



Multi-class Tissue Classification in Colorectal Cancer with Handcrafted and Deep Features

Nicola Altini¹ (✉), Tommaso Maria Marvulli¹, Mariapia Caputo², Eliseo Mattioli³,
Berardino Prencipe¹, Giacomo Donato Cascarano^{1,4}, Antonio Brunetti^{1,4},
Stefania Tommasi², Vitoantonio Bevilacqua^{1,4}, Simona De Summa²,
and Francesco Alfredo Zito³

¹ Department of Electrical and Information Engineering (DEI), Polytechnic University of Bari,
70126 Bari, Italy

nicola.altini@poliba.it

² Molecular Diagnostics and Pharmacogenetics Unit, IRCCS Istituto Tumori Giovanni Paolo II,
Viale Orazio Flacco 65, Bari, Italy

³ Pathology Department, IRCCS Istituto Tumori Giovanni Paolo II, Viale Orazio Flacco 65,
Bari, Italy

⁴ Apulian Bioengineering Srl, Via delle Violette 14, 70026 Modugno, BA, Italy

Abstract. Multi-class tissue classification from histological images is a complex challenge. The gold standard still relies on manual assessment by a trained pathologist, but it is a time-expensive task with issues about intra- and inter-operator variability. The rise of computational models in Digital Pathology has the potential to revolutionize the field. Historically, image classifiers relied on handcrafted feature extraction, combined with statistical classifiers, as Support Vector Machines (SVMs) or Artificial Neural Networks (ANNs). In recent years, there has been a tremendous growth in Deep Learning (DL), for all the image recognition tasks, including, of course, those concerning medical images. Thanks to DL, it is now possible to also learn the process of capturing the most relevant features from the image, easing the design of specialized classification algorithms and improving the performance. An important problem of DL is that it requires tons of training data, which is not easy to obtain in medical domain, since images have to be annotated by expert physicians. In this work, we extensively compared three classes of approaches for the multi-class tissue classification task: (1) extraction of handcrafted features with the adoption of a statistical classifier; (2) extraction of deep features using the transfer learning paradigm, then exploiting SVM or ANN classifiers; (3) fine-tuning of deep classifiers. After a cross-validation on a publicly available dataset, we validated our results on two independent test sets, obtaining an accuracy of 97% and of 77%, respectively. The second test set has been provided by the Pathology Department of IRCCS Istituto Tumori Giovanni Paolo II and has been made publicly available (<http://doi.org/10.5281/zenodo.4785131>).

Keywords: Colorectal cancer · Deep learning · Handcrafted features · Histological tissue classification

N. Altini, T.M. Marvulli, M. Caputo, S. De Summa, F.A. Zito—Equally contributed to this paper.

© Springer Nature Switzerland AG 2021

D.-S. Huang et al. (Eds.): ICIC 2021, LNCS 12836, pp. 512–525, 2021.

https://doi.org/10.1007/978-3-030-84522-3_42

1 Introduction

Colorectal cancer (CRC) is the second cause of death for cancer with mortality ranging almost 35% over the CRC patients [1]. In the last years, new therapeutic approaches have been introduced in the clinical practice but, due to the high mortality, genomic-driven drugs are under evaluation. In particular, the advent of immunotherapy has represented a promising approach for many tumours (e.g., melanoma, non-small cell lung cancer) but results of clinical trials related to CRC have revealed that patients do not benefit from such therapeutical approaches. The chance to molecularly classify this tumour could lead to a better assessment of the regimen to be administered. Many research groups are focusing on these aspects and a multilayer approach could lead in a substantial improvement of the clinical outcomes.

The advent of different computational models allows to perform multilayer analyses including deep study of histological images. Such an approach relies on the automatic assessment of tissue types.

The classical pipeline for building an image classifier involves handcrafted feature extraction and statistical classification. A typical choice was Support Vector Machines (SVMs), or Artificial Neural Networks (ANNs), plus eventual stages of preprocessing and dimensionality reduction.

Linder *et al.* addressed the problem of classification between epithelium and stroma in digitized tumour tissue microarrays (TMAs) [2]. The authors exploited Local Binary Patterns (LBP), together with a contrast measure C (they referred to their union as LBP/ C) as input for their classifier, an SVM. In the end, they compared LBP/ C classifier with those based on Haralick texture features and Gabor filtered images, and the LBP/ C classifier resulted the best model (area under the Receiver Operating Characteristic – ROC – curve was 0.995).

In the context of colorectal cancer histology, it is worth of note the multi-class texture analysis work of Kather *et al.* [3], which combined different features (considering the original RGB images as grey-scale ones), namely: lower-order and higher-order histogram features, Local Binary Patterns (LBP), Grey-level co-occurrence matrix (GLCM), Gabor filters and Perception-like features. As statistical classifiers, they considered: 1-nearest neighbour, linear SVM, radial-basis function SVM and decision trees. Even though they get good performances, by repeating the experiment with the same features, we noted that adopting the red channel leads to better results than using grey-scale images (data not shown). This consideration does not hold after that staining normalization techniques are applied.

