

Supplementary Tables:

Table S1: Creating classification classes for the histopathology reports

Number	Regular expression code	Classification class
1	M-8\d\d\d/0	Non-malignant
2	M-8\d\d\d/1	Non-malignant
3	M-8\d\d\d/2	Malignant
4	M-8\d\d\d/3	Malignant
5	M-8\d\d\d/6	Malignant
6	M-8\d\d\d/9	Non-malignant
7	M-9\d\d\d/0	Non-malignant
8	M-9\d\d\d/1	Non-malignant
9	M-9\d\d\d/2	Malignant
10	M-9\d\d\d/3	Malignant
11	M-9\d\d\d/6	Malignant
12	M-9\d\d\d/9	Non-malignant
13	M-0000/0	No diagnosis

Table S2: Important words according to TFIDF vectorizer

token	weight
keeping cutaneous	9,207129
inflammation benign	9,207129
consistent metastatic	9,207129
consistent metastatic adenocarcinoma	9,207129
consistent poorly	9,207129
consistent poorly differentiated	9,207129
compatible invasive adenocarcinoma	9,207129
compatible infiltrating	9,207129
complex cyst	9,207129
invasive adenocarcinoma	9,207129
invasive component	9,207129
features malignant	9,207129
inflamed hyperplastic	9,207129
inflamed intestinal	9,207129
inflamed invasive	9,207129
inflamed macroscopic	9,207129
inflamed nodular	9,207129
inflamed polypoid	9,207129
inflamed seborrhoeic	9,207129
inflamed stromal	9,207129
inflamed tubular	9,207129
inflamed tumour	9,207129
inflamed ulcer	9,207129
inflamed urothelium	9,207129

inflammation acanthosis	9,207129
inflammation acinar	9,207129
inflammation activity	9,207129
inflammation addition	9,207129
unremarkable mild	9,207129
inflammation aggregates	9,207129
inflammation association	9,207129
inflammation atrophic	9,207129
inflammation atypical	9,207129
inflamed free	9,207129
inflamed focus	9,207129
inflamed foci	9,207129
infiltrative lesion	9,207129
infiltration soft	9,207129
infiltration subepithelial	9,207129
infiltration surrounding	9,207129
unremarkable nasal	9,207129
infiltrative areas	9,207129
infiltrative borders	9,207129
infiltrative carcinoma	9,207129
infiltrative clusters	9,207129
infiltrative edge	9,207129
infiltrative margin	9,207129
inflamed cyst	9,207129
infiltrative margins	9,207129
infiltrative minimum	9,207129
infiltrative moderately	9,207129
unremarkable myometrial	9,207129
infiltrative solid	9,207129
unremarkable mucosal	9,207129
inflamed areas	9,207129
inflamed colonic	9,207129
inflamed contains	9,207129
inflammation base	9,207129
inflammation calcification	9,207129
unremarkable patent	9,207129
inflammation carcinoma	9,207129
inflammation large	9,207129
inflammation left	9,207129
inflammation likely	9,207129
inflammation malignant	9,207129
inflammation microabscess	9,207129
inflammation mitotic	9,207129
inflammation multinucleate	9,207129
inflammation observed	9,207129

inflammation oncocytic	9,207129
inflammation organising	9,207129
inflammation perforation	9,207129
inflammation possibly	9,207129
inflammation presence	9,207129
unremarkable maximum	9,207129
inflammation radiation	9,207129
inflammation reactive	9,207129
inflammation regenerative	9,207129
inflammation see	9,207129
unremarkable mammary	9,207129
inflammation special	9,207129
inflammation subtle	9,207129
inflammation high	9,207129
inflammation hepatic	9,207129
inflammation hemorrhage	9,207129
inflammation desmoplastic	9,207129
inflammation composed	9,207129
inflammation conclusion	9,207129
inflammation congested	9,207129
inflammation consistent	9,207129
inflammation cryptitis	9,207129
inflammation cyst	9,207129
inflammation cystic	9,207129
inflammation dense	9,207129
inflammation denuded	9,207129
inflammation endocervical	9,207129
inflammation haemorrhage	9,207129
inflammation epithelium	9,207129
inflammation evident	9,207129
inflammation extensive	9,207129
inflammation fibrin	9,207129
inflammation fibrois	9,207129
inflammation findings	9,207129
inflammation focus	9,207129
inflammation follicular	9,207129
inflammation fragment	9,207129
infiltration smooth	9,207129
infiltration posterior	9,207129
infiltration not	9,207129
infiltration neoplasm	9,207129
infiltrate obscures	9,207129
infiltrate one	9,207129
infiltrate pathological	9,207129
infiltrate prostatic	9,207129

