Predicting Microsatellite Instability from Pathological Images Using Self-Supervised Learning Methods

Ruyun Ruan
College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China

Shoujin Huang*
College of Health Science and Environmental Engineering, Shenzhen Technology University, Shenzhen, China

Liyilei Su
College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China

Rashid Khan
College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China

Bingding Huang†
College of Big Data and Internet, Shenzhen Technology University, Shenzhen, China
huangbingding@sztu.edu.cn

ABSTRACT

Microsatellite instability (MSI) is a phenotypic consequence of defect deoxyribonucleic acid (DNA) mismatch repair function. The MSI status is a crucial biomarker in assessing a cancer patient's response to immunotherapy. Polymerase Chain Reaction (PCR) testing or immunohistochemical detection were the gold-standard methods for determining MSI. Recent studies have demonstrated that deep learning approaches can rapidly and cost-effectively determine MSI status from standard pathological images. This study employed several self-supervised techniques, including ResNest, Transformer, and other deep learning architectures, on The Cancer Genome Atlas (TCGA) pathological image datasets to detect MSI status. The experimental results demonstrated that all methods achieved an area under the curve (AUC) exceeding 0.90. These results surpass the performance of a single traditional network structure on the datasets, thereby validating the superiority of the self-supervision training methods to predict MSI status from pathological images. Our proposed self-supervised methods effectively bridge the knowledge gap between natural and medical image-pretrained models.

Additional Keywords and Phrases: Microsatellite instability, Self-supervised, Deep learning, Pathological image

CCS CONCEPTS

• **Computing methodologies** → Artificial intelligence; Computer vision; Computer vision problems; Object detection; Artificial intelligence; Computer vision; Computer vision problems; Image segmentation; Artificial intelligence; Computer vision; Computer vision problems; Object identification.

ACM Reference Format:
Ruyun Ruan, Shoujin Huang, Liyilei Su, Rashid Khan, and Bingding Huang. 2023. Predicting Microsatellite Instability from Pathological Images Using Self-Supervised Learning Methods. In *2023 International Conference on Advances in Artificial Intelligence and Applications (AAIA 2023), November 18–20, 2023, Wuhan, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3603273.3635242

1 INTRODUCTION

A microsatellite (MS) position refers to a constantly repeated single nucleotide with a length of 10 to 20 base pairs in the genomic sequence. In colorectal cancer (CRC) and other cancer types, mismatch repair deficiency (dMMR) generates Microsatellite instability (MSI), a specific DNA damage pattern with the successive repeating length of these microsatellite positions differing from the reference genome (becoming shorter or longer). In intermediate and late-stage CRC cancer patients, MSI is frequently linked to a lack of chemotherapeutic response and is a good predictor of immune checkpoint inhibition response [1]. In addition, Lynch syndrome, the most prevalent hereditary cause of CRC, MSI is one of the genetic mechanisms underlying carcinogenesis [2]. Therefore, MSI undoubtedly has a variety of clinical implications. Usually, multiplex polymerase chain reaction (PCR) [3] test, multiplex immunohistochemistry (IHC) panel [4], or next-generation sequencing are required to detect MSI status. Due to the high cost of the tests, these methods might not be available to all patients. Additionally, finding accredited testing facilities can be challenging for cancer patients. Further, the limited accuracy of these tests makes them less than ideal for all the patients. For MSI testing, some studies demonstrate a sensitivity of 100% and a specificity of 61.1% [5] or a higher specificity of 92.5% with a lower sensitivity of 66.7% [6]. On the other hand, predicting MSI via deep learning has outperformed the MSI genetic test in terms of cost and accuracy. Because MSI diagnosis is related to tumor patients' prognosis and therapy, avoiding the additional cost of Hematoxylin and Eosin (H&E) pathological slices, as well as the difficulty of finding accredited institutions can be important for patients. Moreover, pathological images of MSI tumor patients reveal morphological characteristics such as

*Shoujin Huang and Ruyun Ruan contributed equally to this work.
†Corresponding author: Bingding Huang (huangbingding@sztu.edu.cn)

AAIA 2023, November 18–20, 2023, Wuhan, China

ACM ISBN 979-8-4007-0826-8/23/11…$15.00
https://doi.org/10.1145/3603273.3635242

tumor-infiltrating lymphocytes, cultural differentiation, heterogeneous morphology, and poor differentiation [7]. Although these characteristics are difficult to quantify manually, deep learning can be used to train and produce reliable and accurate MSI predictions.

