Novel Computational Pipeline Enables Reliable Diagnosis of Inverted Urothelial Papilloma and Distinguishes It From Urothelial Carcinoma

Wei Shao, PhD[1,2]; Michael Cheng, BS[1,2]; Antonio Lopez-Beltran, MD[3]; Adeboye O. Osunkoya, MD[4]; Jie Zhang, PhD[5]; Liang Cheng, MD, MS[6]; and Kun Huang, PhD[1,2]

DOI https://doi.org/10.1200/CCI.24.00059

ABSTRACT

PURPOSE With the aid of ever-increasing computing resources, many deep learning algorithms have been proposed to aid in diagnostic workup for clinicians. However, existing studies usually selected informative patches from whole-slide images for the training of the deep learning model, requiring labor-intensive labeling efforts. This work aimed to improve diagnostic accuracy through the statistic features extracted from hematoxylin and eosin-stained slides.

METHODS We designed a computational pipeline for the diagnosis of inverted urothelial papilloma (IUP) of the bladder from its cancer mimics using statistical features automatically extracted from whole-slide images. Whole-slide images from 225 cases of common and uncommon urothelial lesions (64 IUPs; 69 inverted urothelial carcinomas [UCInvs], and 92 low-grade urothelial carcinoma [UCLG]) were analyzed.

RESULTS We identified 68 image features in total that were significantly different between IUP and UCInv and 42 image features significantly different between IUP and UCLG. Our method integrated multiple types of image features and achieved high AUCs (the AUCs) of 0.913 and 0.920 for classifying IUP from UCInv and conventional UC, respectively. Moreover, we constructed an ensemble classifier to test the prediction accuracy of IUP from an external validation cohort, which provided a new workflow to diagnose rare cancer subtypes and test the models with limited validation samples.

CONCLUSION Our data suggest that the proposed computational pipeline can robustly and accurately capture histopathologic differences between IUP and other UC subtypes. The proposed workflow and related findings have the potential to expand the clinician's armamentarium for accurate diagnosis of urothelial malignancies and other rare tumors.

ACCOMPANYING CONTENT

Appendix

Accepted January 13, 2025
Published March 13, 2025

JCO Clin Cancer Inform
9:e2400059

INTRODUCTION

It is estimated that bladder cancer accounts for approximately 84,870 new cases and 17,420 deaths in the United States in 2025.[1] Inverted urothelial papilloma (IUP) is an uncommon benign urothelial neoplasm.[2,3] It is diagnostically challenging to distinguish IUP from other bladder tumors, especially inverted variant of urothelial carcinoma (UC), namely, inverted noninvasive urothelial carcinoma (UCInv) and low-grade noninvasive urothelial carcinoma (UCLG).[4-6] Recently, aided by the advances in digitized microscopic imaging and machine learning technology, many computational systems have been built for various histopathology tasks.[7-11] For instance, transfer learning, a machine learning technique, has been applied to quantify histopathologic patterns across 17,355 hematoxylin and eosin (H&E) histopathology images with matched genomic and survival data.[12] Notably, Campanella et al[13] reported a clinical-grade computational pathology framework that was evaluated on a data set of 44,732 whole-slide pathologic images. Moreover, Coudray et al[14] proposed a computational pipeline for the classification between lung adenocarcinoma and lung squamous cell carcinoma and Cheng et al implemented a machine learning pipeline for diagnosing the rare TFE3 Xp11.2 translocation renal cell carcinoma.[15] These studies demonstrated the efficacy of computational pathology in the precise diagnosis of human cancers.

CONTEXT

Key Objective

Inverted urothelial papilloma (IUP) is an uncommon benign urothelial neoplasm that is hard to distinguish from other bladder tumors. This study implements an automatic computational pipeline from histopathologic images that can capture subtle morphological differences between IUP and other bladder tumors and contribute to a potential guideline for IUP diagnosis.

Knowledge Generated

We constructed an automatic, reliable, comprehensive, interpretable, and reproducible whole-slide pathologic image feature extraction pipeline to distinguish IUP from its mimickers. Our methods can facilitate the accurate diagnosis of IUP, which is a benign entity, and its distinction from malignant mimickers, thereby contributing to precision cancer diagnostics.

Relevance

The proposed computational pipeline can robustly and accurately distinguish benign IUP of the urinary bladder from urothelial carcinomas.

Although much progress has been achieved, most of the existing studies are based on the deep learning technology requiring amounts of annotated images to achieve satisfactory results. In the case of IUP, the sample rarity may severely affect the success of a deep learning approach to this disease. On the basis of the above consideration, we implemented an automated workflow that calculated 160 objective features from the histopathologic images. The image features were extracted from the whole slides, which covered not only a large tumor area but also a wide spectrum of cell nucleus morphologies, including nucleus size, texture, shape, and density from the heterogeneous cancer tissue that can help characterize the significant differences between IUP and other types of UC. Specifically, we designed a fully automated pipeline to extract and combine multiple types of statistic descriptors (ie, morphology, topology, and texture features) and built classifiers to distinguish IUP from UCLG and UCInv. We demonstrated that the proposed pipeline can accurately distinguish IUP from UCInv and UCLG. Moreover, we constructed an ensemble classifier to test the prediction accuracy of IUP from an external validation cohort, which provided a new workflow to diagnose rare cancer subtypes and test the models with limited validation samples.

METHODS

Specimens

Two cohorts of 225 H&E-stained whole-slide images from 225 cases of common and uncommon urothelial lesions (64 cases of IUP, 69 cases of UCInv, and 92 cases of UCLG) were analyzed (Appendix Table A1). The sample histology figures of these three types of urothelial lesions (ie, IUP, UCInv, and UCLG) are shown in Figure 1. Before feature extraction, we transformed the color appearance of the images from the external IUP cohort into that used for the Indiana University cohort using a structure-preserving color normalization algorithm.[14] This research was approved by the Institutional Review Board in accordance with the Institutional Committee for the Protection of Human Subjects.

Preprocessing of Pathologic Images

We divided each whole-slide image (40×) into patches (1,024 × 1,024 pixels each) without overlap to facilitate efficient processing. Because the resulted patches may not cover enough tissue, we selected the valid patches with larger than 70% tissue content for further analysis. The tissue content was then calculated as the percentage of nonwhite pixels (at least one of the red, green, and blue [RGB] color channel values was below 200 in the 24-bit RGB color space) in a particular patch.

