# 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

**Journal Article**

**Author(s):**
Fremond, Sarah; Andani, Sonali; Wolf, Jurriaan Barkey; Dijkstra, Jouke; Melsbach, Sinéad; Jobsen, Jan J.; Brinkhuis, Mariel; Roothaan, Suzan; Jurgenliemk-Schulz, Ina; Lutgens, Ludy C.H.W.; Nout, Remi A .; van der Steen-Banasik, Elzbieta M.; de Boer, Stephanie M.; Singh, Naveena; Mileshkin, Linda R.; Mackay, Helen J.; Leary, Alexandra; Nijman, Hans W.; Smit, Vincent T.H.B.M; Creutzberg, Carien L.; Horeweg, Nanda; Koelzer, Viktor H.; Bosse, Tjalling

**Publication date:**
2023-02

**Permanent link:**
https://doi.org/10.3929/ethz-b-000595178

**Originally published in:**
The Lancet Digital Health 5(2), https://doi.org/10.1016/S2589-7500(22)00210-2

# 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

*Sarah Fremond*, Sonali Andani*, Jurriaan Barkey Wolf, Jouke Dijkstra, Sinéad Melsbach, Jan J Jobsen, Mariel Brinkhuis, Suzan Roothaan, Ina Jurgenliemk-Schulz, Ludy C H W Lutgens, Remi A Nout, Elzbieta M van der Steen-Banasik, Stephanie M de Boer, Melanie E Powell, Naveena Singh, Linda R Mileshkin, Helen J Mackay, Alexandra Leary, Hans W Nijman, Vincent T H B M Smit, Carien L Creutzberg, Nanda Horeweg, Viktor H Koelzer†, Tjalling Bosse†*

## Summary

**Background** Endometrial cancer can be molecularly classified into *POLE*$^{mut}$, mismatch repair deficient (MMRd), p53 abnormal (p53abn), and no specific molecular profile (NSMP) subgroups. We aimed to develop an interpretable deep learning pipeline for whole-slide-image-based prediction of the four molecular classes in endometrial cancer (im4MEC), to identify morpho-molecular correlates, and to refine prognostication.

**Methods** This combined analysis included diagnostic haematoxylin and eosin-stained slides and molecular and clinicopathological data from 2028 patients with intermediate-to-high-risk endometrial cancer from the PORTEC-1 (n=466), PORTEC-2 (n=375), and PORTEC-3 (n=393) randomised trials and the TransPORTEC pilot study (n=110), the Medisch Spectrum Twente cohort (n=242), a case series of patients with *POLE*$^{mut}$ endometrial cancer in the Leiden Endometrial Cancer Repository (n=47), and The Cancer Genome Atlas-Uterine Corpus Endometrial Carcinoma cohort (n=395). PORTEC-3 was held out as an independent test set and a four-fold cross validation was performed. Performance was measured with the macro and class-wise area under the receiver operating characteristic curve (AUROC). Whole-slide images were segmented into tiles of 360 µm resized to 224×224 pixels. im4MEC was trained to learn tile-level morphological features with self-supervised learning and to molecularly classify whole-slide images with an attention mechanism. The top 20 tiles with the highest attention scores were reviewed to identify morpho-molecular correlates. Predictions of a nuclear classification deep learning model serve to derive interpretable morphological features. We analysed 5-year recurrence-free survival and explored prognostic refinement by molecular class using the Kaplan-Meier method.

**Findings** im4MEC attained macro-average AUROCs of 0·874 (95% CI 0·856–0·893) on four-fold cross-validation and 0·876 on the independent test set. The class-wise AUROCs were 0·849 for *POLE*$^{mut}$ (n=51), 0·844 for MMRd (n=134), 0·883 for NSMP (n=120), and 0·928 for p53abn (n=88). *POLE*$^{mut}$ and MMRd tiles had a high density of lymphocytes, p53abn tiles had strong nuclear atypia, and the morphology of *POLE*$^{mut}$ and MMRd endometrial cancer overlapped. im4MEC highlighted a low tumour-to-stroma ratio as a potentially novel characteristic feature of the NSMP class. 5-year recurrence-free survival was significantly different between im4MEC predicted molecular classes in PORTEC-3 (log-rank p<0·0001). The ten patients with aggressive p53abn endometrial cancer that was predicted as MMRd showed inflammatory morphology and appeared to have a better prognosis than patients with correctly predicted p53abn endometrial cancer (p=0·30). The four patients with NSMP endometrial cancer that was predicted as p53abn showed higher nuclear atypia and appeared to have a worse prognosis than patients with correctly predicted NSMP (p=0·13). Patients with MMRd endometrial cancer predicted as *POLE*$^{mut}$ had an excellent prognosis, as do those with true *POLE*$^{mut}$ endometrial cancer.

**Interpretation** We present the first interpretable deep learning model, im4MEC, for haematoxylin and eosin-based prediction of molecular endometrial cancer classification. im4MEC robustly identified morpho-molecular correlates and could enable further prognostic refinement of patients with endometrial cancer.

**Funding** The Hanarth Foundation, the Promedica Foundation, and the Swiss Federal Institutes of Technology.



## Introduction

Endorsed by the fifth edition (2020) of the WHO classification of female genital tumours[1] and 2021 guidelines from the European Society of Gynaecological Oncology, the European Society for Radiotherapy and Oncology, and the European Society of Pathology,[2] the introduction

*Lancet Digit Health* 2023; 5: e71–82

Published Online December 7, 2022 https://doi.org/10.1016/S2589-7500(22)00210-2

*Contributed equally and share first authorship

†Contributed equally and share senior authorship

**Department of Pathology** (S Fremond MSc, J Barkey Wolf MSc, S Melsbach MSc, Prof V T H B M Smit PhD, T Bosse PhD), **Department of Vascular and Molecular Imaging** (J Dijkstra PhD), **and Department of Radiation Oncology** (S M de Boer PhD, Prof C L Creutzberg PhD, N Horeweg PhD), **Leiden University Medical Center, Leiden, Netherlands; Department of Computer Science, ETH Zurich, Zurich, Switzerland** (S Andani MSc); **Department of Pathology and Molecular Pathology, University Hospital Zurich, University of Zurich, Zurich, Switzerland** (S Andani, Prof V H Koelzer MD); **Swiss Institute of Bioinformatics, Lausanne, Switzerland** (S Andani); **Department of Radiation Oncology, Medisch Spectrum Twente, Enschede, Netherlands** (J J Jobsen PhD); **Department of Pathology, LabPON, Hengelo, Netherlands** (M Brinkhuis MD, S Roothaan MD); **Department of Radiation Oncology, University Medical Center Utrecht, Utrecht, Netherlands** (I Jurgenliemk-Schulz PhD); **Department of Radiation Oncology, Maastricht University Medical Center+, Maastricht, Netherlands** (L C H W Lutgens PhD);

**Department of Radiation Oncology, Erasmus University Medical Center, Rotterdam, Netherlands** (Prof R A Nout PhD); **Department of Radiation Oncology, Radiotherapiegroep, Arnhem, Netherlands** (E M van der Steen-Banasik MD); **Department of Clinical Oncology** (Prof M E Powell PhD) **and Department of Pathology** (N Singh PhD), **Barts Health NHS Trust, London, UK; Department of Medical Oncology, Peter MacCallum Cancer Center, Melbourne, VIC, Australia** (Prof L R Mileshkin PhD); **Department of Medical Oncology and Hematology, Odette Cancer Center Sunnybrook Health Sciences Center, Toronto, ON, Canada** (Prof H J Mackay PhD); **Medical Oncology Department, Gustave Roussy Institute, Villejuif, France** (A Leary PhD); **Department of Obstetrics and Gynecology, University Medical Center Groningen, Groningen, Netherlands** (Prof H W Nijman PhD)

Correspondence to:
Dr Tjalling Bosse, Department of Pathology, Leiden University Medical Center, Leiden 2300 RC, Netherlands
**t.bosse@lumc.nl**

or

Prof Viktor H Koelzer, Department of Pathology and Molecular Pathology, University Hospital Zurich, University of Zurich, Zurich 8091, Switzerland
**viktor.koelzer@usz.ch**

## Research in context

### Evidence before this study

We searched Google Scholar for papers published from database inception to June 13, 2022, using the search terms ("deep learning" OR "artificial intelligence" OR "AI") AND ("histopathology" OR "histology" OR "slide image") AND ("molecular" OR "genotype" OR "phenotype"), without any date or language restrictions. This search returned 45 500 articles. There is an increasing number of proof-of-concept studies in various cancer types of deep learning models that can predict genetic mutations or molecular alterations by recognising associated morphology on haematoxylin and eosin-stained tumour whole-slide images. In the endometrial cancer field, only two studies have used deep learning for a haematoxylin and eosin-based molecular class prediction task on small cohorts of 456 or 516 patients, with no or very few samples including two histological subtypes for independent testing. Both studies provided insufficient development towards clinical translation for several reasons: the authors did not consider a four-class classification problem but rather one or multiple oversimplified binary tasks; molecular classes were defined by The Cancer Genome Atlas and not by the surrogate marker approach used in endometrial cancer diagnostics; the authors did not have access to structured survival outcome data; and interpretability was missing or limited to visualising image datapoints in a two-dimensional map. Additionally, in both deep learning models, molecular classification was done at the tile level instead of the whole-slide image level.

