

Article

Deep Learning Applied to Raman Spectroscopy for the Detection of Microsatellite Instability/MMR Deficient Colorectal Cancer

Nathan Blake ¹, Riana Gaifulina ¹, Lewis D. Griffin ², Ian M. Bell ³, Manuel Rodriguez-Justo ⁴
and Geraint M. H. Thomas ^{1,*}

¹ Department of Cell and Developmental Biology, University College London, London WC1E 6BT, UK

² Department of Computer Science, University College London, London WC1E 6BT, UK

³ Spectroscopy Products Division, Renishaw PLC, Wotton-under-Edge GL12 8JR, UK

⁴ Department of Research Pathology, Cancer Institute, University College London, London WC1E 6DD, UK

* Correspondence: g.thomas@ucl.ac.uk; Tel.: +44-20-3549-5456

Simple Summary: Colorectal cancer has several disease pathways which have implications for how patients are monitored and treated. One important pathway is caused by deficiencies to genes responsible for repairing pre-cancerous cells. Methods to detect these deficiencies exist, but are not implemented as often as recommended and could be improved. Raman spectroscopy is a technique that could provide such an improvement, having shown potential in other areas of cancer research. The full potential of Raman datasets may be achieved by exploiting modern machine learning models. We evaluated a small colorectal tissue dataset to assess the viability of some common machine learning techniques to detect different colorectal cancer pathways. We find that Raman spectroscopy in conjunction with machine learning could be a viable means of improving screening and potentially diagnostic tools and warrants further research with larger sample sizes.

Abstract: Defective DNA mismatch repair is one pathogenic pathway to colorectal cancer. It is characterised by microsatellite instability which provides a molecular biomarker for its detection. Clinical guidelines for universal testing of this biomarker are not met due to resource limitations; thus, there is interest in developing novel methods for its detection. Raman spectroscopy (RS) is an analytical tool able to interrogate the molecular vibrations of a sample to provide a unique biochemical fingerprint. The resulting datasets are complex and high-dimensional, making them an ideal candidate for deep learning, though this may be limited by small sample sizes. This study investigates the potential of using RS to distinguish between normal, microsatellite stable (MSS) and microsatellite unstable (MSI-H) adenocarcinoma in human colorectal samples and whether deep learning provides any benefit to this end over traditional machine learning models. A 1D convolutional neural network (CNN) was developed to discriminate between healthy, MSI-H and MSS in human tissue and compared to a principal component analysis–linear discriminant analysis (PCA–LDA) and a support vector machine (SVM) model. A nested cross-validation strategy was used to train 30 samples, 10 from each group, with a total of 1490 Raman spectra. The CNN achieved a sensitivity and specificity of 83% and 45% compared to PCA–LDA, which achieved a sensitivity and specificity of 82% and 51%, respectively. These are competitive with existing guidelines, despite the low sample size, speaking to the molecular discriminative power of RS combined with deep learning. A number of biochemical antecedents responsible for this discrimination are also explored, with Raman peaks associated with nucleic acids and collagen being implicated.

Keywords: Raman spectroscopy; deep learning; oncology; microsatellite instability; diagnostics; colorectal cancer



Citation: Blake, N.; Gaifulina, R.; Griffin, L.D.; Bell, I.M.; Rodriguez-Justo, M.; Thomas, G.M.H. Deep Learning Applied to Raman Spectroscopy for the Detection of Microsatellite Instability/MMR Deficient Colorectal Cancer. *Cancers* **2023**, *15*, 1720. <https://doi.org/10.3390/cancers15061720>

Academic Editor: Antonio Randazzo

Received: 13 January 2023

Revised: 8 March 2023

Accepted: 9 March 2023

Published: 11 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Colorectal cancer (CRC) encompasses cancers of the large colon and rectum. It is one of the major malignancies of the world, being the third most commonly occurring and the second most deadly cancer, with an estimated 1.8 million new cases and 881,000 deaths worldwide in 2018 [1]. With a few exceptions, CRC incidence is increasing globally, particularly in developing countries, as shifting dietary and lifestyle factors are likely driving an increase in early onset CRC [2].

There are several known pathological pathways leading to CRC, resulting in heterogeneous presentations, therapies and outcomes. One such pathway is DNA mismatch repair deficient (dMMR) CRC, in which there are pathological alterations to any of a number of MMR genes (MLH1, MSH2, MSH6 or PMS2). This loss of MMR function causes high-level microsatellite instability (MSI-H), characterised by mononucleotide, dinucleotide and trinucleotide tandem repeats. This can occur sporadically or as an inherited trait, as in Lynch syndrome (LS). Hence, the detection of MSI-H is recommended in every case of CRC to screen for LS [3]. The high mutational burden seen in MSI-H tumours also has implications for treatment, providing potential targets for immunotherapy such as immune checkpoint inhibitors [4].

Despite recommendations for universal testing for MSI-H for all CRC cases, resource limitations mean that this cannot always happen and is particularly poor for young adults [5]. Testing for dMMR/MSI-H typically involves either immunohistochemistry (IHC) of the mismatch repair proteins or PCR amplification of consensus microsatellite repeats. Recent developments in machine learning (ML) have led to the possibility of exploiting morphological information in standard H&E slides [4,6]. Such digital pathology techniques require few additional resources and have proven highly accurate in high quality, curated datasets. However, consistent with other domains using modern ML, these promising results do not generalise well when applied to settings or cohorts outside of the context in which they were developed [6]. The literature has thus far focused on H&E stained slides but ML applied to other histological stains, such as IHC, may yield further improvements by exploiting molecular-level information. Raman spectroscopy is a technique which opens the possibility of digitally staining a sample for many biomolecules, without the need for specific sample preparation [7].

Raman spectroscopy (RS) is a technique which interrogates molecular vibrational states through inelastically scattered photons from a sample. As the vibrational modes of any given molecule are unique, it is possible to identify a molecule through its Raman spectrum. A Raman spectrum represents the change in photon wavenumber from a monochromatic light source along the x-axis and the intensity (i.e., number of photons) thus scattered on the y-axis. RS has been successfully applied to discriminate between numerous cancer types in human tissues, most recently to the brain, breast, cervical, colon, lung, nasopharyngeal, prostate, skin and tongue [8]. The applications include early diagnosis, biopsy guidance and intraoperative tumour margin detection. As a molecularly sensitive modality, RS may be able to detect the molecular antecedents of microsatellite stable (MSS) and MSI-H samples.

Despite promising results, RS has yet to become routinely used in the oncology setting for several reasons, including technical limitations in *in vivo* applications, a lack of visibility in the medical literature and a lack of thorough validation of the models on truly independent datasets [9]. Additionally, the inherent complexity and high dimensionality (number of discrete wavenumber points per spectrum) of biomedical RS data, compounded with various sources of noise, such as fluorescence, necessitates the use of mathematical modelling to extract coherent molecular information.

Deep learning (DL) is a family of modelling techniques under the umbrella of ML models. Its ability to capture non-linear relationships make it suited to complex clinical datasets, and it may help unleash the potential of RS. In addition to being increasingly applied to oncology problems in the context of digital histopathology [10], it is now extending into oncological applications of RS [8]. However, DL is notoriously data-intensive, which makes it difficult to apply to smaller medical proof-of-concept studies. Combined with

the large size of DL models, it is susceptible to over-fitting, in which excellent results on a dataset fail to transfer across to general settings. However, with developing techniques such as data augmentation and strict validation practices, it may still be possible to leverage DL even with small datasets.