It was first shown that a deep learning algorithm could predict MSI from a dataset of pathological images. According to Kather et al. [8], deep learning approaches can be applied to determine MSI status from the whole slide image (WSI). They trained one ResNet18 [9] model to segment tumor regions on WSI and another to predict MS status (microsatellite instability or microsatellite stability) in each tumor tile. Each ResNet18 model was pretrained on ImageNet, then the models were trained and validated on various TCGA cohort datasets, with an Area Under Curve (AUC) of 0.77, 0.81, and 0.75 on colorectal, gastric, and endometrial formalin-fixed paraffin-embedded (FFPE) datasets, respectively. Following this work, many researchers applied deep learning methods to predict MSI, but most used CRC datasets. Ke et al. employed a knowledge distillation model to identify patch-level MSI [10]. Kim et al. constructed a model based on DeepLabv3+ with OctaveResNet to employ MSI prediction of CRC [11]. Wang et al. used the classical ResNet model to predict MSI status based on the endometrial cancer (UCEC) dataset [12]. They took patches from each WSI and trained a ResNet18 model to predict the MS status of each patch, then used the patch likelihood histogram to integrate the patch predictions further to infer a patient's MS status. Their approaches achieved AUCs of 0.73. Furthermore, self-supervised learning has recently gained popularity in the computer vision community and has been demonstrated effective in medical imaging segmentation. Venkatakrishnan et al. introduced a self-supervised construction method based on reconstruction and predictive geometric transformation [13]. Zhang et al. introduced a 3D network architecture that leveraged semantic supervision from a large-scale 2D natural image dataset to increase the accuracy and training convergence speed of 3D medical imaging tasks [14].

Inspired by the superior performance of self-supervised learning, we combined several popular self-supervised learning techniques with various neural network structures to predict MS status from WSI images. We compared our model with ImageNet pretrained and random initialization models to explore the advantages of self-supervised learning in MS status prediction. Prior to this study, Saillard et al. [15] showed that using self-supervision to detect dMMR/MSI tumors outperformed ImageNet pre-trained models and obtained an AUC of 0.92 for CRC tumors. However, their research did not compare the outcomes of self-supervised learning methods with random initialization models. Besides, we compared the performance of several self-supervised approaches on the same dataset, further demonstrating the superiority of self-supervised learning methods in such studies. Our best model yielded an AUC of 0.93 for CRC tumor samples. Furthermore, we compared the performance of transformer architecture on datasets with different methods. We found that using the random initialization model, the accuracy of the vision transformer [16] and Swin-transformer [17] models is only 83%. However, the convergence of training and model accuracy could be significantly augmented using self-supervised and ImageNet pretrained models.

2 METHODS

2.1 Workflow Overview

An overview of our proposed method is shown in Figure 1. First, the NCT-CRC-HE-100K [16] dataset was used to train the tumor detection model from WSI. It contains 100,000 non-overlapping picture patches from histological images of human colorectal cancer (CRC) and normal tissue stained with Hematoxylin and Eosin (H&E). All images are 224 * 224 pixels at 0.5 microns per pixel (MPP). The approach proposed by Macenko [17] is used to color-normalize all images. Tissue classes include Adipose (ADI), background (BACK), debris (DEB), lymphocytes (LYM), mucus (MUC), smooth muscle (MUS), normal colon mucosa (NORM), cancer-associated stroma (STR) and colorectal adenocarcinoma epithelium (TUM). Tumor samples contained CRC primary tumor slides and tumor tissue from CRC liver metastases. Normal tissue slides were augmented with non-tumorous regions from gastrectomy specimens. Then, the tumor detection model was used to determine the tumor area in WSI and obtain each slice's MSI or MSS status. Finally, the patient-level MS status was determined through a major voting approach. To evaluate the MS status prediction model, two TCGA datasets [8] TCGA-CRC-DX (250 CRC patients) and TCGA-STAD (315 STAD patients) were used in this study.