### Added value of this study

To our knowledge, this study is the first to use a large-scale haematoxylin and eosin-stained whole-slide-image dataset of 2028 patients with endometrial cancer from three randomised trials and four clinical cohorts to develop an interpretable, deep learning-based model, im4MEC, tailored towards predicting the four-class molecular classification of endometrial cancer. The im4MEC pipeline is novel as it leverages self-supervised learning to learn endometrial-cancer-specific morphological feature representation of tiles in whole-slide images along with an attention-based classification model for whole-slide-image-level prediction. The attention mechanism enabled identification of the most predictive morphological regions and downstream interpretability by cell-level analysis. These design choices helped in the establishment of known and novel morpho-molecular correlates for each molecular class, which were shown to impact clinical outcome. The use of whole hysterectomy specimen slides without human-annotated regions of interest supported the discovery of relevant molecular-class-specific features outside of the tumour region. Future deep-learning-based studies should consider similar unbiased approaches, as novel and potentially relevant features might lie outside of the tumour area.

### Implications of all the available evidence

Our study, with its state-of-the-art model performance and interpretable results, suggests that a deep learning-based model can have a relevant role in endometrial cancer diagnostics. The identification of morpho-molecular correlates and their impact on prognosis advances the evidence towards building an improved risk stratification system in endometrial cancer and unifying the molecular-driven and morphology-driven classification systems. Future work should focus on prospective validation, the further exploration of tumour heterogeneity in endometrial cancer, solving the current difficulty in distinguishing *POLE*$^{mut}$ from mismatch-repair-deficient endometrial cancer, and reproducibility on biopsy specimens.

of the molecular classification has had major implications for the diagnosis and clinical management of endometrial cancer. The new molecularly driven diagnostic algorithm uses clinically applicable surrogate markers for the four molecular classes[3–5] described by The Cancer Genome Atlas (TCGA) Research Network in 2013.[6] *POLE* mutational status is assessed by use of targeted DNA sequencing of exons 9–14 (*POLE*$^{mut}$). Loss of expression of mismatch repair proteins (MMR deficient [MMRd]) by immunohistochemistry is an excellent surrogate for microsatellite instability. Tumours with a high copy number are characterised by *TP53* mutations and can be identified by abnormal, mutant-like cellular tumour antigen p53 expression (p53abn). Tumours with a low copy number are referred to as having no specific molecular profile (NSMP). Endometrial cancers with more than one classifying feature should be classified according to their dominant molecular class by sequential testing of *POLE*, MMR, and p53 status.[1] The molecular endometrial cancer classification has independent prognostic impact: patients with *POLE*$^{mut}$ endometrial cancer have an excellent prognosis, whereas patients with p53abn endometrial cancer have a poor prognosis and patients with MMRd or NSMP endometrial cancer have (stage-dependent) intermediate clinical outcomes.[3–5]

The molecular-based and histology-based classification systems of endometrial cancer are still reported independently because histological subtypes and grades only partly overlap with molecular classes.[3–5] This separation has introduced some controversies in endometrial cancer diagnostics, as molecular classification is now recommended over histological classification, especially for patients considered to have a high risk of recurrence on the basis of clinicopathological factors.[1,2] To unify both classification systems, it is necessary to get an accurate and evidence-based representation of the morphology associated with each molecular endometrial cancer class. Furthermore, improving clinical risk stratification might be possible

when incorporating tumour microenvironmental immune-related features, which have been shown to carry prognostic value independent of molecular class;[7,8] however, how to best capture these features is not currently clear. Establishing the complete set of endometrial cancer morpho-molecular correlates and discovering those that are clinically relevant will thus deepen current understanding of the molecular classification of endometrial cancer and provide opportunities for prognostic refinement.

Deep learning has shown impressive ability to associate complex visual features of tissue organisation with molecular disease characteristics,[9–11] to predict treatment response,[12] and to enable screening for hereditary cancer syndromes[13] from clinically established and cost-effective haematoxylin and eosin-stained whole-slide images of tumours. Two studies have shown proof of concept for the binary prediction of molecular alterations in endometrial cancer from histopathology images.[14,15] Hong and colleagues[14] considered one-versus-all binary classes and reported area under the receiver operating characteristic curve (AUROC) values of 0·66 for *POLE*$^{mut}$, 0·76 for microsatellite instability, 0·87 for high copy number, and 0·65 for low copy number on a small independent test set (n=95) and Wang and colleagues[15] predicted microsatellite instability status with an AUROC of 0·73 on a TCGA internal test set (n=155). However, their deep learning models did not predict the four-class molecular classification of endometrial cancer, had limited interpretability, and did not relate predictions to survival outcomes.

To apply deep learning in pathology, gigapixel whole-slide images are often partitioned into smaller regions called tiles. Many deep learning models rely on tile-based, weakly supervised learning by assigning a molecular class to each tile during training and then aggregating tiles to obtain a prediction at the whole-slide-image level.[9–11,14–16] This approach is built on the assumption that each tile contributes equally towards the prediction at the whole-slide-image level, which is sensitive to spatial morphological heterogeneity within the whole-slide image. To address this limitation, Lu and colleagues[17] proposed a deep learning model with an attention mechanism that identifies tiles containing the most predictive morphological information to classify the whole-slide image. In addition to improving classification performance, an attention-based model can produce interpretable attention heatmaps that highlight the relative importance of tissue subregions for the whole-slide image classification. These deep learning architectures use a deep learning model that represents histology tile images as feature vectors, referred to as a feature extractor model, which is often trained on the non-histology-related ImageNet dataset.[18] Instead, self-supervised learning possesses great potential for training feature extractor models that outperform ImageNet-based models in identifying tile-level, cancer-specific morphological features when trained on large-scale and heterogeneous histology datasets.[19,20]

In this study, we aimed to develop a deep learning pipeline (im4MEC) for the image-based prediction of the molecular classification of endometrial cancer from haematoxylin and eosin-stained whole-slide images and for the identification of known and novel human-interpretable morpho-molecular correlates and the investigation of their prognostic relevance. To this end, we leveraged self-supervised learning to extract endometrial-cancer-specific morphological features, and trained an interpretable attention-based classification model using data from the three PORTEC randomised trials and four clinical cohorts.

## Methods

### Study design and participants

In this combined analysis, we used formalin-fixed paraffin-embedded tumour material and complete molecular and clinicopathological data from three randomised trials and four clinical cohorts (appendix 1 pp 2–4): the randomised PORTEC-1 trial[21] (714 patients with early-stage, intermediate-risk endometrial cancer, recruited in the Netherlands); the randomised PORTEC-2 trial[22] (427 patients with early-stage, high-to-intermediate-risk endometrial cancer; the Netherlands); the randomised PORTEC-3 trial[23] (660 patients with stage I–III, high-risk endometrial cancer; the Netherlands, UK, France, Italy, Canada, Australia, and New Zealand); the retrospective TransPORTEC pilot study[5] (116 patients with high-risk endometrial cancer; the Netherlands, UK, and France); the prospective Medisch Spectrum Twente (MST) cohort[24] (257 patients with high-risk endometrial cancer; the Netherlands); a case series of 48 patients with *POLE*$^{mut}$ endometrial cancer from the Leiden Endometrial Cancer Repository (the Netherlands); and TCGA-Uterine Corpus Endometrial Carcinoma cohort (TCGA-UCEC; n=529), extracted from the cBioPortal for Cancer Genomics.[6,25] Long-term follow-up is available for the PORTEC-1, PORTEC-2, and PORTEC-3 trials. For all patients, except those in TCGA-UCEC, molecular classification had been done by our group[3–5] following the surrogate marker-based diagnostic algorithm of WHO 2020,[1] resulting in *POLE*$^{mut}$, MMRd, p53abn, NSMP, or inconclusive molecular class.

