Original Research

Fast and label-free automated detection of microsatellite status in early colon cancer using artificial intelligence integrated infrared imaging

Klaus Gerwerta, Stephanie Schörnera, Frederik Großerueschkampa, Anna–Lena Kraeftb, David Schuhmacherc, Carlo Sternemann, Inke S. Federd, David Schuhmacherc, Carlo Sternemann, Inke S. Federd, Sarah Wisserd, Celine Lugierneb, Dirk Arnold, Christian Teschendorff, Lothar Mueller, Nina Timmesfeldh, Axel Mosic, Anke Reinacher-Schickb, Andrea Tannapfel, Andrea Tannapfel*

a Center for Protein Diagnostics (PRODI), Department of Biophysics, Ruhr University Bochum, Bochum, Germany
b Department of Haematology, Oncology and Palliative Care, St. Josef-Hospital, Ruhr University Bochum, Bochum, Germany
c Center for Protein Diagnostics (PRODI), Dept. of Bioinformatics, Ruhr University Bochum, Bochum, Germany
d Institut für Pathologie, Ruhr-Universität Bochum, Bochum, Germany
e Oncology, Haematology, Palliative Care Department Asklepios Tumorzentrum Hamburg AK Altona, Hamburg, Germany
f Internal Medicine, Medizinische Klinik St.-Josefs-Hospital, Dortmund, Germany
g Onkologie UnterEms Leer Emden Papenburg, Onkologische Schwerpunktpraxis Leer-Emden, Leer, Germany
h Medical Informatics, Biometry and Epidemiology, Ruhr University Bochum, Bochum, Germany

Received 4 October 2022; received in revised form 14 December 2022; accepted 23 December 2022
Available online 9 January 2023

KEYWORDS
Microsatellite instability; Colon cancer; Artificial intelligence; Deep learning; Infrared imaging; Label-free; Convolutional neural networks

Abstract Purpose: Microsatellite instability (MSI) due to mismatch repair (MMR) defects accounts for 15–20% of colon cancers (CC). MSI testing is currently standard of care in CC with immunohistochemistry of the four MMR proteins representing the gold standard. Instead, label-free quantum cascade laser (QCL) based infrared (IR) imaging combined with artificial intelligence (AI) may classify MSI/microsatellite stability (MSS) in unstained tissue sections user-independently and tissue preserving.

Methods: Paraffin-embedded unstained tissue sections of early CC from patients participating in the multicentre AIO ColoPredict Plus (CPP) 2.0 registry were analysed after dividing into three groups (training, test, and validation). IR images of tissue sections using QCL-IR microscopes were classified by AI (convolutional neural networks [CNN]) using a two-step

* Corresponding author: Institute of Pathology, Ruhr University Bochum, Bürkle-de-la-Camp-Platz 1, Bochum, 44789, Germany.
E-mail address: andrea.tannapfel@pathologie-bochum.de (A. Tannapfel).
1 Both authors contributed equally.

https://doi.org/10.1016/j.ejca.2022.12.026
0959-8049/2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

Microsatellite instability (MSI) is a distinct genomic feature caused by a defective mismatch repair system (dMMR) present in multiple cancers. It occurs in approximately 15% of colorectal cancer (CRC) cases [1]. MSI is stage-dependent, with over 20% detected in stage II and less than 5% in advanced CRC (aCRC), referring to stage IV metastatic CRC. MSI tumours are characterised by poor differentiation, mucinous component, proximal location, female, and older age [2]. MSI tumours also occur in younger patients with Lynch syndrome, the most common form of hereditary colon cancer (CC) [1]. MSI tumours are associated with a high tumour mutational burden deriving from the MSI-driven oncogenic pathway, (over-)expressing highly immunogenic neoantigens, and immune checkpoints. MSI has long been established as a positive prognostic parameter in stage II CRC and has now become an important predictive marker for the efficacy of immune checkpoint inhibition in multiple cancer types including CRC [3,4], setting the new standard of care in this molecular patient subgroup. Immune checkpoint inhibitors are currently also being evaluated in randomised controlled trials in localised CRC both in adjuvant and neoadjuvant settings [1]. Testing of MSI is therefore recommended in CRC of all stages [5].