This proof-of-concept study explores the potential of RS to distinguish between healthy, MSS adenocarcinoma (AC) and MSI-H AC in human tissue. A DL model is developed alongside two traditional ML models commonly used in RS, taking great methodological care not to produce overly optimistic performance estimates. Biochemical antecedents responsible for the DL model's discriminative performance are then inferred. The exploratory nature of this study seeks only to assess whether RS applied to this particular clinical problem merits additional studies, to highlight some of the methodological considerations for such a larger study and to explore potential clinical uses such as screening or diagnosis.

2. Materials and Methods

2.1. Tissue Acquisition and Processing

Formalin fixed paraffin embedded (FFPE) human colon samples were obtained from the UCL/UCLH Biobank for Studying Health and Disease (REC 20/YH/0088). A total of 10 FFPE samples of resection margins of normal colonic mucosa from sporadic CRC cases were obtained along with 10 MSS/MMR proficient samples from the same patients. A total of 10 archival MSI-H samples were also obtained and matched to the sporadic AC samples by cancer stage (T-stage, see Appendix D for details).

From each sample, one section was cut at 8 μm thickness and mounted onto silanised 304L super mirror stainless steel slides for Raman analysis and one 3 μm section was cut and mounted onto standard glass slides for H&E staining. The steel mounted samples were prepared as described by Gaifulina et al. [11]. Steel slides have been shown to improve Raman signal acquisition by up to a factor of four and reduce background signal compared to calcium fluoride (CaF_2), the standard substrates often used in RS [11,12], and are far cheaper.

Unstained tissue sections were immersed in a series of baths to remove paraffin wax. Four successive ten-minute baths in xylene (VWR International Ltd., Lutterworth, UK) with gentle agitation, were followed by a series of rehydration steps in graded ethanol absolute (VWR International Ltd., Lutterworth, UK), followed by a final immersion in distilled water for ten minutes.

The 3 μm sections were subject to standard automated staining and cover-slipping for H&E slides. The histopathology of all samples was confirmed by a resident consultant pathologist. A full breakdown of the patient samples and cancer stages can be found in Appendix D Table A2.

2.2. Raman Spectroscopy

Point spectra were acquired using the Renishaw prototype RA800 series benchtop system with a 785 nm laser (Renishaw plc, Wotton-under-Edge, UK). A total laser intensity of approximately 158 mW was focused onto samples through a 50 \times NA 0.8 objective. A 1500 L/mm grating was used to disperse the light providing a spectral range of 0–2100 cm^{-1} in the low wavenumber range. An integration time of 20 s was used for all measurements. A total of 50 individual spectra were collected from each tissue sample, except for one sample with only 40 spectra, resulting in a total of 1490 across the 3 classes. All spectra were acquired from the glandular mucosal region in normal samples and from confirmed cancerous regions in all cancer samples, located by the resident pathologist prior to Raman measurement.

2.3. Modelling and Cross-Validation Strategy

Cosmic rays were removed using a combination of the width of feature and nearest neighbour methods available in the manufacturer's software. Spectra were visually inspected at the time of acquisition and any saturated spectra were discarded and a new

spectrum obtained from a different region. Each spectrum was standard normal variate (SNV) normalised to have zero mean and unit variance.

Baseline correction was not performed. An initial analysis showed that baseline correction via several methods (references [13–16]) did not improve performance, and significantly altered resulting mean and difference spectra which impacted the biochemical interpretation (details in Appendix A).

A principal component analysis–linear discriminant analysis (PCA–LDA), a support vector machine (SVM) and a convolutional neural network (CNN) were developed using SNV normalised Raman spectra truncated to the range of 400–1800 cm^{-1} . PCA–LDA is one of the most commonly used ML model in biomedical RS. PCA reduces the dimensionality of data, and LDA constructs a linear decision boundary for classification. SVM is an ML model which can construct non-linear decision boundaries by using, amongst others, a radial basis function kernel. These are both traditional ML models. A CNN is a DL model which has become popular in medical imaging. It also constructs non-linear decision boundaries and its invariance to certain data inputs make it robust to irrelevant features in the data. Figure A2 provides an overview of the custom CNN developed for this application. These models represent increasingly complex modelling techniques that can be applied to the data. Often with small datasets, a simpler model such as PCA–LDA will perform best as it does not overfit the data, which complex models such as CNNs are prone to do. However, given the complexity of RS data, simpler models may not be able to capture sufficient nuances to be useful. It is not clear which modelling technique would best facilitate discriminating between disease classes and, therefore, a secondary aim of this study is to compare the performance of this range of models.

These models all have hyperparameters, which can be understood as decisions a researcher may make to optimise performance. However, it is possible to over-optimize these hyperparameters so that a model performs well on the research data, but does not generalise to new data. To mitigate against this risk, a repeated nested cross-validation (CV) strategy was used. This allows for hyperparameters to be optimised in an inner CV loop, and then tested against held-out data in an outer CV loop (Figure 1). Nested CV has been shown to reduce a model’s estimated accuracy by as much as 20% in oncological applications of RS, giving a more realistic assessment of the model’s generalisability [8]. A single metric needs to be selected to optimise: the log loss was chosen as it is a proper scoring metric which utilises distributional information, compared to typical binarised metrics such as accuracy [17].

Additionally, we use an occlusion study in order to determine those regions of a spectrum which the CNN uses during classification. This technique involves sequentially “blinking” out regions of an input and returning a prediction. This occluded prediction is then compared to the whole spectrum prediction. A reduction in prediction suggests that the occluded spectral region contained an important feature for the CNN. This technique has, for instance, been used to highlight diagnostically significant brain regions from MRI scans of Alzheimer’s patients [18].

Each sample has at least 40 Raman spectra. The interest in this application is the overall sample label rather than that of individual spectra. Therefore, sample labels were determined by taking all the spectrum-level disease classifications and using a simple majority voting consensus to return an overall label for the sample. Thus, all spectra are used for model construction but a single prediction per sample was obtained. During voting, any ties were to be broken by classifying to the clinically worst disease label, but there were no ties. Data were split into training/test sets on the basis of samples rather than spectra, ensuring that spectra from the same sample were not present in both training and test (or validation) sets, thus maintaining the independence assumption required for model validation. This has been shown to return far more realistic measures of ML performance [8].

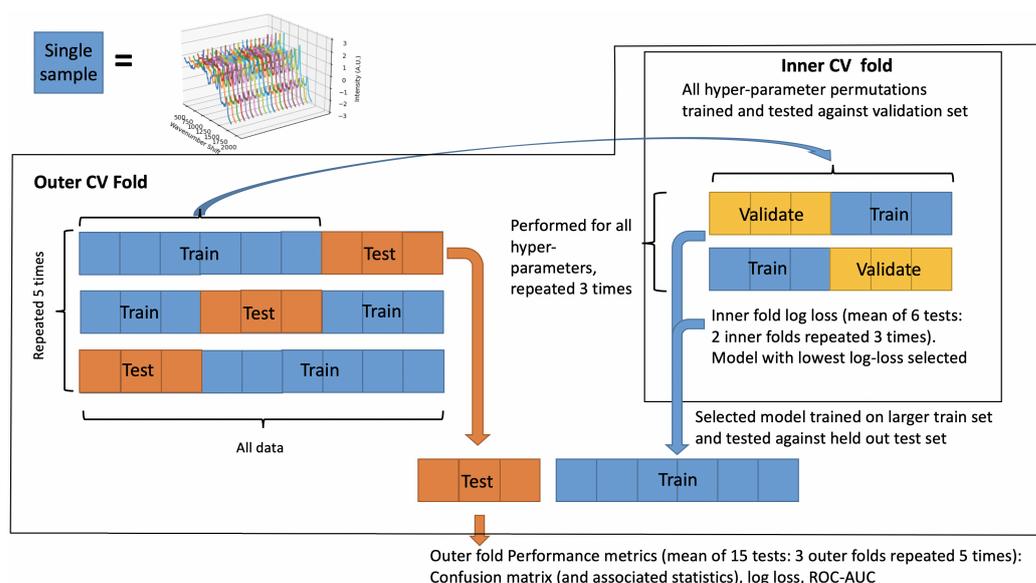


Figure 1. Nested CV strategy.