Cell-Level Morphology and Topology Feature Extraction

The image-level cell *morphology* and *topology* feature extraction pipeline was composed of the following three steps: *nucleus segmentation*, *cell-level feature extraction*, and *aggregation of cell-level features into image-level features*. Specifically, a multithreshold-based unsupervised nucleus segmentation method was used to segment all cells in each whole-slide image.[15] Next, four cell-level morphological features, including the nucleus area (denoted as *Area*),;the major and minor axis length of cell nucleus (denoted as *Major* and *Minor*); the ratio of major axis length to minor axis length (denoted as *Ratio*); and three cell-level topological features characterizing the minimum, maximum, and mean cell neighboring distance (denoted as *distMin*, *distMax*, and *distMean*), were extracted. Here, the neighboring nucleus distance was calculated using the Delaunay triangulation graph, where the vertex set included all cell nuclei and the edge set contained triangle edges linking neighboring nuclei. Then, for each type of cell-level feature, a five-bin

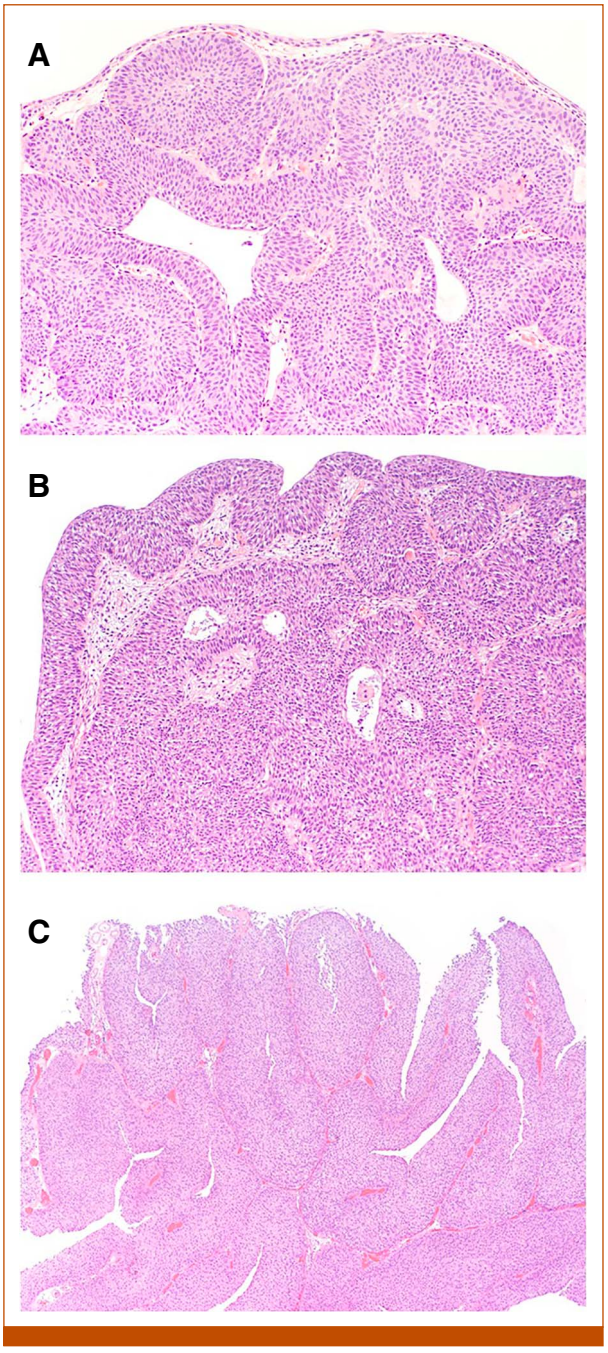


FIG 1. Histopathology of different types of urothelial lesions: (A) IUP, (B) UCInv, and (C) UCLG. IUP, inverted urothelial papilloma; UCInv, inverted urothelial carcinoma; UCLG, low-grade urothelial carcinoma.

histogram and three statistic measurements (ie, mean, standard deviation, and entropy) were calculated to aggregate the nucleus features into image-level nucleus features, and we finally obtained a 32-dimension morphology and a 24-dimension topology feature set for each patient. We used the same naming rule for both cell-level and image-level features. For instance, the feature *ratio_1* represented the percentage of cells with small ratio of major axis length to minor axis length, implying the round-shaped cells, whereas *major_5* referred to the percentage of cells with larger long axes in nuclei, often implying cells with an elongated shape.

Patch-Level Texture Feature Extraction

For each valid patch derived from whole-slide images, we first converted it to a gray-scale image and then extracted the well-established 13-dimensional Haralick features (denoted as *Hara_1*, *Hara_2*, ..., *Hara_13*) from its corresponding gray-level co-occurrence matrix (GLCM).[16] The 13 different Haralick features were derived from four basic GLCM features that can describe the textural information of the divided patches. Following the same aggregation strategy applied on nucleus features, we summarize all patch-level Haralick features into a 104-dimensional image-level feature set.

Multimodal Data Fusion and Classification Via Kernel Combination

We applied linear multikernel-based Support Vector Machine (ie, SVM) to combine multiple types of image features (morphology, topology, and texture features) for discriminating between IUP and UCInv (or UCLG).[21] Suppose that we were given n training samples, let $x_i^1 \in R^{32}$, $x_i^2 \in R^{24}$, and $x_i^3 \in R^{195}$ denote the morphology, topology, and texture features of the i-th sample, respectively, and its corresponding class label was represented by $y_i \in \{-1, 1\}$. The cost function for the multikernel-based SVM could be formulated as

$$\min_{w^m, b, \varepsilon_i} \frac{1}{2} \sum_{m=1}^{3} \beta_m \|w^m\|^2 + C \sum_{i=1}^{n} \varepsilon_i,$$

$$\text{s.t. } y_i \left(\sum_{m=1}^{3} \beta_m \left((w^m)^T \varnothing^{(m)} (x_i^m) + b \right) \right) \geq 1 - \varepsilon_i, \quad (1)$$

$$\varepsilon_i \geq 0, \mathrm{i} = 1, \ldots, \mathrm{n},$$

where w^m, $\varnothing^{(m)}$, and $\beta_m \geq 0$ represent the normal vector to the hyperplane separating the two classes, the kernel-induced mapping function, and the weight value for the m-th type of feature, respectively. Here, we constrained $\beta_1 + \beta_2 + \beta_3 = 1$ and applied a coarse-grid search method to determine the optimal value of β_m.

Feature Weight for Multimodal Learning

Since the linear kernel, that is, $k^m(x_i^m, x^m) = (x_i^m)^T x^m$, was used in the proposed multikernel SVM, Equation 2 could be reformulated as

$$\mathrm{f}(\mathrm{x}) = sign \left(\sum_{i=1}^{n} y_i \alpha_i \sum_{m=1}^{3} \beta_m \left(x_i^m \right)^T x^m + b \right). \quad (2)$$

From Equation 2, the weight for the k-th feature in feature type m, that is, D_k^m could be calculated as

$$D_k^m = \sum_{i=1}^{n} y_i \alpha_i \beta_m \left(x_i^m\right)^k, \quad (3)$$

where $(x_i^m)^k$ represents the k-th feature in x_i^m. Thus, we could rank the importance of each feature by the absolute value of D_k^m.

Ensemble Learning for External Validation

We constructed the ensemble classifier by combining two individual classifiers, that is, IUP versus UCLG and IUP versus UCInv, for classifying the IUP samples on the external IUP cohort. For each of these two classifiers outputting the probability of whether a particular sample was IUP, we averaged these two probability values to generate the final prediction results.

RESULTS

Patient characteristics are listed in Appendix Table A1. On the Indiana University cohort, we conducted two classification tasks, IUP versus UCInv and IUP versus UCLG, to evaluate the classification performance of our method. The external IUP cohort was used to test the generalization capability of our classification model. Because of the lack of control samples (ie, UCInv and UCLG) on the external cohort, we calculated the proportion of correctly identified external IUP samples to test the generalization ability of our model.