2.2 Tumor Prediction Model

ResNet [18] is a neural network model that plays a significant role in computer vision. It has shown promising results in classification, segmentation, and detection tasks with strong robustness and generalization and has been widely employed in medical imaging tasks. ResNest [9] is a ResNet variation that uses split attention and channel attention strategies to improve the effect of the ResNet model. Therefore, ResNest was chosen as our automatic tumor detection model. We first pretrained the model with ImageNet [19], and then the NCT-CRC-HE-100K dataset was used to fine-tune the ResNest model. The tumor detection model was then tested on the whole image to obtain all the tumor positions (each pixel in WSI was classified by 0 or 1, 0 for normal tissue pixels, 1 for tumor tissue pixels). The whole process of tumor prediction is illustrated in Figure 2.

2.3 Patch Embedding and MS Status Prediction

The most significant benefit of self-supervised strategies is that they enable data to provide supervised information to the learning model independently. We, therefore, used the self-supervised algorithms to uncover potential knowledge of the pathological image dataset to improve the MS status prediction precision. The tumor areas of all CRC WSI slides were divided into 224 * 224 patches. We chose ResNeSt50, Swin-transformer, and Vision Transformer as our patch embedding models in this study. Then, we used self-supervised learning algorithms to pretrain the parameters. After pretraining, we removed the output head by self-supervised learning approaches, established a deep neural network connector, and randomly initialized the classification connector parameters to classify the images. Lastly, we fine-tuned the model to conduct supervised learning on the patches, using Mixup [20] and Auto-augment [21] as data augmentation and cross-entropy as the loss function.

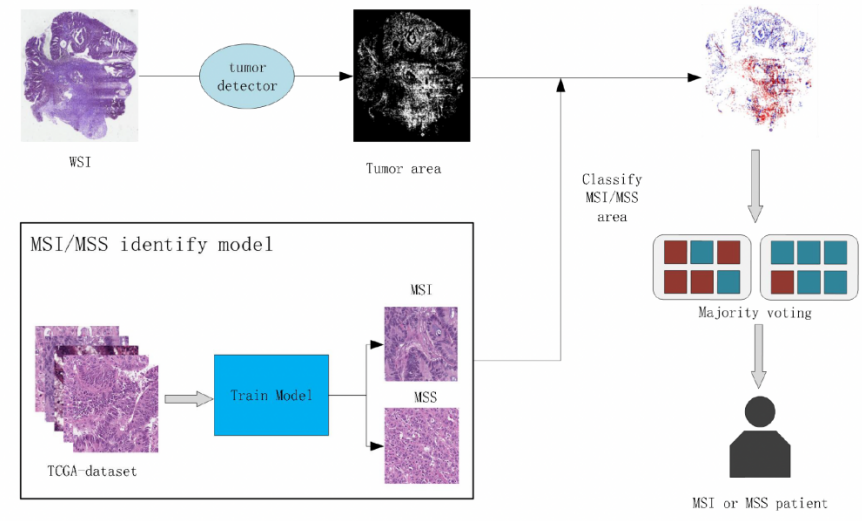

Figure 1: The whole workflow of our approach to predict MS status from histology whole-slide image.