infiltrate sheets	9,207129
infiltrate spiculated	9,207129
infiltrate stroma	9,207129
infiltrate superficial	9,207129
infiltrate together	9,207129
infiltrate trabeculae	9,207129
unremarkable paratubal	9,207129
infiltrated anastomosing	9,207129
infiltrated ductal	9,207129
infiltrated large	9,207129
infiltrated monotonous	9,207129
infiltrated sheet	9,207129
infiltrated short	9,207129
infiltrated single	9,207129
infiltrated spiculated	9,207129
infiltrated variably	9,207129
infiltrates fat	9,207129
infiltrate not	9,207129
infiltrate non	9,207129
unremarkable parenchyma	9,207129
infiltrate deeper	9,207129
infiltrate acute	9,207129
infiltrate adipose	9,207129
infiltrate atypia	9,207129
infiltrate bland	9,207129
infiltrate cd20	9,207129
infiltrate cd3	9,207129
infiltrate comment	9,207129
infiltrate consistent	9,207129
infiltrate consists	9,207129
infiltrate endometrial	9,207129
infiltrate mature	9,207129
infiltrate fused	9,207129
infiltrate glandular	9,207129
infiltrate haemosiderin	9,207129
infiltrate inflammatory	9,207129
infiltrate islands	9,207129
infiltrate isolated	9,207129
infiltrate keeping	9,207129
infiltrate lamina	9,207129
infiltrate mainly	9,207129
infiltrates haemosiderin	9,207129
infiltrates large	9,207129
infiltrates scattered	9,207129
infiltration epidermis	9,207129

unremarkable native	9,207129
infiltration adipose	9,207129
infiltration cannot	9,207129
infiltration carcinoma	9,207129
infiltration clusters	9,207129
infiltration cords	9,207129
infiltration diffuse	9,207129
infiltration discohesive	9,207129
infiltration ductal	9,207129
infiltration glands	9,207129
infiltrating tongues	9,207129
infiltration groups	9,207129
infiltration however	9,207129
infiltration identified	9,207129
infiltration infiltrates	9,207129
infiltration infiltrating	9,207129
infiltration intertrabecular	9,207129
infiltration left	9,207129
infiltration muscularis	9,207129
infiltration myometrium	9,207129
infiltrating tubules	9,207129
infiltrating thickened	9,207129
infiltrates small	9,207129
infiltrating dermal	9,207129
infiltrates stromal	9,207129
unremarkable ovarian	9,207129
infiltrating adjacent	9,207129
infiltrating annular	9,207129
infiltrating bone	9,207129
infiltrating cervical	9,207129
infiltrating cohesive	9,207129
infiltrating collagen	9,207129
unremarkable oral	9,207129
infiltrating glandular	9,207129
unremarkable not	9,207129
infiltrating groups	9,207129
infiltrating high	9,207129
infiltrating lesion	9,207129
unremarkable oesophageal	9,207129
infiltrating malignancy	9,207129
unremarkable nutritional	9,207129
unremarkable numerous	9,207129
infiltrating normal	9,207129
infiltrating pattern	9,207129
inflammation suggesting	9,207129

inflammation superficial	9,207129
inflammation telangiectasia	9,207129
instability detected	9,207129
intact glandular	9,207129
intact invasion	9,207129
intact measures	9,207129