Later works exploited the power of Deep Learning (DL), in particular of Convolutional Neural Networks (CNNs), for classifying histopathological images.

Kather *et al.* employed CNN for performing automating tissue segmentation of Hematoxylin-Eosin (HE) images from 862 whole slide images (WSIs) of The Cancer Genome Atlas (TCGA) cohort. Then, they exploit the output neuron activation in the CNN for calculating a "deep stroma score", which proved to be an independent prognostic factor for overall survival (OS) in a multivariable Cox proportional hazard model [4].

Kassani *et al.* proposed a Computer-Aided Diagnosis (CAD) system, composed of an ensemble of three pre-trained CNNs: VGG-19 [5], MobileNet [6] and DenseNet [7], for binary classification of HE stained histological breast cancer images [8]. They came

to the conclusion that their ensemble performed better than single models and widely adopted machine learning algorithms.

Bychkov *et al.* introduced a DL-based method for directly predicting patient outcome in CRC, without intermediate tissue classification. Their model consists in extracting features from tiles with a pretrained model (VGG-16 [5]), and then applying a LSTM [9] to these features [10].

In this work, we extensively compare three classes of approaches for the multi-class tissue classification task: (1) extraction of handcrafted features with the adoption of a statistical classifier; (2) extraction of deep features using the transfer learning paradigm, then exploiting ANN or SVM classifiers; (3) fine-tuning of deep classifiers. We also proposed a feature combination methodology in which we concatenate the features of different pretrained deep models, and we investigate the effect of dimensionality reduction techniques. We identified the best feature set and classifier to perform inferences on external datasets. We investigated the explainability of the considered models, by looking at t-distributed Stochastic Neighbour Embedding (t-SNE) plots and saliency maps generated by Gradient-weighted Class Activation Mapping (Grad-CAM).

2 Materials

The effort of Kather *et al.* resulted in the development and diffusion of different datasets suitable for multi-class tissue classification [3, 4, 11, 12].

[3, 11] describe the collection of $N = 5.000$ histological images, with size of 150×150 pixels (corresponding to $74 \times 74 \mu\text{m}$).

[4, 12] introduce a dataset of $N = 100.000$ image patches from HE stained histological images of human colorectal cancer (CRC) and normal tissue. Images have size of 224×224 pixels, corresponding to $112 \times 112 \mu\text{m}$. This dataset is the designated train set in their experiments, whereas a dataset of $N = 7.180$ has been used as validation set. We denote the first one with **T** and the latter one with **V1**. For the train set, they provide both the original version and a normalized version exploiting the Macenko's method [13].

In order to harmonize some differences between the class names of the two collections, we considered the following classes:

- **TUM**, which represents tumour epithelium.
- **MUSC_STROMA**, which represents the union of **SIMPLE_STROMA**, as tumour stroma, extra-tumour stroma and smooth muscle, and **COMPLEX_STROMA**, as single tumour cells and/or few immune cells.
- **LYM**, which represents immune-cell conglomerates and sub-mucosal lymphoid follicles.
- **DEBRIS_MUCUS**, which represents necrosis, hemorrhage and mucus.
- **NORM**, which represents normal mucosal glands.
- **ADI**, which represents adipose tissue.
- **BACK**, which represents background.

Starting from the Dataset of [11], **SIMPLE_STROMA** and **COMPLEX_STROMA** have been merged, resulting into a **MUSC_STROMA** class. For the dataset of [12], **DEB**

and MUC classes have been merged, resulting in a DEBRIS_MUCUS class, and MUS and STR classes have been merged, resulting in a MUSC_STROMA class. Of note, the merging procedure has been performed according to class definition of the **T** training dataset. At the end of the merge, our training dataset is reduced obtaining $N = 77.805$ images, keeping half of the images of each of the two combined classes, and maintaining the balance across classes. After the same merge, the external validation set **V1** resulted to have $N = 5.988$ images.

An additional dataset of $N = 5.984$ HE histological image patches, provided by IRCCS Istituto Tumori Giovanni Paolo II, has been used as another independent test set. The institutional Ethic Committee approved the study (Prot n. 780/CE). This dataset, hereinafter denoted with **V2**, has been made publicly available [14]. The class subdivision has been done according to the list mentioned above and classified by an expert pathologist, in order to gain the ground truth of the **V2** dataset. We made our dataset publicly available, in order to ease the development and comparison of computational techniques for CRC histological image analysis.

Some test images from both the **V1** and **V2** datasets can be seen in Fig. 1.