Data augmentation was used to supplement the training of the CNN. This is a technique which creates new spectra by replicating existing spectra and adding noise. This helps train the CNN to ignore those perturbed features, reducing over-fitting. This was achieved by adding Poisson noise to the Raman intensity (before normalisation) and adding a random wavenumber shift of no more than $\pm 6 \text{ cm}^{-1}$ (details in Appendix C). Data augmentation was performed inside the nested CV loops after the data had been split, increasing the training set size by a factor of 8. This technique is a strictly computational method, and does not seek to simulate biomedical Raman spectra.

Analyses were conducted using Python version 3.10 using the Scikit-learn library [19] for the PCA-LDA and SVM models and PyTorch to develop the CNN [20].

Our results are compared to existing diagnostic and screening tests and benchmarks. Current UK guidelines recommend that all CRC patients are offered IHC testing for MMR proteins (a histology-based technique) or MSI-H testing (a genetic-based technique) [21]. If either of these is positive, then subsequent tests are conducted, including genetic testing of germline DNA to detect LS. These tests are two class models, distinguishing in the first instance between dMMR/MSI-H in samples already diagnosed as CRC. To fit into this existing clinical pipeline, two class models were developed using the nested CV strategy described above. The principle performance metrics used to assess IHC and MSI-H testing in the UK guidelines are sensitivity and specificity and these are reported below, alongside the area under the receiver operating characteristic (AUROC) curve.

3. Results

The three-fold CV strategy, repeated five times, returned 15 estimates of the performance of each model, allowing for an average performance to be calculated. It was not possible to construct confidence intervals from these results as each CV fold contains overlapping data, violating the independence assumption required for confidence intervals. In lieu of confidence intervals, standard deviations are reported.

3.1. Spectral Data Analysis

Spectra belonging to the same disease class were averaged and are shown in Figure 2. Visual inspection reveals little appreciable difference, as the general composition of the tissues is similar. Together with the standard deviations of the average spectra, this shows how subtle the biochemical differences are between disease classes. To better contrast these subtleties, a difference spectrum was obtained by subtracting the average MSS spectrum from the average MSI-H spectrum (Figure 3). From this, it is possible to infer some

biochemical differences between the classes. In particular, peaks at 714, 1081, 1302 and 1445 cm^{-1} have tentatively been assigned to lipids. Other peak assignments include 1672 cm^{-1} (cholesterol), 494 cm^{-1} (glycogen, nucleic acids), 529 cm^{-1} (amino acids), 732 cm^{-1} (phosphatidylserine, adenine), 787 cm^{-1} (nucleic acids), 852 cm^{-1} (ring-breathing mode of proline, hydroxyproline, tyrosine), 1003 and 1034 cm^{-1} (phenylalanine, polysaccharides), 1110 cm^{-1} (lipids, proteins), 1366 cm^{-1} (tryptophan, lipids, guanine) and 1583 cm^{-1} (C-C bending mode of phenylalanine). Overall, these indicate differences in nucleic acids, proteins and lipids between the two classes. Band assignments were made using findings contained within the work of Movasaghi et al. [22].

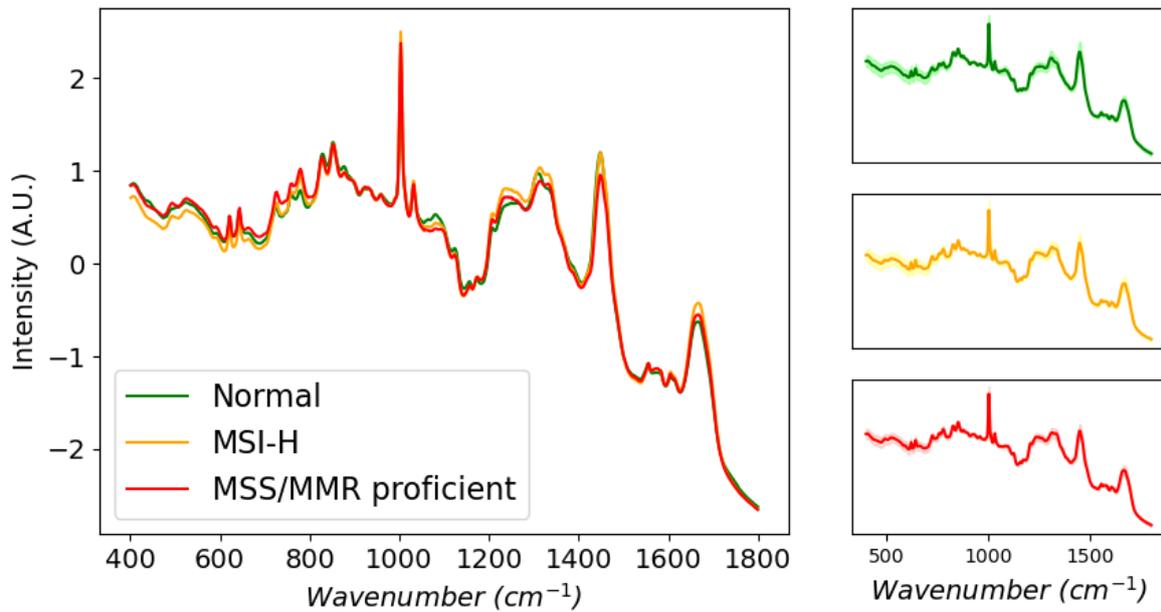


Figure 2. Average normalised spectrum by class. Right panels, average spectra with shaded areas indicating 1 standard deviation.

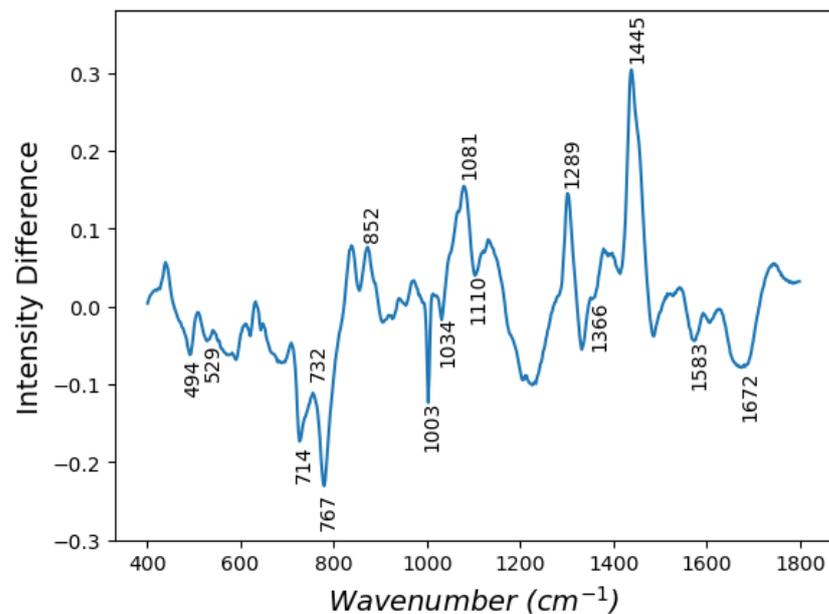


Figure 3. Difference spectrum: MSI-H minus MSS. Numbers indicate peaks mentioned in the text.

