

Article

Not peer-reviewed version

# Prediction of Mismatch Repair Status in Endometrial Cancer from Histological Slide Images Using Various Deep Learning-Based Algorithms

[Mina Umemoto](#), [Tasuku Mariya](#)<sup>\*</sup>, Yuta Nambu, Mai Nagata, Toshihiro Horimai, [Shintaro Sugita](#), [Takayuki Kanaseki](#), Yuka Takenaka, Shota Shinkai, Motoki Matsuura, Masahiro Iwasaki, [Yoshihiko Hirohashi](#), [Tadashi Hasegawa](#), [Toshihiko Torigoe](#), Yuichi Fujino, [Tsuyoshi Saito](#)

Posted Date: 8 March 2024

doi: 10.20944/preprints202403.0474.v1

Keywords: endometrial cancer; deep learning; artificial intelligence; biomarker; mismatch repair; molecular classification; whole slide imaging; digital pathology



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Prediction of Mismatch Repair Status in Endometrial Cancer from Histological Slide Images Using Various Deep Learning-Based Algorithms

Mina Umemoto <sup>1</sup>, Tasuku Mariya <sup>1,\*</sup>, Yuta Nambu <sup>2</sup>, Mai Nagata <sup>2</sup>, Toshihiro Horimai <sup>3</sup>, Shintaro Sugita <sup>4</sup>, Takayuki Kanaseki <sup>5</sup>, Yuka Takenaka <sup>1</sup>, Shota Shinkai <sup>1</sup>, Motoki Matsuura <sup>1</sup>, Masahiro Iwasaki <sup>1</sup>, Yoshihiko Hirohashi <sup>5</sup>, Tadashi Hasegawa <sup>4</sup>, Toshihiko Torigoe <sup>5</sup>, Yuichi Fujino <sup>2</sup> and Tsuyoshi Saito <sup>1</sup>

<sup>1</sup> Department of Obstetrics and Gynecology, Sapporo Medical University School of Medicine, Sapporo 060-8556, Japan; mariya.tasuku@sapmed.ac.jp (T.M.); takenakayuka.skms@gmail.com (Y.T.); saishinsyou@gmail.com (S.S.); motoki.gyne@gmail.com (M.M.); miwasaki@sapmed.ac.jp (M.I.); tsaito@sapmed.ac.jp (T.S.)

<sup>2</sup> Department of Media Architecture, Future University Hakodate, Hakodate 041-8655, Japan; g3221001@fun.ac.jp (Y.N.); g2122046@fun.ac.jp (M.N.); fujino@fun.ac.jp (Y.F.)

<sup>3</sup> Limited Liability Company Gomes Company, Sapporo 004-0875, Japan; t-horimai@gomes-company.co.jp (T.M.)

<sup>4</sup> Department of Surgical Pathology, Sapporo Medical University of Medicine, Sapporo 060-8556, Japan; ssugita@sapmed.ac.jp (S.S.); hasetada@sapmed.ac.jp (T.M.)

<sup>5</sup> Department of Pathology, Sapporo Medical University School of Medicine, Sapporo 060-8556, Japan; kanaseki@sapmed.ac.jp (T.K.); hirohash@sapmed.ac.jp (Y.H.); torigoe@sapmed.ac.jp (T.T.)

\* Correspondence: mariya.tasuku@sapmed.ac.jp; Tel.: +81-11-611-2111, ext. 2692

**Simple Summary:** In the context of endometrial cancer, molecular classification is becoming increasingly significant as the molecular class determines appropriate treatments and the prognosis. However, performing clinical testing for molecular classification in a large number of patients entails significant financial and time costs. Therefore, there is a need to develop a substantial molecular profile screening method for endometrial cancers. The objective of this study was to explore whether the molecular classification of endometrial cancer could be predicted from digital images of hematoxylin and eosin (H&E)-stained slides using deep learning as a screening tool. After making adjustments to the training data set and hyperparameters, we confirmed the feasibility of estimating the mismatch repair status from histological digital images of endometrial cancer. Deep learning was found to be effective for predicting one aspect of the molecular classification from H&E-stained histological digital images.

**Abstract:** The application of deep learning algorithms to predict the molecular profiles of various cancers from digital images of hematoxylin and eosin (H&E)-stained slides has been reported in recent years, mainly in gastric and colon cancers. In this study, we investigated the potential use of H&E-stained endometrial cancer slide images to predict the associated mismatch repair (MMR) status. H&E-stained slide images were collected from 127 cases of the primary lesion of endometrial cancer. After digitization using a Nanozoomer virtual slide scanner (Hamamatsu Photonics), we segmented the scanned images into 5,397 tiles of 512 × 512 pixels. The MMR proteins (PMS2, MSH6) were immunohistochemically stained, classified into MMR proficient/deficient, and annotated for each case and tile. We trained several neural networks, including convolutional and attention-based networks, using tiles annotated with the MMR status. Among the tested networks, ResNet50 exhibited the highest area under the receiver operating characteristic curve (AUROC) of 0.91 for predicting the MMR status. The constructed prediction algorithm may be applicable to other molecular profiles and useful for pre-screening before implementing other, more costly genetic profiling tests.

**Keywords:** endometrial cancer; deep learning; artificial intelligence; biomarker; mismatch repair; molecular classification; whole slide imaging; digital pathology

## 1. Introduction

The lifetime risk of women developing endometrial cancer is approximately 3%, and, over the past 30 years, the overall incidence has increased by 132%, reflecting an increase in risk factors (particularly obesity and aging) [1]. Endometrial cancer is classically classified into two groups—namely, Type I or II tumors [2,3]. Type I endometrial tumors are associated with excess estrogen, obesity, hormone receptor positivity, and abnormalities in hormone receptors. On the other hand, Type II tumors, which are mainly serous, are often observed in older, non-obese women and are considered to have a worse prognosis [2,4].

In recent years, there has been a growing focus on molecular biological classification of endometrial cancers. The classification of endometrial cancer proposed by The Cancer Genome Atlas (TCGA) in 2013—a joint project of the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI)—employed milestone data for molecular classification of endometrial cancer [5]. The TCGA proposed classification into four classes—POLE (ultramutated), MSI (hypermutated), Copy-number low (endometrioid), and Copy-number-high (serous-like)—based on next-generation sequencing (NGS) data obtained from 232 cases of endometrial cancers. Following this TCGA classification, Talhouk et al. developed and verified a modified molecular classification method called "ProMisE" [6,7]. This method replaces the detection of abnormality of the *TP53* gene and microsatellite status, which are dependent on sequencing, with immunohistochemical staining (IHC), making molecular classification of endometrial cancers more clinically accessible. The ProMisE method classifies endometrial cancer into four molecular sub-types: POLEmut, Mismatch Repair Deficient (MMRd), p53abn, and NSMP (No Specific Molecular Profile, p53wt). These four classes correspond to the POLE, MSI, Copy-number low, and Copy-number-high TCGA classes, respectively. In 2020, the World Health Organization (WHO) also recommended a molecular classification for endometrial cancer [8,9]. In 2023, the molecular biology-based classification of the Federation of Obstetrics and Gynecology (FIGO) staging was also demonstrated [10]. Therefore, it is anticipated that the provision of a stable and easy method for molecular profiling of endometrial cancer will become clinically significant in the near future.