See **Online** for appendix 1

For the **cBioPortal for Cancer Genomics** see https://www.cbioportal.org/

Of 2751 patients, we excluded 360 for not having available haematoxylin and eosin-stained tumour slides or material, 177 for having an inconclusive molecular class, and 60 for not having available hysterectomy specimens (appendix 1 p 5). For patients with a conclusive molecular class, one representative haematoxylin and eosin-stained slide of the endometrial cancer hysterectomy specimen was selected by a gynaecological pathologist (TB; appendix 1 p 2). Slides were digitised at 40× magnification on a 3DHISTECH P250 digital scanner (Budapest, Hungary; 0·19 µm/pixel) and a

3DHISTECH P1000 (0·24 µm/pixel). We excluded three patients due to persistent blurriness of their scans, 60 patients from TCGA-UCEC with low-magnification scans, 33 patients with slide images containing no tumour, and 30 patients (25 from TCGA-UCEC, two from PORTEC-1, two from PORTEC-2, and one from MST) with slide images of very poor tissue quality (appendix 1 p 5). 2028 patients with endometrial cancer were included in the final analysis.

The entire PORTEC-3 randomised trial of 393 patients with endometrial cancer, consisting of 51 patients with *POLE*$^{mut}$, 134 with MMRd, 120 with NSMP, and 88 with p53abn, was kept as an independent test set (appendix 1 p 6). All patients from the PORTEC-1 (n=466), PORTEC-2 (n=375), TransPORTEC pilot (n=110), MST (n=242), *POLE*$^{mut}$ Leiden Endometrial Cancer Repository (n=47), and TCGA-UCEC (n=395) cohorts (n=1635) were used to train the feature extractor model using self-supervised learning. To then train the classification model, TCGA-UCEC was excluded because molecular classification in TCGA-UCEC did not use the surrogate marker-based approach. Therefore, the supervised training set consisted of 1240 patients (141 with *POLE*$^{mut}$ endometrial cancer, 326 with MMRd, 611 with NSMP, and 162 with p53abn). We reduced class imbalance in training by randomly downsampling the 611 NSMP whole-slide images to match the number of 326 MMRd whole-slide images, stratified by cohort and scanner type (resulting in 955 whole-slide images). A k-fold cross-validation split was done, stratified by molecular class, cohort, and scanner type, to determine the best model configuration and show model robustness on multiple validation sets. Due to the relative scarcity of *POLE*$^{mut}$ endometrial cancer, cross-validation was limited to four folds. Mean and SD or 95% CI of the macro-average AUROC of the four-fold cross-validation experiments were calculated. In ablation studies, we examined whether the inclusion of the self-supervised learning model in the pipeline improved the mean macro-average AUROC compared with the use of ImageNet[18] pre-trained weights. The model configuration and hyperparameters that achieved the highest mean macro-average AUROC were subsequently used for final training on the combined folds of 955 whole-slide images. To measure final performance, the model was tested on the independent test set from PORTEC-3 (n=393) and the macro-average and class-wise AUROC, precision, recall, F1 score, accuracy, specificity, negative predictive value, and the confusion matrix were calculated.

### Deep learning pipeline

The proposed im4MEC deep learning model (figure 1) consists of: pre-processing the whole-slide image (figure 1A); training the feature extractor model with contrastive self-supervised learning,[26] MoCo-v2, to learn tile-level morphological features (figure 1B); extracting all tile-level feature vectors from the self-supervised learning encoder (figure 1C); and training and inference of the attention-based classification model based on the proposal of Lu and colleagues[17] to molecularly classify whole-slide images from all tile-level feature vectors (figure 1D). Our deep learning pipeline included nuclear segmentation and classification, which was done by use of HoVer-Net[27] (trained on a curated endometrial cancer tile image dataset; appendix 1 p 9), to derive human-interpretable, cell-level morpho-molecular features from the top 20 attended tiles, and a support vector machine trained to measure relative contributions of HoVer-Net-based features to the profiling of the molecular classes (figure 1E). All deep learning models were implemented in PyTorch (version 1.11). The support vector machine model was used from the Python scikit-learn library (version 1.1.1).

During pre-processing, the tissue of each whole-slide image was automatically segmented by use of regular Otsu thresholding. The tissue region of each whole-slide image was cut into non-overlapping square tiles of 360 µm at 40× magnification and resized to 224×224 pixels (figure 1A). This tile size was determined empirically (appendix 1 pp 6–7). Each whole-slide image in the training set (n=1240) contained a mean average of 2600 tiles. Tiles with very little or no tissue were excluded by applying a minimum threshold of 20 to the median value of each 8-bit RGB channel.

To train the feature extractor model with self-supervised learning, a dataset of image tiles was curated by randomly sampling tiles from each whole-slide image. To ensure the equitable representation of each molecular class, the optimal number of tiles to sample was determined for each whole-slide image depending on its molecular class (appendix 1 pp 6–7). This approach yielded a dataset of 1 170 931 tiles that were evenly distributed across the molecular classes, on which MoCo-v2 was trained for 300 epochs on three NVIDIA RTX 6000 (Santa Clara, CA, USA) graphics processing units, with a ResNet-50 encoder and a projection head as originally described (appendix 1 p 8).[26] Experiments with a ResNet-34 encoder (appendix 1 p 7), which is shallower than the ResNet-50, and a fixed number of tiles per whole-slide image yielded a lower performance than the initial approach (appendix 1 pp 6–7).

The tile-level feature vectors from the whole-slide images were extracted from the self-supervised learning ResNet-50 encoder at the last layer, resulting in a feature vector of size 2048 for each tile (figure 1C). Thereafter, the attention-based slide classifier (figure 1D) was trained for 100 epochs on a single graphics processing unit (appendix 1 p 8). The architecture of this attention model was based on an architecture proposed by Lu and colleagues[17] with the secondary clustering objective removed. Ablation studies showed that this clustering objective did not improve the AUROC for our task (appendix 1 p 7).

