Journal of Pathology Informatics 14 (2023) 100155

Contents lists available at ScienceDirect

Journal of Pathology Informatics

journal homepage: www.elsevier.com/locate/jpi

Differentiation of urothelial carcinoma in histopathology images using deep learning and visualization

Aniruddha Mundhada [a,*], Sandhya Sundaram [a,*], Ramakrishnan Swaminathan [b], Lawrence D' Cruze [a], Satyavratan Govindarajan [b], Navaneethakrishna Makaram [b]

[a] Department of Pathology, Sri Ramachandra Institute of Higher Education and Research, Chennai, Tamilnadu, India
[b] Biomedical Engineering, Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, Tamilnadu, India

ARTICLE INFO

Keywords:
Convolutional neural network
Deep learning
Urothelial carcinoma
Transurethral resection of bladder tumor
Whole slide image
Visualization

ABSTRACT

Artificial Intelligence is a tool poised to transform healthcare, with use in diagnostics and therapeutics. The widespread use of digital pathology has been due to the advent of whole slide imaging. Cheaper storage for digital images, along with unprecedented progress in artificial intelligence, have paved the synergy of these two fields. This has pushed the limits of traditional diagnosis using light microscopy, from a more subjective to a more objective method of looking at cases, incorporating grading too. The grading of histopathological images of urothelial carcinoma of the urinary bladder is important with direct implications for surgical management and prognosis.

In this study, the aim is to classify urothelial carcinoma into low and high grade based on the WHO 2016 classification. The hematoxylin and eosin-stained transurethral resection of bladder tumor (TURBT) samples of both low and high grade non-invasive papillary urothelial carcinoma were digitally scanned. Patches were extracted from these whole slide images to feed into a deep learning (Convolution Neural Network: CNN) model. Patches were segregated if they had tumor tissue and only included for model training if a threshold of 90% of tumor tissue per patch was seen. Various parameters of the deep learning model, known as hyperparameters, were optimized to get the best accuracy for grading or classification into low- and high-grade urothelial carcinoma. The model was robust with an overall accuracy of 90% after hyperparameter tuning. Visualization in the form of a class activation map using Grad-CAM was done. This indicates that such a model can be used as a companion diagnostic tool for grading of urothelial carcinoma. The probable causes of this accuracy are summarized along with the limitations of this study and future work possible.

Background

Urothelial cell carcinoma is the most common cancer of the urinary bladder comprising about 90% of all cancers of the bladder. Most cases occur in males over 50 years of age; more in Caucasian males. Etiologic factors include genetic as well as environmental factors; the latter includes exposure to smoking, petrochemicals, aniline dyes, auramines, phenacetin, and cyclophosphamide.[1]

Genetic pathways are mainly 2 separate mechanisms – FGFR3-associated pathway and the TP53-associated pathway. Low-grade non-invasive urothelial carcinomas are associated with activating mutations in FGFR3 while high-grade urothelial carcinomas involve mutations in the TP53 or CDKN2A gene. Parasites such as *Schistosoma haematobium* (urinary blood fluke) is also noted to cause urothelial carcinoma.[2]

LGUC and HGUC non-invasive papillary carcinoma – those are differentiated on architecture and cytological features by histopathological examination. LGUC have delicate papillae and relatively orderly cellular arrangement at low power, and mild nuclear irregularity, mild pleomorphism, and few or minimal mitosis at high power. On the other hand, HGUC at low-power magnification shows architectural features like fused solid papillae with disorderly cell arrangement, with high-power magnification showing nuclear atypia, pleomorphic and irregular nuclei, irregular prominent nucleoli, and numerous mitoses. There is 5% of high-grade component to be present to call the specimen HGUC.

The grading of urothelial carcinoma is important due to the following reasons:

1. It is one of the most important parameters for prognostic significance according to the European Organization for Research and Treatment of Cancer (EORTC) and World Health Organization (WHO).[3]
2. No immunohistochemistry is available for differentiating the grades of urothelial carcinoma.

There are many parameters that are helpful in predicting the risk of recurrence like proliferation rate measured by different methods,[4] mutations

* Corresponding author at: Department of Pathology, SRIHER Chennai, Chennai 600116, Tamilnadu, India.
E-mail addresses: m0619010@sriramachandra.edu.in (A. Mundhada), sandhyas@sriramachandra.edu.in (S. Sundaram).

http://dx.doi.org/10.1016/j.jpi.2022.100155
Received 26 July 2022; Received in revised form 16 October 2022; Accepted 2 November 2022
Available online 08 November 2022

in FGFR3 and PIK3CA associated with superficial low-grade urothelial carcinomas, while PTEN and p53 mutations are related to high-grade urothelial carcinoma. The most common cytogenetic abnormality in non-invasive urothelial carcinoma is partial or complete loss of chromosome 9.[5]

The urinary bladder is a hollow viscus acting as a temporary reservoir of urine for storage before voiding. To contract during micturition, the bladder wall contains specialized smooth muscle – known as detrusor muscle. High-grade urothelial carcinoma can invade the muscle and hence, requires more aggressive surgical management and frequent follow-ups.

The WHO 2016 classification of tumors of the urothelial tract lists the consensus classification used in the WHO 2016 blue book after clinical validation.[3] It divides the non-invasive papillary urothelial carcinoma into low and high grade. The classification is a diagnostic challenge for many pathologists – with few biomarkers to aid in the diagnosis. This study focuses on distinguishing these as a 2-class problem. The diagnosis to distinguish low- and high grade is challenging for pathologists and using AI to classify the 2 will provide an objective screening method for diagnosis.

Artificial intelligence is a science of developing machines capable of making intelligent decisions. This branch of computer science evolved with the rise of computational capabilities and labelled datasets.

Traditional image processing involves multitude of steps which are manually done – this is possible if the images are less in number. Features which are important can be chosen in each given image for image classification done through traditional methods. As the number of classes to classify increases, feature extraction becomes more and more cumbersome.