3.2. Two-Class Model

Current UK guidelines recommend that all CRC patients are offered IHC testing for MMR proteins or MSI-H testing [21]. If either of these is positive, then subsequent tests are conducted, including genetic testing of germline DNA to detect LS. These tests are two-class models, distinguishing between dMMR/MSI-H in samples already diagnosed as CRC. In this section, a two-class model was developed to this end, consisting of samples that have already been diagnosed as CRC but requiring discrimination between MSI-H and MSS. The principle performance metrics used to assess IHC and MSI-H testing in the UK guidelines are sensitivity and specificity, and they are reported below, alongside the area under the receiver operating characteristic (AUROC) curve. Results from all three ML models are shown in Table 1.

Table 1. Two-class models: mean sensitivity, specificity and AUROC across all folds ± 1 standard deviation.

	PCA-LDA	SVM	CNN
Sensitivity	70.0% \pm 36.1	85.6% \pm 21.0	73.0% \pm 10.0
Specificity	62.8% \pm 27.5	32.8% \pm 15.7	48.9% \pm 12.5
AUROC	0.65 \pm 0.21	0.71 \pm 0.16	0.75 \pm 0.15

The SVM model performs best in terms of sensitivity at 85.6%, but trades heavily for this with a low specificity of 32.8%, while the PCA-LDA is best for specificity at 62.8%. The CNN performance is between these two traditional ML models. The large standard deviations are likely an artefact of the small sample size. This makes it difficult to draw any conclusions regarding the superiority of any of the models. Despite the CNN being the largest model, and therefore more prone to over-fitting, it returns the lowest variance, indicating a more stable model.

Binary ML models typically do not give a classification but a prediction, a numerical output between 0 and 1, where in this case 0 represents MSI-H and 1 MSS. To calculate sensitivities and specificities, a threshold needs to be applied to the model outputs. A separate, and much larger, study would be required to calibrate the optimal balance of sensitivity and specificity desirable for this application. In lieu of such calibration, the sensitivities and specificities reported use the standard threshold of 0.5. An ROC curve summarises performance over a range of thresholds, giving an indication of how the models may perform under different calibration conditions. The CNN achieves the best AUROC at 0.75. This is often considered a good performance, though this is context-dependent (Figure 4).

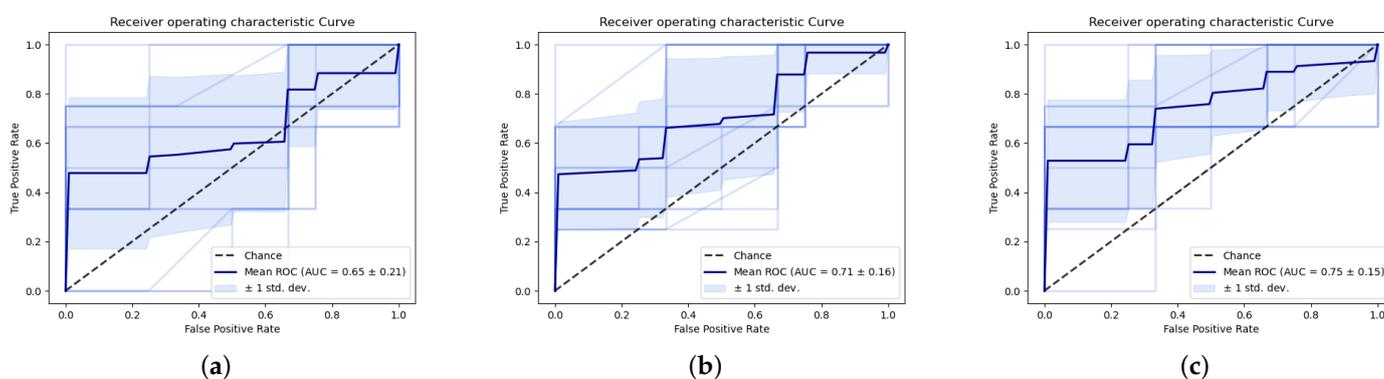


Figure 4. Receiver Operating characteristic curve for (a) PCA-LDA (b), SVM (c), CNN. Bold lines indicate mean ROC, pale lines performance for individual folds and shaded area 1 standard deviation.

3.3. Occlusion Study

In Section 3.1, we explored the mean and difference spectra in an attempt to find the biomolecular markers which distinguish LS from AC. However, it is not necessarily

the case that a model learns to use those particular features to make its prediction. To elucidate this information, we use an occlusion study. This technique involves sequentially “blanking” out regions of an input and returning a prediction. This occluded prediction is then compared to the whole spectrum prediction. A reduction in prediction suggests that the occluded spectral region contained an important feature for the CNN. This technique has, for instance, been used to highlight diagnostically significant brain regions from MRI scans of Alzheimer’s patients [18].

The region between 680 and 1020 cm^{-1} seems to be diagnostically significant to the CNN (Figure 5). In particular, the regions between 680 and 710 cm^{-1} (associated with the ring breathing modes of DNA, C-S bond of methionine and C-N bond of phospholipids and adenine), 800–830 cm^{-1} (uracil-based ring breathing mode, O-P-O stretching and C₅-O-P-O-C₃ phosphodiester bands of RNA, PO₂⁻ stretch of nucleic acids as well as C-C stretching in collagen and proline and hydroxyproline) and 870–900 cm^{-1} (C-C stretching of collagen and C-O-C skeletal mode of monosaccharides, disaccharides and adenine) [22]. Overall, these regions are largely associated with nucleic acids, consistent with findings according to which MSI-H cancers tend to be diploid rather than aneuploid, as well as collagen. The latter is consistent with recent findings according to which *COL11A1* mutations, affecting non-fibrillar collagen expression in the extracellular matrix of MSS colonic and ovarian tissues, may be a useful oncological biomarker [23].

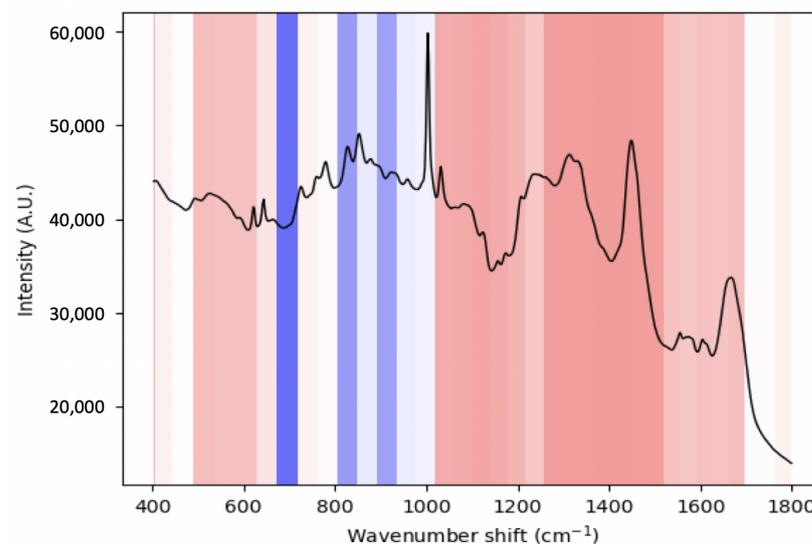


Figure 5. Occlusion study: Blue indicates drops in performance due to occlusion. The stronger the shade, the larger the drop in performance.