Based on these molecular classification results, one of the most crucial therapeutic agents to consider for classification-matched treatment is the immune checkpoint inhibitor (ICI). ICIs are being investigated and gaining interest for various type of tumors, including endometrial cancer [11]. Programmed death receptor-1 (PD-1) is an immune checkpoint molecule expressed on activated T-cells, with programmed cell death ligand 1 (PD-L1) being a representative ligand [12]. PD-1 and PD-L1 inhibitors accelerate cancer cell elimination, mainly mediated through cytotoxic T-cells. One of the accepted surrogate markers for the effectiveness of ICIs is deficient mismatch repair (dMMR) and the resulting microsatellite instability (MSI) [13]. According to a cross-organ analysis of solid tumors, endometrial cancer had the highest frequency of MSI, occurring in 17% [14].

Commonly used methods for determining dMMR/MSI status are based on polymerase chain reaction (PCR) [15,16], IHC for MMR proteins [17], and NGS [18,19]. The use of IHC for MMR status classification involves examining the expression of MutL homolog 1 (MLH1), MutS homolog 2 (MSH2), MutS homolog 6 (MSH6), and post-meiotic segregation increase 2 (PMS2) [17]. In endometrial cancer, the high detection rate of dMMR underscores the utmost importance of immunologic profiles [14,20]; however, testing for detailed molecular profiles (including MMR status) in every endometrial cancer patient can be expensive and time-consuming, which could complicate the course of treatment. Therefore, it is necessary to develop alternative molecular classification methods in order to reduce the associated financial and time costs.

As a novel classification method, we focused on approaches using machine learning. In the field of healthcare, deep learning has already been demonstrated to be useful for the classification of medical images. Originally, deep learning emerged as a prominent sub-field of machine learning [21]. There have been many reports on the effective use of deep learning approaches for image classification in clinical use, such as magnetic resonance imaging (MRI) of the brain [22], retinal images [23], and computed tomography (CT) of the lungs [24], among others [25]. Unlike conventional machine learning methods, deep learning relies on deep neural networks, which mimic

the operation of the neurons in the human brain. Deep learning networks can automatically extract the significant features necessary for the corresponding learning tasks with minimal human effort [21].

Among the various deep learning algorithms, convolutional neural networks (CNNs) have been most commonly used [26]. Each convolutional layer extracts different information and, through stacking multiple convolutional layers, the network can progressively extract more complex and abstract features. The activation function in the middle of the convolutional layers enhances the network's ability to handle non-linear problems and adapt to different distributions [27]. The network may gain features that humans may not be consciously aware of; however, many of these features can be challenging to articulate. Some reports [28–30] in the field of cancer have suggested that hematoxylin and eosin (H&E) staining can be used to predict genetic alterations and features of the tumor microenvironment without the need for further laboratory testing. In particular, determining microsatellite instability (MSI) status by deep learning analysis of H&E-stained slides has been described in several reports focused on gastric cancer [31] and colon cancer [32–36]. In endometrial cancer, Hong et al. [28] attempted a comprehensive assessment using deep learning for the detection of histological subtypes and genetic alterations, achieving an area under the receiver operating characteristics curve (AUROC) ranging from 0.613 to 0.969 despite variations in the assessment criteria. Additionally, Fremont et al. [30] similarly attempted to carry out decision making through the use of deep learning approaches and to visualize the histological features specific to molecular classifications in endometrial cancer.

Thus, the use of artificial intelligence—particularly deep learning—for medical image analysis has been rapidly expanding [25,37]. Therefore, we considered the potential application of deep learning to address issues related to endometrial cancer. In this study, we examined the utility of CNNs and novel attention-based networks for prediction of MMR status of endometrial cancer.

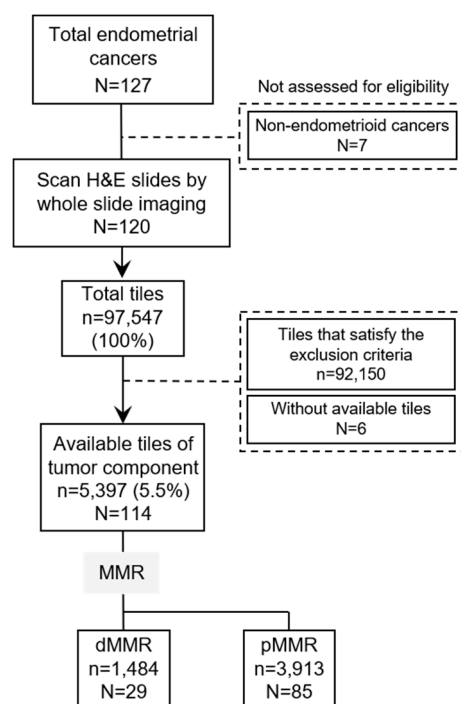
## 2. Materials and Methods

### 2.1. Ethical Compliance

According to the guidelines of the Declaration of Helsinki, informed consent was acquired through an opt-out form on the website of Sapporo Medical University. The Sapporo Medical University Hospital's Institutional Review Board granted approval for this study under permission number 332-158.

### 2.2. Patients and Specimens

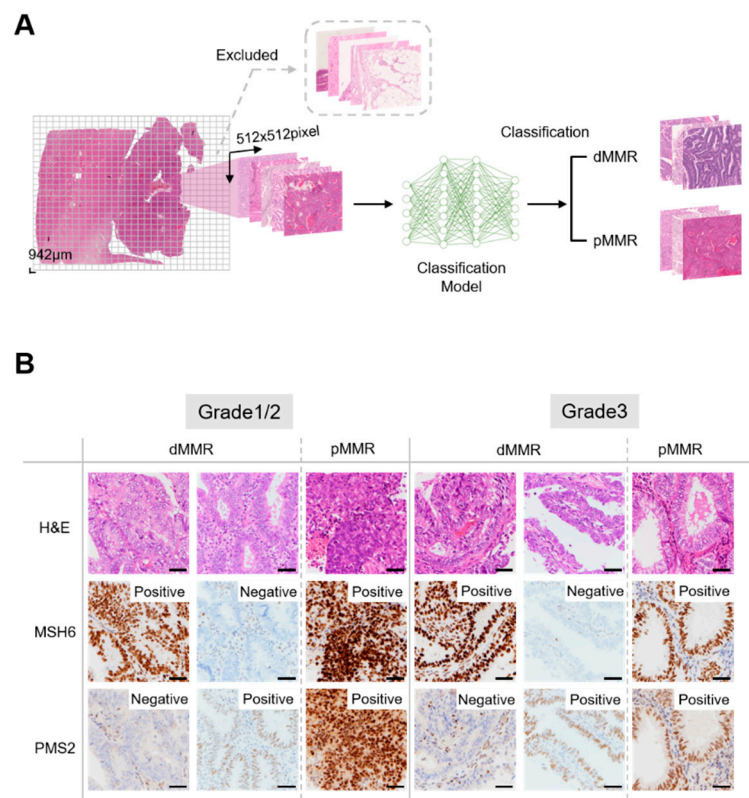
For this study, formalin-fixed paraffin-embedded (FFPE) tumor samples from Sapporo Medical University Hospital were used. Surgical specimens were obtained from patients with a primary site of endometrial cancer. A pathologist and a gynecologic oncologist chose representative slides of endometrial cancer resection specimens that were stained with H&E. Of the 127 patients with endometrial cancer treated from 2005 to 2009 in total, we excluded 7 patients with non-endometrial cancer and 6 patients without sufficiently available tumor component tiles (Figure 1).