Machine learning is a branch of AI that deals with systems to learn from data and make decisions based on earlier patterns after data analysis.[6] A subset of machine learning is deep learning that is inspired by the way the human brain makes a decision. The algorithm is inspired from a neuron and a simplest deep learning model is the perceptron. This has only a single layer of neurons or nodes capable of processing information. More complex algorithm has more nodes and more layers in it that deals with more data in multiple layers, hence the term deep learning. The way this system works is by recognizing features which are defining characteristics of an image.[7]

The most important difference between deep learning and traditional methods is its performance as data increases. Deep learning algorithms need a large amount of labelled data to train and validate and hence, function well if the dataset is big.[8]

Materials and methods

Patient selection

A total of 20 non-muscle invasive bladder cancer (NMIBC) specimens from 20 patients from SRMC, Chennai were included. All patients underwent a transurethral resection of bladder tumor (TURBT) procedure between January 2020 and December 2021. The Institutional Review Board of the Sri Ramachandra Institutional Ethics Committee, Sri Ramachandra Medical College, Chennai has granted approval to conduct this research study. These specimens were obtained with the aim to classify patients according to WHO 2016 guidelines for non-muscle invasive bladder cancer. Hematoxylin and eosin-stained slides from 4 μm thick sections from formalin fixed paraffin embedded (FFPE) tissue blocks were digitized using the Morphle DigiPath 6T Scanner (Morphle Labs, Bangalore, India). The digitized sections were exported with 40x resolution, resulting in an in-plane resolution of 0.22 μm/pixel. (Fig. 1 shows the overall workflow).

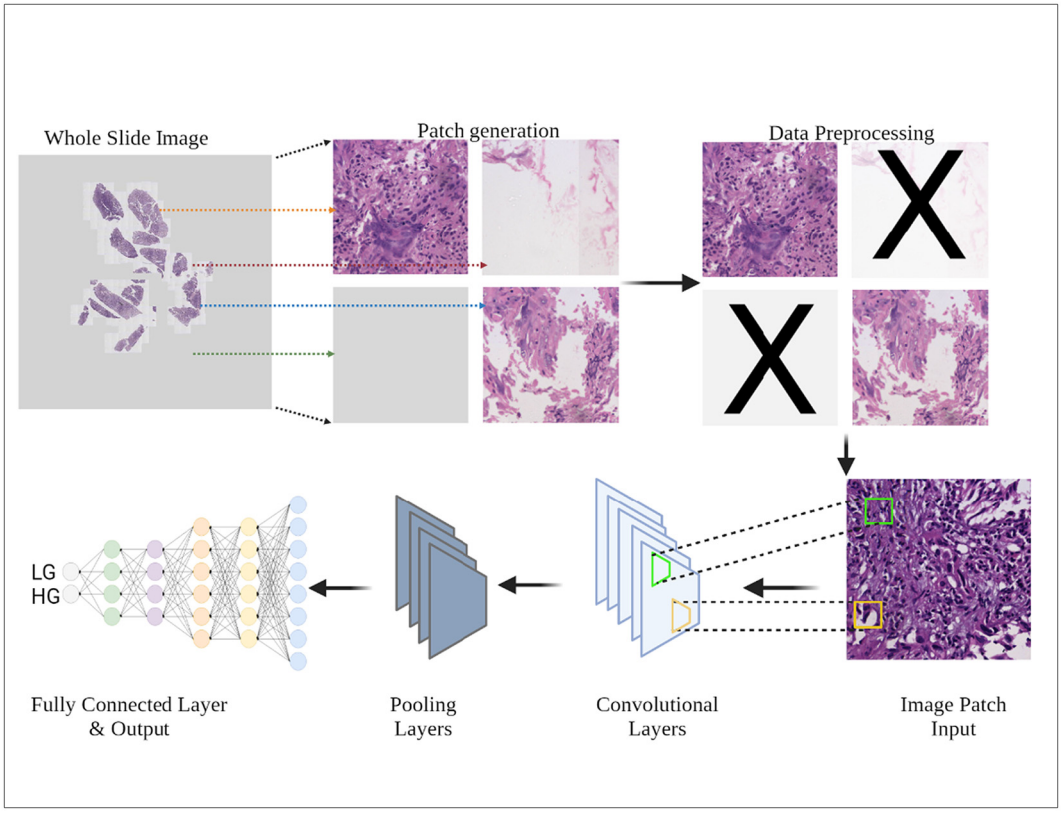
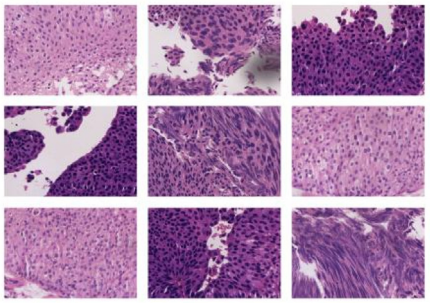
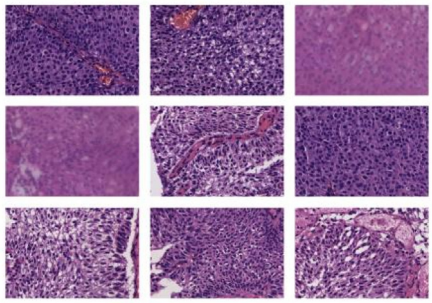


Fig 1. Classification workflow: H&E-stained histopathology slide of the TURBT sample was scanned – patches were extracted using a Python code from this whole slide image (WSI). Patches with less than 90% tumor tissue was discarded. Manual annotation was done for all the patches and artifacts separated out – generating about 60 000 patches of each class. This is known as data cleaning or data preprocessing. The urothelium regions as determined by the data preprocessing and cleaning and manual annotation were then split into 70:15:15:: training:validation:testing ratio, and fed into a ImageNet pre-trained 16-layer VGG architecture. The convolution layers extracted relevant features from these patches and fed them into a pooling layer that reduces the number of features to make it easier to compute and avoid overfitting and generalization. These are then used in the fully connected layer for giving an output.



A. Low grade urothelial carcinoma: patches of tumor cells at low magnification showing tall, thin, fingerlike, and branching papillary architecture with orderly cellular arrangement, characteristic of low grade urothelial carcinoma.