3.4. Three-Class Model

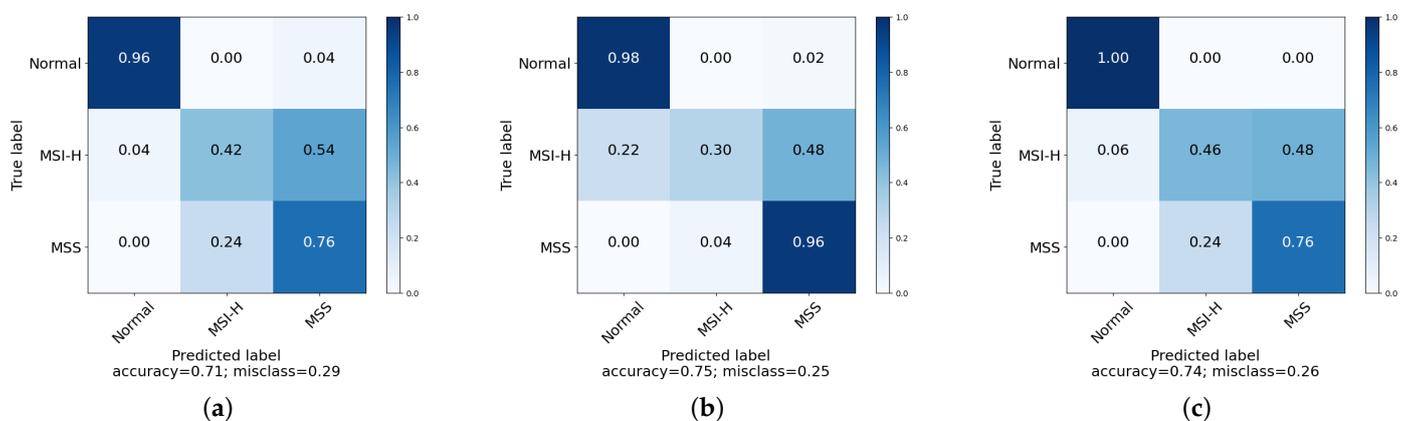
Although current practice is to test for dMMR/MSI-H only on confirmed cases of CRC, RS allows for the possibility of screening all suspected CRC samples by using a three-class model at first inspection. If the performance of a model is sufficiently discriminatory between non-cancerous tissue, MSS AC and MSI-H AC, this would bypass the need for the two-step approach currently in practice, whereby samples are first tested for CRC and if positive then undergo a further test to discriminate between MSI-H and MSS. Such a streamlined process may help ameliorate the afore-mentioned lack of surveillance in this regard [5]. To assess this possibility, a three-class model was trained using the same CV strategy outlined in Section 2.3.

Sensitivity, specificity and ROC curves are only defined for binary classification, often implemented as one-vs.-all in multiclass tasks. However, the log loss is reported here (Table 2), to remain consistent with prior results, alongside the more intuitive accuracy and confusion matrices (Figure 6).

Table 2. Three class models: mean log loss and accuracy across all folds \pm 1 standard deviation.

	PCA-LDA	SVM	CNN
Log Loss	0.66 \pm 0.17	0.80 \pm 0.07	0.54 \pm 0.17
Accuracy	71.3% \pm 8.8	74.7% \pm 7.2	74.0% \pm 12.0

The three models return similar accuracies, though the corresponding confusion matrices show that they achieve this by different means; while all models are able to well separate healthy tissues from diseased, the sub-division of the AC cases is more difficult. The SVM does well in correctly identifying MSS with only 4% error, but misclassifies 22% of MSI-H samples as healthy. PCA-LDA and the CNN perform less well discriminating between MSI-H and MSS with 24% error, but only misclassify AC tissue as healthy 4% and 6% of the time, respectively.

**Figure 6.** Confusion matrix for (a) PCA-LDA (b), SVM (c), CNN.

The log loss measures the performance of predictions rather than classifications: it compares the probability of belonging to a disease class, rather than a definitive disease label. The lower the log loss, the closer a prediction is to the true disease class. This allows for a more subtle interpretation of performance. The lower log loss of the CNN indicates that this model gives more conservative predictions compared to the other models, particularly the SVM, meaning it is less likely to confidently give incorrect classifications.

4. Discussion

A direct comparison with existing testing (i.e., IHC or MSI testing) is difficult as various studies have used different thresholds and definitions for these tests and used different methods for the gold standard genetic testing. The resulting study's heterogeneity means no meta-analysis is available to provide a statistically pooled estimate of performance [21]. For MSI testing, sensitivities range from 100% to 66.7% [24–26] with specificities from 92.5% to 61.1% [24,26]. For IHC testing, sensitivities range from 100% to 80.8% [25,27], and specificities range from 91.9% to 80.5% [27,28].

False positives from both methods have been related to the tissue-processing method. Formalin fixation is known to fragment DNA, causes issues with epitope access due to excessive cross-linking and alters native proteins. These biochemical alterations have also been shown to alter Raman spectra, similarly afflicting the technique's diagnostic potential [29]. However, RS can easily bypass these alterations by taking spectra from fresh tissues or minimally processed samples, though the degree of any improvement in performance would need empirical corroboration.

The results in this study sit in the lower range of the above sensitivities and specificities, indicating some potential to compete with molecular techniques, but quite some improvement would be needed to establish superiority. Other screening methods based on

familial history are also used to identify at-risk patients. The Amsterdam II criteria achieve a sensitivity and specificity of 72% and 78% respectively; the Bethesda protocol 94% and 25%, against which these study's results are competitive.

DL has also been explored for MSI-H prediction performed on H&E stained samples [4,6]. These take CNNs designed for image classification, which tend to be very large models and require more data to train. These results tend to be given in terms of AUROC, which on average range from 0.77 to 0.93 during internal validation and from 0.60 to 0.89 when applied to an external dataset. This indicates that morphology alone can distinguish MSI-H, a hitherto unused biomarker, though the generalisability of the DL models to external datasets needs improving before clinical adoption. The results of this study show the potential of RS to discriminate between MSS and MSI-H tumours based on molecular information alone. RS allows for the possibility of combining biomolecular and morphological information. While typical image classification takes images as 3D inputs, with red, green and blue channels, a Raman hyperspectral image has several hundreds of channels, which contain biomolecular information.

This study is too small for clinical utility. However, it does demonstrate the predictive capability of RS to this particular clinical problem and therefore motivates the development of larger studies. This study has also confirmed the suitability of CNNs to modelling such data, which has at the least displayed non-inferiority to simpler modelling techniques traditionally used in RS. This would likely improve with larger sample sizes, as DL requires many examples in order to achieve its best possible performance. Consistent with this, AUROC scores assessing the performance of DL to predict MSI-H from H&E slides have been shown to be positively correlated with sample size [6]. Indeed, a related technique to RS, infrared (IR) spectroscopy, has been used to predict MSI-H and achieved a sensitivity and specificity of 100% and 93%, respectively, with a sample size of 100 patients [30].

Another limitation of this study is that the normal, healthy samples were taken from the same patients as the MSS/MMR proficient samples, due to the ethical constraints of obtaining samples from truly healthy colon biopsies. This means that these two classes are not truly independent and there is a risk that the three-class model is consequently biased. Additionally, the three-class model suffers from the fact that the normal samples have been taken from ostensibly healthy regions from patients with confirmed disease. These may well harbour sub-clinical oncogenic mutations that have yet to manifest morphologically, but that may have biochemical antecedents detectable by RS. Hence, the extent to which these samples represent truly healthy tissue is debatable. This problem is common to many oncology RS studies [8]. It has been argued that this approach, called "paired sampling", reduces interference by individual differences [31], as seen with traditional statistical hypothesis testing. However, it is not clear that ML models similarly benefit from this effect. Conversely, sampling from the same patients likely hinders the generalisability of models trained in this manner when deployed on truly independent samples. The two-class model is unaffected by this potential bias as it did not use the healthy samples. Another limitation is that this study only took resected tissue and no biopsied tissue. It has been shown that models trained to detect MSI-H with H&E samples on one type suffer when trying to predict tissue taken by the other method [32]. Finally, consensus agreement between pathologists is the gold standard for labelling histology slides for ML, but this study only obtained single pathologist labelling.