**Figure 1.** Study flow diagram. In total, 114 eligible patients were investigated in this study, and we used 5397 tiles as the study data set. N = number of patients, n = number of tiles, dMMR = deficient MMR, pMMR = proficient MMR, WSI = whole slide imaging.

2.3. Immunohistochemistry Staining and Evaluation of MMR Status

FFPE tumor tissues were cut into 4 µm slices, and Target Retrieval Solution at pH 9 (DAKO, Glostrup, Denmark) was used for epitope retrieval. The tissues were then stained with rabbit anti-MSH6 monoclonal antibody (clone, EP51; DAKO) and mouse anti-PMS2 monoclonal antibody (clone, ES05; DAKO), which were used to detect MMR proteins in the tissues. The slides then underwent incubation with a secondary antibody. Subsequently, the slides underwent hematoxylin counterstaining, followed by rinsing, alcohol dehydration, and cover-slipping with mounting medium. Two gynecologists and one pathologist evaluated the resulting IHC and MMR status. As previously reported, negative staining for MSH6 corresponds to a lack of MSH2 and/or MSH6 proteins, as the stability of MSH6 depends on MSH2 [38]. In the same way, PMS2 staining covers the protein expression of PMS2 and/or MLH1. Therefore, if either PMS2 or MSH6 expression was deficient, it was determined as dMMR and, if not, it was determined as proficient MMR (pMMR) (Figure 2B). In total, 29 patients were classified as dMMR, while 85 patients were classified as pMMR (Figure 1).



**Figure 2.** Overview of this study and evaluation of immunohistochemistry findings for MSH6 and PMS2 in endometrial cancer. (A) Overview of data preparation and model construction. Whole slide images were cut into non-overlapping square tiles of 512 pixels at 5× magnification. Tiles that met the exclusion criteria were excluded (Supplementary Figure 1A,B), and only eligible tiles (Supplementary Figure 1C) were used as the data set. For each tile, the classification model was used to perform binary classification of mismatch repair (MMR) status. (B) Evaluation of immunohistochemistry findings for MSH6 and PMS2. Cases with a loss of expression in either PMS2 or MSH6 were classified as deficient mismatch repair (dMMR), while those without such loss were classified as proficient mismatch repair (pMMR). Bar = 100 μm.

2.4. Pre-Processing of Whole Slide Images

The H&E slides were then digitized using a Nanozoomer whole slide scanner (Hamamatsu Photonics, Japan). Each whole slide image (WSI) was divided into non-overlapping square tiles of 942 μm at a magnification of 5×, each with dimensions of 512 × 512 pixels (Figure 2A). On average, each WSI was divided into 813 tiles, and processing WSIs from 120 cases of endometrioid cancer resulted in the creation of 97,547 tiles.

We first constructed an image exclusion program, in which we specifically conducted the following pre-processing steps using OpenCV: (i) Excluding edge tiles with different numbers of pixels in height and width; and (ii) Converting the tile to HSV format and binarizing the tile through treating pixels that matched the specified pink color range from (100, 50, 50) to (179, 255, 255) as white (255) and pixels that did not match as black (0). The program calculated the average value of the pink color area and excluded it if it was greater than 25 (i.e., if there was a large amount of pink color within one tile). In total, 38,699 tiles were excluded automatically using the constructed program. Furthermore, we manually excluded 53,451 tiles without sufficient tumor component. The exclusion criteria were specified as follows: tiles in which more than 25% of the tile area consists of non-tumor components (e.g., stroma), tiles containing irrelevant contaminants within the slide, tiles with folding due to poor tissue extension during sample preparation and air trapping, and tiles with artifacts during scanning. Supplementary Figure 1A shows an overview of the tiles exclusion process through the program and manual inspection. The total number of excluded tiles (Supplementary Figure 1B)

amounted to 92,150, while eligible tiles (Supplementary Figure 1C) amounted to 5,397, accounting for 5.5% of the total number of divided tiles. Supplementary Table 1 details the number of tiles and characteristics for each patient.

### 2.5. Hardware and Software Libraries Used

The experiments were carried out with Python (version 3.8.10), making use of the following packages: torch (version 2.0.0), torchvision (0.15.1), numpy (version 1.24.1), scikit-learn (version 1.2.2), matplotlib (version 3.7.1), and timm (version 0.6.13). Model development and evaluation were performed on a workstation with GeForce RTX 3080 (NVIDIA, Santa Clara, CA) graphics processing units, a Ryzen Threadripper 3960X (24 cores, 3.8 GHz) central processing unit (Advanced Micro Devices, Santa Clara, CA), and 256 GB of memory.

### 2.6. Data Split and Training Data Preparation

The useful tiles were divided into separate data sets for training, validation, and testing. The data set cases were randomly split into training, validation, and test sets for each prediction task, such that tiles from the same patient were contained in only one of these sets. This approach ensured that the test data set was independent from the training process, allowing for a patient-level split. The split ratio for training:validation:testing was set at 70%:15%:15%.

### 2.7. Classification Model Construction Using Convolutional Neural Networks

Construction of the CNN-based binary classification model for MMR status, pMMR or dMMR, was conducted using pre-trained CNN models through torchvision in the Pytorch library, including GoogLeNet [39], VGG19 [40], ResNet50 [41], ResNet101 [41], wideResNet101-2 [42], and EfficientNet-B7 [43]. We constructed a model that input a non-overlapping image tile of size  $512 \times 512$  pixels at a resolution of  $1.84 \mu\text{m}/\text{pixel}$  and output a tile-level probability for MMR status. We fine-tuned the pre-trained models in torchvision using the prepared training data set and validated the results using the validation data set, following the provided instructions. The trainable parameters were fine-tuned using a stochastic gradient descent optimization method, and we examined the conditions for data pre-processing and the hyperparameters needed for model training. To address the imbalance in the number of tiles in each class, we down-sampled the larger class of pMMR, randomly reducing cases to align with the smaller class in terms of slide numbers. The detailed results of the down-sampling process are presented in Supplementary Table 2. We also examined changes in model performance with data augmentation. We conducted the following four patterns of data augmentation: (i) No data expansion (original tile), (ii) original tile with added  $90^\circ$  and  $270^\circ$  rotations (resulting in three times the data), (iii) original tile with added vertical and horizontal flips (resulting in four times the data), and (iv) original tile with both rotations (as in ii) and flips (as in iii) (resulting in six times the data). Furthermore, we examined the conditions for the hyperparameters, provisionally using ResNet50 [41] for the validation network. For the hyperparameters, we changed the batch size (8, 16, 32), number of epochs (30, 60, 90, 120), and learning rate ( $1e-2$ ,  $1e-3$ ,  $1e-4$ ).