B. High grade urothelial carcinoma: patches of tumor cells at low magnification showing complex, fused papillary architecture with marked loss of polarity,, characteristic of high grade urothelial carcinoma.

Fig 2. H & E, 100x. Patches of LGUC & HGUC of size 1024 × 1024 pixels.

Exclusion criteria

Diagnosis of other urothelial pathologies and muscle invasive urothelial carcinoma and resection specimens were excluded.

Data collection

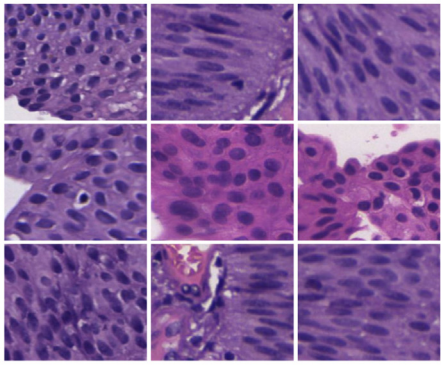
The scanned whole slide image (WSI) ranged from 500 MB to 8.8 GB in size. The scanning was done on newly prepared hematoxylin and eosin-stained tissue sections. Care was taken to remove the dirt and scan only clean slides. This ensured that any artifacts related to slide preparation, staining, and scanning were kept to a minimum.

Patch generation and image preprocessing

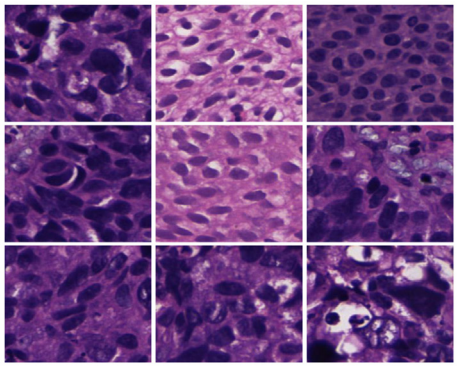
The whole slide images were in tiff format with artifacts also present. Patches were generated using python code of size 256 × 256 pixels with zero overlap of area. This resulted in 134 500 and 144 500 patches of low- and high-grade papillary urothelial carcinoma. This data was then processed to remove the whitespace – a cut-off of 90% of tumor tissue present in patches. Remaining patches were manually checked for any stain deposit, blurring, scanning artifacts, and pure stroma. This resulted in 64 700 and 67 500 patches of high- and low-grade urothelial carcinoma. The training:validation:testing is done on 70:15:15 split of the dataset. This corresponds to 92 540:19 830:19 830 image patches. (Fig 2 & 3 show the different patches. Fig 4 shows the different tissue artifacts).

Data split

All digitized images were split into training, testing, and validation sets containing 70%, 15%, and 15% of all patches amongst both the classes. The split was done at the patient level. The patients were segregated at random into 70:15:15:: training:validation:testing ratio. This ensured a random split with tiles or patches being only in training, validation, or testing sets.



A. Low grade urothelial carcinoma: Patches of tumor cells at high magnification with relatively ordered arrangement with mild loss of polarity, mild nuclear pleomorphism and atypia is noted. Nucleoli are inconspicuous and mitosis is hardly seen.



B. High grade urothelial carcinoma: Patches of tumor cells at high magnification showing loss of polarity with disordered cellular arrangement, nuclear overlapping, pleomorphism, prominent nucleoli and mitotic figures are noted.

Fig 3. H & E, 400x. Patches of LGUC & HGUC of size 256 × 256 pixels.

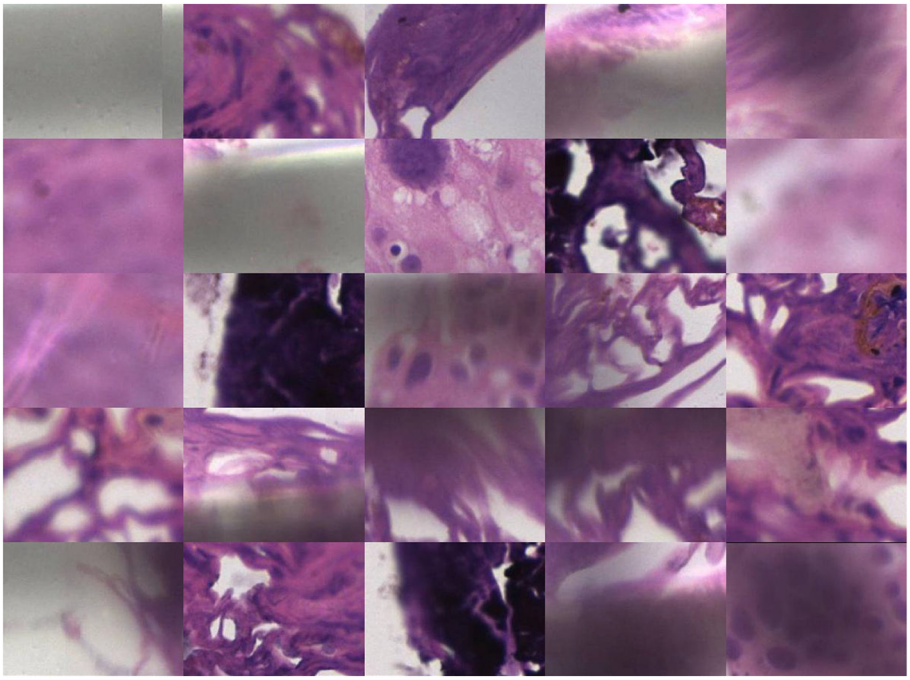


Fig 4. Patches of various artifacts of size 256 × 256 pixels showing blank spaces within tumor papillae, pure stroma with few or no cellularity, stain deposits, overstained or improperly stained sections, blurred out-of-focus images, and thicker sections leading to poor image analysis.

Care was taken that images were not overlapping that may affect the accuracy adversely.

Different variations tried

Patches of 1024 × 1024 pixels were generated – this had a better inter-observer correlation amongst pathologists for the diagnosis. This led to generation of 12 450 and 13 440 patches of low- and high-grade papillary urothelial carcinomas.

