2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)
April 3-7, 2020, Iowa City, Iowa, USA

MICROSATELLITE INSTABILITY PREDICTION OF UTERINE CORPUS ENDOMETRIAL CARCINOMA BASED ON H&E HISTOLOGY WHOLE-SLIDE IMAGING

Tongxin Wang[a,b,⋆] *Weijia Lu*[a,⋆] *Fan Yang*[a,⋆] *Li Liu*[c] *Zhongyi Dong*[d]
Weimin Tang[e] *Jia Chang*[e] *Wenjing Huan*[e] *Kun Huang*[f,g] *Jianhua Yao*[a]

[a] Tencent AI Lab, Shenzhen, China
[b] Indiana University Bloomington, Bloomington, USA
[c] Hepatology Unit and Department of Infectious Diseases, Nanfang Hospital, Southern Medical University, Guangzhou, China
[d] Department of Radiation Oncology, Nanfang Hospital, Southern Medical University, Guangzhou, China
[e] Tencent Healthcare, Shenzhen, China
[f] Indiana University School of Medicine, Indianapolis, USA
[g] Regenstrief Institute, Indianapolis, IN, USA
[⋆] These authors contributed equally to this work

ABSTRACT

Microsatellite instability is an important clinical marker for various types of cancers and is related to patients' prognosis and response to immunotherapy. Currently, identifying microsatellite status relies on genetic tests, which are not widely accessible for every patient. We propose a novel pipeline to predict MSI directly from histology slides which represent the gold standard for cancer diagnosis and are ubiquitously available for cancer patients. Our method outperformed existing method on the uterine corpus endometrial carcinoma cohort in The Cancer Genome Atlas (AUC 0.73 *vs.* 0.56).

***Index Terms*—** Microsatellite instability, histopathology, deep residual learning

1. INTRODUCTION

Microsatellite instability (MSI) is a form of genetic hypermutation caused by defects in the DNA mismatch repair system. Identifying the microsatellite status (MS status) of a patient is clinically important. MSI has been shown to be a positive prognostic factor in colorectal cancer [1], and the revised Bethesda guidelines have included MSI testing in diagnosis of Lynch syndrome for some colorectal cancer patient cohorts [2]. Patients with MSI tumor in gastrointestinal cancer have also been shown to be more responsive to immune-enhancing therapies [3]. With the popularization of immunotherapy, identification of the MS status of tumors may be crucial to determine the treatment plan for patients. However, measuring MSI requires additional genetic tests, which may not be available for every patient.

Since pathological slides are the foundation and gold standard for identifying and diagnosing carcinomas, and are ubiquitous available for cancer patients, there have been several attempts to predict MS status using pathological features. Jenkins et al. [2] utilized a wide range of pathological features scored by pathologists and developed a linear regression model (MsPath) to predict MSI based on these features. Hyde et al. [4] further expanded the pathological features used in MsPath and achieved superior performance. Recently, Kather et al. [3] utilized a deep learning framework to predict MSI extent of patches sampled from WSIs. Patch-wise predictions were then integrated by majority voting to estimate MSI at patient-level, representing the state-of-the-art practice in predicting MSI in gastrointestinal cancer from histological slides.

However, most existing methods rely on domain experts to extract the pathological features [2, 4], which are not feasible for large-scale applications. Although deep learning was used to infer patch-wise MS status automatically in [3], the results of patch-wise prediction were not fully utilized by taking advantage of machine learning algorithms. Moreover, most previous studies focused on Colon Adenocarcinoma (COAD) and Stomach Adenocarcinoma (STAD), despite the prevalence of microsatellite instable (MSI) tumors in Uterine Corpus Endometrial Carcinoma (UCEC). According to [1], in The Cancer Genome Atlas (TCGA), MSI are detected in 31.37% of patients in the UCEC cohort, while MSI only happens in 19.72% of the patients in the COAD cohort and in 19.09% of the patients in the STAD cohort. So, there is a need for new methods to predict patient-wise MS status in UCEC from histological slides more effectively.

978-1-5386-9330-8/20/$31.00 ©2020 IEEE

In this paper, we aim to predict MS status of patients directly from H&E-stained histological images. The proposed method utilized deep residual learning to identify MSI-related patches in WSIs and further used machine learning algorithms to integrate the information of patches for patient-level predictions. Our method outperformed state-of-the-art pipeline in UCEC cohort in TCGA.

2. METHODS

2.1. Method overview

In this study, an H&E-stained histology whole-slide image (WSI) was acquired for the examined tumor from each patient. The MS status of each patient was represented as a score (MS-SC) ranging from 0 to 50. Following the suggestions of domain experts, we further divided the MS-SCs into two groups, microsatellite instable (MSI) and microsatellite stable (MSS) using the threshold of 10, where MSI tumors were represented by higher MS-SCs.