### 2.8. Classification Model Construction Using Attention Networks and Our API-Net-Based Model

We also verified the performance differences between CNNs and attention-based networks, such as Vision Transformer (ViT) [44]. We selected pre-trained ViT models from the torchvision models in the Pytorch library, as mentioned above. The hyperparameters and data set were similarly chosen as for the CNNs mentioned above. We examined two ViT models—ViT\_b16 and ViT\_b32—in this study. Additionally, we examined the model of the modified network based on API-Net [45]. This modified network is a class-aware visualization and classification technique that employs attention mechanisms, which we developed for cytopathological classification and feature extraction. This API-Net-based model takes pairs of images as input and learns the embeddings of input features and representative embeddings, called prototypes, for each MMR class. We used the existing API-Net to estimate attention vectors. Given an unknown image, the classification model predicts classes

through comparing the unknown images to prototypes, recognizing their similarity for the determination of classes.

2.9. Evaluation of Constructed Model Performance

The following calculated parameters were used as indicators of model performance: Accuracy =  $(TP + TN)/(TP + FP + FN + TN)$ ; Precision =  $TP/(TP + FP)$ ; Recall =  $TP/(TP + FN)$ ; and F-score =  $2 * \text{precision} * \text{recall}/(\text{precision} + \text{recall})$ . TP, TN, FP, and FN represent the number of true positive, true negative, false positive, and false negative tiles, respectively. A receiver operating characteristic (ROC) curve is a probability curve for classification of problems at various threshold settings. The ROC curve was plotted using the TPR against the FPR, where TPR is on the y-axis and FPR is on the x-axis. The AUROC represents the area under the ROC curve.

3. Results

3.1. Pre-Processing of Data Set before Model Training

Table 1 shows the results of the data set pre-processing before model training. First, we examined the ratio of the number of tiles between data sets. Supplementary Table 2 shows the number of tiles regarding different data set ratios. When the ratio of the number of tiles was not adjusted for predicting the MMR status, the pMMR class had approximately 2.6 times the amount of data as the dMMR class. Specifically, there were 1484 tiles for dMMR and 3913 tiles for pMMR.

Next, we created a data set using down-sampling in order to match the number of slides for pMMR to the lower slide count of dMMR and compared it with the original ratio without adjustment. In the down-sampled data set, the number of tiles for dMMR remained unchanged (at 1,484); while, for pMMR, it was 1,484. Consequently, the original ratio seemed to be better when examining accuracy values exclusively. As a result, in the original ratio, the classification results were biased toward pMMR, which had a larger number of tiles. In other words, due to the increase in FN and TN, the recall rate was 0.09 and the precision rate was 0.55, both of which are low. Compared to the original data set, the down-sampled data set exhibited superior overall performance. Therefore, it was revealed that the classification performance improved in the data set with down-sampling, even though much of the pMMR training data were excluded. Additionally, we examined the effect of training on a data set that had undergone data augmentation through flipping and rotating processes, but no improvement in performance was observed. We used ResNet50 for consideration of these effects on the data set pre-processing.

**Table 1.** Results of metrics concerning different pre-processing of data sets and various hyperparameter configurations. We examined the ratio of the number of tiles between data sets, data augmentation, and hyperparameters.

			Accuracy	Precision	Recall	F-score	AUROC
Pre-processing	Ratio	Original	0.74	0.55	0.09	0.15	0.74
		Down-sampled	0.80	0.76	0.88	0.81	0.89
	Data augmentation	None	0.80	0.76	0.88	0.81	0.89
		Rotate	0.76	0.70	0.90	0.79	0.87
		Flip	0.75	0.73	0.77	0.75	0.84
		Rotate and flip	0.77	0.77	0.76	0.77	0.86
Hyper-parameter	Batch	8	0.80	0.76	0.88	0.81	0.89
		16	0.72	0.70	0.74	0.72	0.80
		32	0.69	0.70	0.67	0.68	0.77
	Epoch	30	0.80	0.76	0.88	0.81	0.89

	60	0.78	0.75	0.84	0.79	0.88
	90	0.80	0.79	0.81	0.80	0.88
	120	0.75	0.69	0.89	0.78	0.87
Learning rate	1e-2	0.80	0.76	0.88	0.81	0.89
	1e-3	0.70	0.68	0.75	0.71	0.78
	1e-4	0.69	0.64	0.86	0.74	0.78

3.2. Validation of Model Performance in Various Hyperparameter and Classification Models

Table 1 shows the hyperparameter tuning results. To predict the MMR status, we used a down-sampled data set with good performance and adjusted the hyperparameters. Compared to batch sizes of 16 or 32, a batch size of 8 showed good accuracy results; therefore, we selected the batch size of 8. Regarding the number of epochs, we adopted 30 epochs as, in our study, the model prediction performance in the validation data set presented high values up to epoch 30, regardless of whether a greater number of epochs was considered. Regarding the learning rate, we chose 1e-2, which performed better than even lower values. We conducted a hyperparameter search for the API-Net-based model and adopted a learning rate of 1e-3, which yielded the highest accuracy.

Table 2 shows the results regarding the differences among the classification models. We examined the differences among the classification models without changing hyperparameters between CNNs and ViT. We utilized pre-trained networks available in Pytorch and focused on validating the classical networks commonly used for CNNs. In all CNNs, the AUROC exceeded 0.8, with particularly high values (0.89) observed for ResNet50 and ResNet101. On the other hand, in the comparison and evaluation of attention methods, satisfactory performance was not achieved with any ViT model (AUROC: ViT\_B16, 0.62; ViT\_B32, 0.76). However, regarding our developed API-based network, the model achieved sufficient performance, with 0.81 accuracy and 0.89 AUROC.