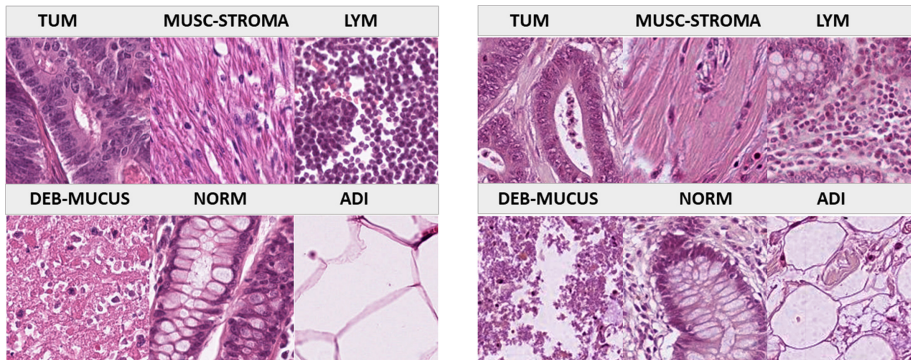


Fig. 1. Test dataset example patches for each class. Left: **V1** dataset; right: **V2** dataset. All images have been pre-processed with Macenko's method.

3 Methods

3.1 Image Features

Different features can be extracted from single channel histogram of an image. In [3], the authors only considered the grey-scale version of the image, but also other color channels may be considered. For HE images, red channel can be more informative.

According to the convention used in [3], we can consider two sets of features from the histogram: a histogram-lower, which contains mean, variance, skewness and kurtosis index, and a histogram-higher, composed of the image moments from 5th to 11th.

Another set of features used was Local Binary Patterns (LBP). An LBP operator considers the probability of occurrence of all the possible binary patterns that can arise

from a neighbourhood of predefined shape and size. A neighbourhood of eight equally spaced points arranged along a circle of radius 1 pixel has been considered. The resulting histogram was reduced to the 38 rotationally-invariant Fourier features proposed by [15]; these are frequently used for histological texture analysis. To extract this set of features it is possible to use the MATLAB tool from Center for Machine Vision and Signal Analysis (CMVS) available at the link <http://www.cse.oulu.fi/CMV/Downloads/LBPMatlab> [16, 17].

Kather *et al.* also considered the Grey-level co-occurrence matrix (GLCM); in particular, they considered four directions (0° , 45° , 90° and 135°) and five displacement vectors (from 1 to 5 pixels). To make this texture descriptor invariant with respect to rotation, the GLCMs obtained from all four directions were averaged for each displacement vector. From each of the resulting co-occurrence matrices the following four global statistics were extracted: contrast, correlation, energy and homogeneity, as described by Haralick *et al.* in [18], thereby obtaining 20 features for each input image.

As the latest set of features, Kather *et al.* considered Perception-like features, that included features based on image perception. Tamura *et al.* in [19] showed that the human visual system discriminates texture through several specific attributes that were later on refined and tested by Bianconi *et al.*; the features considered in [3] were the following five: coarseness, contrast, directionality, line-likeness and roughness [20].

This procedure leads to the extraction of a feature vector with 74 elements.

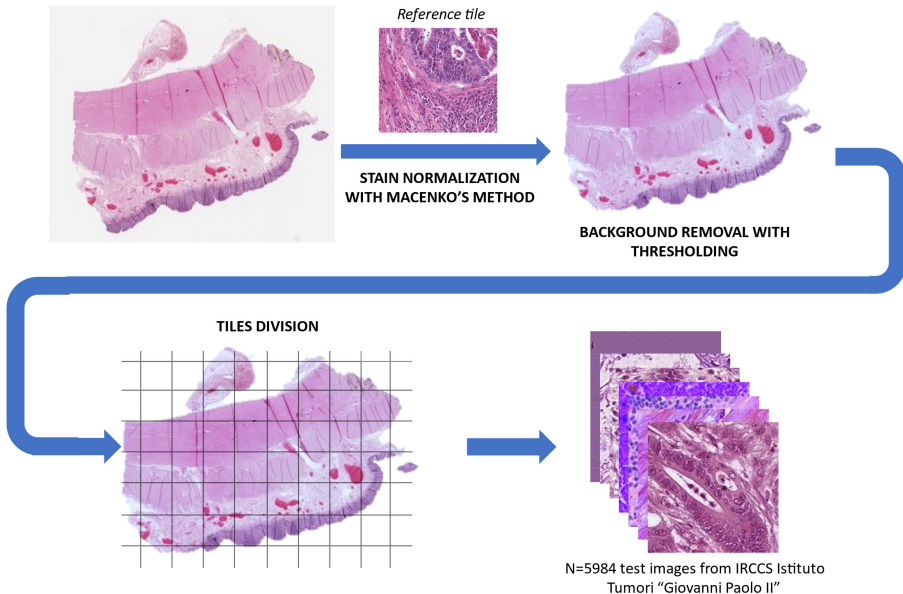


Fig. 2. Stain normalization with Macenko's method [13] and tiling, analogously to Kather *et al.* [4, 12]. This procedure has been followed to generate the test patch images for our dataset.