Workflow for Machine Learning–Aided Pathologic Image Analysis

Given the input whole slide imaging from the Indiana University cohort, a pathologic image analysis pipeline (Fig 2) was used to extract the nucleus morphology, topology, and texture features. Specifically, after segmenting the nucleus in the whole-slide pathologic image, four cell-level morphological features were extracted with this pipeline (Appendix Fig A1), including the nucleus area (denoted as area), the major and minor axis length of cell nucleus (denoted as major and minor), the ratio of major axis length to minor axis (ratio), and three cell-level topological features (Appendix Fig A2) characterizing the minimum, maximum, and mean cell neighboring distance (denoted as distMin, distMax, and distMean). As to texture features, a 13-dimension Haralick feature[17] (Appendix Fig A3) was extracted from each valid patch (details in the Methods section) in the histopathologic images. Next, a five-bin histogram and three statistic measurements, that is, mean, standard deviation, and entropy, were used to aggregate the nucleus and texture features into 160-dimension image-level features (details in the Methods section).

After the feature extraction step, we applied the multikernel combination method to effectively integrate the image-level morphology, topology, and texture features. The kernel combination method could be naturally embedded into the conventional SVM without extra steps (details in the Method section) for discriminating between IUP and UCInv or UCLG and validating on an external IUP data set.

Extracted Image Feature Showed Significant Differences Between IUP and Other UC Subtypes

We applied the Mann–Whitney U test to identify individual image features that were capable of distinguishing between IUP and UCLG (or UCInv) after multiple testing corrections (false discovery rate < 0.01; Fig 3). A significant feature was reported as under-represented if its median value in IUP was lower than that in UCInv (or UCLG). Otherwise, it was designated as an over-represented feature.

We identified 68 image features in total that were significantly different between IUP and UCInv (Fig 3A) and 42 image features significantly different between IUP and UCLG (Fig 3B). For the morphology features, we found that ratio_bin2 and ratio_bin3 were over-represented (Figs 3A and 3B), whereas ratio_bin1 was under-represented in IUP cohorts compared with the UCInv cohort (Fig 3A). Here, the features from ratio_bin1 to ratio_bin5 directly described the percentages of the nuclei whose shape changed from round to elongated, indicating that IUP samples contain more cells with a larger major to minor axis ratio. In addition, we observed that area_bin1 and area_bin2 were over-represented in IUP when compared with UCInv, whereas area_bin4 and area_bin5 were under-represented (Fig 3A). These results suggested that IUP contained more small-size cells when compared with UCInv.

Moreover, it is of great interest to find that all extracted topology features, except for distMin_bin2, showed significant differences between IUP and UCInv (or UCLG). Among these features, distMin_bin1, distMax_bin1, and distMean_bin1, reflecting a small cell neighboring distance, were over-represented in IUP, whereas the remaining topology features were under-represented in IUP (Figs 3A and 3B). These results indicated that the cells in IUP were more likely clustered together than those in UCInv and UCLG.

As for the Haralick texture features, we found that Hara_2_bin4 was under-represented in IUP (Figs 3A and 3B). Actually, Hara_2 characterized the local neighborhood variation of the images, indicating that the local appearance of IUP was more uniform than UCInv and UCLG.

Moreover, we found that Hara_5_bin5 was over-represented in IUP compared with UCLG (Fig 3B). Contrary to Hara_2, Hara_5 was used to describe the local homogeneity of the image and image features from Hara_5_bin1 to Hara_5_bin5 described the local image patterns ranging from heterogeneous to homogenous. Therefore, both the significant features of Hara_2 and Hara_5 indicated that the variation of local appearance in IUP was small in comparison with that in UCInv and UCLG.

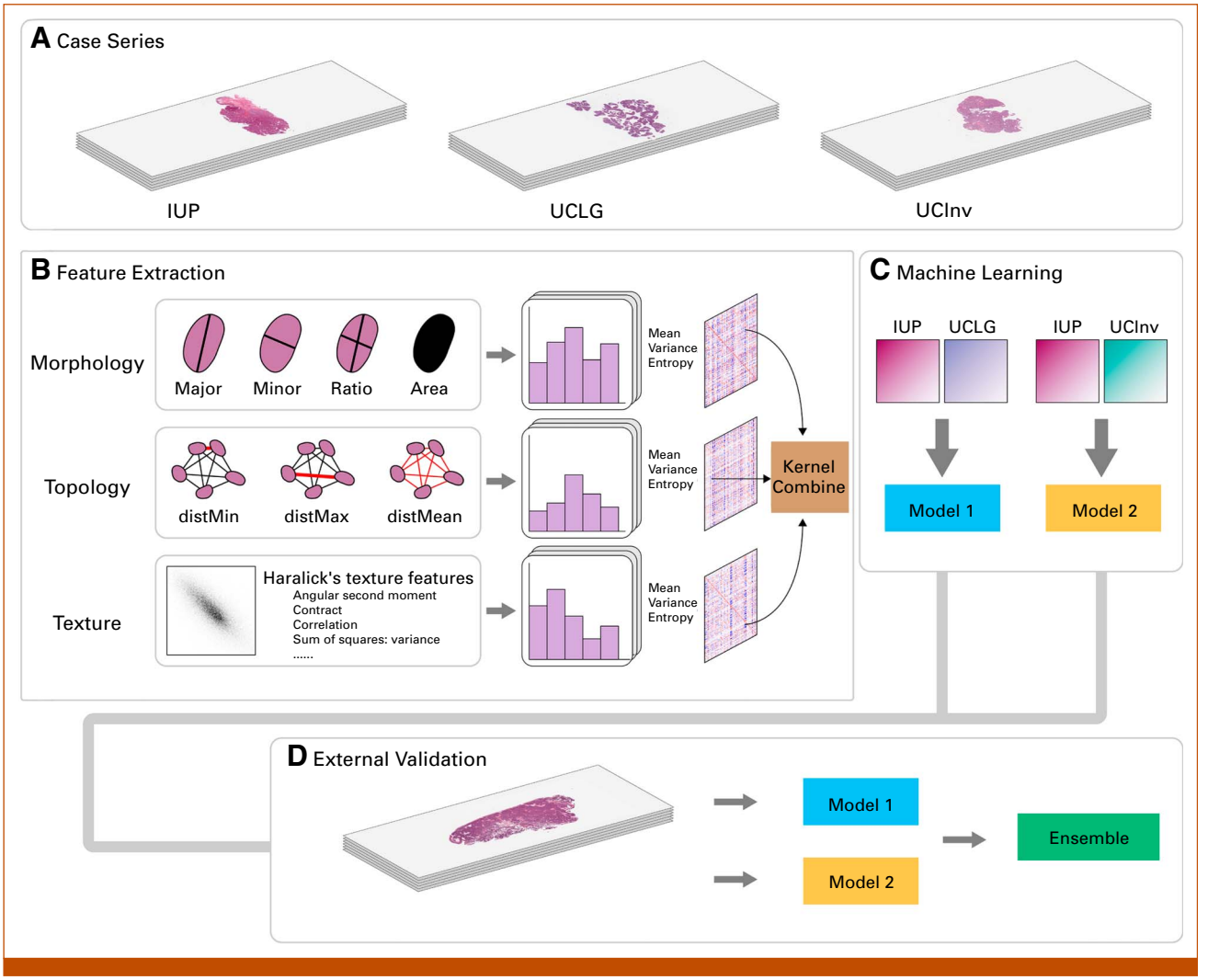


FIG 2. The machine learning workflow to extract image features and construct models. (A) The whole-slide pathologic images from the Indiana University cohort collected with 47 IUP, 69 UCInv, and 92 UCLG cases. (B) For each segmented nucleus, four morphology features and three topology features were extracted. For each valid patch, 13 Haralick texture features were extracted. Then, each cell-level or patch-level feature was aggregated to eight image-level feature using five-bin histogram and three distribution statistics (mean, variance, and entropy). Finally, the multikernel combination strategy was used to integrate the three types of image-level features. (C) On the basis of the combination of the extracted image-level feature, machine learning models were applied to discriminate between IUP and UCInv (or UCLG). (D) An ensemble strategy was applied to combine the IUP versus UCInv and IUP versus UCLG classification models to predict the IUP samples on the external validation cohort. IUP, inverted urothelial papilloma; UCInv, inverted urothelial carcinoma; UCLG, low-grade urothelial carcinoma.