Annotations of tumoral areas

All exported images were annotated using manual annotation by experienced pathologists. All regions with tumor tissue were delineated and annotated, and non-atypical urothelium and extensive regions of fibrovascular tissue, tissue folds, mechanical damage such as cauterization artifacts, blank areas arising from TURBT extraction or sectioning, and out-of-focus regions, were annotated. However, these regions were not meticulously annotated. Patches with poor staining or less than 90% of tumor tissue were removed from the dataset. All annotations were checked by a specialized uropathologist. All cases were previously graded by 3 experienced pathologists with experience in reporting uropathology cases. Briefly, the grade of the tumors was initially assessed according to the histopathological pattern using the WHO 2016 grading system, followed by a subsequent consensus reading in case of disagreement. Several slides were studied for TURBT sample per patient for making a diagnosis.

Classification network

The urothelium regions as determined by the pathologist were used for the training of the classification network. ImageNet pretrained 16-layer VGG architecture was optimized for the grading of bladder tissue.

Training of the convolutional neural network

In this study, NVIDIA Quadro P2200 GPU with CUDA cores was used.

VGG16 was developed by Visual Geometry Group. The VGG16 model pretrained for generic image classification, was adapted to use for our classification problem. VGG16 has 16 layers; with a filter size of 8 × 8 was used.[9] The final layer of the network was removed and replaced with a single fully connected layer for our 2-class classification problem. While fixing the ImageNet feature maps of the earlier layers, the additional fully connected layer was trained on our dataset using gradient descent. The input to the network were 1024 × 1024 size RGB pixel patches and 256 × 256 size RGB pixel patches from the whole slide images. The Keras framework (v2.4.0) with Tensorflow (v2.7.0) was used as the backend. The softmax function was used as the activation function of the fully connected layer. Adam optimizer was used, with a learning rate of 0.001.

On the test set, the accuracy was assessed. The accuracy for the whole test set was then calculated. The graph showed the gain in accuracy during training and the decrease in loss, which mainly suggests the confidence by which the model is classifying the patches into one of the 2 classes.

Results

The regions segmented by the expert annotation by a trained uropathologist resulted in a total of 128 000 patches from both classes for the classification network. WSI were used to annotate; and then, images were split into patches using MATLAB R2021b. Annotation was done according to the WHO 2016 grading scheme for urothelial carcinoma.

The patch extraction after digitizing the slides yielded a total of 132 200 image patches at 256 × 256 pixel patch size distributed as 67 500 and 64 700 patches of low- and high-grade urothelial carcinoma [approximately, four billion two hundred sixty million annotated pixels each, of which only 51 million pixels contained non-atypical urothelium].

A confusion matrix shows the number of patches which were classified as true positives and true negatives. In the 1024 × 1024 patches, sensitivity was 88% and specificity 87%. Precision was 84% and F1 score was 86% (Table 1). In the 256 × 256 patches, sensitivity was 70% and specificity 76%. Precision was 70% and F1 score was 70% (Table 2).

Patch size: Hyperparameters were set, and the patches were fed into the model. Accuracy was 84% for LGUC and 91% for HGUC using

Table 1
Confusion matrix shows high true positives and true negatives with few false positives (1171) and false negatives (820). This gives a low type 1 and type 2 error with high accuracy. Confusion matrix is a summary of prediction results on a classification problem that shows true positives and true negatives.

Confusion matrix for 1024 × 1024 pixel patches		Prediction	
		LGUC	HGUC
Class	LGUC	6250	1171
	HGUC	820	8091

Table 2
Confusion matrix shows the number of true positives and true negatives with higher false positives (1171) and false negatives (821). This gives a lower accuracy than 256 × 256 pixel patches. Confusion matrix is a summary of prediction results on a classification problem that shows true positives and true negatives.

Confusion matrix for 256 × 256 pixel patches		Prediction	
		LGUC	HGUC
Class	LGUC	4244	1740
	HGUC	1820	5740

Table 3
Accuracy of classifier using 1024 × 1024 pixel patches with learning rate & batch size. Accuracy is higher at 90% with learning rate 0.001. Accuracy is higher at 90% with Patch size 16. The model saturates after epoch 30.

Hyperparameter		Accuracy
Learning rate (Batch size 16)	0.001	**90**
	0.01	78
Batch size (Learning rate 0.001)	16	**90**
	32	88

1024x1024 pixels patches with learning rate 0.001, epochs 50, and batch size 16 (Table 3) (Fig. 5).

The accuracy falls to 78% and 70% for HGUC and LGUC using 256 × 256 pixel patches (Table 4). This may be due to the architecture of the tissue, which is an important feature in grading seen at low magnification and is unable to be discerned by the model when so small patches are used. Cross-entropy loss was used to train the network since it's a classification problem. Loss value is also more which indicates the model's prediction was worse as compared to 1024 × 1024 pixel patches. If the model's prediction is perfect, the loss is zero; otherwise, the loss is greater.

Batch size: The accuracy for LGUC and HGUC dropped to 84% and 88%, respectively, with batch size 32 (Table 3).

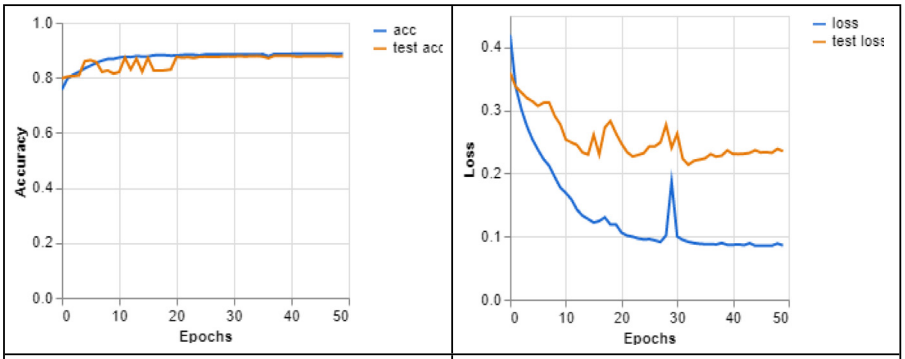
Learning rate: Accuracy depends on learning rate which is the length of strides an algorithm takes while sifting through the image for feature learning. The accuracy falls to 81% for LGUC and 84% for HGUC using learning rate of 0.01 (Table 3) (Fig. 6).