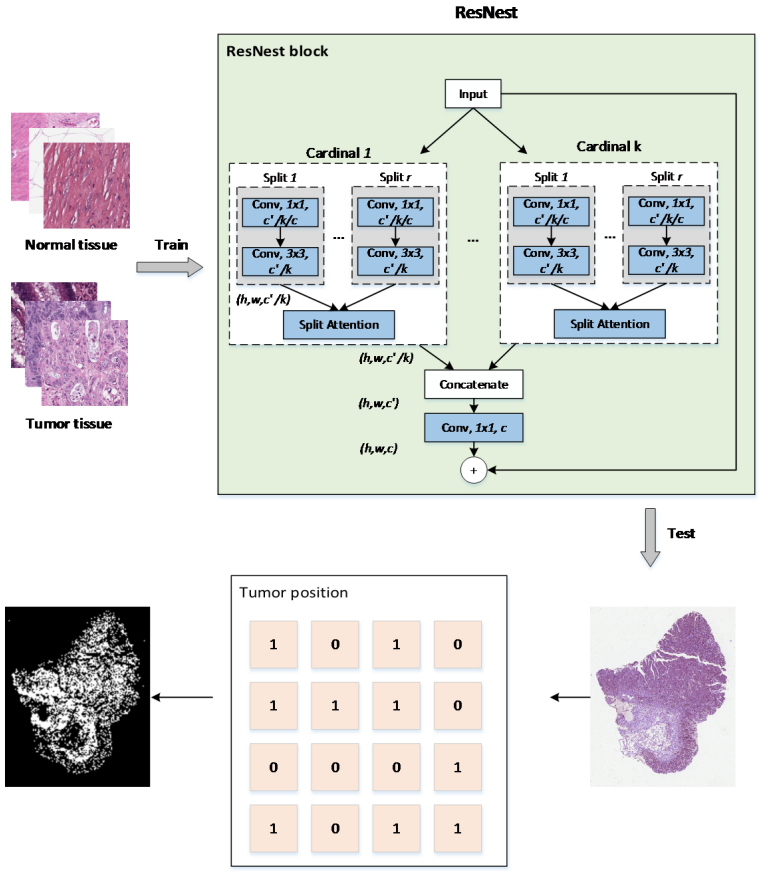

Figure 2: The tumor detection process. First, the tumor tissue and normal tissue WSIs in the training dataset were trained using ResNest network. Then given a WSI in the test dataset, the model can predict the tumor positions in WSI (tumor: 1, normal: 0).

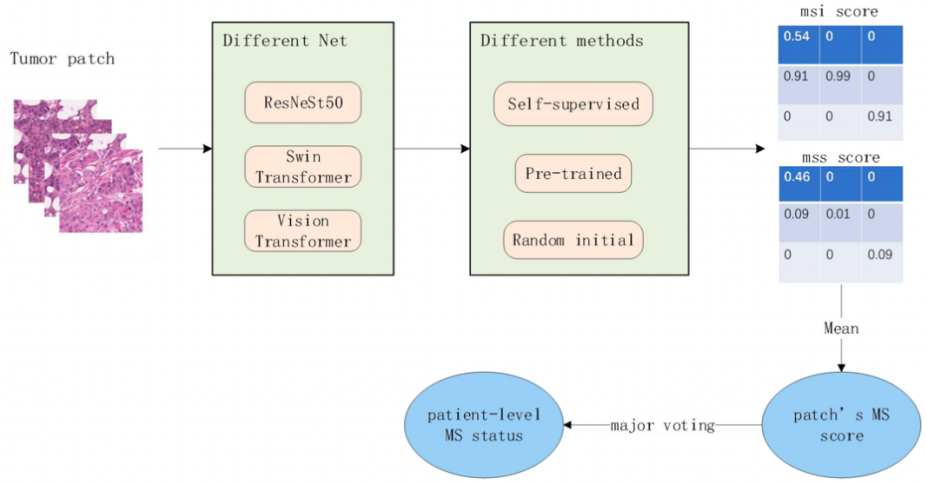

Figure 3: The overall workflow to predict patient-level MS status from tumor patch. First, the tumor patches were embedded by three different networks (ResNest50, Swin-Transformer and Vision Transformer). Then different fine-tune methods were tested to derive MS score for each tumor patch. Finally, the patient-level MS status was predicted using a major voting approach.