Machine Learning Models on the Basis of Image Features Enabled More Reliable Diagnosis of IUP

We first trained and evaluated the proposed method on the Indiana University cohort. During the experiments, we randomly partitioned the samples into five folds, and one partition was used as testing data, whereas the remaining four were used for model training. For the training data set, we split 25% of samples to tune model parameters. We repeated the five-fold split 10 times, and the mean, the standard deviation of the area under the receiver operating characteristic curve (AUC), and the accuracy were reported.

We first tested our method on IUP versus UCInv (Fig 4A) and IUP versus UCLG (Fig 4B) classification tasks by the combination of different types of features, compared with the methods using only individual feature types. The results showed that using the multikernel combination method to integrate morphology, topology, and texture features consistently achieved more accurate discrimination between IUP and UCInv or UCLG. Specifically, for distinguishing IUP from UCInv (Figs 4A and 5), our multikernel combination method achieved an AUC of 0.913, an accuracy of 0.863, and an F1 score of 0.874, whereas the best AUC, accuracy, and F1 score using only a single type of feature were 0.807, 0.747, and 0.774, respectively.

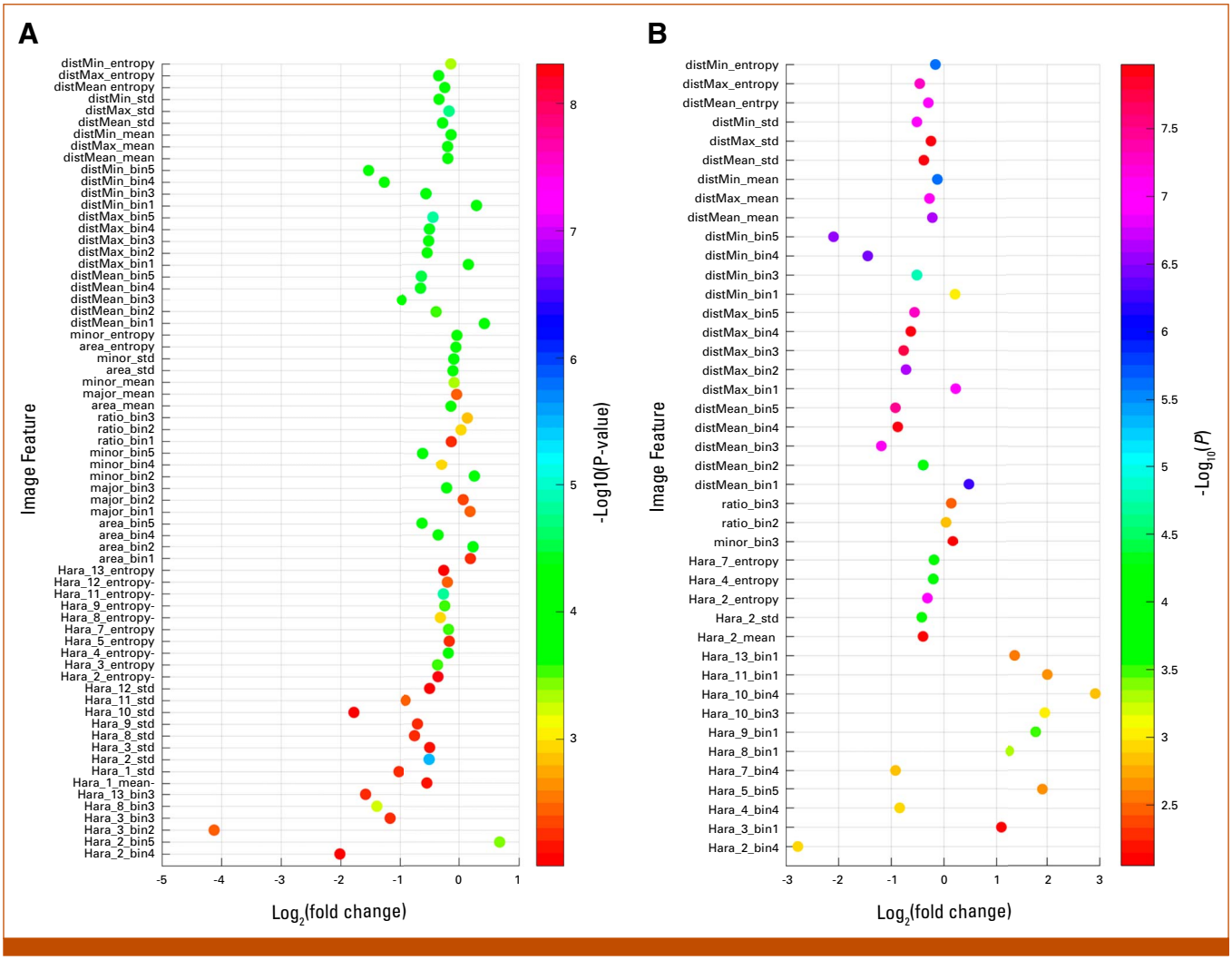


FIG 3. Compared individual image features between IUP and another noninvasive urothelial carcinoma by the Mann–Whitney *U* test. Univariate image feature analysis (A) between IUP and UCInv and (B) between IUP and UCLG. The foldchange was calculated by the ratio of the median feature values between IUP and UCInv (or UCLG). Multiple comparison correction was performed, and an adjusted *P* value <.01 was considered statistically significant. IUP, inverted urothelial papilloma; UCInv, inverted urothelial carcinoma; UCLG, low-grade urothelial carcinoma.

Similarly, for classifying IUP from UCLG (Figs 4B and 5), our multikernel combination method achieved an AUC of 0.920, an accuracy of 0.876, and an F1 score of 0.901, whereas the best AUC, accuracy, and F1 score on individual feature types were only 0.816, 0.748, and 0.759, respectively. Moreover, the confusion matrices shown in Figures 5A and Fig 5B also indicated that our method can effectively identify patients with IUP in different classification tasks (ie, IUP *v* UCInv and IUP *v* UCLG).

In addition, we also compared the results of the proposed method with the direct combination method that concatenated all types of features into a long vector, followed by the linear SVM for classification, which is referred as concatenation in Figure 3. It is clear from the results that our kernel combination method significantly outperformed the feature concatenation method on both IUP versus UCInv (*t*-test, AUC: *P* value = .05, accuracy: *P* value = .004) and IUP versus UCLG (*t*-test, AUC: *P* value = .03, accuracy: *P* value = .0051) classification tasks (Figs 4A and 4B).

Moreover, similar to our workflow only requiring whole-slide level label for model training, we compared our classification methods with two deep learning algorithms, that is, MIL[13] and DeepPATH[14] (Fig 4C). We observed that our method could consistently achieve significant higher AUC than these competing methods on both IUP versus UCInv and IUP versus UCLG tasks. One possible reason was that the deep learning–based algorithm typically required training on large data sets, whereas IUP was a urothelial neoplasm subtype with scarce samples, and thus, it might be difficult to collect enough IUP to train a deep learning model.