Epochs: The curve reached a steady state after around 30 epochs. This is due to the small number of images in the dataset. No considerable differences were observed on changing the epoch size.

Areas other than urothelium were either:
(i) blurred regions and stain deposits;
(ii) artifacts caused by cauterization or crush artifacts;
(iii) solid nests of benign urothelium in the lamina propria known as Von Brunn's nests.

Grad-CAM is a class activation map that highlights the regions of the image the deep learning algorithm has deemed important in classification. It is an abstract visualization of the features learnt during training and produces a coarse heat map showing important features in red.[10]

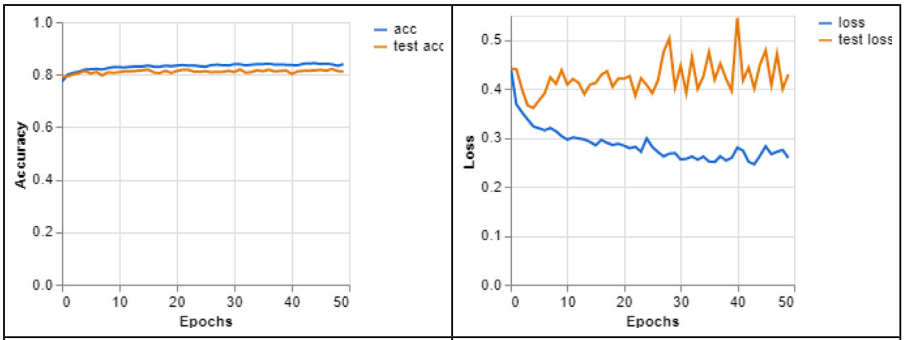
In our study, the model was trained on ImageNet and then Grad-CAM technique was applied on the final layer of convolution in VGG16 (Fig. 7). This pretrained model was used based on transfer learning for visualization in urothelial carcinoma images. The heat maps show the cellular areas of the tumor as consistently being marked as features which are important. The stroma, blood vessels, and blank areas also are colored as blue or green which represents the least important regions. However, there are notable



A. The model gets saturated after 25 epochs. The learning rate of 0.001 is the size of the update steps the model takes to train itself. With such a low learning rate, the model takes more time to learn about minute features and is more precise. Accuracy per class: LGUC 0.84 HGUC 0.91

B. This is evident in the validation loss which falls to 0.2 after 25 epochs. Loss is a number indicating how bad the model's prediction was on a single example. If the model's prediction is perfect, the loss is zero; otherwise, the loss is greater. The batch size is 16 meaning that 16 images are processed at one point of time, and this was the best result after trying out with batch size 8 and 32.

Fig 5. Results on 1024 × 1024 pixel patches; with learning rate 0.001, epochs 50, and batch size 16.



A: The higher learning rate, which is 10 times higher than the previous rate of 0.001, means the model takes longer strides and skips features for learning but is faster. Accuracy per class LGUC 0.81, HGUC 0.84.	*B:This is evident in loss value which is more here. Accuracy per class LGUC 0.81, HGUC 0.84.*

Fig 6. Results on 1024 × 1024 pixel patches; with learning rate 0.01, epochs 50, and batch size 16.

inaccuracies in such a depiction with few images showing regular uniform distribution of low-grade urothelial carcinoma cells being shown in blue. This representation lends an insight into how the model learns, and how it misclassifies artifacts into one of the 2 classes. Our results also show that Grad-CAM is an intuitive visual tool that helps to give an idea of the model's working. In this study, the highlighted area is the area of interest that the model is classifying as tumor regions through Grad-CAM (Fig. 8).

Discussion

We describe a deep learning-based system for grading of urothelial cancers in this research study. In the initial stages, urothelial tumors are mostly asymptomatic. However, in later stages, these may present with difficulty in micturition and hematuria. On cystoscopy, the tumors have a varied appearance, and it is possible to understand the tumor location, size, and appearance of the tumor.[11] On fluorescence cystoscopy, it is possible to delineate flat neoplastic lesions using photosensitizers such as 5-aminolevilunic acid or hexaminolevulinate acid after intravesical administered. The lesion is sampled and sent for histopathological examination. During transurethral resection, surrounding muscle is sampled to check for muscle invasion. The histopathological features suggestive of LGUC include, almost orderly arrangement of cells showing mild loss of polarity with mild nuclear pleomorphism, smooth nuclear margin, inconspicuous nucleoli, and minimal or no mitotic figures. Histopathological features of HGUC are fused papillary architecture, marked irregularities in nuclear contours, moderate nuclear pleomorphism with visible nucleoli, and atypical mitotic figures.