3.2 Stain Normalization

Stain Normalization is necessary due to the pre-analytical bias specific to different laboratories; it can lead to miscalculation of images by ANN or CNN. Techniques for handling stain color variation can be grouped into two categories: stain color augmentation, which mimics a vast assortment of realistic stain variations during training and stain color normalization, which intends to match training and test color distributions for the sake of reducing stain variation [21].

In order to normalize the images coming from different datasets, we exploited the Macenko's normalization method [13], as reported by Kather *et al.* [4, 12], allowing comparability across different datasets.

The procedure adopted for the stain normalization is depicted in Fig. 2.

3.3 Deep Learning Models

Deep Learning refers to the adoption of hierarchical models to process data, extracting representations with multiple levels of abstraction [22]. Convolutional Neural Network (CNN) have a prominent role in image recognition problems. A huge amount of literature data regarding the construction of DL-based classifiers for images [5, 23–29]. Some example of application in histological images include classification of breast biopsy HE images [30], semantic segmentation, detection and instance segmentation of glomeruli from kidney biopsies [31, 32].

An important concern about CNN is that training a network from scratch requires tons of data. One interesting possibility is that offered by transfer learning, which is a methodology for training models by using data which is more easily collected compared to the data of the problem under consideration. Refer to [33] for a comprehensive survey of the transfer learning paradigm, here we will consider models pre-trained on ImageNet as feature extractors for histological images, as done also in [10, 34–38]. The paradigm of DL-based transfer learning has led to the term Deep Transfer Learning [39]. It has been noted that, although histopathological images are different from RGB images of everyday life, they share common basic structures as edges and arcs [40]. Earlier layers of CNN capture this kind of elementary patterns, so transfer learning may be useful also for digital pathology images.

One potential drawback of deep feature extractor is the high dimensionality. Cascianelli *et al.* attempted to solve this problem by considering different technique of dimensionality reduction [38]. We investigated the combinations of deep features extracted by pretrained models, also considering different levels of compression, after having applied Principal Component Analysis (PCA). In particular, we concatenated the features coming from the ResNet18, GoogleNet and ResNet50 models, obtaining a feature set of 3584 elements. Then, different numbers of features, ranging from 128 to 3584, have been considered for training our classifiers. To ensure that deep features are relevant for the problem under consideration, we compared them to smaller sets of handcrafted features. In particular, we checked: (1) that they tend to represent similar tissue types into defined regions of the feature space, by considering a 2D scatter plot after having applied t-SNE [41] on the deep and handcrafted features; (2) that they lead to the training of an accurate model, without overfitting problems; (3) the saliency

maps highlighted by Grad-CAM [42]. t-SNE can both capture the local structure of high dimensional data and reveal global structure at several scales (e.g. the presence of clusters), as image features in this case. Grad-CAM is a class-discriminative localization technique for CNN-based models to make them more transparent by producing a visual explanation.

We considered three different topologies of deep networks: ResNet18, ResNet50 [28] and GoogLeNet [25]. For each architecture, we compared the ImageNet [43] pretrained version (the network is working only as feature extractor in this case) with the fine-tuned version on our data.

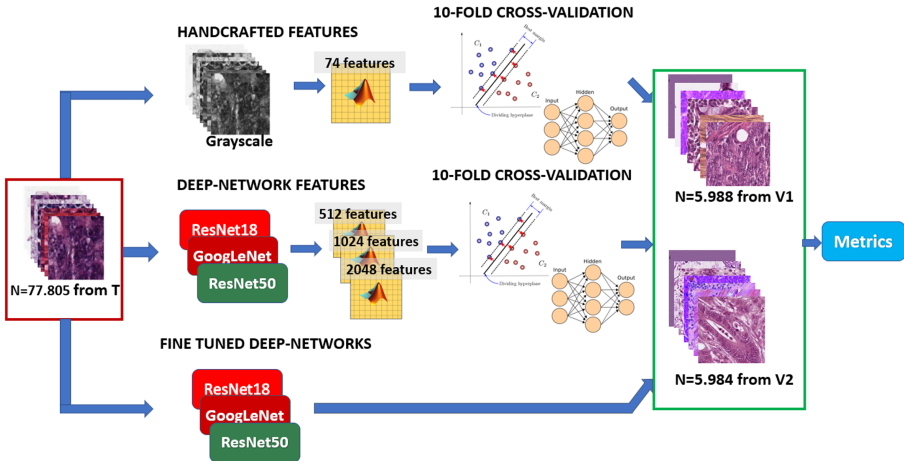


Fig. 3. Training procedure. Starting from a subset of the dataset T , we compared three kinds of models. 10-fold Cross-validation was performed to find the best model. **Validation procedure.** We externally validated the models found as best from internal cross-validation on two datasets: $V1$ and $V2$. T refers to the Training set from Kather *et al.*; $V1$ stands for Test set from Kather *et al.*; $V2$ refers to the Test set from IRCCS Istituto Tumori Giovanni Paolo II.