Two standard reference methods, immunohistochemistry (IHC) and fragment length analysis by polymerase chain reaction (PCR), are used to screen for MSI. IHC analysis of the MMR proteins MLH1, MSH2, MSH6, and PMS2 is the gold standard in clinical routine [6]. Next-generation sequencing is an alternative test for MSI with approval by the Federal Drug Administration for MSI determination in aCRC. Considering its crucial positive predictive value for checkpoint inhibitor therapy, misdiagnosis of MSI, which has been reported in up to 10% for IHC, may result in substantial disadvantages for the individual patient [7]. Furthermore, immunohistochemical staining leaves tissue unusable for further molecular analysis.

With the advances in computing power and the possibility of utilising big data, artificial intelligence (AI)-assisted digital pathology has gained significant attraction [8–11]. MSI was detected from haematoxylin–eosin (H&E)-stained sections applying a deep learning approach requiring very large patient cohorts for training, showing impressive validation [12]. However, all methods for MSI detection so far depend on variously modified tissue — rendering the sample inaccessible for further molecular characterisation.

Infrared (IR) imaging is an emerging label-free technology for unstained tissue sections. A quantum cascade laser (QCL) based IR microscope records spatially resolved infrared spectra providing 427 molecular-specific channels. Previously, precise, operator-independent identification of cancer, subtypes, grading, and molecular alterations in unstained tissue sections has been reported [13–15]. Feasibility of IR-based MSI detection using a machine learning approach has previously been demonstrated by our group using a random forest classifier trained on 40 patients and tested on 60 patients [16]. Although a very good area under receiver operating characteristic curve (AUROC) could be achieved on the small collective, an internal validation on a larger cohort from the multicentre AI (Arbeitsgemeinschaft internistische Onkologie) ColoPredict Plus (CPP) 2.0 registry showed that the robustness and generalisability was not given. To overcome these limitations, the presented study aimed to test and validate an IR imaging AI-based approach to classify MSI vs. microsatellite stability (MSS) in early CC tissue samples from the AIO CPP 2.0 registry. This approach now allows us for the first time a robust detection of MSI in CC by IR imaging.
2. Material and methods

2.1. Patients

The multicentre AIO CPP 2.0 registry trial (DRKS00004305) captures clinical data and tissue samples of patients with early CC in 188 trial centres throughout Germany [17]. All study participants provided informed consent, and the study protocol was approved by the institutional review board (IRB) of the Ruhr-University Bochum (registration number: 4453-12 & 20-6830) and by local IRBs of participating sites. The sample study is based on feasibility studies conducted with IR imaging and a statistical evaluation for the expected length of the 95% confidence interval (CI) for the sensitivity and specificity using the Agresti–Coull method [16,18]. The study design and report of results followed the STAndards for Reporting Diagnostic accuracy studies (STARD) guidelines as described in Supplementary Table 1.

According to the study protocol, prerequisites for the selection of patients were age above 18, CC stage I, II, and III, written consent, and known microsatellite status. All samples are surgical specimens, and MSI status was assessed centrally using immunohistochemistry in combination with PCR-based fragment length analysis as described recently [19]. The selection aimed to balance parameters of age, sex, localisation, BRAF, and RAS mutational status. Randomly assigned training (optimisation of the neural network) and testing (model selection) samples were chosen with a greater proportion of MSI cases to stabilise the training process with a more balanced dataset. Samples for the validation cohort were randomly selected and therefore show a natural occurrence of MSI cases. The full study flowchart is illustrated in Supplementary Figure 1.