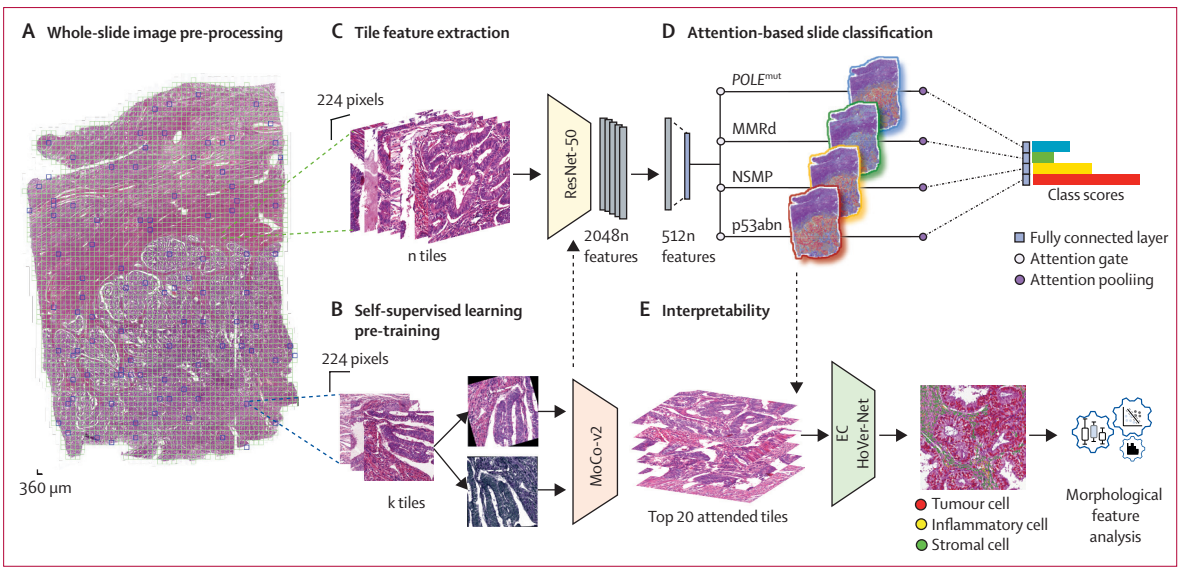


***Figure 1:*** **im4MEC deep learning pipeline**

For a high-resolution version of this image see appendix 2. (A) Whole-slide images were segmented and cut into non-overlapping square tiles of 360 µm at 40× magnification and resized to 224 × 224 pixels. (B) The optimal number of tiles were sampled from each whole-slide image to build a training dataset for the MoCo-v2 self-supervised learning model. (C) Features were extracted from all tiles of the whole-slide image by use of the self-supervised learning encoder, ResNet-50, at the last layer resulting in features of size 2048. (D) The model was trained to molecularly classify the whole-slide image, assigning attention scores to each tile and molecular class. Attention heatmaps are displayed from low attention (blue) to high attention (red). (E) The top 20 attended tiles were extracted from the predicted attention branch only. The predictions of HoVer-Net, a nuclear segmentation and classification deep learning model trained on an endometrial cancer tile image dataset, were used to compute counts of the three cell types and size and shape of the tumour nuclei (appendix 1 p 29). Subsequent analyses described these morphological features in association with the molecular classes and measured their relative feature importance with a support vector machine. MMRd=mismatch repair protein deficient. NSMP=no specific molecular profile. p53abn=abnormal cellular tumour antigen p53 expression.

See Online for appendix 2

A qualitative assessment of the attention heatmap (appendix 1 p 8) of the predicted molecular class and its top 20 tiles with the highest attention scores assigned by im4MEC was done for all PORTEC-3 patients by an expert gynaecological pathologist (TB). The spatial distribution of attention from the correctly classified patients was quantitatively analysed, and statistical significance was measured with an unpaired *t* test for each molecular class across three annotated regions: tumour, invasive border, and normal myometrium (appendix 1 p 8).

The top 20 tiles with the highest attention scores for each predicted molecular class were extracted. HoVer-Net architecture[27] was trained from scratch on a hand-curated endometrial cancer tile image dataset to predict the nuclear contour and three cell types on these tiles: inflammatory cells, stromal cells, and tumour cells (appendix 1 pp 9–11). Five morphological features were computed from the HoVer-Net predictions: counts of the three cell types and nuclear size and shape of the tumour cells averaged across the top 20 attended tiles per whole-slide image. The nuclear size used the nuclear surface area in µm, the nuclear shape used the non-convexity score because it describes the non-compactness of a particle, and together these features provided a proxy for nuclear atypia (appendix 1 p 12). A support vector machine with a linear kernel was trained on the five HoVer-Net-based morphological features to extract the support vector machine weights, a measure of relative feature importance. Input features were normalised after verifying non-collinearity with Pearson correlation and a cutoff of 0·5 (appendix 1 pp 12–13).

## Statistical analysis

We tested the associations between misclassification and the relative size of tumour tissue, HistoQC-based[28] slide quality with an unpaired *t* test, or the presence of secondary molecular class labels by Fisher's exact test. The statistical significance of any differences within the correctly classified cases or image-based molecular classes was measured pairwise with an unpaired *t* test for each of the five HoVer-Net-based morphological features. Additionally, the association between model predictions and endometrial cancer clinicopathological data (ie, histological subtypes, grade, and stage) was statistically tested with Fisher's exact test (appendix 1 p 13). Statistical analyses were performed with Python SciPy library (version 1.5.2) and statistical significance was accepted for p values of less than 0·05. The analysis of recurrence-free survival at 5 years (recurrences included local-regional and distant events) by true and image-based molecular classification was done according to the Kaplan-Meier method, with subgroups compared with the log-rank test. Prognostic refinement was also explored by selecting the predicted image-based molecular classes among one true molecular class.

Survival analyses were done by use of the R software (version 3.6.3); statistical significance was accepted for two-sided p values of less than 0·05.

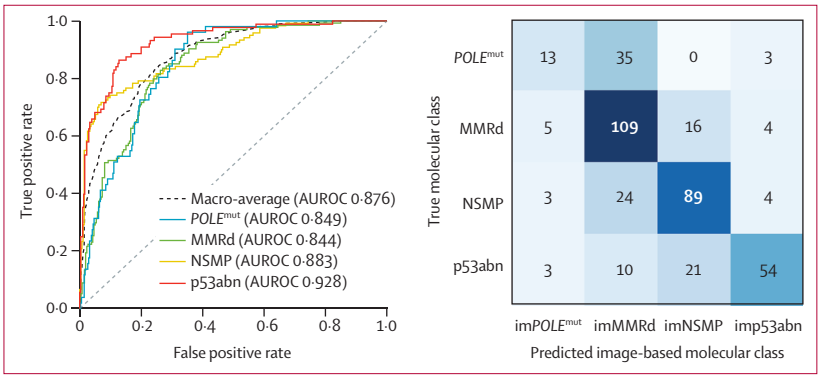


***Figure 2:*** **Model performance on the PORTEC-3 independent test set (n=393)**
(A) The macro-average and class-wise one-versus-rest ROC curves. The true positive rate represents sensitivity and the false positive rate represents 1 minus the specificity. (B) The confusion matrix. AUROC=area under the receiver operating characteristic curve. MMRd=mismatch repair protein deficient. NSMP=no specific molecular profile. p53abn=abnormal cellular tumor antigen p53 expression.


## Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.


## Results

The im4MEC model attained a mean macro-average AUROC of 0·874 (95% CI 0·856–0·893) on the four-fold cross-validation. Ablation studies showed that the inclusion of the self-supervised learning model in the pipeline improved the mean macro-average AUROC by 0·131 compared with the use of ImageNet pre-trained weights (appendix 1 p 7).

On the independent test set (PORTEC-3; n=393), im4MEC achieved a macro-average AUROC of 0·876 (figure 2A). The class-wise AUROCs were 0·849 for *POLE*$^{mut}$ (n=51), 0·844 for MMRd (n=134), 0·883 for NSMP (n=120), and 0·928 for p53abn (n=88; figure 2A). We report macro-average and class-wise precision, recall, F1 score, accuracy, specificity, and negative predictive value in appendix 1 (p 14). The confusion matrix indicated

