prostate Tissue Grading with Deep Quantum Measurement Ordinal Regression”, https://arxiv.org/abs/2103.03188

14. Jim ́enez del Toro, O., Atzori, M., Ot alora, S., Andersson, M., Eur ́en, K.,Hedlund, M., R onnquist, P., M uller, H., “Convolutional neural networks for an automatic classification of prostate tissue slides with high-grade Gleason score”,

15. Medical Imaging 2017: Digital Pathology 10140, 101400O(2017). https://doi.org/10.1117/12.225571

16. Yang B, Xiao Z, “A Multi-Channel and Multi-Spatial Attention Convolutional Neural Network for Prostate Cancer ISUP Grading”, Applied Sciences. 2021, 11(10):4321. https://doi.org/10.3390/app11104321

17. https://www.kaggle.com/c/prostate-cancer-grade-assessment (accessed on 20 March 2021)

18. Mun, Y., Paik, I., Shin, SJ. *et al*, “Yet Another Automated Gleason Grading System (YAAGGS) by weakly supervised deep learning”, *npj Digit. Med.* **4,** 99 (2021). https://doi.org/10.1038/s41746-021-00469-6

19. Renshaw, A. A. et al, “Accurate Gleason grading of prostatic adenocarcinoma in prostate needle biopsies by general pathologists”, Arch. Pathol. Lab. Med. 127,1007–8,https://doi.org/10.1043/1543-2165(2003) 127

20. Jiang, Nan & Hou, Yaqing & Zhou, Dongsheng & Wang, Pengfei & Zhang, Jianxin & Zhang, Qiang. (2021), “Weakly Supervised Gleason Grading of Prostate Cancer Slides using Graph Neural Network”, 426-434. 10.5220/0010264804260434.

21. Wang, J., Chen, R. J., Lu, M. Y., Baras, A. S., and Mahmood, F, “Weakly supervised prostate tma classification via graph convolutional networks”, 2019, arXiv:Computer Vision and Pattern Recognition.

22. Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K.R. M., Ozenberger, B., Ellrott, K., Sander, C., Stuart,J. M., Chang, K., Creighton, C. J., et al, “The cancer genome atlas pan-cancer analysis project”, Nature Genetics, 2013, 45(10):1113–1120.

23. Bulten, W., Balkenhol, M., Belinga, JJ.A. et al, “Artificial intelligence assistance significantly improves Gleason grading of prostate biopsies by pathologists”, Mod Pathol 34, 660–671 (2021). https://doi.org/10.1038/s41379-020-0640-y

24. Bulten W, Pinckaers H, van Boven H, Vink R, de Bel T, van Ginneken B, et al, “Automated deep-learning system for Gleason grading of prostate cancer using biopsies: a diagnostic study”, Lancet Oncol. 2020,21:233–41.

25. Jimenez-del Toroab, O., et al., “Convolutional neural networks for an automatic classification of prostate tissue slides with high-grade Gleason score”, In: Proceedings of SPIE, vol. 10140 (2017). 101400O–1.

26. Hongming Xu, Sunho Park, and Tae Hyun Hwang, “Automatic classification of prostate cancer gleason scores from digitized whole slide tissue biopsies”, In 17th International Workshop on Data Mining in Bioinformatics, 2018.

27. A. Tabesh, M. Teverovskiy, H. Pang, V. Kumar, D. Verbel, A. Kotsianti, et al., "Multi Feature prostate cancer diagnosis and gleason grading of histological images", IEEE Transactions on Medical Imaging, vol. 26, no. 10, pp. 1366- 1378, 2007

.https://www.kaggle.com/c/prostate-cancer-grade-assessment

28. Kaggle.com. PANDA 16x128x128 tiles. https://www.kaggle.com/c/prostate-cancer-grade- assessment/discussion/146855. [Accessed 20 March 2021].

29. Kaggle.com. PANDA concat tile pooling starter [0.79 LB]. https://www.kaggle.com/iafoss/panda-concat-tile- pooling-starter-0-79-lb. [Accessed 20 March2021].

30. Ho, T.K.K., Gwak, J., Prakash, O., Song, J.I., Park, C.M., “Utilizing pretrained deep learning models for automated pulmonary tuberculosis

detection using chest radiography", In: Asian Conference on Intelligent Information and Database Systems, pp. 395–403. Springer (2019)

31. Luz, E., Silva, P., Silva, R. et al, "Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images", Res. Biomed. Eng. (2021). https://doi.org/10.1007/s42600-021-00151-6
32. Wang, J., Liu, Q., Xie, H., Yang, Z., Zhou, H, "Boosted EfficientNet: Detection of Lymph Node Metastases in Breast Cancer Using Convolutional Neural Networks", Cancers 2021, 13, 661. https://doi.org/10.3390/cancers13040661