2.2. Sample processing

Tissue samples were submitted from participating sites and collaborating institutes of pathology as formalin-fixed, paraffin-embedded (FFPE) samples and therefore passed through different, locally established, preanalytic procedures. For IR imaging measurements using the QCL-based IR microscope, 7 µm tissue sections were mounted on IR-suitable polyethylene terephthalate frame slides and measured as described previously [20]. Subsequently, samples were H&E stained and digitised using an Olympus slide scanner (Olympus VS120).

2.3. Infrared imaging

The QCL-based microscope was used in the standard transmission configuration (Supplementary Figure 2) [15]. The light sources are QCLs implemented in the platform Spero QT (Daylight Solutions, San Diego, CA, USA). This QCL platform analyses tissue sections within 30 min including AI-based classification. Detailed measurement parameters are provided in Supplementary Table 2, and study workflow is illustrated in Supplementary Figure 3.

2.4. Deep learning algorithms

For classification of the IR images, we used convolutional neural networks (CNN) in a two-step setup: during step 1, cancer areas were detected within the IR images, while in step 2, MSI/MSS was classified within the detected tumour tissue (Fig. 1).

The step 1 CNN is an in-house developed weakly supervised extension of a U-net architecture, named comparative segmentation network (CompSegNet) [20,21], resulting in pixel-precise cancer detection. The CompSegNet was trained using subregions from the whole slide IR images that were assigned by an experienced pathologist (A.T.). The CompSegNet training was conducted on these manually selected and binary labelled subregions of size 256×256 pixels (px) (Supplementary Info 1). Each region that was labelled as cancer contained a minimum area of 20% cancer. Cancer areas identified in step 1 were then used for the step 2 CNN – a VGG classification network (see Fig. 1) – for the training of the MSI/MSS classifier [22]. The technical parameters for the deep learning workflow are presented in Supplementary Table 3. For the application of the CNNs on whole slide IR images, the detection/classification was performed for each subregion in the whole slide IR image and merged for a whole slide result.

2.5. Statistical analysis

The statistical evaluation was carried out in two levels: on the single input subregions and on the result for the whole slides (averaged for all subregions) on patient level. The subregion-based evaluation serves as assessment of the technical AI approach, whereas the whole slide evaluation reflects the performance relevant for clinical application (details in Supplementary Info 2).

The performance of the classification process was evaluated using AUROC and area under precision recall curve (AUPRC) defining MSI samples as positive and MSS samples as negative. Here, 95% robust pointwise CI for AUROC and AUPRC were generated by bootstrapping. In addition, sensitivity, specificity, accuracy, and F1 score (definition in Supplementary Info 4) were determined by selecting a cut-off value for the prediction (range: 0 to 1) balancing sensitivity and specificity on the test cohort. Additionally, dice score was determined for step 1 cancer detection using 25 randomly selected samples from the validation cohort.
3. Results

3.1. Patients

As shown, 633 patients were selected from the AIO CPP 2.0. Four samples were excluded before the analysis after quality check of the IR measurements. The step 1 cancer detection cohort therefore consisted of 629 patients as shown in Supplementary Table 4. A total of 10,942 subregions were selected manually from the IR measurements (training: 4955, test: 2348, validation: 3639) representing cancer and cancer-free regions and used for training, model selection, and statistical evaluation.

The step 2 MSI/MSS classifier sub-cohort included 547 patients (patients with tissue sections containing cancerous regions from the abovementioned cohort) as presented in Table 1 and Supplementary Figure 1. Patients with only normal tissue sections (n = 80) or samples with insufficient cancer material for further analysis (n = 2) were excluded. The characteristics are equally distributed except MSI for methodological reasons (see Material and methods section).