4 Experimental Results

We considered three types of experiments: (1) training of ANN and SVM classifiers after handcrafted feature extraction; (2) training of ANN and SVM classifiers after deep feature extraction with models pretrained on ImageNet; (3) fine-tuning of deep classifiers. The workflow is depicted in Fig. 3. For the ANN and SVM trained after handcrafted feature extraction or pretrained deep feature extraction, we made a 10-fold cross validation (90% train, 10% test for each iteration) on the train dataset T , after having pre-processed it as described in Sect. 2.

Then, we exploited the best classifier for each category for testing it on the validation datasets $V1$ and $V2$. Performances reported in Table 1, Table 2, Table 3 and Table 4 are assessed in terms of accuracy.

Table 1. Results of 10-fold cross-validation on the train dataset **T**. Performances for SVM and ANN are expressed as accuracy: mean \pm std.

Feature set	# Features	SVM	ANN
Handcrafted Grey	74	94.64 \pm 1.69%	94.59 \pm 0.92%
Handcrafted Red	74	83.23 \pm 0.85%	85.24 \pm 0.08%
Pretrained ResNet18	512	97.91 \pm 0.14%	97.15 \pm 0.03%
Pretrained GoogleNet	1024	94.41 \pm 0.23%	93.92 \pm 0.05%
Pretrained ResNet50	2048	97.61 \pm 0.08%	98.04 \pm 0.03%

Table 2. Results on the **V1** dataset. Performances are reported as accuracy measure.

Feature set	# Features	Best SVM	Best ANN
Handcrafted Grey	74	85.97%	86.84%
Handcrafted Red	74	83.76%	81.66%
Pretrained ResNet18	512	95.61%	94.56%
Pretrained GoogleNet	1024	90.36%	90.20%
Pretrained ResNet50	2048	95.11%	95.24%
Fine-tuned ResNet18	512	97.06%	
Fine-tuned GoogleNet	1024	96.99%	
Fine-tuned ResNet50	2048	97.26%	

Table 3. Results on the **V2** dataset. Performances are reported as accuracy measure.

Feature set	# Features	Best SVM	Best ANN
Handcrafted Grey	74	36.48%	24.62%
Handcrafted Red	74	9.71%	19.72%
Pretrained ResNet18	512	77.19%	71.19%
Pretrained GoogleNet	1024	62.47%	60.80%
Pretrained ResNet50	2048	75.94%	71.59%
Fine-tuned ResNet18	512	66.34%	
Fine-tuned GoogleNet	1024	68.82%	
Fine-tuned ResNet50	2048	72.31%	

Table 4. Proposed methodology. Feature set is given by the concatenation of pretrained ResNet18, GoogleNet, ResNet50, considering different numbers of principal components after the PCA. Results are shown on both the **V1** and **V2** datasets. Percentages represent accuracies.

# Features	128	256	512	1024	3584
ANN on V1	94.17%	94.47%	93.33%	94.42%	95.94%
SVM on V1	95.72%	96.34%	96.37%	95.24%	95.86%
ANN on V2	58.02%	64.43%	63.15%	67.67%	73.15%
SVM on V2	64.10%	64.15%	62.85%	61.56%	76.36%

For the best classifier of each category (handcrafted features, pretrained deep features, finetuned deep model), we computed the confusion matrix to assess how errors are distributed across the different classes. Confusion matrices are reported in Tables 5, 6 and 7.

Table 5. Confusion matrix on the **V2** dataset for the best handcrafted model.

TRUE CLASS	TUM	44	25	0	32	0	0	2
	MUSC-STROMA	292	1782	0	702	1	4	94
	LYM	4	15	0	48	0	1	4
	DEBRIS-MUCUS	46	82	0	329	0	1	27
	NORM	1492	386	0	462	16	0	32
	ADI	0	41	0	7	0	12	1
	BACK	0	0	0	0	0	0	0
		TUM	MUSC-STROMA	LYM	DEBRIS-MUCUS	NORM	ADI	BACK
PREDICTED CLASS								

4.1 Discussion and Explainability

Looking at the confusion matrices, we observed that handcrafted features are not able to well generalize on our dataset, whilst deep features are better suited for the task. In particular, the model trained with handcrafted features is not able to recognize any LYM tissue from our **V2** dataset. For the proposed method which combines features of different deep architectures, we showed that PCA could be a useful tool for reducing dimensionality without incurring in a decrease of accuracy. Among the pretrained models on the **V1** dataset, the proposed methodology slightly outperforms the best pretrained model alone, ResNet18, using also less features. For the SVM classifiers on the **V1** dataset, using more than 256 features after PCA does not result in measurable improvements.