An overview of our proposed method is shown in Fig. 1, which consists of two parts. First, patches were sampled from each WSI and a deep neural network was trained to predict the MS status of each patch. Then, PAtch Likelihood HIstogram (PALHI) was used to further integrate the prediction of patches to infer the MS status of a patient. Details of each step are introduced in the following sections.

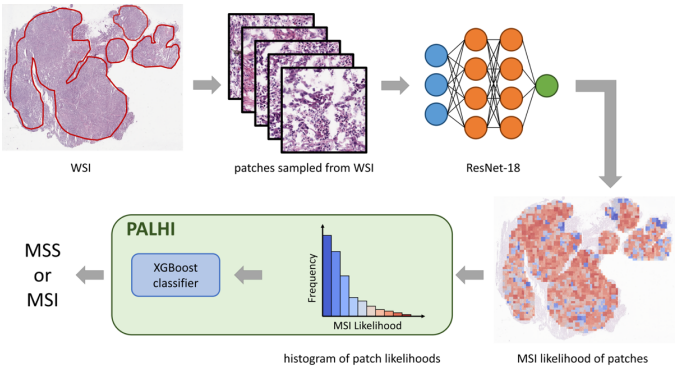


Fig. 1: Workflow to predict MS status from H&E histology whole-slide imaging.

2.2. Patch-wise MSI prediction

For each WSI, tiled patches were sampled from cancerous regions. Patches with size of 512 $\times$ 512 were sampled at 20$\times$ magnification level with spatial resolution of 0.5 μm/pixel. The label of each patch was determined by the MS status of its corresponding patient and a residual learning convolution network (ResNet-18) pre-trained on ImageNet was fine-tuned using the extracted training patches [3, 5]. Binary cross-entropy (BCE) loss (Eq. 1) was optimized when training the network with mini-batch gradient descent algorithm.

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{n=1}^{N}[y_n log x_n + (1-y_n)log(1-x_n)] \quad (1)$$

N is the size of each mini-batch. y_n is usually a binary label (*i. e.* $y_n \in \{0, 1\}$), where 0 denotes MSS and 1 denotes MSI. $x_n \in (0, 1)$ is the output of the network after a sigmoid activation, representing the estimated MSI likelihood of a patch.

In the perspective of information theory, given the true distribution p and the estimated probability distribution q, cross entropy represents the average number of bits needed to identify events drawn from these two distributions, where the coding scheme is optimized for the estimated distribution q. Under this framework, the cross entropy could be represented as $-E_p(log(q))$. Here, p can be either a discrete distribution or a continuous distribution. Similarly, in our study, MS status represented as y_n is Eq. 1 could be either defined as $\in \{0, 1\}$, representing MSS or MSI, or $\in [0, 1]$, representing the continuous variation of extent of MSI. We refer to the first definition of y_n as "hard label" and the second as "soft label". Since the MS-SC varies continuously from 0 to 50 with 10 as the boundary between MSS and MSI, we scaled MS-SC to $[0, 1]$ using Eq. 2 to generate "soft label" $\widetilde{\text{MS-SC}}$. Soft labels were further used in optimizing the loss function in Eq. 1 for training the patch classifier.

$$\widetilde{\text{MS-SC}} = \begin{cases} 0.5\text{MS-SC}/10 & \text{MS-SC} < 10 \\ \dfrac{0.5(\text{MS-SC} - 10)}{40} + 0.5 & \text{MS-SC} \geq 10 \end{cases} \quad (2)$$

After the training of ResNet-18 converged, we calculated the network output of each patch with sigmoid activation, which represented the estimated MSI probability.

2.3. Patient-wise MSI prediction

Since our goal is to determine the MS status for cancer patients, we need to integrate MSI likelihood estimation of patches to infer the MS status at patient-level. PALHI was used to predict the MS status of patients from patch-wise MSI likelihood. For each patient, PALHI generated a histogram by dividing 0 to 1 with nb bins of equal widths using the MSI likelihood estimation of patches. This histogram can represent the distribution of patch likelihoods, which can be used to capture the characteristics related to MS status of a WSI. Patient-level MS status inference can be further made using the histograms as features. XGBoost [6], a powerful and efficient gradient tree boosting based classifier was then trained using the generated histograms to distinguish MSI and MSS patients.

To show the superiority of our method, we compared PALHI with two other patient-wise MSI prediction methods. One is majority voting (Majority) used by Kather et

al. [3], which represents the state-of-the-art method in predicting MSI directly from pathological slides. The other one is Bag of Words (BoW), which was inspired by document classification problems.