3.2. Label-free cancer detection

The step 1 CNN achieved an AUROC of 1.0 (95% CI 1.000) for whole slide classification, thus, forming the basis for the MSI/MSS classifier. All parameters (AUROC, AUPRC, F1 score, sensitivity, specificity, and accuracy) achieved a value of 1.0 on the whole slide level for cancer classification (see Fig. 2A and B), and the cancer detector reaches a dice score of 0.5. As shown in Fig. 2, the detection of cancer and cancer-associated tissue performs very precisely. Whole slide images illustrating the step 1 cancer detection are shown in Fig. 2C with pixel-precise detection illustrated in green. A cancer-free sample (NT = no tumour) is shown for comparison.

3.3. MSI/MSS classification

The step 2 MSI/MSS classifier was trained on cancer subregions detected by step 1. 1500 epochs (training curves in Supplementary Figure 4) were trained, creating an equal number of classifier models. The best
performing model from epoch 1000 was selected based on the AUROC value of the test set. Fig. 3 summarises the statistical results and illustrates the application on whole slide images. The subregion-based evaluation performance was comparable to the whole slide evaluation, with AUROC values of 0.87 (95% CI 0.78–0.97) and 0.90 (95% CI 0.83–0.97) and AUPRC values of 0.89 (95% CI 0.80–0.98) and 0.74 (95% CI 0.59–0.89) obtained for test and validation, respectively, in the whole slide setting. Classification on the validation cohort had 85% sensitivity and 84% specificity. ROC curves, PR curves, and statistical parameters characterising the three cohorts are provided in Supplementary Figure 5 and 6, including a detailed analysis of the 95% CI. Steps 1 and 2 are summarised in Fig. 3C illustrating the process of cancer detection and subsequent MSI/MSS classification within the cancer regions. For the whole slide results of step 2, green represents high predictive values for MSI, whereas purple represents high predictive values for MSS.

#### 4. Discussion

At present, two standard reference methods, IHC and PCR, are recommended for the detection of MSI in CRC. IHC is considered the gold standard in routine practice followed by other molecular approaches such as PCR when IHC is equivocal. Next-generation sequencing has been reported to exhibit high concordance with IHC and PCR and has recently been approved by the Federal Drug Administration in aCRC [24]. All techniques listed above use formalin-fixed and stained tumour tissue, thereby limiting its availability for further tissue-based analyses. MSI has long been known as a positive prognostic marker in early CRC. Most guidelines now recommend testing for MSI in patients with low-risk stage II cancer to avoid the use of ineffective fluoropyrimidines as adjuvant treatment in this setting [5,30]. Recently, MSI has emerged as a crucial predictive biomarker for the efficacy of immune checkpoint inhibition (e.g. pembrolizumab) which is now regarded standard of care in MSI aCRC and is also evaluated in early CRC in clinical trials [3]. The diagnosis of MSI must therefore be highly reliable. Presently, misdiagnosis by IHC has been reported to be as high as 10% resulting in a substantial disadvantage for the individual patient [7]. With the rapid IR imaging technique (30 min turn-around time), we were able to reliably distinguish MSI from MSS in the tissue of patients with early CC achieving high sensitivity and specificity, comparable to the present gold standard (IHC) [7,25]. With this novel technique tissue samples are kept available for further molecular assessment.

We developed and validated a novel automated and label-free digital pathology approach to distinguish MSI from MSS reliably and rapidly in CC tissue samples. FFPE samples from a total of 547 patients from the German AIO CPP registry trial were analysed with centrally determined MSI/MSS status. These tissue samples were initially processed by local standards at different institutes of pathology by investigators, thus

---

**Table 1**

Patient cohort for MSI/MSS classification.