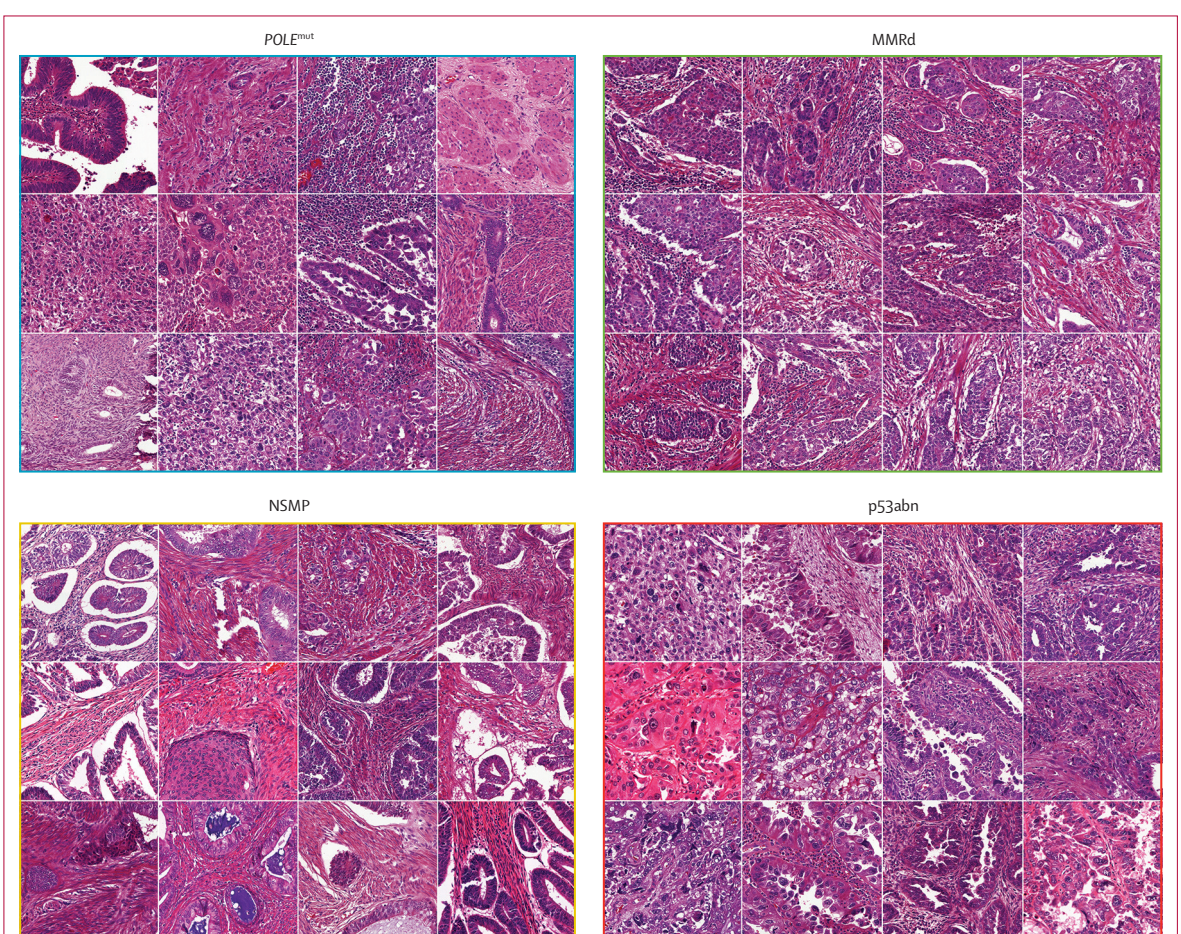


***Figure 3:*** **Image tile galleries showing morpho-molecular correlates**
See Online for appendix 3
For a high-resolution version of this image, see appendix 3. For each molecular class, we present an image gallery of the top one tile with the highest attention score of the 12 concordant cases with the highest class probability scores in PORTEC-3. MMRd=mismatch repair protein deficient. NSMP=no specific molecular profile. p53abn=abnormal cellular tumour antigen p53 expression.

concordant and some discordant cases between the true molecular classes and the predicted image-based molecular classes (referred to by the prefix im; figure 2B). The discordant cases, such as the 35 patients with *POLE*$^{mut}$ endometrial cancer predicted as being imMMRd endometrial cancer, were further investigated. We found no evidence of associations between misclassification and low tumour surface (appendix 1 p 14), HistoQC-based slide quality (appendix 1 p 15), or the presence of secondary molecular class labels (appendix 1 pp 15–16). The concordant cases were analysed for identification of endometrial cancer morpho-molecular correlates.

Quantitative evaluation of the attention spatial distribution of the concordant cases for each molecular class showed that the model mainly focused on the tumour region rather than the invasive border and normal myometrium (appendix 1 pp 16–19). The model showed a significantly higher attention towards the normal myometrium for concordant *POLE*$^{mut}$ cases than for the other concordant cases (p=0·0063), where a strong representation of *POLE*$^{mut}$-specific immune-related features,[7,8] such as lymphocytic aggregates or tertiary lymphoid structures, were found (appendix 1 pp 18–19).

Image galleries of the top 20 highly attended tiles from concordant patients showed key prototypical and novel morpho-molecular correlates (appendix 1 pp 25–28); in figure 3 we present the top one tile with the highest attention score of the patients with the highest predicted confidence score for each molecular class. *POLE*$^{mut}$ tiles highlighted solid tumour growth, a high density of tumour-infiltrating and peritumoural lymphocytes, and the presence of scattered tumour giant cells. MMRd tiles displayed similar traits, with a high density of lymphocytes and mainly solid tumour growth; however, some tiles also had a glandular architecture. NSMP tiles predominantly showed glands with smooth luminal borders, mild nuclear atypia, a low density of lymphocytes, and focal squamous differentiation. Although these features in NSMP endometrial cancer are consistent with expert interpretation (by TB), we also identified a low tumour-to-stroma ratio. p53abn tiles displayed a low density of lymphocytes, strong nuclear atypia, a high tumour-to-stroma ratio, and tumoural areas with both solid and glandular architecture, in which the glands typically showed a ragged luminal surface.

We analysed the top 20 attended tiles of the concordant cases using interpretable cell-type predictions from HoVer-Net to investigate the composition of each molecular class at the resolution of a single cell (appendix 1 pp 19–20; appendix 1 pp 25–28). Concordant *POLE*$^{mut}$ and MMRd cases showed significantly more inflammatory cells than NSMP and p53abn cases. *POLE*$^{mut}$ cases also displayed a significantly larger size of tumour nuclei compared with MMRd and NSMP cases. Concordant NSMP cases showed significantly more stromal cells and fewer tumour cells than *POLE*$^{mut}$, MMRd, and p53abn cases, implicating a low tumour-to-stroma ratio as a novel feature of this class. Concordant p53abn cases showed a stronger tumour nuclear atypia, based on nuclear size and shape, than *POLE*$^{mut}$, MMRd, and NSMP cases.

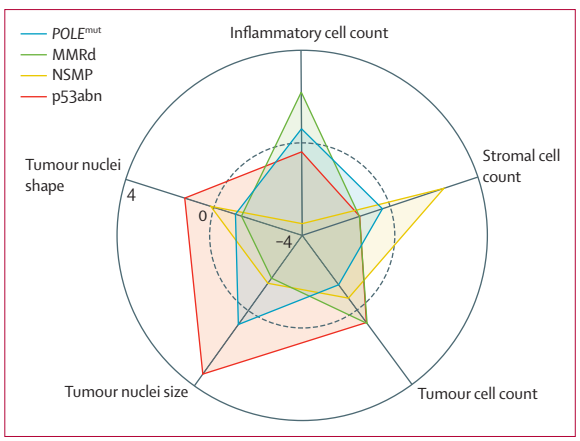


***Figure 4:*** **Support vector machine-based morphological feature importance of concordant cases in PORTEC-3**
MMRd=mismatch repair protein deficient. NSMP=no specific molecular profile. p53abn=abnormal cellular tumour antigen p53 expression.

The subsequent training of a support vector machine on concordant cases showed the relative contribution of each morphological feature to the profile of the molecular class (figure 4). Our data indicated that the inflammatory cell count positively contributed towards the prediction of concordant *POLE*$^{mut}$ cases and MMRd cases and negatively contributed to the prediction of concordant NSMP and p53abn cases. The stromal cell count positively contributed to the prediction of concordant NSMP cases (indicative of a low tumour-to-stroma ratio) and tumour nuclear atypia strongly contributed to the prediction of concordant p53abn cases, with the size of tumour nuclei having a larger impact than nuclei shape. Similar morpho-molecular correlates and their relative contributions were observed when extending the analysis to the image-based molecular classes, indicating that predictive morphological features were provided by the im4MEC predictions and highlighted some degree of heterogeneity within each molecular class (appendix 1 pp 20–21).