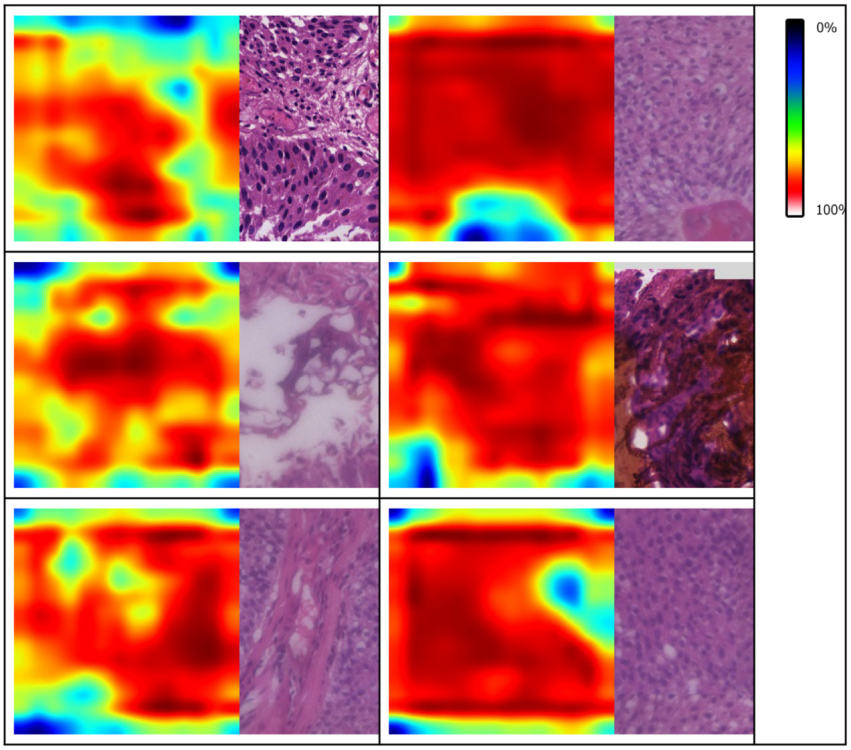


Fig 7. Grad-CAM visualization showing LGUC, artifacts, and HGUC being shown with tumor area in red (color bar on the side representing tumor area in red).

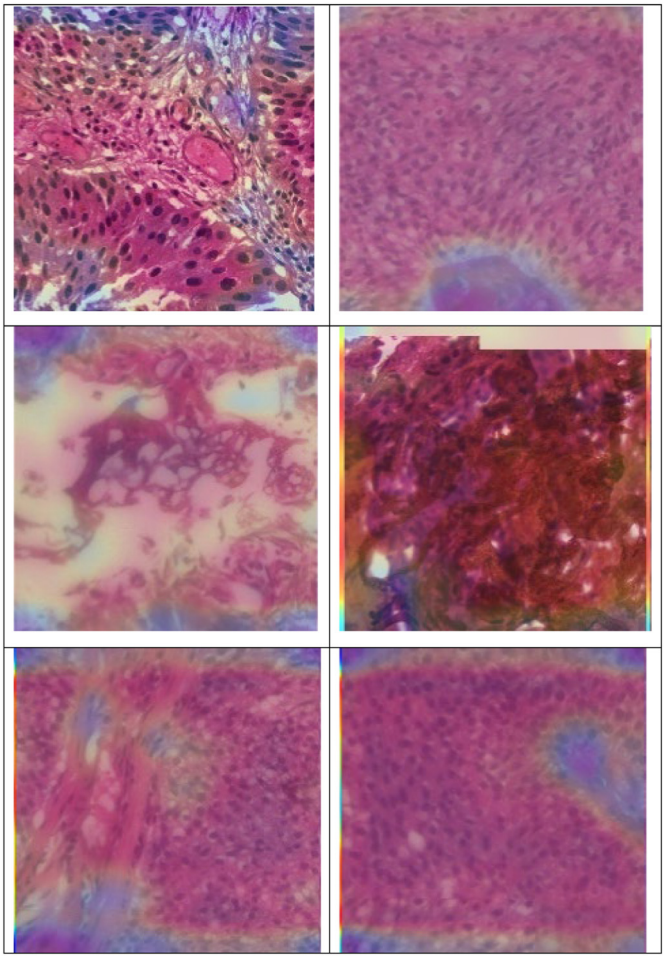


Fig 8. An overlay of hematoxylin and eosin-stained slides of urothelium with Grad-CAM visualization showing red areas as tumor tissue.

Table 4
At 1024 × 1024 pixels, the best accuracy of classification of high-grade urothelial carcinoma at 91% and LGUC at 89% – overall accuracy is 90%. These are the results after repeated testing of the tunable parameters, namely epochs, learning rate, and batch size.

Accuracy	HGUC	LGUC
1024 × 1024 pixel patches	**91**	**89**
256 × 256 pixel patches	78	70

This problem can be tackled in 2 ways:

(1) By traditional image processing, which has many applications in digital pathology, such as Ki67 quantification and nucleus segmentation in tumor cells based on certain parameters like size and hyperchromasia.[4]
(2) By artificial intelligence, where the task is done by either one of these approaches: (i) traditional machine learning algorithms like Support Vector Machines (SVM) and Random Forest, or (ii) deep neural network algorithms like CNNs.

Although traditional ML algorithms might get some decent results for some problems, convolutional neural networks (CNNs) truly shine in processing and classifying images in the most complex problems. This is because of the very nature of the technology which is best suited to the task of identifying the features in images. Input images pass through the layers of the neural network to learn their features layer-by-layer. The learning of these features depends on the number of layers of the network, the learning rate, and the number of epochs. The last layer of the neural network usually acts as the classifier that outputs the class label.[6,9]

Annotated data is scarce, especially for histopathology images. One solution is the use of transfer learning – use of pretrained model on a different dataset for a different objective. Our study used this approach. A recent approach is to use weakly supervised deep learning where a large quantity of raw data is labelled and used for training a deep learning model. This approach has been proven to give a good predictive accuracy.[7,12]

The study demonstrates a unique method of grading bladder cancers in a reliable and repeatable manner. There are very few studies that discuss the application of AI in urothelial tumors.[13] The model network focused on non-invasive low- and high-grade papillary urothelial carcinoma was trained. A pretrained VGG-16 model on ImageNet dataset was used. The automated grading was then matched to 2 expert pathologists' grading. Machine learning approaches were utilized in 2 previous studies to automate the grading of urinary bladder cancer using histopathological images from areas of interest. This system is more objective for grading bladder cancer that differs from previous research in that it requires less human input and so has higher repeatability. This model shows results of grading that are consistent with those of the other previously published research. According to Soukup et al., agreement for the categorization of low- and high-grade tumors ranged from 65% to 88%, with kappa values between 0.30 and 0.73.[14]

This study had kappa values of 0.7 for LGUC corresponding to a substantial agreement among the pathologists. Kappa value was 0.9 for HGUC corresponding to an almost perfect agreement.