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Training</th>
<th>Test</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSI + MSS</td>
<td>MSI</td>
<td>MSS</td>
<td>MSI</td>
</tr>
<tr>
<td>N (% of train/test/vali)^a</td>
<td>547</td>
<td>142 (43)</td>
<td>189 (57)</td>
<td>30 (43)</td>
</tr>
<tr>
<td>Sex f/m in %b</td>
<td>50/50</td>
<td>64/36</td>
<td>106/64</td>
<td>57/43</td>
</tr>
<tr>
<td>Age Mean ± SD (years)</td>
<td>69 ± 12</td>
<td>71 ± 14</td>
<td>68 ± 12</td>
<td>73 ± 12</td>
</tr>
<tr>
<td>Location left (%b)</td>
<td>211 (39)</td>
<td>30 (21)</td>
<td>98 (52)</td>
<td>6 (20)</td>
</tr>
<tr>
<td>Right (%b)</td>
<td>331 (61)</td>
<td>112 (79)</td>
<td>90 (48)</td>
<td>24 (80)</td>
</tr>
<tr>
<td>UICC I (%b)</td>
<td>13 (2)</td>
<td>9 (6)</td>
<td>1 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>II (%b)</td>
<td>152 (28)</td>
<td>64 (45)</td>
<td>37 (20)</td>
<td>14 (47)</td>
</tr>
<tr>
<td>III (0%)</td>
<td>382 (70)</td>
<td>69 (49)</td>
<td>151 (80)</td>
<td>14 (47)</td>
</tr>
<tr>
<td>Grading G1 (%)</td>
<td>21 (4)</td>
<td>8 (6)</td>
<td>3 (2)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>G2 (%)</td>
<td>343 (63)</td>
<td>59 (42)</td>
<td>144 (76)</td>
<td>11 (37)</td>
</tr>
<tr>
<td>G3 (%)</td>
<td>160 (29)</td>
<td>67 (47)</td>
<td>36 (19)</td>
<td>16 (53)</td>
</tr>
<tr>
<td>G4 (%)</td>
<td>21 (4)</td>
<td>4 (6)</td>
<td>6 (3)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>unknown (%b)</td>
<td>152 (28)</td>
<td>64 (45)</td>
<td>37 (20)</td>
<td>14 (47)</td>
</tr>
</tbody>
</table>

MSI, microsatellite instability; MSS, microsatellite stability; SD, standard deviation. The numbers for the total cohort are printed in bold.

^a Percentage calculated for training (train)/test/validation (vali) cohorts.

^b Percentage calculated column-wise and are printed in italics.
increasing variation in the dataset and therefore adding to the robustness of the method. This includes slight variations in the thickness of the tissue sections, which did not affect the classification results. In the first step, the approach was able to classify cancer with an AUROC of 1.0 (95% CI 1.00–1.00) and a dice score of 0.5 in the validation cohort of 218 patients. MSI vs. MSS could be correctly classified with an AUROC value of 0.9 (95% CI 0.83–0.97) and an AUPRC value of 0.74 (95% CI 0.59–0.89) in a validation cohort of 147 patients. The AUROC in the validation cohort increased compared to the test cohort, which may be due to a limited number of MSI cases in the validation cohort following the natural occurrence of MSI. Overall, the
modified architecture of the VGG network exhibited a stable training behaviour. Our approach could thus also be reliably used in a multicentre setting and has the potential to be transferred to institutions and caregivers globally in the future.

The label-free digital pathology approach described here can detect MSI vs. MSS with sensitivity and specificity comparable to the approach, which has been previously described on H&E-stained images [12,26,27]. Compared to the results from Yamashita et al., the
The label-free approach achieves similar accuracy without the need for staining or other processing of the sample. The H&E-based approaches by Echle et al. showed best results on a 5500-patient cohort with an AUROC of 0.92. In their study, training sets consisting of 500 patients reached an average AUROC of only 0.82. The label-free approach described here achieved an AUROC of 0.9 (95% CI 0.83–0.97) on an even smaller 331-patient training dataset, emphasizing the information density that IR imaging can retrieve in unstained samples. This is an advantage of IR imaging regarding smaller molecular or clinical subgroups as well as potentially detecting novel predictive biomarkers, e.g., from randomised trials with limited numbers of participants.

This label-free approach enables the detection of areas of interest in unstained tumour material, making, beyond the immediate clinical application, tumour heterogeneity analysis, elucidation of underlying pathways, and provision of enriched material for multi-omics approaches possible [13,28,29]. In contrast to other stain-based methods, the label-free IR approach allows pixel-precise dissection on the same sample slide the detection has been performed on.