We examined the association between the im4MEC classification and endometrial cancer clinicopathological data (appendix 1 pp 21–22). We found that concordant *POLE*$^{mut}$ cases were associated with high-grade endometrial cancer (p=0·0004), that concordant MMRd cases were associated with an endometrioid histological subtype (p=0·0023), and that the only five low-grade, endometrioid *POLE*$^{mut}$ cases were predicted as imMMRd. Concordant NSMP cases were associated with a low grade, endometrioid subtype (p=0·062), whereas concordant p53abn cases were associated with a high grade subtype (p<0·0001). The only four low-grade

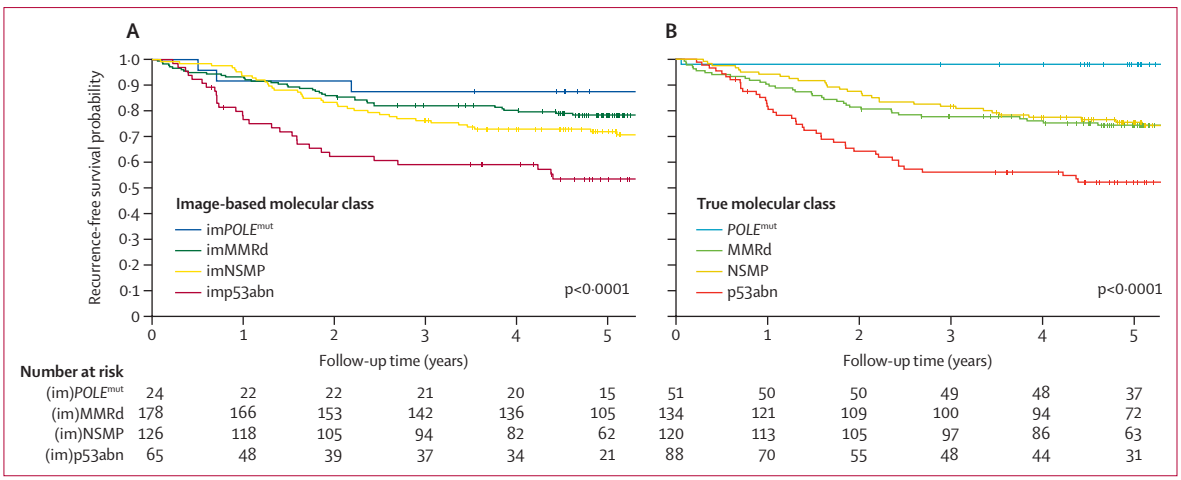


***Figure 5:*** **Recurrence-free survival at 5 years in PORTEC-3**
Stratified by image-based molecular class (A) and true molecular class (B). The plus signs (+) on each curve represent censored patients. MMRd=mismatch repair protein deficient. NSMP=no specific molecular profile. p53abn=abnormal cellular tumour antigen p53 expression.

endometrioid p53abn cases were classified as imNSMP. When extending the scope to the image-based molecular classes, the associations between predictions and histological subtypes or grades were found to be consistent (appendix 1 pp 20–21). We did not find any correlation between model discordance and stage (appendix 1 p 22).

The image-based molecular classification by im4MEC showed a strong prognostic value in the PORTEC-3 independent test set, a randomised trial of patients with high risk endometrial cancer (p<0·0001; figure 5A). The observed 5-year recurrence-free survival outcomes were consistent with the true molecular classification (figure 5B), with the best prognosis observed in patients with predicted im*POLE*$^{mut}$ endometrial cancer, intermediate prognoses observed in patients with predicted imMMRd or imNSMP endometrial cancer, and poor prognosis observed in patients with predicted imp53abn endometrial cancer. The im*POLE*$^{mut}$ class had a slightly worse prognosis than the true *POLE*$^{mut}$ class. This discrepancy was caused by the three patients with p53abn endometrial cancer and the three patients with NSMP endometrial cancer who were predicted as im*POLE*$^{mut}$ and not by the five patients with MMRd endometrial cancer who were predicted as im*POLE*$^{mut}$ and showed no recurrence (appendix 1 p 23).

The interpretability of our deep learning pipeline was leveraged to explore the prognostic relevance of the morpho-molecular correlates among the discordant cases (figure 6). The ten patients with p53abn endometrial cancer predicted as imMMRd did not show prototypical morphological features of p53abn endometrial cancer and instead showed MMRd endometrial cancer-like morphological features, including high counts of inflammatory cells and mild tumour nuclear atypia, which might have led to the discordance (figure 6A). Patients in this specific discordant subset of the p53abn class with MMRd-like morphology appeared to have a better prognosis than the concordant p53abn patients, although this difference was not statistically significant (p=0·30; figure 6A). Similarly, the three patients with p53abn endometrial cancer that was classified as im*POLE*$^{mut}$ had a substantially higher count of inflammatory cells than concordant p53abn patients, although this finding did not translate into the good prognosis that is observed in patients with true *POLE*$^{mut}$ endometrial cancer (appendix 1 p 23). The four patients with NSMP endometrial cancer predicted as imp53abn showed higher nuclear atypia (in terms of tumour nuclei size) and a higher tumour-to-stroma ratio, which are both prototypical features of p53abn endometrial cancer, compared with NSMP concordant patients (figure 6B). NSMP discordant patients with p53abn-like morphology appeared to have a worse prognosis than NSMP concordant patients, similar to the poor outcome of typical p53abn endometrial cancer, although this difference was not statistically significant (p=0·13; figure 6B). Similar observations were found for the 16 patients with MMRd endometrial cancer predicted as imNSMP endometrial cancer, who showed a lower count of inflammatory cells and a lower tumour-to-stroma ratio, with a similar prognosis as the MMRd concordant patients (appendix 1 p 24). Regarding the discordance between *POLE*$^{mut}$ and MMRd classes, im4MEC identified a set of five patients with MMRd endometrial cancer predicted as im*POLE*$^{mut}$ who had an excellent prognosis (figure 6C). A morphological analysis of the respective discordant calls between these two classes indicated that predicted *POLE*$^{mut}$ slides had larger tumour nuclei than MMRd concordant slides but cellular composition was similar, which might explain the model's confusion (figure 6C; appendix 1 p 24).

## Discussion

To our knowledge, we present the first interpretable deep learning pipeline that can predict the four-class molecular

endometrial cancer classification from digitised haematoxylin and eosin-stained whole-slide images. Our deep learning model, im4MEC, was built, validated, and tested in 2028 patients with molecularly classified endometrial cancer from three randomised trials[3,4,21–23] and four clinical cohorts.[5,6,24] We combined self-supervised learning[26] with an attention-based, slide-level classification model[17] for learning endometrial-cancer-specific morphological