Next, we summarized the top five important features of each feature type according to their weights in our multikernel learning model (details in the Methods section) for IUP

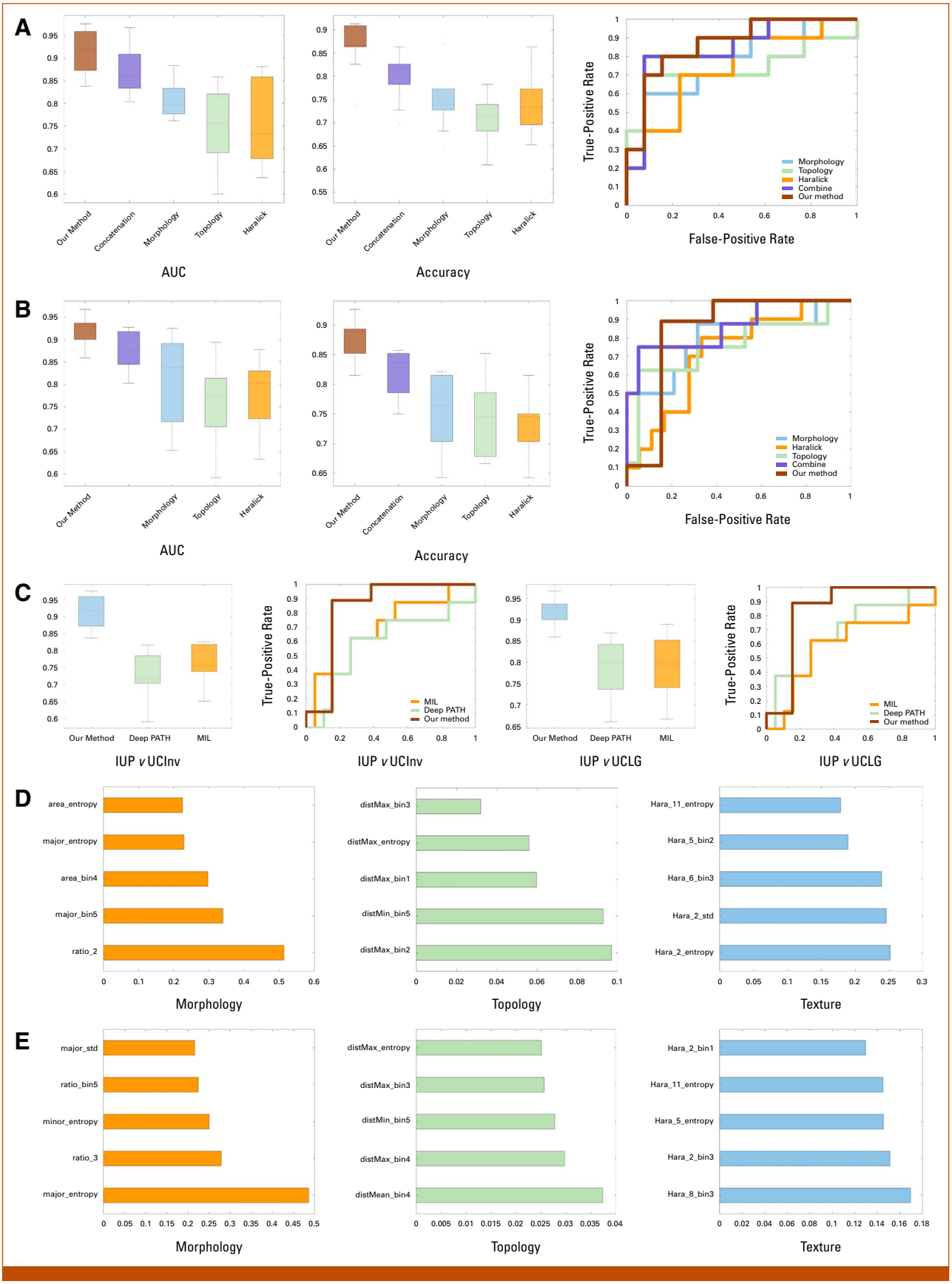


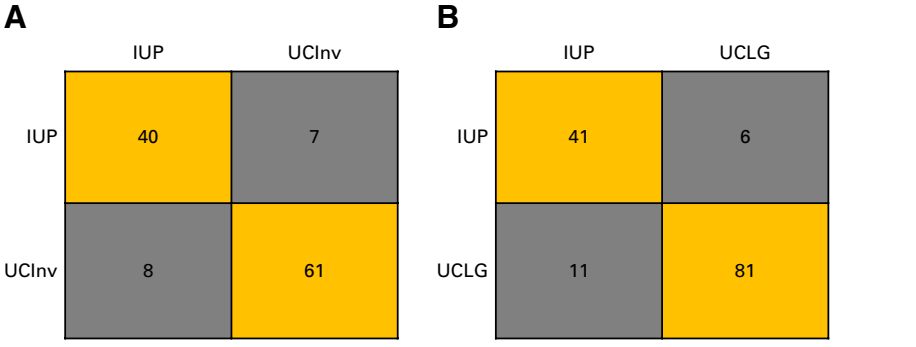
FIG 4. The classification results of different machine learning models. Comparison of different methods on (A) IUP versus UCInv classification task and (B) IUP versus UCLG classification task. (C) The AUC comparison between our proposed method and two weakly supervised deep learning algorithms (ie, DeepPATH and MIL). (D) The top-five important features of (continued on following page)

FIG 4. (Continued). each feature type produced from our kernel combination method on IUP versus UCInv classification task. (E) The top-five important features of each feature type produced from our kernel combination method on IUP versus UCLG classification task. IUP, inverted urothelial papilloma; UCInv, inverted urothelial carcinoma; UCLG, low-grade urothelial carcinoma.

versus UCInv (Fig 4D) and IUP versus UCLG tasks (Fig 4E). Specifically, for the morphologic features, ratio_bin2 and ratio_bin3 were identified by both univariate analysis (Fig 3) and our machine learning model (Figs 4D and 4E). Further investigation of the histogram for ratio feature (Fig 6A) indicated that the ratio_bin2 and ratio_bin3 values in IUP were larger than those in UCInv and UCLG, suggesting that IUP samples contained more elongated cells with a larger major to minor axis ratio. As for the topological feature, distMax_entropy was identified on both IUP versus UCInv and IUP versus UCLG tasks with higher weights; we also plotted the histograms for the feature distMax to show its distribution difference among different cancer cohorts (Fig 6B). Both the feature histograms and representative image patches showed that the cells in IUP samples were clustered more densely (larger distMax_bin1 value) than IUP and UCLG. Finally, the texture Hara_2 feature characterizing the local variation of the input image was also identified by univariate analysis (Fig 3) and our multikernel combination model (Figs 4D and 4E). The representative samples and their corresponding histograms (Fig 6C) demonstrated that the local appearance of IUP samples was more homogenous (ie, larger Hara_2_bin2 value) than those in UCInv and UCLG.