We employed the soft label to constrain the cross entropy to avoid overfitting. We configured the model's L2 regularization to 1e-5 and established a batch size of 128. The model's performance was then evaluated on a test dataset. Like the training process, the whole slide images from the test dataset were segmented into patches of 224 * 224 pixels. The model then computed a prediction score for MS status for each patch, embedding this score in the corresponding location within the image. Subsequently, the patient-level MS status was predicted using a majority voting approach based on the prediction scores from each method. Starting from the tumor patch, the MS status prediction process is depicted in Figure 3.

3 RESULTS

3.1 Evaluation Metrics

In this work, we mainly use the area under the curve (AUC) value to evaluate the performance of our approach to determine MS status. With MSI status as positive classes and MSS status as negative classes, the AUC is the area under the ROC curve that shows the relationship between the true positive rate (TPR) and the false positive rate (FPR). TPR and FPR are based on True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). The TPR and FPR metrics are calculated as follows:

$$TPR = \frac{TP}{TP + FN} \tag{1}$$

$$FPR = \frac{FP}{FP + TN} \tag{2}$$

3.2 Tumor Prediction Result

In the tumor detection step, we used the ResNest50 model as the training model, as it achieved the highest accuracy in the validation set, and it has fewer network parameters, faster reasoning speed, and lower computing costs than the Swin-Transformer model and other networks (data not shown). As shown in Figure 4, when the training epoch number reached 240, the ResNest50 model achieved an accuracy of 96.89%, a precision of 95.78%, a recall of 95.47%, and an F1-score of 95.44% on the NCT-CRC-HE-100K validation dataset. Furthermore, we found that when the epoch number reached 250, the tumor prediction accuracy increased to 99.82%,

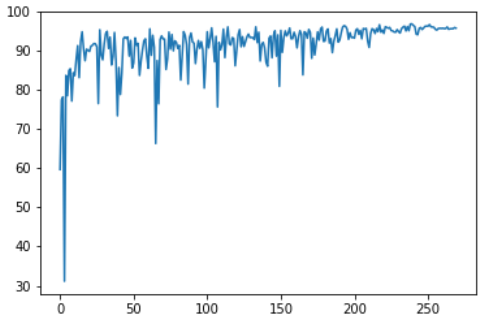

Figure 4: The accuracy of tumor prediction increases when the number of training epochs increases.

3.3 MS Status Prediction Result

In the MS status prediction step, we combined the most popular self-supervised fine-tune algorithms with the ResNest50, Vision Transformer, and Swin-Transformer, respectively. These self-supervised algorithms included MoCoV2 [17], SimSiam [22], SwAV [23], Rotation Prediction [24], MoCoV3 [25] and SimMIM [26]. Then, we fine-tuned each self-supervised learning approach using tiny learning rates. We compared the three fine-tuning strategies (parameter random initialization, pretrained using ImageNet, and self-supervised learning approaches) to train all network models on

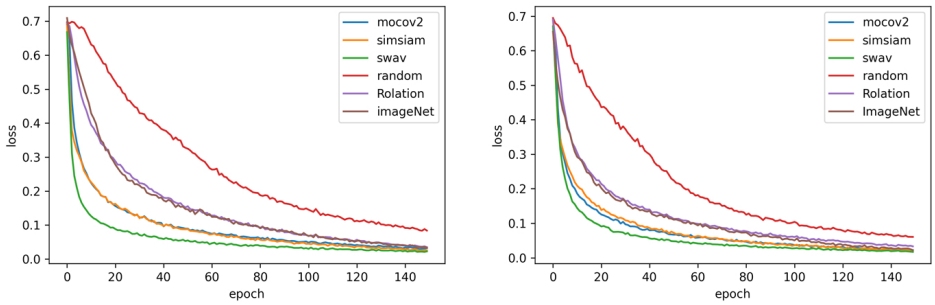

Figure 5: Training convergence speed of different methods on TCGA-CRC dataset (left) and TCGA-STAD- dataset (right).