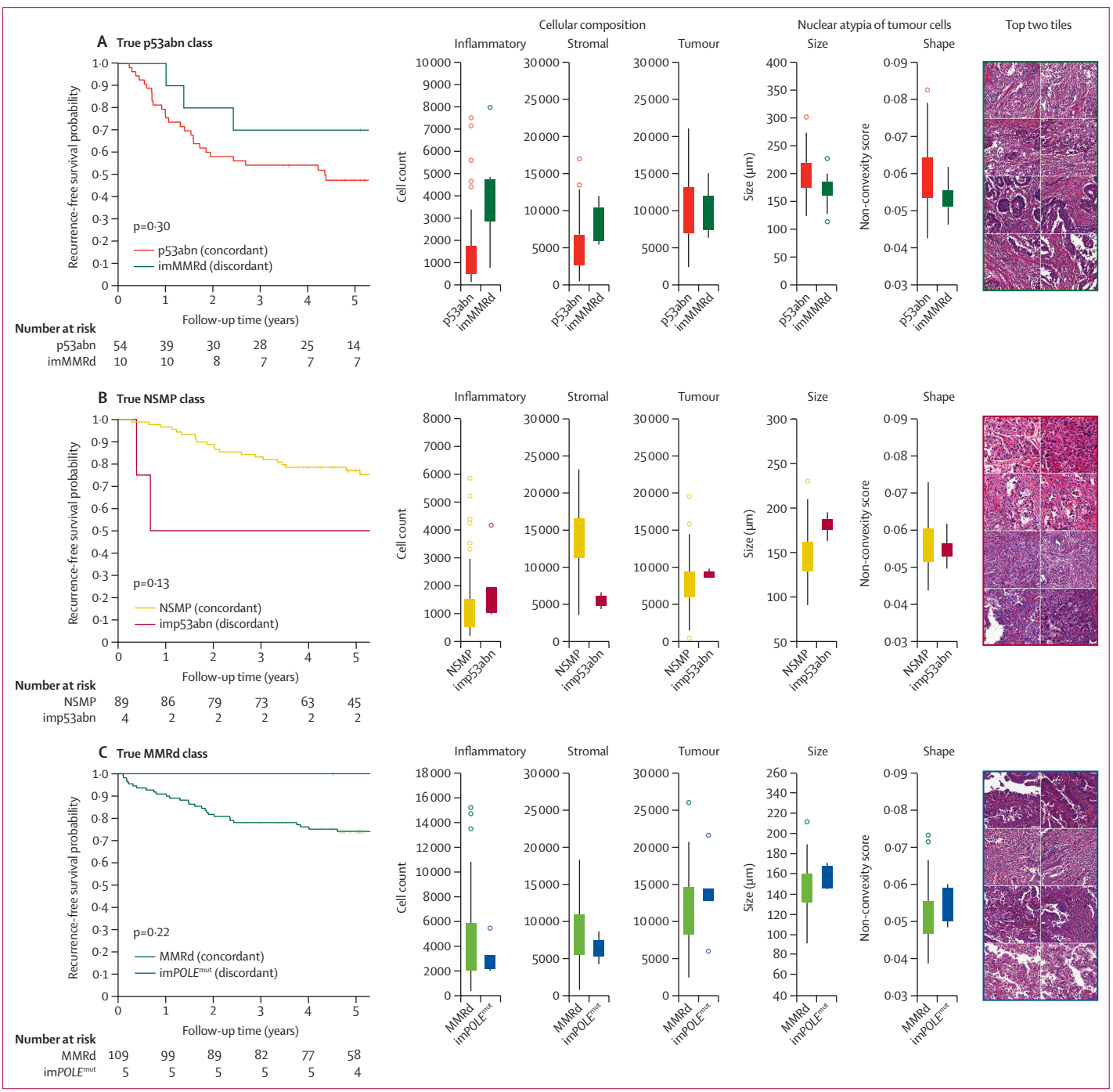


***Figure 6:*** **Discordant image-based predictions by prognosis and morphology in PORTEC-3**
5-year recurrence-free survival and HoVer-Net-based morphological features stratified by concordant and discordant cases and the top two tiles of the top four discordant cases for the comparisons of concordant p53abn cases (n=54) versus discordant imMMRd cases (n=10) in the true p53abn class (A), concordant NSMP cases (n=89) versus discordant imp53abn cases (n=4) in the true NSMP class (B), and concordant MMRd cases (n=109) versus discordant im*POLE*$^{mut}$ cases (n=5) in the true MMRd class (C). The plus signs (+) on each curve represent censored patients. High-resolution histology images can be found in appendix 1 (pp 30–32). MMRd=mismatch repair protein deficient. NSMP=no specific molecular profile. p53abn=abnormal cellular tumour antigen p53 expression.

features associated with each molecular class. The im4MEC model obtained a high performance that is comparable to other four-class molecular classification tasks of similar complexity, such as the image-based consensus molecular subtype classification of colorectal cancer.[10] Additionally, im4MEC identified human-interpretable morpho-molecular correlates that could improve prognostication.

By using the strength of our deep learning pipeline and an independent test set of 393 patients with high-risk endometrial cancer, long-term follow-up, and complete clinicopathological data from the PORTEC-3 randomised trial, we made concordant and discordant cases between true and image-based molecular classes interpretable. The attention mechanism and cell-level quantitative analyses revealed insights into endometrial cancer-specific morpho-molecular correlates that are visible on haematoxylin and eosin-stained slides and go beyond classic histological subtyping or grading variables. Furthermore, our analyses provide insights into intra-class heterogeneity within endometrial cancer molecular subgroups, which could be exploited to refine prognostication. Our data corroborated established prototypical morpho-molecular correlates in endometrial cancer, such as a high lymphocyte density in the *POLE*$^{mut}$ and MMRd classes[7] and strong nuclear atypia in p53abn endometrial cancer.[29] Our results suggested that it is possible to stratify a specific subset of patients with an MMRd-like morphology and a better prognosis than the typically aggressive p53abn endometrial cancer and to stratify a subset of patients with NSMP endometrial cancer of p53abn-like morphology and a poor prognosis. Our work also showed a low tumour-to-stroma ratio within NSMP endometrial cancer slides, which constitutes a novel morphological feature and might improve biological understanding of this poorly defined molecular class. Furthermore, the spatial distribution of highly attended regions not only demonstrated that the tumour centre was informative, but it also showed that predictive morphological features were present in the invasive border and even in the normal myometrium, which could be related to the presence of microenvironment immune-related features. This finding was particularly true for predicting *POLE*$^{mut}$ whole-slide images displaying lymphocytic aggregates or tertiary lymphoid structures within the normal myometrium, a specific phenotypical trait that has only recently been linked to this molecular class.[8] These findings exemplify the fact that the use of attention-based deep learning models without predefined annotated regions can reveal biologically relevant morphological information in adjacent non-tumoural regions, which might also be of interest in other cancer types.[10]

This study has some limitations. First, im4MEC was trained on patients with early-stage, intermediate-to-high-risk endometrial cancer[3,5,21,22,24,30] and independently tested on patients with high-risk endometrial cancer.[4,23] Thus, the subsequently identified morpho-molecular correlates should be interpreted in the context of high-risk endometrial cancer. Second, no data on patient ethnicity were available. Third, the cell-level morphological features used for interpretability were limited to the counts of inflammatory, stromal, and tumour cells and the size and shape of tumour nuclei. These features could be expanded by increasing cell-level granularity in the nuclear classification deep learning model, or by adding a layer of higher-order histopathological features, such as architectural growth pattern (solid, villoglandular, or papillary) and type of invasion (infiltrative or pushing invasion). Fourth, the difficulty im4MEC had in separating *POLE*$^{mut}$ endometrial cancer from MMRd endometrial cancer on haematoxylin and eosin-stained whole-slide images could have been caused by the relative scarcity of *POLE*$^{mut}$ in the training set. However, it is also plausible that these tumours simply have highly overlapping phenotypes as a result of their shared molecular characteristic of a high mutational load. Finally, our analyses indicated the possibility of prognostic refinement for a defined subset of patients with endometrial cancer within each molecular class. The clinical relevance of these findings requires prospective validation and our results should be further investigated by use of interpretable deep-learning-based models tailored to predict survival outcomes.

With the goal of implementing im4MEC in clinical practice, future work should address the generalisability of im4MEC's performance on large cohorts with diverse morphology, especially those of patients with non-endometrioid endometrial cancer, as these patients were under-represented in our training set. Next, the utility of im4MEC to delineate tumour heterogeneity should be tested and verified by predicting molecular class on consecutive haematoxylin and eosin-stained whole-slide images and tumour blocks from the same patient. Understanding the reproducibility of model performance and the identified clinically relevant morphological features on preoperative biopsies or metastases might be of additional interest. The impact of discordant predictions on clinical management and treatment decisions should be further investigated, especially for the most biologically aggressive p53abn and *POLE*$^{mut}$ endometrial cancers.

We envision the initial clinical application of im4MEC to evolve into two possible directions that could provide a haematoxylin and eosin-based molecular profiling of patients with endometrial cancer. First, im4MEC could be positioned as a potentially low-cost pre-screening tool to identify occurrences of p53abn endometrial cancer for further confirmatory immunohistochemistry or molecular testing. This approach would be particularly attractive in patients with low-risk and early-stage, low-grade endometrial cancer, for whom molecular classification is not routinely done.[2] Second, im4MEC could be used with immunohistochemistry to determine MMRd

status, which is recommended for all newly diagnosed patients with endometrial cancer.[2] This second approach would result in a three-class (*POLE*$^{mut}$, p53abn, and NSMP) model, im3MEC, which would achieve more accurate haematoxylin and eosin-based predictions than a four-class model by resolving the challenge of distinguishing *POLE*$^{mut}$ from MMRd endometrial cancer.

Taken together, we have proposed a robust deep learning model that has enabled the image-based molecular classification and prognostic refinement of endometrial cancer by use of standard diagnostic histology sections in a digital pathology workflow. Prospective validation in well designed and statistically robust studies will be crucial to support clinical translation.

#### Contributors

VHK and TB conceived the study. SF, SA, JBW, JD, CLC, NH, VHK, and TB participated in designing the study and reviewing the literature. NH, VHK, and TB supervised the study. SF, SM, and NH collected the data. SF, SA, and JBW contributed to code writing, model training, and data analysis. JBW prepared the code repository. NH, VHK, and TB contributed to expert review and data interpretation. SF, SA, JBW, JD, VTHBMS, NH, and VHK, and TB designed the figures. SF and SA drafted the manuscript. SF, SA, JBW, JD, NH, VHK, and TB revised the manuscript. JJJ, MB, SR, IJ-S, LCHWL, RAN, EMvdS-B, SMdB, MEP, NS, LRM, HJM, AL, and HWN contributed to providing data and tumour material used in this study. All authors had access to the result data presented in the final manuscript and all authors read and approved the final manuscript. Verification and access to raw data was limited to SF, SA, JBW, NH, VHK, and TB due general data protection regulations. The decision to submit was made by all authors. SF and SA contributed equally to this work and share first authorship. VHK and TB contributed equally to this work and share senior authorship.