Validating the Ensemble Model on the External IUP Cohort

After applying our method on the Indiana University cohort, two machine learning models were obtained, that is, IUP versus UCInv and IUP versus UCLG. To improve the model generalizability, we constructed an ensemble classifier (described in the Methods section) by combining the above two models for identifying IUP samples in the external IUP cohort. Because of the difficulty in obtaining UCLG and UCInv samples, our external data set only consisted of 17 IUP samples. We therefore calculated the proportion of correctly identified IUP as the metric for accuracy of the learned model. The comparison between accuracies for identifying IUP in the Indiana University cohort and external validation cohort (Appendix Fig A4A) indicated that while the IUP identification accuracy on the external cohort was slightly lower than the performance on the Indiana University cohort, it was still as high as 88.2%. Therefore, it clearly demonstrated the robustness and generality of the learned model, implying the generalizability of our approach. Moreover, as shown in Appendix Fig A4B, except for the texture Haralick and topology features, the ensemble



C

Classification Task	Our Method	Direct Combination	Morphology	Topology	Haralick
IUP *v* UCInv	0.874 (0.026)	0.837 (0.041)	0.774 (0.061)	0.739 (0.047)	0.717 (0.062)
IUP *v* UCLG	0.901 (0.037)	0.851 (0.039)	0.759 (0.048)	0.743 (0.051)	0.689 (0.046)

FIG 5. Confusion matrix for (A) the IUP versus UCInv classification task and (B) the IUP versus UCLG classification task. (C) Comparison of the F1 Score for different classification tasks. IUP, inverted urothelial papilloma; UCInv, inverted urothelial carcinoma; UCLG, low-grade urothelial carcinoma.

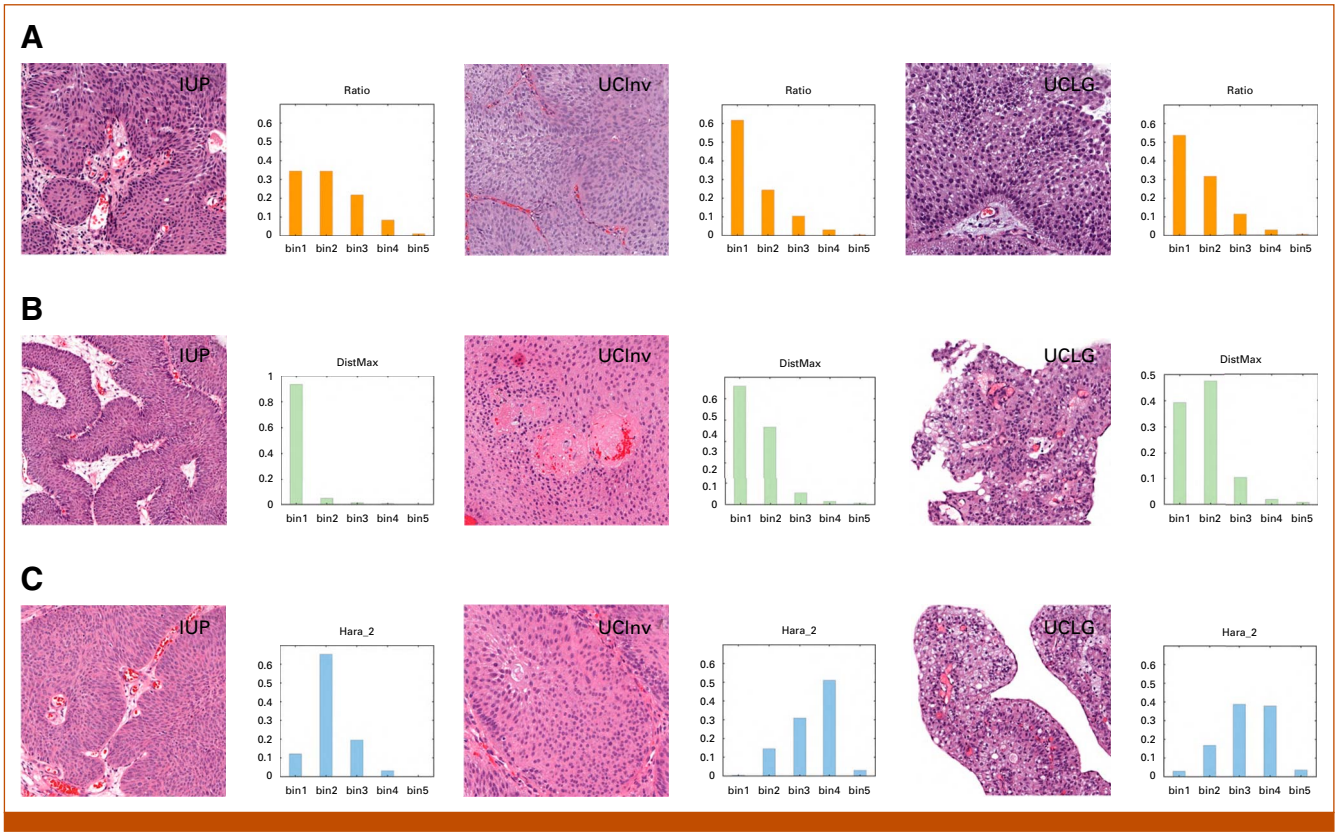


FIG 6. (A) The representative image samples and their corresponding histograms on Ratio feature among different cohorts. (B) The representative image samples and their corresponding histograms on the distMax feature among different cohorts. (C) The representative image samples and their corresponding histograms on the Hara_2 feature among different cohorts.

classifier could better identify the external cohort IUP samples with high accuracy.

DISCUSSION

IUP is an unusual neoplasm of the urinary tract that is considered as a benign tumor. Microscopically, it is challenging for general surgical pathologists to distinguish IUP from other noninvasive UC, that is, UCInv and UCLG, since they exhibit similar histopathologic appearance.[2-6] In this investigation, we designed a computational pipeline for the identification of IUP using image features extracted from whole slide imaging. To the best of our knowledge, this is the first study to provide an accurate, comprehensive. and interpretable machine learning pipeline to classify IUP from other noninvasive UC on the basis of the machine learning model.

This study had several strengths and advantages. First, analyzing histopathologic images is one of the most difficult machine learning tasks, hindered by the large size of the microscopy images. Studies usually sliced whole slide imagings into amounts of small patches, followed by the selection of most informative patches for machine learning model construction. Intuitively, it required labor-sensitive patch-level annotation and different choices of patch size can increase the uncertainties of the model performances. In the current investigation, our method was assessed on the whole slide imaging level, which was a more robust and reproducible method that can reduce the expert annotation efforts.

Second, with the help of ever-increasing computing resources, many computational histopathologic systems have been proposed to extract different types of histopathologic image features to help diagnose human cancers.[18-27] However, most existing research focused on using single type of image feature for cancer diagnosis although recent studies have shown that different biomarkers may provide complementary information for the diagnosis of human cancer. In this study, we extracted multiple types of image features, that is, morphology, topology, and texture features from each whole slide imaging, and applied the multikernel combination strategy to fuse multimodal data for the diagnosis of IUP. The experimental results, as shown in Figure 3, clearly indicated that our method can achieve the mean AUC of 0.913 and 0.920 for IUP versus UCInv and IUP versus UCLG distinction tasks, respectively, which were significantly superior to the results when using even the best individual type of image feature. These results demonstrated the effectiveness of integrating different types of features in distinguishing IUP and other noninvasive UC subtypes. In addition, it is noteworthy that our proposed method could achieve higher AUC than two deep learning–based whole slide imaging classification algorithms

for both IUP versus UCInv and IUP versus UCLG classification tasks. The growth and success of deep learning approaches can be attributed to the availability of a large number of training samples. Nonetheless, the diagnosis of rare UC subtypes using whole slide imaging faced the challenges of high dimensionality and limited sample size, which may lead to overfitting and inaccurate classification for deep learning–based algorithms. Our data suggested that the proposed feature extraction and classification pipelines were more suitable for handling whole slide imaging classification problems when the available training data were not sufficient to train the deep learning models.