Table 6. Confusion matrix on the **V2** dataset for the best pre-trained deep features model.

TRUE CLASS	TUM	100	0	1	0	2	0	0
	MUSC-STROMA	234	2047	32	377	118	21	46
	LYM	1	3	41	9	15	0	3
	DEBRIS-MUCUS	27	14	14	401	12	0	17
	NORM	129	36	57	147	2011	0	8
	ADI	0	8	0	28	6	19	0
	BACK	0	0	0	0	0	0	0
		TUM	MUSC-STROMA	LYM	DEBRIS-MUCUS	NORM	ADI	BACK
PREDICTED CLASS								

Table 7. Confusion matrix on the **V2** dataset for the best deep fine-tuned model.

TRUE CLASS	TUM	101	1	0	1	0	0	0
	MUSC-STROMA	370	1856	37	434	41	56	81
	LYM	2	13	23	13	11	0	10
	DEBRIS-MUCUS	8	14	3	423	11	1	25
	NORM	226	65	56	133	1890	2	16
	ADI	0	1	0	26	0	34	0
	BACK	0	0	0	0	0	0	0
		TUM	MUSC-STROMA	LYM	DEBRIS-MUCUS	NORM	ADI	BACK
PREDICTED CLASS								

We observed that frequent misclassification errors involved NORMAL and MUSC-STROMA patches which are predicted as TUMOUR or DEBRIS-MUCUS.

In order to assess the explainability of the obtained results, we considered different techniques. First, we looked at the t-SNE embeddings, to understand if deep features, also those obtained by pre-training on ImageNet, are meaningful for the problem under consideration. Figure 4a displayed that clusters are much better defined from the **V1** dataset. It is important to highlight that they considered tiles clearly belonging to only one class, whereas we also allowed the inclusion of patches more difficult to be classified.

The presence of a sub-cluster of TUM tiles can be seen within the MUSC_STROMA cluster. As stated above, MUSC_STROMA derives from the merging of simple and complex stroma classes, the latter including also sparse tumor cells. Thus, the TUM sub-cluster and the misclassification could be explained by both the class definition and, from a biological perspective, the fact that tumor tissue invades the surrounding stroma. Moreover, it could be observed in Fig. 4b that NORM cluster includes DEBRIS_MUCUS sub-cluster. Such a result makes sense because in this case mucus containing exfoliated

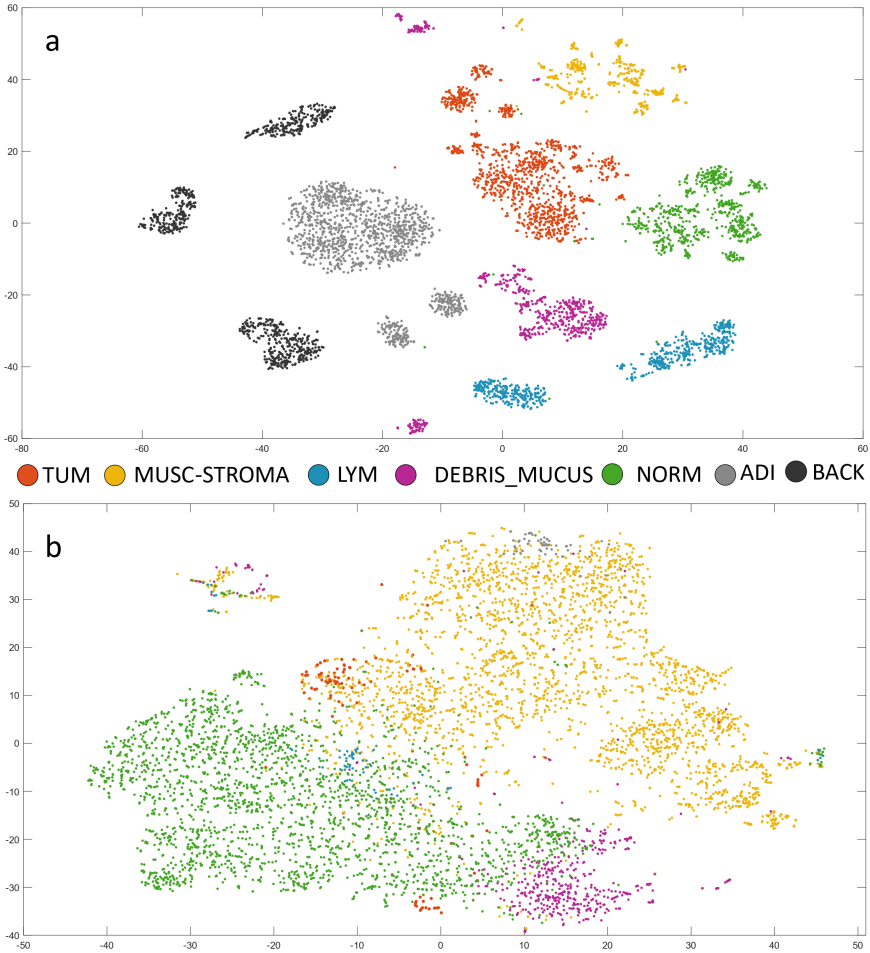


Fig. 4. t-SNE of the best classifier features; a) Fine-tuned ResNet50 features t-SNE of V1 dataset b) Pretrained ResNet18 features t-SNE of V2 dataset.

epithelial cells is mainly produced by the glands of the normal tissue component at the periphery of the tissue sample.