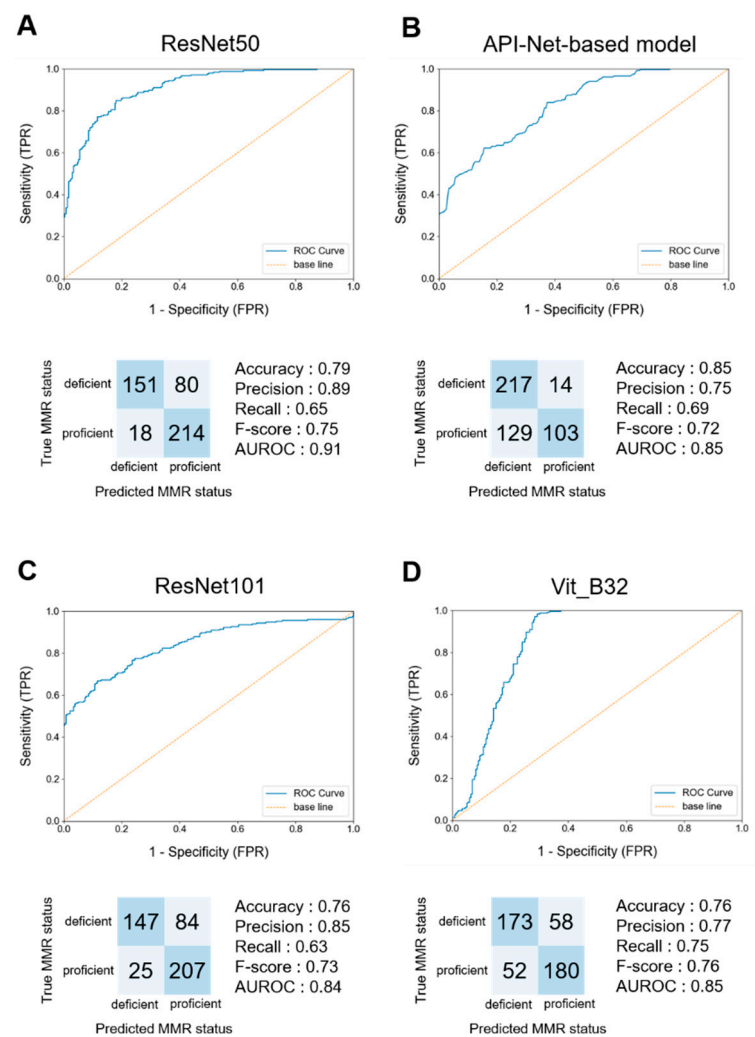
**Table 2.** Results of metrics concerning various classification models. We examined the differences in performance among classification models using CNN or attention mechanism.

		Accuracy	Precision	Recall	F-score	AUROC
Convolutional neural network	GoogLeNet	0.74	0.72	0.79	0.75	0.83
	VGG_19_BN	0.79	0.86	0.68	0.76	0.85
	ResNet50	0.80	0.76	0.88	0.81	0.89
	ResNet101	0.81	0.78	0.88	0.82	0.89
	wideResNet101-2	0.77	0.88	0.62	0.73	0.88
	EfficientNet-B7	0.74	0.77	0.68	0.72	0.81
Attention mechanism	ViT_B16	0.57	0.59	0.43	0.50	0.62
	ViT_B32	0.67	0.61	0.89	0.73	0.76
	API-Net-based model	0.81	0.81	0.81	0.81	0.89

3.3. Performance of the Models for Unseen Test Data Set

Next, we examined the generalization performance of the model through adapting it to untested data sets. Figure 3 shows the results regarding the performance of the best-performing model on the test data set, while Supplementary Figure 2 shows the performance of all models validated on the test data set. We investigated the best methods for pre-processing of the data set, hyperparameters, and classification models, as described above. As a result, we adopted a combination of down-sampling in the ratio between the two classes. For CNNs, with the combination of batch size = 8, epochs = 30, and learning rate = 1e-2, utilizing ResNet50 or ResNet101 yielded the best performance.

Among the attention methods, our API-Net-based model achieved the best accuracy when compared to the other pre-trained ViT models.



**Figure 3.** Model performance on test data set at the per-tile level. (A) Receiver operating characteristic (ROC) curves and confusion matrix using ResNet50. ResNet50 had the highest accuracy among the CNN models. (B) ROC curves and confusion matrix using API-Net-based model. The API-Net-based model had the highest accuracy among the models using the attention mechanism. (C) ROC curves and confusion matrix using ResNet101. ResNet101 had the second-highest accuracy among the CNN models. (D) ROC curves and confusion matrix using Vit\_B32. Vit\_B32 had the second-highest accuracy among the models using the attention mechanism. TPR = True positive rate, FPR = False positive rate.

#### 4. Discussion

The incidence of endometrial cancer is increasing worldwide [46] and, considering the rising importance of molecular biological tests, we need to think about future approaches to diagnosis and treatment in this context. The PORTEC-RAINBO [47] trial is one of the largest clinical trials investigating genotype-matched therapy for endometrial cancer, which aims to improve clinical outcomes and reduce the toxicity of unnecessary treatments in patients with endometrial cancer through molecularly directed adjuvant therapy strategies. One of the RAINBO trials, the MMRd-GREEN trial, enrolled patients with dMMR endometrial cancer at stage II with significant lymphovascular space invasion (LVSI) or stage III, mismatch repair-deficient endometrial cancer. It then compared a group receiving adjuvant radiotherapy with concurrent and adjuvant durvalumab

for one year with a group receiving radiotherapy alone. In this trial, IHC was used for the determination of dMMR, as performed in the present study. Assessment of MMR status will become increasingly important in the future.

Both MSI and MMR status can serve as surrogate markers for selecting candidates to receive ICI and refer to similar biological entities; however, there are various methods used for their detection in clinical laboratories. In the MSI test, five microsatellite regions (BAT-25, BAT-26, MONO-27, NR-21, and NR-24) in DNA obtained from tumor and normal tissue from the same patient are amplified using PCR [15]. Tumors are classified as high microsatellite instability (MSI-H) if two of the five microsatellite markers present a length difference between the tumor and normal samples; low microsatellite instability (MSI-L) if only one microsatellite marker presents a length difference; and microsatellite stable (MSS) if no length difference is observed. In addition, there are NGS methods that specifically target only microsatellite regions and that evaluate MMR function as part of comprehensive cancer genome profiling approaches. When targeting microsatellite regions only, the length of a total of 18 microsatellite marker regions is measured through NGS, and MSI-H is diagnosed when 33% or more of the markers present instability [18].

Regarding MMR status, it has been reported [48] that there is a concordance rate of 90% or higher between IHC staining and MSI testing in colorectal cancer; however, another report [49] has suggested lower concordance rates in other types of cancer. In the evaluation of immunohistochemistry staining and MSI testing for endometrial cancer, the overall concordance rate was 93.3% and, in cases that were discordant, the reason was promotor hypermethylation of MLH1 [50]. Moreover, in endometrial cancer, although specific discrepancies are observed in the dMMR sub-group, IHC results are considered a better predictive factor for MMR gene status than determination using PCR [51].

Although IHC was performed for assessment in the current study, it is important to recognize that MSI testing has limitations that should be understood. When the DNA extraction quantity is low or the DNA quality is poor, there is a 14% probability that the test cannot provide an accurate evaluation. Furthermore, if the purity of the tumor cells in the sample is less than 30%, the results are likely to be false negative [52].

We previously constructed a model using the TCGA data set annotated with MSI, following a similar approach to that used in the present study; however, the accuracy of determination was low (data not shown). Therefore, while there is a certain degree of agreement between the assessment methods, differences in assessment methods could lead to variations in the obtained results.

While most previous investigations in medical imaging classification have used CNNs, a combined analysis of the PORTEC randomized trial and a clinical cohort conducted by Fremont et al. [30] used attention-based models for class classification. Traditionally, CNNs have been widely used for image classification tasks; however, the introduction of the attention mechanism [53] has allowed for more accurate execution. CNNs capture relationships between adjacent pixels in images and recognize the content being displayed through structures called convolutional layers in the architecture. However, a disadvantage of CNNs is that they are influenced by elements other than the intended target, such as background objects.