Third, our study provided an interpretable framework to generate hypotheses for clinically relevant biomarkers. For instance, the over-representation of three image features (ie,,distMax_bin1, distMin_bin1, and distMean_bin1) indicated that more cells in IUP samples are clumped together than those in UCInv and UCLG. In addition, we also found that the IUP samples have more spindle-shaped cells with higher major axis to minor axis ratio.

Our study had a few limitations. First, because of the limitation of the data source, the classification model has only been validated on a small external data set, but we were still able to achieve good accuracy to demonstrate the strong generalization ability of our proposed workflow. Second, we applied a simple but effective kernel combination strategy to fuse different types of histopathologic image features for IUP diagnosis, and we would expect to see more advanced machine learning studies that can better combine multiple types of image features for performance improvement. Another important limitation was that we only investigated the classification between noninvasive UC (either UCInv or UCLG) and IUP and did not test the ability of the proposed method to discriminate IUP from other UC subtypes. Future studies should include other UC subtypes to test if the proposed method can still successfully identify IUP.

In summary, we constructed an automatic, reliable, comprehensive, interpretable, and reproducible whole slide imaging feature extraction pipeline and used these extracted features to develop a machine learning model to distinguish IUP from its mimickers. The experimental results demonstrated that the histopathology image classifiers on the basis of quantitative features can successfully identify IUP with a high accuracy on the external validation set. Our methods can facilitate the accurate diagnosis of IUP, which is a benign entity and its distinction from malignant mimickers, thereby contributing to precision cancer diagnostics.

AFFILIATIONS

[1]Department of Medicine, Indiana University School of Medicine, Indianapolis, IN
[2]Regenstrief Institute, Indianapolis, IN
[3]Department of Morphological Sciences, University of Cordoba Medical School, Cordoba, Spain
[4]Departments of Pathology and Urology, Emory University School of Medicine, Atlanta, GA
[5]Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN
[6]Department of Pathology and Laboratory Medicine, Department of Surgery (Urology), Brown University Warren Alpert Medical School, the Legorreta Cancer Center at Brown University, and Brown University Health, Providence, RI

CORRESPONDING AUTHOR

Liang Cheng, MD, MS; e-mail: liang_cheng@yahoo.com.

EQUAL CONTRIBUTION

W.S. and M.C. contributed equally to this work as cofirst authors.

SUPPORT

Supported in part by the Indiana University Precision Health Initiative (to K.H.), the NIH U54AG065181 grant (to K.H., J.Z., W.S.), and the National Science Foundation of China (No. 61902183) to W.S.

AUTHOR CONTRIBUTIONS

Conception and design: Wei Shao, Jie Zhang, Liang Cheng, Kun Huang
Financial support: Liang Cheng
Administrative support: Liang Cheng
Provision of study materials or patients: Wei Shao, Antonio Lopez-Beltran, Adeboye O. Osunkoya, Liang Cheng
Collection and assembly of data: Liang Cheng
Data analysis and interpretation: Michael Cheng, Antonio Lopez-Beltran, Adeboye O. Osunkoya, Liang Cheng
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/cci/author-center.
Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Jie Zhang
Stock and Other Ownership Interests: Monument Biosciences

Kun Huang
Research Funding: Merck (Inst), Lilly (Inst)
Travel, Accommodations, Expenses: Merck

No other potential conflicts of interest were reported.

REFERENCES

1. Siegel RL, Kratzer TB, Giaquinto AN, et al: Cancer statistics, 2025. CA Cancer J Clin 75:10-45, 2025
2. Cheng L, Lopez-Beltran A, Montironi R: Bladder Pathology (ed 2). Hoboken, NJ, Wiley-Blackwell, 2025
3. Hodges KB, Lopez-Beltran A, Maclennan GT, et al: Urothelial lesions with inverted growth patterns: Histogenesis, molecular genetic findings, differential diagnosis and clinical management. BJU Int 107:532-537, 2011
4. Montironi R, Cheng L, Lopez-Beltran A, et al: Inverted (endophytic) noninvasive lesions and neoplasms of the urothelium: The Cinderella group has yet to be fully exploited. Eur Urol 59:225-230, 2011
5. Lopez-Beltran A, Henriques V, Montironi R, et al: Variants and new entities of bladder cancer. Histopathology 74:77-96, 2019
6. Cheng L, Davidson DD, Wang M, et al: Telomerase reverse transcriptase (TERT) promoter mutation analysis of benign, malignant and reactive urothelial lesions reveals a subpopulation of inverted papilloma with immortalizing genetic change. Histopathology 69:107-113, 2016
7. Kleppe A, Skrede OJ, De Raedt S, et al: Designing deep learning studies in cancer diagnostics. Nat Rev Cancer 21:199-211, 2021
8. Song AH, Jaume G, Williamson DFK, et al: Artificial intelligence for digital and computational pathology. Nat Rev Bioeng 1:930-949, 2023
9. George RS, Htoo A, Cheng M, et al: Artificial intelligence in prostate cancer: Definitions, current research, and future directions. Urol Oncol 40:262-270, 2022
10. Wu Y, Cheng M, Huang S, et al: Recent advances of deep learning for computational histopathology: Principles and applications. Cancers (Basel) 14:1199, 2022
11. van der Laak J, Litjens G, Ciompi F: Deep learning in histopathology: The path to the clinic. Nat Med 27:775-784, 2021
12. Fu Y, Jung AW, Torne RV, et al: Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer 1:800-810, 2020
13. Campanella G, Hanna MG, Geneslaw L, et al: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 25:1301-1309, 2019
14. Coudray N, Ocampo PS, Sakellaropoulos T, et al: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 24:1559-1567, 2018
15. Cheng J, Han Z, Mehra R, et al: Computational analysis of pathological images enables a better diagnosis of TFE3 Xp11.2 translocation renal cell carcinoma. Nat Commun 11:1778, 2020
16. Shao W, Han Z, Cheng J, et al: Integrative analysis of pathological images and multi-dimensional genomic data for early-stage cancer prognosis. IEEE Trans Med Imaging 39:99-110, 2020
17. Haralick RM: Statistical and structural approaches to texture. Proc IEEE 67:786-804, 1979
18. Moen E, Bannon D, Kudo T, et al: Deep learning for cellular image analysis. Nat Methods 16:1233-1246, 2019
19. Adamson AS, Welch HG: Machine learning and the cancer-diagnosis problem—No gold standard. N Engl J Med 381:2285-2287, 2019
20. Park SH, Han K, Jang HY, et al: Methods for clinical evaluation of artificial intelligence algorithms for medical diagnosis. Radiology 306:20-31, 2023
21. Corti C, Cobanaj M, Dee EC, et al: Artificial intelligence in cancer research and precision medicine: Applications, limitations and priorities to drive transformation in the delivery of equitable and unbiased care. Cancer Treat Rev 112:102498, 2023
22. Lopez-Beltran A, Cimadamore A, Montironi R, et al: Molecular pathology of urothelial carcinoma. Hum Pathol 113:67-83, 2021
23. Lopez-Beltran A, Cookson MS, Guercio BJ, et al: Advances in diagnosis and treatment of bladder cancer. BMJ 384:e076743, 2024
24. Swanson K, Wu E, Zhang A, et al: From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell 186:1772-1791, 2023
25. Haug CJ, Drazen JM: Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med 388:1201-1208, 2023
26. Zhu L, Pan J, Mou W, et al: Harnessing artificial intelligence for prostate cancer management. Cell Rep Med 5:101506, 2024
27. Riaz IB, Harmon S, Chen Z, et al: Applications of artificial intelligence in prostate cancer care: A path to enhanced efficiency and outcomes. Am Soc Clin Oncol Educ Book 44:e438516, 2024