5. Conclusions

This study is the first of its kind to carry out a preliminary investigation into the use of RS as a clinical diagnostic tool in discriminating MSS AC from MSI-H AC in FFPE colonic tissues. From a very small sample size (10 samples per group), promising results were achieved with the use of ML models, which show that a reasonable degree of discrimination is possible from samples that appear to be spectrally very similar. This is the first proof-of-principle study of its kind that is both label-free and has a rapid sample turnaround time.

Through post hoc analysis of the DL model, diagnostically relevant molecular biomarkers have been implicated, which may distinguish MSS from MSI-H, with nucleic acids and collagen being particularly pertinent. The DL model was able to achieve equivalent performance with screening methods based on familial history, despite the low sample size, though this is not yet competitive with molecular testing.

Author Contributions: Conceptualisation, R.G. and M.R.-J.; methodology, R.G. and N.B.; software, N.B., I.M.B. and L.D.G.; validation, N.B., I.M.B. and L.D.G.; formal analysis, N.B.; investigation, R.G.; resources, R.G.; data curation, N.B. and R.G.; writing—original draft preparation, N.B. and R.G.; writing—review and editing, M.R.-J., I.M.B. and L.D.G.; visualisation, N.B.; supervision, G.M.H.T. and L.D.G.; project administration, R.G.; funding acquisition, G.M.H.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by EPSRC Ph.D. Studentship to N.B., grant number EP/R513143/1, M.R.-J. is partly funded by UCLH/UCL BRC and R.G. was funded by the UCL Impact Ph.D. Scheme.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of UCL/UCLH Biobank for Studying Health and Disease Renewal 2020, REC reference: 20/YH/0088, IRAS project ID: 272816, date: 15 May 2020.

Informed Consent Statement: Project approved by Biobank Ethics Research Committee, this gives access to samples without individual consent provided the data remain anonymised (as is the case).

Data Availability Statement: The data and models presented in this study are available upon request from the corresponding author (G.M.H.T.).

Acknowledgments: We would like to acknowledge the UCL/UCLH Biobank for Studying Health and Disease for providing all the tissue samples used within this study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AC	Adenocarcinoma
AUROC	Area Under Receiver Operating Curve
CNN	Convolutional Neural Network
CRC	Colorectal Cancer
CV	Cross-Validation
DL	Deep Learning
FFPE	Formalin Fixed Paraffin Embedded
IHC	Immunohistochemistry
LDA	Linear Discriminant Analysis
LS	Lynch Syndrome
ML	Machine Learning
MMR	Mismatch Repair
MSI-H	Microsatellite Instability (High)
MSS	Microsatellite Stability
PCA	Principle Component Analysis
PCR	Polymerase Chain Reaction
ROC	Receiver Operating Curve
RS	Raman Spectroscopy
SNV	Standard Normal Variate
SVM	Support Vector Machine

Appendix A. Baseline Correction Experiments

The number of baseline correction techniques is too high to exhaustively search, and there is no a priori reason to think that any one technique will work better than another for any given application. Hence, we investigate a few techniques, which are well established in RS and are relatively easy to perform. The first is the famous “Mod Polyfit” method of Lieber and Mahadevan-Jansen, in which a modified polynomial is iteratively fit to a spectrum using a least-squares-based polynomial curve fitting function which ignores Raman peaks [13]. This requires selecting a single parameter: the order of the polynomial to fit. An extension of this is “Improved Modified Multi-Polynomial Fitting” (I Mod Polyfit) [14]. This was designed to improve fluorescence background removal by being able to account for signal noise distortion and the influence of large Raman peaks. This method also has the parameter of polynomial order. Zhang et al. developed an adaptive iteratively reweighted penalised least squares method which is similar to the previous methods but has the desirable trait of not needing any parameters [15]. Finally, extended multiplicative scattering correction (EMSC) was developed to correct for additive baseline effects, multiplicative scaling effects and interference in near infrared (NIR) spectroscopy, but has been found to be useful in Raman applications for fluorescence removal [16].

We took these four techniques, two of which extend over a range of polynomial orders, and applied them to the dataset. The polynomial order ranged from 1 to 5. Higher orders were not explored due to a mathematical artefact known as Runge’s phenomenon which causes increasingly severe distortions to the tails of Raman spectra as the order of the polynomial increases.

Appendix A.1. Methods

A total of 13 baseline correction methods were evaluated; one with no correction, five using Mod Polyfit with order of 1–5, I Mod Polyfit with order of 1–5, Zhang’s method and EMSC. These were analysed using the three models with all other hyperparameters held constant, allowing only the baseline method to vary. The 5 × 3-fold CV strategy was used for a total of 15 folds per correction method. The mean accuracy and SD of each method was calculated.

This study was conducted before the nested CV and so the hyperparameters used here may not be the optimal ones found during nested CV. The hyperparameters for each model were PCA–LDA: PCs = 20, SVM: C = 10, $\gamma = 0.01$, CNN: learning rate = 0.001, batch size = 32. All other hyperparameters are as stated in the main text.

Appendix A.2. Results

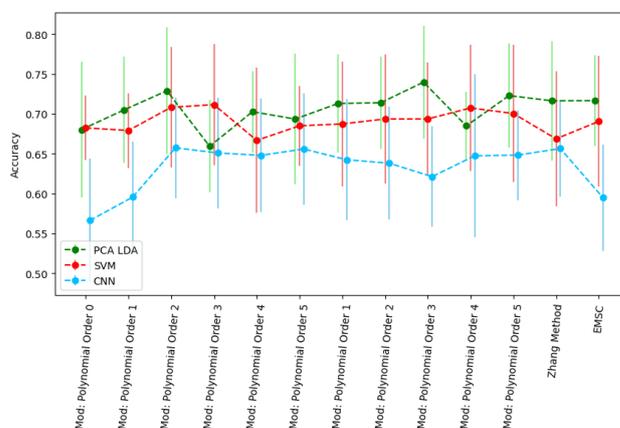


Figure A1. Model Accuracy by baseline correction method: Lynch dataset. Mean value over 15 folds and ± 1 SD bars.

Appendix A.3. Conclusions

Although there is some suggestion that the CNN performs best with second-order modpoly fit, the variance of the accuracies is such that no conclusion of superiority can be made. The RS literature specific to oncological applications regarding baseline improvement is ambivalent, with some studies directly comparing correction vs. no correction finding no correction better [33], and others finding correction better for subsequent model performance [34,35]. In lieu of empirical confirmation, we choose the most parsimonious model: no baseline correction.

Appendix B. Custom Convolutional Neural Network

A custom built CNN was developed. Figure A2 describes the overall model architecture including three convolutional layers and two fully connected layers. Table A1 shows the hyperparameters used. Nested CV was used to choose the learning rate and batch size using grid search over the ranges $[10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}]$ and $[32, 64, 128, 256, 512, 1032]$, respectively. A drop out rate of 0.2 was applied to the fully connected layers and early stopping applied after five epochs of no test score improvement to a maximum of 30 epochs to mitigate over-fitting.