The reason for the minor approximately 10% inaccuracy in our study can be hypothesized. A major reason is the presence of few areas of low-grade urothelial carcinoma in high-grade urothelial carcinoma. However, a predominantly high-grade urothelial carcinoma gives slightly better accuracy of 91% as compared to 89% in low-grade urothelial carcinoma. This is reflected in the kappa values as well, demonstrating that architecture of the tissue is appreciable at low power. This could be because LGUC samples can encompass some areas of normal urothelium and papilloma like architecture also, and hence represents a more heterogeneous class than HGUC. The visualization using Grad-CAM lends an insight into how the model delineates the various heterogenous elements in the image.

Other studies have used different network architectures with good results – Inception v3 has been used to detect epithelial tumors in stomach and colon.[15]

Dimensionality reduction approaches have been tried to better understand the features learnt and how they affect the working of the algorithm – one example of a study using t-SNE for neuropathological tissue sample.[16]

Applications of AI are varied, including prognostication of patients with urothelial cancer. If clubbed with data from gene expression, AI has the capability to be more robust and predict recurrence of the tumor as well.[17] Use of AI is also proven useful in breast pathology,[18] bone pathology,[19] prostatic pathology,[20] lung pathology,[21] and predicting cancers of unknown primary.[22] AI is making in-roads into other fields of medicine as well – most notably radiology,[23] dermatology,[24] and neurology.[25]

Grad-Cam has been used in multiple instances in histopathology and other visual-based medicine fields showing a heatmap for adenocarcinoma lung[26] and adenocarcinoma colon classification in histology slides, and in colonoscopy images to detect colorectal cancer[27] and in fundus images to detect diabetic retinopathy.[28]

A similar study by Jansen et al. used smaller patches (224 × 224 pixels) for a 3-class classification. The segmentation network used is a U-Net that detected more urothelium than what was annotated. The authors provide probable reasons for the discrepant nondiagnostic urothelium too, including artifacts. Both the segmentation and the classification network were independent of each other. The magnification of the patches exported were at 20x. In our study, we have used 2 different image patch sizes 256 × 256 and 1024 × 1024 pixel patches. They reported an accuracy of 76% and 71% for LGUC and HGUC.[29]

Limitations & future scope

Inter-institutional comparison is a better test for the practical utility of such algorithms to classify the carcinoma in a real-world scenario. The use of an independent AI screening test without any oversight raises the ethical issue of responsibility in a case of wrong diagnosis.[30] Very high accuracy will be useful for deployment of this as a screening tool in pathology labs – provided the other infrastructure capabilities are met, like computational and file storage requirements.[31]

Conclusions

This study demonstrates a computational model to classify between low- and high-grade urothelial carcinoma. Total 20 samples of TURBT were collected and digitized using a whole slide scanner. The WSI were then split into patches of 256 × 256 pixels giving total patches of 65 000 of each class, and 1024 × 1024 pixels giving 13 000 patches of each class. These patches were used to train a deep learning model trained in Python after tuning the hyperparameters.

Best accuracy for classification after many iterations and hyperparameter tuning was 91% for high-grade urothelial carcinoma and 89% for low-grade urothelial carcinoma. A good average accuracy of 90% demonstrates the effectiveness of this model to be used as a companion diagnostic tool, especially in ambiguous cases. The visualization using Grad-CAM provides an insight on how the model delineates the various heterogenous elements in an image.

This study is relevant clinically as the high-grade urothelial carcinoma requires aggressive management with frequent periodic follow-ups after radical cystectomy. The classification is a diagnostic challenge for many pathologists – and few biomarkers to reliably distinguish between the 2.

This is one of the earliest studies to the best of our knowledge, of using deep learning in grading of urothelial carcinoma. Currently, most deep learning studies in histopathology focus on classification with many reporting good accuracy, however, the actual working of such an algorithm leading to the final decision is obscure.

Segmentation can be tried on the images to delineate the tumor areas and other classes – stroma, muscle, lymphatics, and blood vessels. The inclusion of life expectancy data can also enhance the predictions about 5-year survival rate and prognosis of the patient. One more application is the prediction of germline mutations based on occult morphological clues picked up by deep learning algorithms. Such features learned will lead to better understanding of these and their morphology on hematoxylin and eosin-stained histopathology slides.

The advent of cheaper storage, easy to deploy and use AI tools will help cut costs and make it economical and practical for computational pathology use in day-to-day practice of pathologists[32,33] and even in pathology education.[34]