Table 1: The AUC values of different combined approaches on TCGA-CRC-DX and TCGA- STAD datasets.

Method	TCGA-CRC-DX	TCGA-STAD
MoCoV2-ResNet50	0.90	0.89
SimSiam-ResNet50	0.89	0.82
SwAV-ResNest50	0.93	0.86
Rotation Prediction-ResNet50	0.88	0.81
MoCoV3-ViT	0.90	0.83
SimMIM-Swin Transformer	0.91	0.82

Table 2: Comparison AUC values of our best model with the other two previous methods.

Method	TCGA-CRC-DX	TCGA-STAD
Kather et al. [8]	0.77	0.81
Sailard et al. [15]	0.92	0.83
Our best model (SwAV-ResNest50)	0.93	0.86

the same dataset. We found that, compared to the other models, the convergence speed of the self-supervised training model was significantly improved on the same dataset (Figure 5). SwAV has the fastest convergence speed among all self-supervised methods.

We combined the ResNest50 model with the MoCoV2, SimSiam, SwAV, and Rotation Prediction algorithms, the MoCoV3 algorithm with the Vision Transformer model, and the SimMIM algorithm with the Swin-Transformer model. The TCGA-CRC-DX and TCGA-STAD datasets were used to train and test all the combination approaches. Table 1 shows the AUC score for each combination approach. Finally, on the same TCGA-CRC-DX and TCGA-STAD datasets at the patch level, the combination of SwAV and ResNest50 model achieved the best result with an AUC of 0.93, outperforming the results reported in Kather *et al.* [8] and Saillard *et al.* [15] (Table 2).

Furthermore, we found that the Vision Transformer and Swin-Transformer models could only achieve an accuracy score of 0.83 when utilizing random weight initialization. However, using a self-supervised method or ImageNet pretrained model weights significantly improved the model's accuracy and reduced the convergence time (see Table 3). Given the substantial disparity between natural and pathological images, the model pretrained on ImageNet fails to outperform the self-trained model on TCGA pathological image datasets. Direct migration might adversely affect the model prediction performance. Overall, our research uncovered that the self-supervised model could notably improve the performance in MSI status prediction.

4 CONCLUSIONS

In this work, we compared the performance of the same model trained by different self-supervised algorithms to identify the MS status on the TCGA pathological imaging datasets. The experiments showed that self-supervised algorithms have superior performance of AUC value on these datasets. Compared to models initialized randomly or pre-trained by ImageNet, the models trained by different self-supervised algorithms all demonstrated superior performance and rapid convergence. The training AUC value of all self-supervised methods reached a remarkable 0.90, with the best result peaking at 0.93. These results demonstrated the self-supervised algorithm's superior suitability for MS detection from pathological images compared to conventional approaches.

Table 3: AUC results of our best self-supervised method with pre-trained model and random initialization model.

Method	TCGA-CRC-DX	TCGA-STAD
ImageNet pretrained model	0.91	0.84
Random initialization model	0.90	0.82
Our best model (SwAV-ResNest50)	0.93	0.86

ACKNOWLEDGMENTS

This research was funded by the Project of the Educational Commission of Guangdong Province of China (No. 2022ZDJS113) and the School-Enterprise Graduate Student Cooperation Fund of Shenzhen Technology University.