The determination of MSI is also clinically relevant in multiple other solid tumours such as gastric or endometrial cancer. Various other molecular alterations are of impact on clinical management (RAS, BRAF, HER2), which have different prevalence in MSI vs. MSS cancers. Therefore, this reliable MSI/MSS classification method may also serve as a screening method to identify patients for subsequent molecular analyses.

5. Conclusion

This label-free digital pathology approach for MSI/MSS classification is a novel and valid tool for stratification in precision oncology. The analysis of unstained tissue sections moreover allows targeted, subsequent molecular analysis on microdissectable, predefined areas of interest. Overall, this method has the potential to become an applicable and valuable diagnostic tool for stratification beyond the scope of known biomarkers for tumour-agnostic assessments.

Credit author statement


Conflict of interest statement

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dirk Arnold reports Consultation/Advisory role from AstraZeneca, Bristol Myers Squibb, Merck Sharp & Dohme, Sanofi (Genzyme), Bayer, Terumo, Roche, Hexal, Samsung Bioepis, Pierre Fabre Pharma, CRA International, Ketchum, IQVIA; he reports speaker’s engagement from AstraZeneca, Bristol Myers Squibb, Merck Sharp & Dohme, Sanofi (Genzyme), Bayer, Eli Lilly, Terumo, Boston Scientific, Amgen, Roche, Ipsen, Merck (Serono), Samsung Bioepis, Pierre Fabre Pharma, Servier, PharmaCept, PRMA Consulting, Tactics MD LLC, WebMD Health Corp, From Research to Practice, Aptitude Health, art tempi media, ACE Oncology, Imexed, streamitup Germany, MedAhead (Austria), Clinical Care Options (CCO); he reports Local PI for Bristol Myers Squibb, Pierre Fabre Pharma and Coordinating PI of OncoLytics; he reports grant funding from AbbVie; he reports DSMB chair of Sanofi (Genzyme); he reports Steering Committee Member of Roche. Anke Reina cher-Schick reports Consulting or Advisory Role from Amgen, Merck Serono, Roche Pharma AG, Bristol-Myers Squibb, MSD, AstraZeneca, Pierre Fabre; she reports Honoria from Amgen, Merck Serono, Roche Pharma AG, Bristol-Meyers Squibb, MSD, MCI Group, AstraZeneca; she reports research funding for her institution from Celgene (Inst), Roche Pharma AG (Inst), Ipsen, Amgen, Alexion Pharmaceuticals, AstraZeneca, Lilly, Servier, AIO-Studien, Rafael Pharmaceuticals, ERYTECH Pharma, BioNTech; she reports Travel/Accommodations/Expenses from Roche. Andrea Tannapfel reports Honoraria from Amgen, Roche, Merck Serono, Bristol-Meyers Squibb, MSD, AstraZeneca; she reports research funding from
Roche, Amgen, BioNTech. All remaining authors have declared no conflicts of interest.

**Funding**

This work was supported by the Ministry of Culture and Science (MKW) of the State of North-Rhine Westphalia, Germany [grant number: 111.08.03.05–133974] and by Roche Pharma AG to the Colo-Predict Plus 2.0 registry (no grant number). Part of this research was conducted within the Slide2Mol project funded by the Computational Life Sciences program of the German Federal Ministry of Education and Research, grant number 031L0264.

**Data sharing**

The spectral and medical data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgements**

The authors thank the patients participating in the CPP registry study and their families for allowing us to realise this research work. The authors further thank the staff involved at the CPP study sites all over Germany for their contribution. The authors thank Dr. Claus Küpper and Melina Helfrich for their support in data acquisition, Lidia Janota for her assistance in sample preparation, and Dr. Oliver Overheu for his valuable help in describing the cohort.

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2022.12.026.

**References**