In BoW, for each patch, MSI likelihood estimated in the previous step was first truncated to a certain precision and then represented as a string. After this transformation, each WSI can be treated as a document, or a bag of words, with terms of patch-level MSI likelihoods. A patient-level classifier can be then trained using the frequency patterns of such terms to predict the MS status at patient level. We took advantage of term frequency inverse document frequency (tf-idf) [7] to represent the frequency patterns and capture the importance of patch-level MSI likelihoods to WSIs. As one of the most popular term-weighting schemes, tf-idf conforms to the intuition where patches with likelihood only appear in a small set of patients might be more informative than common patches. After tf-idf transformation, we followed a widely used pipeline in information retrieval [8] and used a Naive Bayes classifier for patient-level MS status classification.

3. RESULTS AND DISCUSSION

3.1. Results on the UCEC cohort

A UCEC cohort with 516 patients were obtained from TCGA, which contains 388 MSS patients and 128 MSI patients. For each patient, one flash frozen WSI and the corresponding MS-SC were used in this study. Each WSI was manually annotated by pathologists to highlight the regions of carcinoma. An MS status label was assigned to each WSI based on the MS-SC and the threshold of 10. An average of 224 patches were sampled from each WSI, where the maximum number of patches sampled from a single WSI was 1149 and the minimum was 4. The label of each patch was determined by its corresponding WSI. We used 361 patients for training and the rest 155 patients for testing, where the percentage of MSI patients in training and testing set were the same.

ResNet-18 pre-trained on ImageNet was first fine-tuned using patches in the training set, where two kinds of labels ("hard label" and "soft label") were used in optimizing the loss function, resulting in two patch discriminators for the UCEC cohort. Combining with the following patient-level prediction methods (PALHI, BoW, and Majority), six different pipelines were tested for comparison (*i.e.* PALHI_hard, PALHI_soft, BoW_hard, BoW_soft, Majority_hard, and Majority_soft). The performance of each method was evaluated using area under the receiver operating characteristic curve (AUC), and the training and testing results are shown in Table 1.

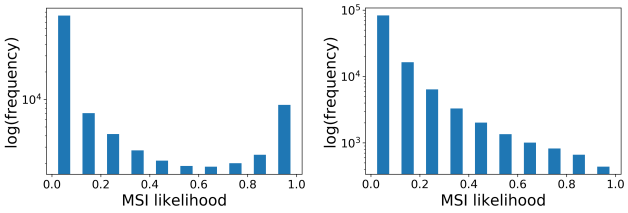
From Table 1, we observe that PALHI_soft outperformed all the other methods in the UCEC cohort, including the state-of-the-art pipeline (Majority_hard and Majority_soft) [3]. Table 1 shows that performance of patient-level prediction could be improved by taking advantage of machine learning methods to integrate the patch-wise prediction results.

Table 1: Results on TCGA UCEC cohort

Method	Training AUC	Testing AUC
PALHI_hard	0.99	0.65
PALHI_soft	0.93	**0.73**
BoW_hard	0.99	0.68
BoW_soft	0.94	0.71
Majority_hard	0.98	0.60
Majority_soft	0.82	0.56

3.2. Training with "soft labels"

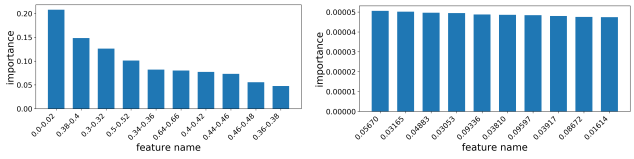
From Table 1, we observe that the performance of PALHI was improved by utilizing "soft labels" during the training of patch-wise predictors. Fig. 2 compares the distribution of estimated patch-wise MSI likelihoods for networks trained using "hard labels" and "soft labels".



(a) hard label (b) soft label

Fig. 2: Estimated patch MSI likelihood distributions

Since "hard labels" only consists of 0 (MSS) and 1 (MSI), the inferred likelihoods from the patch-level network tend to concentrate on regions near 0 and 1 (Fig. 2a) due to the supervision given during training. However, since the likelihood corresponds to the extent of MSI, patches with high MSI likelihoods should occur more rarely that those with low MSI likelihoods intuitively. So, the distribution skewed to 0 and 1 in the "hard labels" results might be the artifacts of training with binary labels. By introducing soft labels in the patch-level training, we observe that the distribution of MSI likelihoods varies monotonously (Fig. 2b), which faithfully captured the underlying distribution of patch likelihoods.