REFERENCES

[1] Echle, A., *et al.*, Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. Gastroenterology 159(4), 1406-+ (2020).
[2] Boland, C.R. and A. Goel, Microsatellite Instability in Colorectal Cancer. Gastroenterology 138(6), 2073-U87 (2010).
[3] Boland, C.R., *et al.*, A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer. Cancer Research 58(22), 5248-5257 (1998).
[4] Kawakami, H., A. Zaanan, and F.A. Sinicrope, Microsatellite Instability Testing and Its Role in the Man-agement of Colorectal Cancer. Current Treatment Options in Oncology 16(7) (2015).
[5] Poynter, J.N., *et al.*, Molecular Characterization of MSI-H Colorectal Cancer by MLH1 Promoter Methylation, Immunohistochemistry, and Mismatch Repair Germline Mutation Screening. Cancer Epidemiology Biomarkers & Prevention 17(11), 3208-3215 (2008).
[6] Barnetson, R.A., *et al.*, Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. New England Journal of Medicine 354(26), 2751-2763 (2006).
[7] De Smedt, L., *et al.*, Microsatellite instable vs stable colon carcinomas: analysis of tumour heterogeneity, inflammation and angiogenesis. British Journal of Cancer 113(3), 500-509 (2015).
[8] Kather, J.N., *et al.*, Deep learning can predict microsatellite instability directly from histology in gastro-intestinal cancer. Nature Medicine 25(7), 1054-+ (2019).
[9] Zhang, H., *et al.*, ResNeSt: Split-Attention Networks. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2735-45 (2022).
[10] Ke, J., *et al.* Identifying patch-level MSI from histological images of Colorectal Cancer by a Knowledge Distillation Model. in IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM). Electr Network (2020).
[11] Kim, H.-R., *et al.* Colorectal Cancer Image Segmentation and Classification with Deep Neural Network Based on Information Theory. in IEEE International Conference on Bioinformatics and Biomedicine (IEEE BIBM). Electr Network (2020).
[12] Wang, T., *et al.* MICROSATELLITE INSTABILITY PREDICTION OF UTERINE CORPUS ENDOMETRIAL CARCINOMA BASED ON H&E HISTOLOGY WHOLE-SLIDE IMAGING. in IEEE 17th International Symposium on Biomedical Imaging (ISBI). Iowa, IA (2020).
[13] Venkatakrishnan, A.R., *et al.*, Self-Supervised Out-of-Distribution Detection in Brain CT Scans. abs/2011.05428 (2020).
[14] Zhang, S., *et al.*, Advancing 3D Medical Image Analysis with Variable Dimension Transform based Supervised 3D Pre-training. abs/2201.01426 (2022).
[15] Saillard, C., *et al.* Self-supervised learning improves dMMR/MSI detection from histology slides across multiple cancers. in COMPAY@MICCAI (2021).
[16] Kather, J.N., *et al.*, Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. Plos Medicine 16(1), (2019).
[17] Chen, X., *et al.*, Improved Baselines with Momentum Contrastive Learning. abs/2003.04297 (2020).
[18] Kaiming, H., *et al.*, Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-8, (2016).
[19] Deng, J., *et al.* ImageNet: A Large-Scale Hierarchical Image Database. in IEEE-Computer-Society Conference on Computer Vision and Pattern Recognition Workshops. Miami Beach, FL (2009).
[20] Zhang, H., *et al.*, mixup: Beyond Empirical Risk Minimization. abs/1710.09412 (2018).
[21] Cubuk, E.D., *et al.*, AutoAugment: Learning Augmentation Policies from Data. abs/1805.09501 (2018).
[22] Chen, X., K. He, and S.O.C. Ieee Comp. Exploring Simple Siamese Representation Learning. in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Electr Network (2021).
[23] Caron, M., *et al.*, Unsupervised learning of visual features by contrasting cluster assignments. arXiv (2020).
[24] Gidaris, S., P. Singh, and N. Komodakis, Unsupervised Representation Learning by Predicting Image Rotations. arXiv (2018).
[25] Chen, X., *et al.* An Empirical Study of Training Self-Supervised Vision Transformers. in 18th IEEE/CVF International Conference on Computer Vision (ICCV). Electr Network (2021).
[26] Zhenda, X., *et al.*, SimMIM: a Simple Framework for Masked Image Modeling. IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9643-53 (2022).