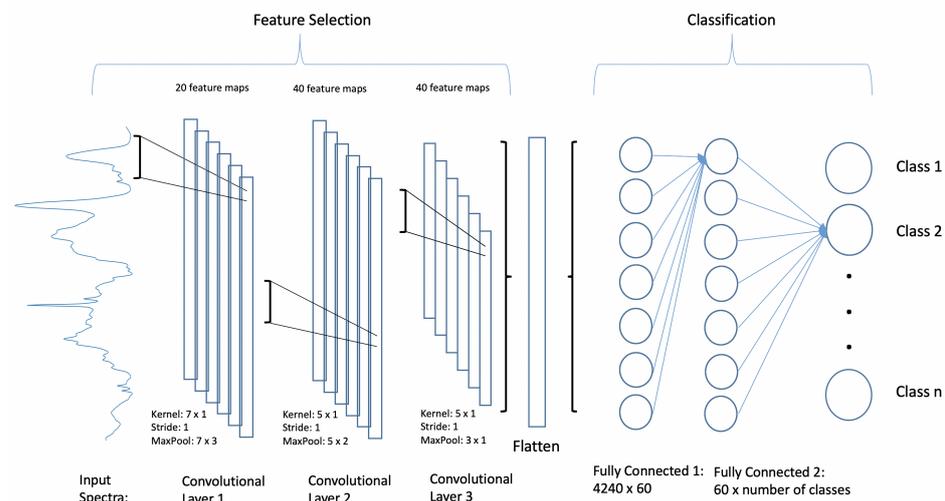


Figure A2. CNN architecture.

Table A1. CNN hyperparameters.

Hyperparameter	Value
Learning Rate	0.0001
Batch Size	256
Drop Out Rate	0.2
Early Stopping	5 epochs
Optimiser	ADAM ($\beta_1 = 0.9, \beta_2 = 0.999$)
Loss Function	Cross-Entropy

For the PCA-LDA model, the number of principal components to retain was searched between 2 and 30 during nested CV, with 15 components being selected. For the SVM, the C parameter (which controls the degree of misclassification tolerated during training) was searched over the ranges $[10^{-2}, 10^{-1}, 10^0, 10^1, 10^2]$: the γ parameter (which controls the extent of curvature of the decision boundary in feature space) was searched over the ranges $[10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0]$, with $C = 0.1$ and $\gamma = 0.001$ being selected during nested CV.

Appendix C. Augmentation Process

Data were augmented by adding two perturbations to the training data: Poisson noise was added, and the wavenumber axis was randomly shifted.

Appendix C.1. Poisson Noise

Poisson noise was added in a particular manner in order to more closely resemble how noise is generated by the spectrometer. The wavelike properties of photons means that their wavelengths are not equally distributed across the CCD pixels, which are of fixed size. Therefore, to express Raman intensities in terms of raw electron count, and not the usual photons/cm⁻¹, we need to correct for the dispersion of light. We can calculate this dispersion factor, d ,

$$d_i = \frac{x_{i-1} - x_{i+1}}{2} \quad (\text{A1})$$

For the i th wavenumber of a given spectrum x . The spectrum was then be multiplied by this dispersion factor to give the electron count per wavenumber pixel in the CCD. The spectrum is then adjusted so that its maximum intensity was 10,000 and Poisson noise added according to this scaled signal intensity at each wavenumber.

Appendix C.2. Wavenumber Shifting

Wavenumber perturbation was achieved by shifting the entire wavenumber axis, represented by a vector. This involves shifting every element in the vector up to a specified amount, which varied from -3 to 3 . Each element in the vector corresponds to approximately 2 cm^{-1} ; thus, the induced shifting varies from -6 to 6 cm^{-1} .

Appendix D. Sample Characteristics

Table A2. Breakdown of patient samples used to build the ML models for the classification of normal (N), sporadic adenocarcinoma (AC) and Lynch syndrome (LS) patients.

Sample ID	Sample Type	TNM Stage	Tumour Grade
LS1	Resection cancer	T2 N0 M0	Mod. Diff.
LS2	Resection cancer	T2 N0 M0	Mod. Diff.
LS3	Resection cancer	T2 N0 M0	Mod. Diff.
LS4	Resection cancer	T3 N0 M0	Poor diff.
LS5	Resection cancer	T3 N0 M0	Mod. Diff.
LS6	Resection cancer	T3 N0 M0	Mod. Diff.
LS7	Resection cancer	T3.N1.Mx	Poor diff.
LS8	Resection cancer	T3 N1 M0	Poor diff.
LS9	Resection cancer	T4 N0 M0	Mod. Diff.
LS10	Resection cancer	T4 N1 M0	Mod. Diff.
AC1	Resection cancer	T2 N2 M0	Mod. Diff.
N1	Normal	-	-
AC2	Resection cancer	T2 N0 M0	Mod. Diff.
N2	Normal	-	-
AC3	Resection cancer	T2 N0 M0	Mod. Diff.
N3	Normal	-	-
AC4	Resection cancer	T3 N1 M0	Mod. Diff.
N4	Normal	-	-
AC5	Resection cancer	T3 N3 M0	Mod. Diff.
N5	Normal	-	-
AC6	Resection cancer	T3 N0 M0	Mod. Diff.
N6	Normal	-	-
AC7	Resection cancer	T3 N0 M0	Mod. Diff.
N7	Normal	-	-
AC8	Resection cancer	T3 N0 M0	Mod. Diff.
N8	Normal	-	-
AC9	Resection cancer	T4 N2 M0	Mod. Diff.
N9	Normal	-	-
AC10	Resection cancer	T4 N0 M1	Poor diff.
N10	Normal	-	-