#### Declaration of interests

HWN and JJJ declare grants from Merck, NH declares grants from Varian, and VHK declares research funding from Roche, unrelated to the subject of this manuscript. All other authors declare no competing interests.

#### Data sharing

The tumour material and datasets generated during or analysed in this study are not publicly available due to restrictions by privacy laws. Data and tumour material from PORTEC-1, PORTEC-2, PORTEC-3, MST, the TransPORTEC pilot study, and the case series of patients with *POLE*$^{mut}$ endometrial cancer are held by the Gynecological Oncology Research Group of the Department of Radiation Oncology and the Department of Pathology of Leiden University Medical Center, Leiden, the Netherlands. Data and material from PORTEC-3 and the TransPORTEC pilot study are available to members of the international TransPORTEC Research Consortium. Requests for sharing of all data and material should be addressed to the corresponding author within 15 years of the date of publication of this Article and include a scientific proposal. Depending on the specific research proposal, the TransPORTEC consortium (PORTEC-3 and TransPORTEC pilot study) or coauthor of this paper CLC (PORTEC-1, PORTEC-2, MST) will determine when, for how long, for which specific purposes, and under which conditions the requested data can be made available, subject to ethical consent. TCGA-UCEC data are publicly available via the cBioPortal for Cancer Genomics at https://www.cbioportal.org/study/clinicalData?id=ucec_tcga_pan_can_atlas_2018. Code is available at https://github.com/AIRMEC/im4MEC.

#### Acknowledgments

The PORTEC-1, PORTEC-2, and PORTEC-3 trials were supported by grants from the Dutch Cancer Society (CKTO 90–01, CKTO 2001–04, and CKTO 2006–04, respectively), and HWN, JJJ, and NH received grants from the Dutch Cancer Society unrelated to the subject of this manuscript. This study was supported by a translational research project grant from the Hanarth Foundation and the Swiss Federal Institutes of Technology (strategic focus area of personalised health and related technologies; 2021–367) and a grant from the Promedica Foundation (F-87701–41–01) during the conduct of the study. We thank the participants, investigators, and pathologists who recruited patients and collected samples in the PORTEC-1, PORTEC-2, and PORTEC-3 randomised trials. We thank the PORTEC-3 study group and the TransPORTEC Research Consortium for the establishment of the TransPORTEC biobank and JJJ and investigators of the prospective MST cohort. We thank the Netherlands Cancer Institute for the use of their 3DHISTECH P1000 scanner, and Tessa Rutten and Natalja ter Haar, Leiden University Medical Center, Leiden, the Netherlands, for excellent technical support. We express our gratitude towards Gunnar Rätsch, ETH Zurich, Zurich, Switzerland, for providing valuable feedback, thesis supervision, and sharing his biomedical expertise during the course of the project.

#### References

1 International Agency for Research on Cancer, WHO. WHO classification of tumours. 5th edition. Female genital tumours. In: Herrington CS, ed. Lyon, France: International Agency for Research on Cancer, 2020.

2 Concin N, Matias-Guiu X, Vergote I, et al. ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma. *Int J Gynecol Cancer* 2021; **31:** 12–39.

3 Stelloo E, Nout RA, Osse EM, et al. Improved risk assessment by integrating molecular and clinicopathological factors in early-stage endometrial cancer—combined analysis of the PORTEC cohorts. *Clin Cancer Res* 2016; **22:** 4215–24.

4 León-Castillo A, de Boer SM, Powell ME, et al. Molecular classification of the PORTEC-3 trial for high-risk endometrial cancer: impact on prognosis and benefit from adjuvant therapy. *J Clin Oncol* 2020; **38:** 3388–97.

5 Stelloo E, Bosse T, Nout RA, et al. Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; a TransPORTEC initiative. *Mod Pathol* 2015; **28:** 836–44.

6 Kandoth C, Schultz N, Cherniack AD, et al. Integrated genomic characterization of endometrial carcinoma. *Nature* 2013; **497:** 67–73.

7 Horeweg N, de Bruyn M, Nout RA, et al. Prognostic integrated image-based immune and molecular profiling in early-stage endometrial cancer. *Cancer Immunol Res* 2020; **8:** 1508–19.

8 Horeweg N, Workel HH, Loiero D, et al. Tertiary lymphoid structures critical for prognosis in endometrial cancer patients. *Nat Commun* 2022; **13:** 1373.

9 Yamashita R, Long J, Longacre T, et al. Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. *Lancet Oncol* 2021; **22:** 132–41.

10 Sirinukunwattana K, Domingo E, Richman SD, et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* 2021; **70:** 544–54.

11 Naik N, Madani A, Esteva A, et al. Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains. *Nat Commun* 2020; **11:** 5727.

12 Feng L, Liu Z, Li C, et al. Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit Health* 2022; **4:** e8–17.

13 Echle A, Ghaffari Laleh N, Quirke P, 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.

14 Hong R, Liu W, DeLair D, Razavian N, Fenyö D. Predicting endometrial cancer subtypes and molecular features from histopathology images using multi-resolution deep learning models. *Cell Rep Med* 2021; **2:** 100400.

15 Wang T, Lu W, Yang F, et al. Microsatellite instability prediction of uterine corpus endometrial carcinoma based on H&E histology whole-slide imaging. 2020 17th IEEE International Symposium on Biomedical Imaging; April 3–7, 2020 (abstr 1289–92).

16 Qu H, Zhou M, Yan Z, et al. Genetic mutation and biological pathway prediction based on whole slide images in breast carcinoma using deep learning. *NPJ Precis Oncol* 2021; **5:** 87.

17 Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng* 2021; **5:** 555–70.

18 Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015; **115:** 211–52.

19 Schirris Y, Gavves E, Nederlof I, Horlings HM, Teuwen J. DeepSMILE: contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. *Med Image Anal* 2022; **79:** 102464.

20 Ciga O, Xu T, Martel AL. Self supervised contrastive learning for digital histopathology. *Mach Learn Appl* 2022; **7:** 100198.

21 Creutzberg CL, van Putten WLJ, Koper PCN, et al. Surgery and postoperative radiotherapy versus surgery alone for patients with stage-1 endometrial carcinoma: multicentre randomised trial. *Lancet* 2000; **355:** 1404–11.

22 Nout RA, Smit VTHBM, Putter H, et al. Vaginal brachytherapy versus pelvic external beam radiotherapy for patients with endometrial cancer of high-intermediate risk (PORTEC-2): an open-label, non-inferiority, randomised trial. *Lancet* 2010; **375:** 816–23.

23 de Boer SM, Powell ME, Mileshkin L, et al. Adjuvant chemoradiotherapy versus radiotherapy alone for women with high-risk endometrial cancer (PORTEC-3): final results of an international, open-label, multicentre, randomised, phase 3 trial. *Lancet Oncol* 2018; **19:** 295–309.

24 Jobsen JJ, Naudin Ten Cate L, Lybeert MLM, et al. Outcome of endometrial cancer stage IIIA with adnexa or serosal involvement only. *Obstet Gynecol Int* 2011; **2011:** 962518.

25 Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; **2:** 401–04.

26 Chen X, Fan H, Girshick R, He K. Improved baselines with momentum contrastive learning. *arXiv* 2020; published online March 9. https://arxiv.org/abs/2003.04297 (preprint).

27 Graham S, Vu QD, Raza SEA, et al. HoVer-Net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* 2019; **58:** 101563.

28 Janowczyk A, Zuo R, Gilmore H, Feldman M, Madabhushi A. HistoQC: an open-source quality control tool for digital pathology slides. *JCO Clin Cancer Inform* 2019; **3:** 1–7.

29 Kang EY, Wiebe NJ, Aubrey C, et al. Selection of endometrial carcinomas for p53 immunohistochemistry based on nuclear features. *J Pathol Clin Res* 2022; **8:** 19–32.

30 Jobsen JJ, Lybeert MLM, van der Steen-Banasik EM, et al. Multicenter cohort study on treatment results and risk factors in stage II endometrial carcinoma. *Int J Gynecol Cancer* 2008; **18:** 1071–78.