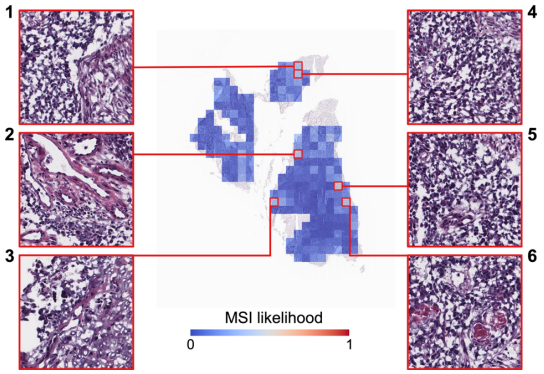
(a) PALHI (b) BoW

Fig. 3: Top important features selected by BoW and PALHI with patch-wise predictor trained with "soft labels"

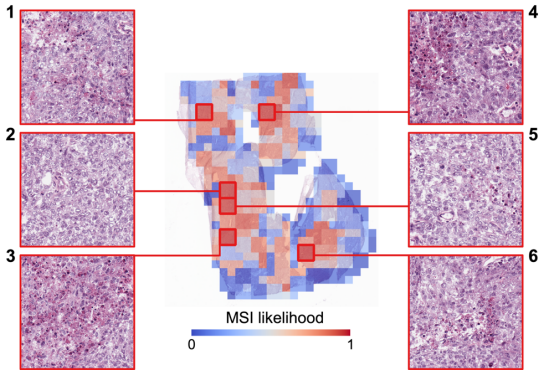
Using the likelihoods estimated by the patch discriminator trained using soft labels, a wide range of patch likelihoods

were selected by the XGBoost classifier in the PALHI method (Fig. 3a), resulting in the superior performance in patient-level prediction. However, in the BoW method, although tf-idf was designed to reduce the influence of common patches across patients and focus on the patient-specific patches, it overly averaged the importance of different likelihoods when the estimated likelihoods were already adjusted by training with soft labels, making the following Naive Bayes classifier unable to identify important patient-level features (Fig. 3b).

3.3. High MSI region identification



(a) MSS WSI



(b) MSI WSI

Fig. 4: Patches with top MSI likelihoods in WSIs

Our proposed method can also identify tumor regions corresponding to high MSI likelihoods that provide important pathological insights. For example, Fig. 4 shows a MSS WSI and a MSI WSI in the UCEC cohort with patches of top 6 highest estimated MSI likelihoods zoomed for details. We observe that top MSI patches in MSS and MSI WSIs have different patterns. Specifically, high MSI patches in MSS WSIs show heterogeneous structures with malignant cells and connective tissues (Fig. 4a1, 2). Moreover, high MSI patches in MSI WSIs also show signs of poorly differentiated cancer cells (Fig. 4b). Different patterns in the identified high MSI likelihood regions between MSS and MSI WSIs may be related to the underlying relations between MSI and histology, which is worth further investigation in the future.

4. CONCLUSION

We proposed a pipeline to predict the MS status of patients based on H&E-stained WSIs. MSI likelihood of patches sampled from WSIs was first estimated by fine-tuning a patch discriminator pretrained on ImageNet. "Soft labels" were utilized in the training process to faithfully capture the patch-wise MSI likelihood distribution. PALHI was then used to integrate the information of patches and predict patient-level MS status. Our pipeline outperformed existing state-of-the-art method (AUC = 0.56) on TCGA UCEC cohort with the testing AUC of 0.73.

5. REFERENCES

[1] Russell Bonneville, Melanie A Krook, Esko A Kautto, et al., "Landscape of microsatellite instability across 39 cancer types," *JCO precision oncology*, vol. 1, pp. 1–15, 2017.

[2] Mark A Jenkins, Shinichi Hayashi, Anne-Marie Oshea, et al., "Pathology features in bethesda guidelines predict colorectal cancer microsatellite instability: a population-based study," *Gastroenterology*, vol. 133, no. 1, pp. 48–56, 2007.

[3] Jakob Nikolas Kather, Alexander T Pearson, Niels Halama, et al., "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer," *Nature medicine*, p. 1, 2019.

[4] Angela Hyde, Daniel Fontaine, Susan Stuckless, et al., "A histology-based model for predicting microsatellite instability in colorectal cancers," *The American journal of surgical pathology*, vol. 34, no. 12, pp. 1820–1829, 2010.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[6] Tianqi Chen and Carlos Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. ACM, 2016, pp. 785–794.

[7] Juan Ramos, "Using tf-idf to determine word relevance in document queries," in *Proceedings of the first instructional conference on machine learning*. Piscataway, NJ, 2003, vol. 242, pp. 133–142.

[8] Thorsten Joachims, "A probabilistic analysis of the rocchio algorithm with tfidf for text categorization.," Tech. Rep., Carnegie-mellon univ pittsburgh pa dept of computer science, 1996.