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)] [[PubMed](#)]
2. Siegel, R.L.; Torre, L.A.; Soerjomataram, I.; Hayes, R.B.; Bray, F.; Weber, T.K.; Jemal, A. Global patterns and trends in colorectal cancer incidence in young adults. *Gut* **2019**, *68*, 2179–2185. [[CrossRef](#)] [[PubMed](#)]
3. Cerretelli, G.; Ager, A.; Arends, M.J.; Frayling, I.M. Molecular pathology of Lynch syndrome. *J. Pathol.* **2020**, *250*, 518–531. [[CrossRef](#)] [[PubMed](#)]
4. Bilal, M.; Nimir, M.; Snead, D.; Taylor, G.S.; Rajpoot, N. Role of AI and digital pathology for colorectal immuno-oncology. *Br. J. Cancer* **2022**, *128*, 3–11. [[CrossRef](#)]
5. Shaikh, T.; Handorf, E.A.; Meyer, J.E.; Hall, M.J.; Esnaola, N.F. Mismatch repair deficiency testing in patients with colorectal cancer and nonadherence to testing guidelines in young adults. *JAMA Oncol.* **2018**, *4*, e173580. [[CrossRef](#)]
6. Hildebrand, L.A.; Pierce, C.J.; Dennis, M.; Paracha, M.; Maoz, A. Artificial intelligence for histology-based detection of microsatellite instability and prediction of response to immunotherapy in colorectal cancer. *Cancers* **2021**, *13*, 391. [[CrossRef](#)]
7. Gaifulina, R.; Maher, A.T.; Kendall, C.; Nelson, J.; Rodriguez-Justo, M.; Lau, K.; Thomas, G.M. Label-free Raman spectroscopic imaging to extract morphological and chemical information from a formalin-fixed, paraffin-embedded rat colon tissue section. *Int. J. Exp. Pathol.* **2016**, *97*, 337–350. [[CrossRef](#)]
8. Blake, N.; Gaifulina, R.; Griffin, L.D.; Bell, I.M.; Thomas, G.M.H. Machine Learning of Raman Spectroscopy Data for Classifying Cancers: A Review of the Recent Literature. *Diagnostics* **2022**, *12*, 1491. [[CrossRef](#)]
9. Santos, I.P.; Barroso, E.M.; Schut, T.C.B.; Caspers, P.J.; van Lanschot, C.G.; Choi, D.H.; Van Der Kamp, M.F.; Smits, R.W.; Van Doorn, R.; Verdijk, R.M.; et al. Raman spectroscopy for cancer detection and cancer surgery guidance: Translation to the clinics. *Analyst* **2017**, *142*, 3025–3047. [[CrossRef](#)]
10. Bera, K.; Schalper, K.A.; Rimm, D.L.; Velcheti, V.; Madabhushi, A. Artificial intelligence in digital pathology—New tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **2019**, *16*, 703–715. [[CrossRef](#)]
11. Gaifulina, R.; Caruana, D.J.; Oukrif, D.; Guppy, N.J.; Culley, S.; Brown, R.; Bell, I.; Rodriguez-Justo, M.; Lau, K.; Thomas, G.M. Rapid and complete paraffin removal from human tissue sections delivers enhanced Raman spectroscopic and histopathological analysis. *Analyst* **2020**, *145*, 1499–1510. [[CrossRef](#)]
12. Lewis, A.T.; Gaifulina, R.; Isabelle, M.; Dorney, J.; Woods, M.L.; Lloyd, G.R.; Lau, K.; Rodriguez-Justo, M.; Kendall, C.; Stone, N.; et al. Mirrored stainless steel substrate provides improved signal for Raman spectroscopy of tissue and cells. *J. Raman Spectrosc.* **2017**, *48*, 119–125. [[CrossRef](#)]
13. Lieber, C.A.; Mahadevan-Jansen, A. Automated method for subtraction of fluorescence from biological Raman spectra. *Appl. Spectrosc.* **2003**, *57*, 1363–1367. [[CrossRef](#)] [[PubMed](#)]
14. Zhao, J.; Lui, H.; McLean, D.I.; Zeng, H. Automated autofluorescence background subtraction algorithm for biomedical Raman spectroscopy. *Appl. Spectrosc.* **2007**, *61*, 1225–1232. [[CrossRef](#)] [[PubMed](#)]
15. Zhang, Z.M.; Chen, S.; Liang, Y.Z. Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst* **2010**, *135*, 1138–1146. [[CrossRef](#)]
16. Afseth, N.K.; Kohler, A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. *Chemom. Intell. Lab. Syst.* **2012**, *117*, 92–99. [[CrossRef](#)]
17. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [[CrossRef](#)]
18. Rieke, J.; Eitel, F.; Weygandt, M.; Haynes, J.D.; Ritter, K. Visualizing convolutional networks for MRI-based diagnosis of Alzheimer’s disease. In *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 24–31.
19. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
20. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 8024–8035.
21. Snowsill, T.; Coelho, H.; Huxley, N.; Jones-Hughes, T.; Briscoe, S.; Frayling, I.M.; Hyde, C. Molecular testing for Lynch syndrome in people with colorectal cancer: Systematic reviews and economic evaluation. *Health Technol. Assess.* **2017**, *21*, 1–280. [[CrossRef](#)]
22. Movasaghi, Z.; Rehman, S.; Rehman, I.U. Raman spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **2007**, *42*, 493–541. [[CrossRef](#)]
23. Li, H.; Sun, L.; Zhuang, Y.; Tian, C.; Yan, F.; Zhang, Z.; Hu, Y.; Liu, P. Molecular mechanisms and differences in lynch syndrome developing into colorectal cancer and endometrial cancer based on gene expression, methylation, and mutation analysis. *Cancer Causes Control* **2022**, *33*, 489–501. [[CrossRef](#)] [[PubMed](#)]
24. Poynter, J.N.; Siegmund, K.D.; Weisenberger, D.J.; Long, T.I.; Thibodeau, S.N.; Lindor, N.; Young, J.; Jenkins, M.A.; Hopper, J.L.; Baron, J.A.; et al. Molecular characterization of MSI-H colorectal cancer by MLHI promoter methylation, immunohistochemistry, and mismatch repair germline mutation screening. *Cancer Epidemiol. Biomark. Prev.* **2008**, *17*, 3208–3215. [[CrossRef](#)] [[PubMed](#)]
25. Shia, J.; Holck, S.; DePetris, G.; Greenon, J.K.; Klimstra, D.S. Lynch syndrome-associated neoplasms: A discussion on histopathology and immunohistochemistry. *Fam. Cancer* **2013**, *12*, 241–260. [[CrossRef](#)] [[PubMed](#)]

26. Barnetson, R.A.; Tenesa, A.; Farrington, S.M.; Nicholl, I.D.; Cetnarskyj, R.; Porteous, M.E.; Campbell, H.; Dunlop, M.G. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. *N. Engl. J. Med.* **2006**, *354*, 2751–2763. [[CrossRef](#)] [[PubMed](#)]
27. Southey, M.C.; Jenkins, M.A.; Mead, L.; Whitty, J.; Trivett, M.; Tesoriero, A.A.; Smith, L.D.; Jennings, K.; Grubb, G.; Royce, S.G.; et al. Use of molecular tumor characteristics to prioritize mismatch repair gene testing in early-onset colorectal cancer. *J. Clin. Oncol.* **2005**, *23*, 6524–6532. [[CrossRef](#)] [[PubMed](#)]
28. Limburg, P.J.; Harmsen, W.S.; Chen, H.H.; Gallinger, S.; Haile, R.W.; Baron, J.A.; Casey, G.; Woods, M.O.; Thibodeau, S.N.; Lindor, N.M. Prevalence of alterations in DNA mismatch repair genes in patients with young-onset colorectal cancer. *Clin. Gastroenterol. Hepatol.* **2011**, *9*, 497–502. [[CrossRef](#)]
29. Faolain, E.O.; Hunter, M.B.; Byrne, J.M.; Kelehan, P.; McNamara, M.; Byrne, H.J.; Lyng, F.M. A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy. *Vib. Spectrosc.* **2005**, *38*, 121–127. [[CrossRef](#)]
30. Kallenbach-Thieltges, A.; Großerueschkamp, F.; Jütte, H.; Kuepper, C.; Reinacher-Schick, A.; Tannapfel, A.; Gerwert, K. Label-free, automated classification of microsatellite status in colorectal cancer by infrared imaging. *Sci. Rep.* **2020**, *10*, 10161. [[CrossRef](#)]
31. Ma, D.; Shang, L.; Tang, J.; Bao, Y.; Fu, J.; Yin, J. Classifying breast cancer tissue by Raman spectroscopy with one-dimensional convolutional neural network. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *256*, 119732. [[CrossRef](#)]
32. Echle, A.; Grabsch, H.I.; Quirke, P.; van den Brandt, P.A.; West, N.P.; Hutchins, G.G.; Heij, L.R.; Tan, X.; Richman, S.D.; Krause, J.; et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. *Gastroenterology* **2020**, *159*, 1406–1416. [[CrossRef](#)]
33. Lee, W.; Lenferink, A.; Otto, C.; Offerhaus, H. Classifying Raman spectra of extracellular vesicles based on convolutional neural networks for prostate cancer detection. *J. Raman Spectrosc.* **2020**, *51*, 293–300. [[CrossRef](#)]
34. Yan, H.; Yu, M.; Xia, J.; Zhu, L.; Zhang, T.; Zhu, Z.; Sun, G. Diverse Region-Based CNN for Tongue Squamous Cell Carcinoma Classification With Raman Spectroscopy. *IEEE Access* **2020**, *8*, 127313–127328. [[CrossRef](#)]
35. Wu, X.; Li, S.; Xu, Q.; Yan, X.; Fu, Q.; Fu, X.; Fang, X.; Zhang, Y. Rapid and accurate identification of colon cancer by Raman spectroscopy coupled with convolutional neural networks. *Jpn. J. Appl. Phys.* **2021**, *60*, 067001. [[CrossRef](#)]

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