APPENDIX

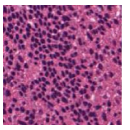
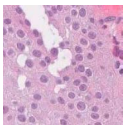
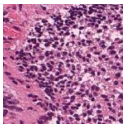
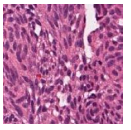
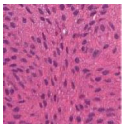
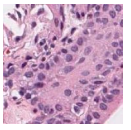
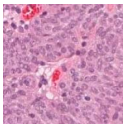
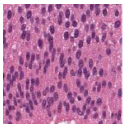
Feature Name	Description	Patch With Small Value	Patch With Large Value
Area	Size of nucleus		
Major	Major axis length		
Minor	Minor axis length		
Ratio	Major to minor ratio		

FIG A1. Illustrations of the four nucleus-level morphology features.

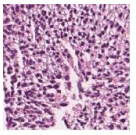
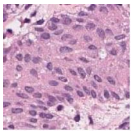
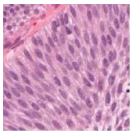
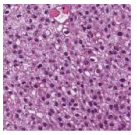
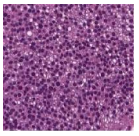
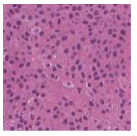
Feature Name	Description	Patch With Small Value	Patch With Large Value
distMean	Mean distance to neighbors		
distMin	Minimum distance to neighbors		
distMax	Maximum distance to neighbors		

FIG A2. Illustrations of the three nucleus-level topology features.

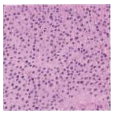
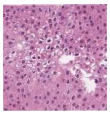
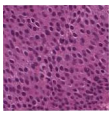
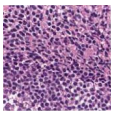
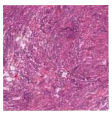
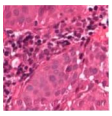
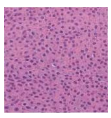
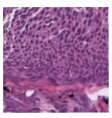
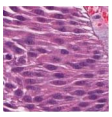
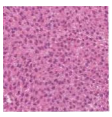
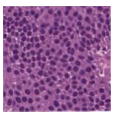
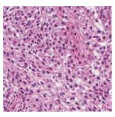
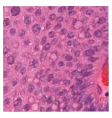
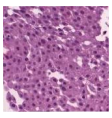
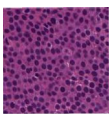
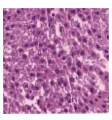
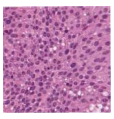
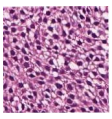
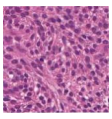
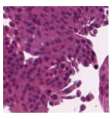
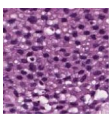
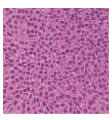
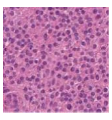
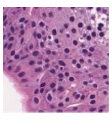
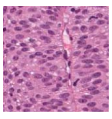
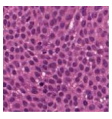
Feature Name	Description	Patch With Small Value	Patch With Large Value
Hara_1	Image global homogeneity		
Hara_2	Local neighborhood variation.		
Hara_3	Dependency of neighboring pixels		
Hara_4	Variance of gray levels		
Hara_5	Local neighborhood homogeneity		
Hara_6	Average sum of gray levels		
Hara_7	Variance of sum of gray levels		
Hara_8	Entropy of sum of gray levels		
Hara_9	Entropy of gray levels		
Hara_10	Variance of difference in gray levels		
Hara_11	Entropy of difference in gray levels		
Hara_12	Mutual information of gray levels		
Hara_13	Difference between joint entropy of gray levels		

FIG A3. Illustrations of the 13 patch-level Haralick texture features.

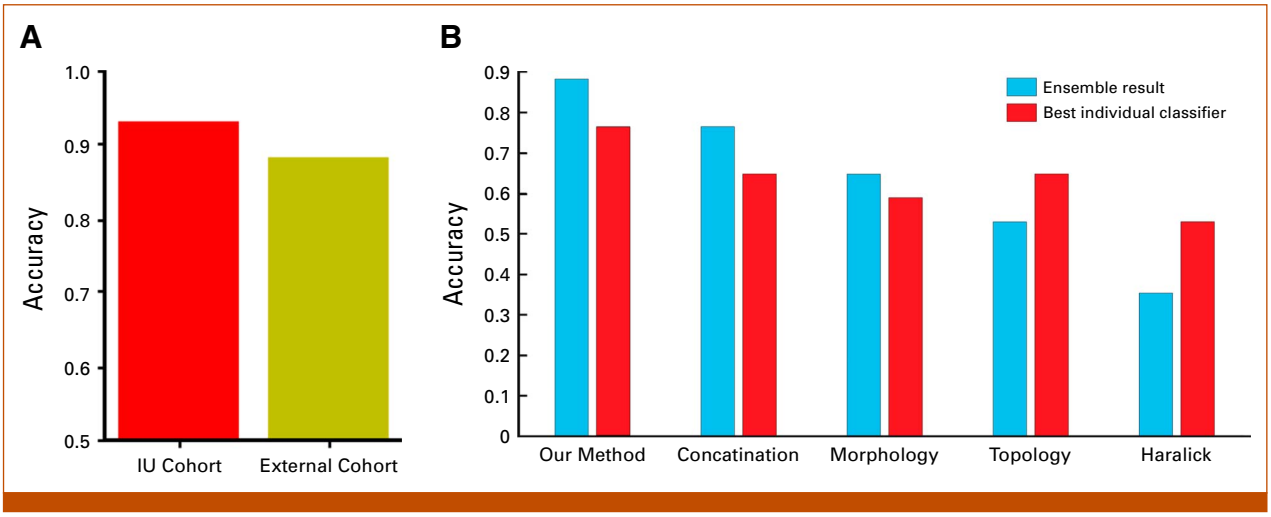


FIG A4. (A) The comparison of accuracies for identifying IUP samples on the IU and the external cohorts by applying the ensemble learning method. (B) Comparisons between individual and ensemble classification accuracy on H&E whole-slide pathologic images from the external validation cohort. H&E, hematoxylin and eosin; IU, Indiana University; IUP, inverted urothelial papilloma.

TABLE A1. Demographic Information of Two Whole-Slide Image Cohorts

Cohort	Tumor Type	Specimen	No.	Male/Female	Mean Age, years (range)
Indiana University cohort	IUP	TURB	47	33/14	63 (21-99)
	UCInv	TURB	69	50/19	64 (23-94)
	UCLG	TURB	92	71/21	69 (44-97)
External cohort	IUP	TURB	17	13/4	60 (34-87)

Abbreviations: IUP, inverted urothelial papilloma; TURB, transurethral resection of bladder; UCInv, inverted Ta urothelial carcinoma; UCLG, low-grade Ta urothelial carcinoma.