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Burger M, Catto JWF, Dalbagni G, et al. Epidemiology and risk factors of urothelial bladder cancer. Eur Urol 2013;63(2):234–241. https://doi.org/10.1016/J.EURURO.2012.07.033.
2. Letaiová S, Medveová A, Ovíková A, et al. Bladder cancer, a review of the environmental risk factors. Environ Health Global Access Sci Source 2012;11(SUPPL1). https://doi.org/10.1186/1476-069X-11-S1-S11.
3. Wang G, McKenney JK. Urinary bladder pathology: World Health Organization classification and American Joint Committee on Cancer Staging Update. Archiv Pathol Lab Med 2019;143(5):571–577. https://doi.org/10.5858/ARPA.2017-0539-RA.
4. Stålhammar G, Robertson S, Wedlund L, et al. Digital image analysis of Ki67 in hot spots is superior to both manual Ki67 and mitotic counts in breast cancer. Histopathology 2018;72(6):974–989. https://doi.org/10.1111/HIS.13452.
5. Sjödahl G, Lauss M, Gudjonsson S, et al. A systematic study of gene mutations in urothelial carcinoma; inactivating mutations in tsc2 and pik3r1. PLoS ONE 2011;6(4). https://doi.org/10.1371/JOURNAL.PONE.0018583.
6. Sodhi P, Awasthi N, Sharma V. Introduction to machine learning and its basic application in Python. SSRN Elect J 2019. https://doi.org/10.2139/SSRN.3323796.
7. Zhou ZH. A brief introduction to weakly supervised learning. National Science Review. Oxford University Press; 2018. p. 44–53. https://doi.org/10.1093/nsr/nwx106.
8. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 2019;25(8):1301–1309. https://doi.org/10.1038/s41591-019-0508-1.
9. VGG16 - Convolutional Network for Classification and Detection. (n.d.). Retrieved December 31, 2021, from https://neurohive.io/en/popular-networks/vgg16/
10. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (n.d.). Grad-cam: Visual explanations from deep networks via gradient-based localization. Openaccess.Thecvf.Com. Retrieved January 26, 2022, from http://openaccess.thecvf.com/content_iccv_2017/html/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.html
11. Herr HW. Does cystoscopy correlate with the histology of recurrent papillary tumours of the bladder? BJU Int 2001;88(7):683–685. https://doi.org/10.1046/J.1464-4096.2001.02396.X.
12. Campanella G, Silva VWK, Fuchs TJ. Terabyte-Scale Deep Multiple Instance Learning for Classification and Localization in Pathology. http://arxiv.org/abs/1805.06983 2018.
13. Jansen I, Lucas M, Bosschieter J, et al. Automated detection and grading of non–muscle-invasive urothelial cell carcinoma of the bladder. Am J Pathol 2020;190(7):1483–1490. https://doi.org/10.1016/J.AJPATH.2020.03.013.
14. Soukup V, Čapoun O, Cohen D, et al. Prognostic performance and reproducibility of the 1973 and 2004/2016 World Health Organization grading classification systems in non-muscle-invasive bladder cancer: a European Association of Urology Non-muscle Invasive Bladder Cancer Guidelines Panel systematic review. Eur Urol 2017;72(5):801–813. https://doi.org/10.1016/J.EURURO.2017.04.015.
15. Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep learning models for histopathological classification of gastric and colonic epithelial tumours. Scient Rep 2020;10(1):1-11. https://doi.org/10.1038/s41598-020-58467-9.
16. Faust K, Xie Q, Han D, et al. Visualizing histopathologic deep learning classification and anomaly detection using nonlinear feature space dimensionality reduction. BMC Bioinform 2018;19(1). https://doi.org/10.1186/S12859-018-2184-4.
17. Trebeschi S, Bodalal Z, van Dijk N, et al. Development of a prognostic AI-monitor for metastatic urothelial cancer patients receiving immunotherapy. Front Oncol 2021;11. https://doi.org/10.3389/FONC.2021.637804/FULL.
18. Duggento A, Conti A, Mauriello A, Guerrisi M, Toschi N. Deep computational pathology in breast cancer. Semin Cancer Biol 2021;72:226–237. https://doi.org/10.1016/J.SEMCANCER.2020.08.006.
19. Tao Y, Huang X, Tan Y, et al. Qualitative histopathological classification of primary bone tumors using deep learning: a pilot study. Front Oncol 2021;11. https://doi.org/10.3389/FONC.2021.735739.
20. Bulten, W., Pinckaers, H., Van Boven, H., Vink, R., De Bel, T., Van Ginneken, B., Van Der Laak, J., Hulsbergen-Van De Kaa, C., & Litjens, G. (n.d.). Automated Gleason Grading of Prostate Biopsies using Deep Learning.
21. Wang, S., Chen, A., Yang, L., Cai, L., Xie, Y., reports, J. F.-S., & 2018, undefined. (n.d.). Comprehensive analysis of lung cancer pathology images to discover tumor shape and boundary features that predict survival outcome. Nature.Com. Retrieved January 2, 2022, from https://www.nature.com/articles/s41598-018-27707-4
22. Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. Nature 2021;594(7861):106–110. https://doi.org/10.1038/s41586-021-03512-4.
23. Jha, S., & Topol, E. (n.d.). Adapting to artificial intelligence: radiologists and pathologists as information specialists. JAMA, 316, 2353–2354
24. Al-Janabi S, Huisman A, Vink A, et al. Whole slide images for primary diagnostics in dermatopathology: a feasibility study. J Clin Pathol 2012;65(2):152–158. https://doi.org/10.1136/JCLINPATH-2011-200277.
25. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev 2012;36(4):1140–1152. https://doi.org/10.1016/j.neubiorev.2012.01.004.
26. He J, Shang L, Ji H, Zhang XL. Deep learning features for lung adenocarcinoma classification with tissue pathology images. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10637 LNCS; 2017. p. 742–751. https://doi.org/10.1007/978-3-319-70093-9_79.
27. Hwang Y, Lee HH, Park C, et al. Improved classification and localization approach to small bowel capsule endoscopy using convolutional neural network. Digest Endosc 2021;33(4):598–607. https://doi.org/10.1111/DEN.13787.
28. Daanouni O, Cherradi B, Tmiri A. Automatic detection of diabetic retinopathy using custom CNN and Grad-CAM. Adv Intel Syst Comput 2021;1188:15–26. https://doi.org/10.1007/978-981-15-6048-4_2.
29. Jansen I, Lucas M, Bosschieter J, et al. Automated detection and grading of non-muscle-invasive urothelial cell carcinoma of the bladder. Am J Pathol 2020;190(7):1483–1490. https://doi.org/10.1016/j.ajpath.2020.03.013.
30. Sorell T, Rajpoot N, Verrill C. Ethical issues in computational pathology. J Med Ethics 2021. https://doi.org/10.1136/MEDETHICS-2020-107024.

31. Rakha EA, Toss M, Shiino S, et al. Current and future applications of artificial intelligence in pathology: a clinical perspective. J Clin Pathol 2021;74(7):409–414. https://doi.org/10.1136/JCLINPATH-2020-206908.
32. Al-Janabi S, Huisman A, Nap M, Clarijs R, Van Diest PJ. Whole slide images as a platform for initial diagnostics in histopathology in a medium-sized routine laboratory. J Clin Pathol 2012;65(12):1107–1111. https://doi.org/10.1136/JCLINPATH-2012-200878.
33. Evans AJ, Vajpeyi R, Henry M, Chetty R. Establishment of a remote diagnostic histopathology service using whole slide imaging (digital pathology). J Clin Pathol 2021;74(7):421–424. https://doi.org/10.1136/JCLINPATH-2020-206762.
34. Arora A, Arora A. Pathology training in the age of artificial intelligence. J Clin Pathol 2021;74(2):73–75. https://doi.org/10.1136/JCLINPATH-2020-207110.