On the other hand, the attention mechanism, originally developed primarily for natural language processing, has also been proven to be useful in the field of image recognition. In the task of image recognition, a technology derived from Transformers [54] incorporating attention, known as Vision Transformer, has emerged [44]. ViT models are capable of visualizing how much attention is paid to which areas within an image. Unlike CNNs, pure ViT does not include convolutional structures and is composed solely of the attention mechanism, although it can also be used in conjunction with CNNs. Through identifying areas of interest within images using the attention mechanism, we believe that the accuracy of recognition can be improved, thus addressing the disadvantage of CNNs when combined with attention mechanisms.

A further structural difference is that CNNs rely on fixed local receptive fields in the early layers, while ViTs use self-attention to aggregate global information in the early layers [55]. We compared the performance of ViT and an original constructed API-Net-based model incorporating a CNN

structure internally, as a model utilizing an attention mechanism. In this study, for the models utilizing the attention mechanism, the accuracy of ViT on the test data set was lower compared to the other networks (Table 2). The reason for the lower accuracy of ViT could be attributed to the fact that, unlike other networks, ViT does not use convolution in its internal structure. This suggests that incorporating a CNN architecture could be beneficial for pathological image diagnosis of endometrial cancer. Additionally, in the comparison of CNNs, ResNet showed higher accuracy. Considering that ResNet is used within the API-Net-based model as the classification backbone, ResNet can be considered highly useful in this context. The potential for further performance improvement through combining CNNs with attention mechanisms may be considered for the molecular classification of cancers.

H&E-stained slides are the most widely used method in the clinical context for pathologists to confirm the histological type of endometrial cancer. In this study, we confirmed that the MMR status of endometrial cancer could be predicted from H&E-stained slides using deep learning. To the best of our knowledge, there are only a few studies [27,28,30,56] worldwide that have tested this concept in endometrial cancer. ResNet, which performed particularly well in the present study, has also been used in several previous studies [27,30,31,34,56] in the field of medical imaging, although the number of layers and target organs differed.

For example, in colorectal cancer, the use of artificial intelligence (AI) in the colorectal cancer diagnostic algorithm is expected to reduce testing costs and avoid treatment-related expenses [57]. A strategy using high-sensitivity AI followed by a high-specificity panel is expected to achieve the most significant cost reduction (about USD 400 million, or 12.9%), compared to a strategy using NGS alone [57]. Meanwhile, a strategy using only high-specificity AI may achieve the highest diagnostic accuracy (97%) and the shortest time to initiation of treatment [57]. This report was based on cost assumptions for colorectal cancer from 2017 to 2020 in the United States. Although it is necessary to assess how much of a cost reduction can be achieved in other contexts, the use of a similar approach for endometrial cancer may have the potential to save time and costs. Additionally, the use of AI-based approaches to assist the decision making of oncologists in treating cancer has the potential to allow optimal treatment to be provided to cancer patients sooner [58].

## 5. Conclusions

Molecular classification has been playing an increasingly vital role in treatment strategies for endometrial cancer. Therefore, it is crucial to incorporate additional user-friendly screening tools that can identify patients requiring further laboratory testing, thereby saving both time and costs. Our study demonstrated the potential of developing AI-based solutions, which are capable of easily and rapidly estimating molecular classifications from H&E-stained slides in clinical practice.

**Supplementary Materials:** The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Figure S1: Overview of process for excluding tiles and examples of excluded tiles and eligible tiles. Figure S2: Performance of all models on test data set at per-tile level. Table S1: Number of tiles and characteristics for each patient. Table S2: Number of tiles regarding different data set ratios for mismatch repair (MMR) status prediction.

**Author Contributions:** Conceptualization, Mina Umemoto and Tasuku Mariya; Data curation, Mina Umemoto; Funding acquisition, Tasuku Mariya; Investigation, Mina Umemoto, Tasuku Mariya, Shintaro Sugita, Takayuki Kanaseki, Yuka Takenaka and Shota Shinkai; Methodology, Mina Umemoto and Tasuku Mariya; Project administration, Tasuku Mariya; Software, Yuta Nambu, Mai Nagata, Toshihiro Horimai and Shota Shinkai; Supervision, Tasuku Mariya, Motoki Matsuura, Masahiro Iwasaki, Yoshihiko Hirohashi, Tadashi Hasegawa, Toshihiko Torigoe, Yuichi Fujino and Tsuyoshi Saito; Validation, Mina Umemoto, Tasuku Mariya, Yuta Nambu, Mai Nagata and Toshihiro Horimai; Visualization, Mina Umemoto; Writing – original draft, Mina Umemoto; Writing – review & editing, Tasuku Mariya.

**Funding:** This research was funded by the Northern Advancement Center for Science & Technology.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of The Sapporo Medical University Hospital (332-158, 13 January 2022).

**Informed Consent Statement:** Informed consent was acquired through an opt-out form on the website.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Crosbie, E.J.; Kitson, S.J.; McAlpine, J.N.; Mukhopadhyay, A.; Powell, M.E.; Singh, N. Endometrial cancer. *Lancet* **2022**, *399*, 1412-1428, doi:10.1016/S0140-6736(22)00323-3.
2. Bokhman, J.V. Two pathogenetic types of endometrial carcinoma. *Gynecol Oncol* **1983**, *15*, 10-17, doi:10.1016/0090-8258(83)90111-7.
3. Sherman, M.E. Theories of endometrial carcinogenesis: a multidisciplinary approach. *Mod Pathol* **2000**, *13*, 295-308, doi:10.1038/modpathol.3880051.
4. Setiawan, V.W.; Yang, H.P.; Pike, M.C.; McCann, S.E.; Yu, H.; Xiang, Y.B.; Wolk, A.; Wentzensen, N.; Weiss, N.S.; Webb, P.M.; et al. Type I and II endometrial cancers: have they different risk factors? *J Clin Oncol* **2013**, *31*, 2607-2618, doi:10.1200/jco.2012.48.2596.
5. Levine, D.A.; Getz, G.; Gabriel, S.B.; Cibulskis, K.; Lander, E.; Sivachenko, A.; Sougnez, C.; Lawrence, M.; Kandoth, C.; Dooling, D.; et al. Integrated genomic characterization of endometrial carcinoma. *Nature* **2013**, *497*, 67-73, doi:10.1038/nature12113.
6. Talhouk, A.; McConechy, M.K.; Leung, S.; Yang, W.; Lum, A.; Senz, J.; Boyd, N.; Pike, J.; Anglesio, M.; Kwon, J.S.; et al. Confirmation of ProMisE: A simple, genomics-based clinical classifier for endometrial cancer. *Cancer* **2017**, *123*, 802-813, doi:10.1002/cncr.30496.
7. Talhouk, A.; McConechy, M.K.; Leung, S.; Li-Chang, H.H.; Kwon, J.S.; Melnyk, N.; Yang, W.; Senz, J.; Boyd, N.; Karnezis, A.N.; et al. A clinically applicable molecular-based classification for endometrial cancers. *Br J Cancer* **2015**, *113*, 299-310, doi:10.1038/bjc.2015.190.
8. board, W.H.O.c.o.t.e.; World Health, O. *Female genital tumours*; World Health Organization, International Agency for Research on Cancer: 2020.
9. Vermij, L.; Smit, V.; Nout, R.; Bosse, T. Incorporation of molecular characteristics into endometrial cancer management. *Histopathology* **2020**, *76*, 52-63, doi:10.1111/his.14015.
10. Berek, J.S.; Matias-Guiu, X.; Creutzberg, C.; Fotopoulou, C.; Gaffney, D.; Kehoe, S.; Lindemann, K.; Mutch, D.; Concin, N. FIGO staging of endometrial cancer: 2023. *Int J Gynaecol Obstet* **2023**, doi:10.1002/ijgo.14923.
11. Shiravand, Y.; Khodadadi, F.; Kashani, S.M.A.; Hosseini-Fard, S.R.; Hosseini, S.; Sadeghirad, H.; Ladwa, R.; O'Byrne, K.; Kulasinghe, A. Immune Checkpoint Inhibitors in Cancer Therapy. *Curr Oncol* **2022**, *29*, 3044-3060, doi:10.3390/curroncol29050247.
12. Freeman, G.J.; Long, A.J.; Iwai, Y.; Bourque, K.; Chernova, T.; Nishimura, H.; Fitz, L.J.; Malenkovich, N.; Okazaki, T.; Byrne, M.C.; et al. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J Exp Med* **2000**, *192*, 1027-1034, doi:10.1084/jem.192.7.1027.
13. Pirš, B.; Škof, E.; Smrkolj, V.; Smrkolj, Š. Overview of Immune Checkpoint Inhibitors in Gynecological Cancer Treatment. *Cancers (Basel)* **2022**, *14*, doi:10.3390/cancers14030631.
14. Le, D.T.; Durham, J.N.; Smith, K.N.; Wang, H.; Bartlett, B.R.; Aulakh, L.K.; Lu, S.; Kemberling, H.; Wilt, C.; Luber, B.S.; et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **2017**, *357*, 409-413, doi:10.1126/science.aan6733.
15. Boland, C.R.; Thibodeau, S.N.; Hamilton, S.R.; Sidransky, D.; Eshleman, J.R.; Burt, R.W.; Meltzer, S.J.; Rodriguez-Bigas, M.A.; Fodde, R.; Ranzani, G.N.; 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 Res* **1998**, *58*, 5248-5257.
16. Umar, A.; Boland, C.R.; Terdiman, J.P.; Syngal, S.; de la Chapelle, A.; Rüschoff, J.; Fishel, R.; Lindor, N.M.; Burgart, L.J.; Hamelin, R.; et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* **2004**, *96*, 261-268, doi:10.1093/jnci/djh034.
17. Shia, J.; Holck, S.; Depetris, G.; Greenson, J.K.; Klimstra, D.S. Lynch syndrome-associated neoplasms: a discussion on histopathology and immunohistochemistry. *Fam Cancer* **2013**, *12*, 241-260, doi:10.1007/s10689-013-9612-4.

18. Hempelmann, J.A.; Scroggins, S.M.; Pritchard, C.C.; Salipante, S.J. MSIplus for Integrated Colorectal Cancer Molecular Testing by Next-Generation Sequencing. *J Mol Diagn* **2015**, *17*, 705-714, doi:10.1016/j.jmoldx.2015.05.008.
19. Frampton, G.M.; Fichtenholtz, A.; Otto, G.A.; Wang, K.; Downing, S.R.; He, J.; Schnall-Levin, M.; White, J.; Sanford, E.M.; An, P.; et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol* **2013**, *31*, 1023-1031, doi:10.1038/nbt.2696.
20. Lainé, A.; Gonzalez-Lopez, A.M.; Hasan, U.; Ohkuma, R.; Ray-Coquard, I. Immune Environment and Immunotherapy in Endometrial Carcinoma and Cervical Tumors. *Cancers (Basel)* **2023**, *15*, doi:10.3390/cancers15072042.
21. Janiesch, C.; Zschech, P.; Heinrich, K. *Machine learning and deep learning*; 2021.
22. Sarraf, S.; Tofighi, G. Classification of Alzheimer's Disease using fMRI Data and Deep Learning Convolutional Neural Networks. **2016**, arXiv:1603.08631, doi:10.48550/arXiv.1603.08631.
23. Gao, X.; Lin, S.; Wong, T.Y. Automatic Feature Learning to Grade Nuclear Cataracts Based on Deep Learning. *IEEE transactions on bio-medical engineering* **2015**, *62*, doi:10.1109/TBME.2015.2444389.
24. Anthimopoulos, M.; Christodoulidis, S.; Ebner, L.; Christe, A.; Mougiakakou, S. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Trans Med Imaging* **2016**, *35*, 1207-1216, doi:10.1109/tmi.2016.2535865.
25. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.; van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med Image Anal* **2017**, *42*, 60-88, doi:10.1016/j.media.2017.07.005.
26. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data* **2021**, *8*, 53, doi:10.1186/s40537-021-00444-8.
27. Zhang, Y.; Chen, S.; Wang, Y.; Li, J.; Xu, K.; Chen, J.; Zhao, J. Deep learning-based methods for classification of microsatellite instability in endometrial cancer from HE-stained pathological images. *J Cancer Res Clin Oncol* **2023**, *149*, 8877-8888, doi:10.1007/s00432-023-04838-4.
28. Hong, R.; Liu, W.; DeLair, D.; Razavian, N.; Fenyo, D. Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models. *Cell Rep Med* **2021**, *2*, 100400, doi:10.1016/j.xcrm.2021.100400.
29. Naik, N.; Madani, A.; Esteva, A.; Keskar, N.S.; Press, M.F.; Ruderman, D.; Agus, D.B.; Socher, R. Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains. *Nat Commun* **2020**, *11*, 5727, doi:10.1038/s41467-020-19334-3.
30. Fremont, S.; Andani, S.; Barkey Wolf, J.; Dijkstra, J.; Melsbach, S.; Jobsen, J.J.; Brinkhuis, M.; Roothaan, S.; Jurgenliemk-Schulz, I.; Lutgens, L.; et al. Interpretable deep learning model to predict the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images: a combined analysis of the PORTEC randomised trials and clinical cohorts. *Lancet Digit Health* **2023**, *5*, e71-e82, doi:10.1016/S2589-7500(22)00210-2.
31. Kather, J.N.; Pearson, A.T.; Halama, N.; Jager, D.; Krause, J.; Loosen, S.H.; Marx, A.; Boor, P.; Tacke, F.; Neumann, U.P.; et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med* **2019**, *25*, 1054-1056, doi:10.1038/s41591-019-0462-y.
32. Saillard, C.; Dubois, R.; Tchita, O.; Loiseau, N.; Garcia, T.; Adriansen, A.; Carpentier, S.; Reyre, J.; Enea, D.; von Loga, K.; et al. Validation of MSIntuit as an AI-based pre-screening tool for MSI detection from colorectal cancer histology slides. *Nat Commun* **2023**, *14*, 6695, doi:10.1038/s41467-023-42453-6.
33. Yamashita, R.; Long, J.; Longacre, T.; Peng, L.; Berry, G.; Martin, B.; Higgins, J.; Rubin, D.L.; Shen, J. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *Lancet Oncol* **2021**, *22*, 132-141, doi:10.1016/s1470-2045(20)30535-0.
34. Echle, A.; Ghaffari Laleh, N.; Quirke, P.; Grabsch, H.I.; Muti, H.S.; Saldanha, O.L.; Brockmoeller, S.F.; van den Brandt, P.A.; Hutchins, G.G.A.; Richman, S.D.; et al. Artificial intelligence for detection of microsatellite instability in colorectal cancer-a multicentric analysis of a pre-screening tool for clinical application. *ESMO Open* **2022**, *7*, 100400, doi:10.1016/j.esmoop.2022.100400.
35. Sirinukunwattana, K.; Domingo, E.; Richman, S.D.; Redmond, K.L.; Blake, A.; Verrill, C.; Leedham, S.J.; Chatzipili, A.; Hardy, C.; Whalley, C.M.; et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* **2021**, *70*, 544-554, doi:10.1136/gutjnl-2019-319866.

36. Kim, K.; Lee, K.; Cho, S.; Kang, D.U.; Park, S.; Kang, Y.; Kim, H.; Choe, G.; Moon, K.C.; Lee, K.S.; et al. PAIP 2020: Microsatellite instability prediction in colorectal cancer. *Med Image Anal* **2023**, *89*, 102886, doi:10.1016/j.media.2023.102886.
37. Li, M.; Jiang, Y.; Zhang, Y.; Zhu, H. Medical image analysis using deep learning algorithms. *Front Public Health* **2023**, *11*, 1273253, doi:10.3389/fpubh.2023.1273253.
38. Shia, J.; Tang, L.H.; Vakiani, E.; Guillem, J.G.; Stadler, Z.K.; Soslow, R.A.; Katabi, N.; Weiser, M.R.; Paty, P.B.; Temple, L.K.; et al. Immunohistochemistry as first-line screening for detecting colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome: a 2-antibody panel may be as predictive as a 4-antibody panel. *Am J Surg Pathol* **2009**, *33*, 1639-1645, doi:10.1097/PAS.0b013e3181b15aa2.
39. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. *Going deeper with convolutions*; 2015; pp. 1-9.
40. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **1409.1556** **2014**.
41. He, K.; Zhang, X.; Ren, S.; Sun, J. *Deep Residual Learning for Image Recognition*; 2016; pp. 770-778.
42. Zagoruyko, S.; Komodakis, N. Wide Residual Networks. **2016**.
43. Tan, M.; Le, Q. *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*; 2019.
44. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*; 2020.
45. Zhuang, P.; Wang, Y.; Qiao, Y. Learning Attentive Pairwise Interaction for Fine-Grained Classification. *Proceedings of the AAAI Conference on Artificial Intelligence* **2020**, *34*, 13130-13137, doi:10.1609/aaai.v34i07.7016.
46. Gu, B.; Shang, X.; Yan, M.; Li, X.; Wang, W.; Wang, Q.; Zhang, C. Variations in incidence and mortality rates of endometrial cancer at the global, regional, and national levels, 1990-2019. *Gynecol Oncol* **2021**, *161*, 573-580, doi:10.1016/j.ygyno.2021.01.036.
47. Consortium, R.R. Refining adjuvant treatment in endometrial cancer based on molecular features: the RAINBO clinical trial program. *Int J Gynecol Cancer* **2022**, *33*, 109-117, doi:10.1136/ijgc-2022-004039.
48. Loughrey, M.B.; McGrath, J.; Coleman, H.G.; Bankhead, P.; Maxwell, P.; McGready, C.; Bingham, V.; Humphries, M.P.; Craig, S.G.; McQuaid, S.; et al. Identifying mismatch repair-deficient colon cancer: near-perfect concordance between immunohistochemistry and microsatellite instability testing in a large, population-based series. *Histopathology* **2021**, *78*, 401-413, doi:10.1111/his.14233.
49. Wang, Y.; Shi, C.; Eisenberg, R.; Vnencak-Jones, C.L. Differences in Microsatellite Instability Profiles between Endometrioid and Colorectal Cancers: A Potential Cause for False-Negative Results? *J Mol Diagn* **2017**, *19*, 57-64, doi:10.1016/j.jmoldx.2016.07.008.
50. McConechy, M.K.; Talhouk, A.; Li-Chang, H.H.; Leung, S.; Huntsman, D.G.; Gilks, C.B.; McAlpine, J.N. Detection of DNA mismatch repair (MMR) deficiencies by immunohistochemistry can effectively diagnose the microsatellite instability (MSI) phenotype in endometrial carcinomas. *Gynecol Oncol* **2015**, *137*, 306-310, doi:10.1016/j.ygyno.2015.01.541.
51. Vermij, L.; León-Castillo, A.; Singh, N.; Powell, M.E.; Edmondson, R.J.; Genestie, C.; Khaw, P.; Pyman, J.; McLachlin, C.M.; Ghatage, P.; et al. p53 immunohistochemistry in endometrial cancer: clinical and molecular correlates in the PORTEC-3 trial. *Mod Pathol* **2022**, *35*, 1475-1483, doi:10.1038/s41379-022-01102-x.
52. Da Cruz Paula, A.; DeLair, D.F.; Ferrando, L.; Fix, D.J.; Soslow, R.A.; Park, K.J.; Chiang, S.; Reis-Filho, J.S.; Zehir, A.; Donoghue, M.T.A.; et al. Genetic and molecular subtype heterogeneity in newly diagnosed early- and advanced-stage endometrial cancer. *Gynecol Oncol* **2021**, *161*, 535-544, doi:10.1016/j.ygyno.2021.02.015.
53. Zhang, Y.; Wang, J.; Gorriz, J.M.; Wang, S. Deep Learning and Vision Transformer for Medical Image Analysis. *J Imaging* **2023**, *9*, doi:10.3390/jimaging9070147.
54. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. **2017**.
55. Raghu, M.; Unterthiner, T.; Kornblith, S.; Zhang, C.; Dosovitskiy, A. *Do Vision Transformers See Like Convolutional Neural Networks?*; 2021.
56. Wang, T.; Lu, W.; Yang, F.; Liu, L.; Dong, Z.; Tang, W.; Chang, J.; Huan, W.; Huang, K.; Yao, J. *Microsatellite Instability Prediction of Uterine Corpus Endometrial Carcinoma Based on H&E Histology Whole-Slide Imaging*; 2020.

57. Kacew, A.J.; Strohhahn, G.W.; Saulsberry, L.; Laiteerapong, N.; Cipriani, N.A.; Kather, J.N.; Pearson, A.T. Artificial Intelligence Can Cut Costs While Maintaining Accuracy in Colorectal Cancer Genotyping. *Front Oncol* **2021**, *11*, 630953, doi:10.3389/fonc.2021.630953.
58. Tran, K.A.; Kondrashova, O.; Bradley, A.; Williams, E.D.; Pearson, J.V.; Waddell, N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med* **2021**, *13*, 152, doi:10.1186/s13073-021-00968-x.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.