Then, we tried to see the activations of the fine-tuned deep models exploiting Grad-CAM method [42]. We can see from Fig. 5a-c and Fig. 5e-g the highlighted regions of sample images from V1 and V2 datasets. Figure 5d and Fig. 5h represent patches which have not been included into V2 dataset since they were not clearly classifiable. In particular, Fig. 5d contains both MUSC_STROMA and TUM classes, whereas Fig. 5h contains both DEBRIS_MUCUS and NORM.

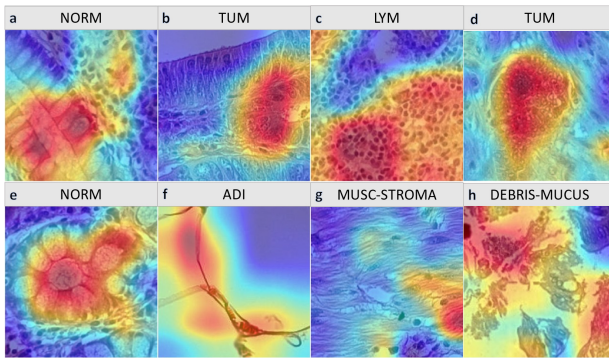


Fig. 5. Grad-CAM of Deep fine-tuned classifiers on the test set; **a, c, f**, from dataset V1 and **b, g, e** from dataset V2; **d, h** are patches not belonging to only one class. Labels are those in output of the classifier.

5 Conclusions and Future Works

In this work, three different methods have been compared for multi-class histology tissue classification in CRC. The most promising approach resulted to be to extract pretrained ResNet18 deep features from tiles combined with classification through SVM; in this way the classifier is able to generalize well on external datasets with good accuracy.

We also investigated explainability of our trained deep models observing that some misclassification issues are related to the biology of CRC. The multi-class tissue classification is a useful task in CRC histology, in particular to exploit a multi-layer approach including genomic data (mutational and transcriptional status).

The present paper could be considered a proof-of-concept because the multi-class tissue classification of digital histological images could, not only be extended to other malignancies, but also be considered as the preliminary step to explore, e.g., the relationship between the tumor, its microenvironment and genomic features.

Acknowledgments. This research was funded by Italian Apulian Region “Tecnopolo per la medicina di precisione”, CUP B84I18000540002.

References

1. Siegel, R.L., et al.: Colorectal cancer statistics, 2020. *CA. Cancer J. Clin.* **70**, 145–164 (2020). <https://doi.org/10.3322/caac.21601>
2. Linder, N., et al.: Identification of tumor epithelium and stroma in tissue microarrays using texture analysis. *Diagn. Pathol.* **7**, 22 (2012). <https://doi.org/10.1186/1746-1596-7-22>
3. Kather, J.N., et al.: Multi-class texture analysis in colorectal cancer histology. *Sci. Rep.* **6**, 1–11 (2016). <https://doi.org/10.1038/srep27988>
4. Kather, J.N., et al.: Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. *PLoS Med.* **16**, 1–22 (2019). <https://doi.org/10.1371/journal.pmed.1002730>

5. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, 1–14 (2014)
6. Howard, A.G., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv Prepr. [arXiv:1704.04861](https://arxiv.org/abs/1704.04861) (2017)
7. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269. IEEE (2017). <https://doi.org/10.1109/CVPR.2017.243>
8. Kassani, S.H., Kassani, P.H., Wesolowski, M.J., Schneider, K.A., Deters, R.: Classification of histopathological biopsy images using ensemble of deep learning networks. In: CASCON 2019 Proc. - Conf. Cent. Adv. Stud. Collab. Res. - Proc. 29th Annu. Int. Conf. Comput. Sci. Softw. Eng., pp. 92–99 (2020)
9. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. *Neural Comput.* **9**, 1735–1780 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
10. Bychkov, D., et al.: Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci. Rep.* **8**, 1–11 (2018). <https://doi.org/10.1038/s41598-018-21758-3>
11. Kather, J.N., et al.: Collection of textures in colorectal cancer histology (2016). <https://doi.org/10.5281/zenodo.53169>
12. Kather, J.N., Halama, N., Marx, A.: 100,000 histological images of human colorectal cancer and healthy tissue (2018). <https://doi.org/10.5281/zenodo.1214456>
13. Macenko, M., et al.: A method for normalizing histology slides for quantitative analysis. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1107–1110 (2009). <https://doi.org/10.1109/ISBI.2009.5193250>
14. Altini, N., et al.: Pathologist’s annotated image tiles for multi- class tissue classification in colorectal cancer (2021). <https://doi.org/10.5281/zenodo.4785131>
15. Ahonen, T., Matas, J., He, C., Pietikäinen, M.: Rotation invariant image description with local binary pattern histogram fourier features. In: Salberg, A.-B., Hardeberg, J.Y., Jenssen, R. (eds.) SCIA 2009. LNCS, vol. 5575, pp. 61–70. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02230-2_7
16. Ojala, T., Pietikäinen, M., Maenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 971–987 (2002). <https://doi.org/10.1109/TPAMI.2002.1017623>
17. Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Computer vision using local binary patterns. Presented at the (2011). https://doi.org/10.1007/978-0-85729-748-8_14
18. Haralick, R.M., Dinstein, I., Shanmugam, K.: Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **SMC-3**, 610–621 (1973). <https://doi.org/10.1109/TSMC.1973.4309314>
19. Tamura, H., Mori, S., Yamawaki, T.: Textural features corresponding to visual perception. *IEEE Trans. Syst. Man Cybern.* **8**, 460–473 (1978). <https://doi.org/10.1109/TSMC.1978.4309999>
20. Bianconi, F., Álvarez-Larrán, A., Fernández, A.: Discrimination between tumour epithelium and stroma via perception-based features. *Neurocomputing* **154**, 119–126 (2015). <https://doi.org/10.1016/j.neucom.2014.12.012>
21. Tellez, D., et al.: Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med. Image Anal.* **58**, 101544 (2019). doi:<https://doi.org/10.1016/j.media.2019.101544>
22. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**, 436–444 (2015). <https://doi.org/10.1038/nature14539>
23. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE.* **86**, 2278–2324 (1998). <https://doi.org/10.1109/5.726791>
24. Krizhevsky, A., Sutskever, I., Hinton, G.E.: 2012 AlexNet. *Adv. Neural Inf. Process. Syst.* (2012). <https://doi.org/10.1016/j.protcy.2014.09.007>

25. Zeng, G., He, Y., Yu, Z., Yang, X., Yang, R., Zhang, L.: InceptionNet/GoogLeNet - going deeper with convolutions. *CVPR* **91**, 2322–2330 (2016). <https://doi.org/10.1002/jctb.4820>
26. He, K., Girshick, R., Dollár, P.: Rethinking ImageNet Pre-training, pp. 1–10 (2018)
27. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 1904–1916 (2015). <https://doi.org/10.1109/TPAMI.2015.2389824>
28. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>.
29. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
30. Araujo, T., et al.: Classification of breast cancer histology images using convolutional neural networks. *PLoS ONE* **12**, 1–14 (2017). <https://doi.org/10.1371/journal.pone.0177544>
31. Altini, N., et al.: Semantic segmentation framework for glomeruli detection and classification in kidney histological sections. *Electronics*. **9**, 503 (2020). <https://doi.org/10.3390/electronics9030503>
32. Altini, N., et al.: A deep learning instance segmentation approach for global glomerulosclerosis assessment in donor kidney biopsies. *Electronics* **9**, 1768 (2020). <https://doi.org/10.3390/electronics9111768>
33. Weiss, K., Khoshgoftaar, T.M., Wang, D.: A survey of transfer learning. *J. Big Data* **3**(1), 1–40 (2016). <https://doi.org/10.1186/s40537-016-0043-6>
34. Campanella, G., et al.: Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019). <https://doi.org/10.1038/s41591-019-0508-1>.
35. Schmauch, B., et al.: A deep learning model to predict RNA-Seq expression of tumours from whole slide images. *Nat. Commun.* **11**, 3877 (2020). <https://doi.org/10.1038/s41467-020-17678-4>
36. Fu, J., Singhrao, K., Cao, M., Yu, V., Santhanam, A.P., Yang, Y., Guo, M., Raldow, A.C., Ruan, D., Lewis, J.H.: Generation of abdominal synthetic CTs from 0.35T MR images using generative adversarial networks for MR-only liver radiotherapy. *Biomed. Phys. Eng. Express*. **6** (2020). <https://doi.org/10.1088/2057-1976/ab6e1f>.
37. Levy-jurgenson, A.: Spatial Transcriptomics Inferred from Pathology Whole-Slide Images Links Tumor Heterogeneity to Survival in Breast and Lung Cancer. 1–16 (2020).
38. Cascianelli, S., et al.: Dimensionality reduction strategies for CNN-based classification of histopathological images. In: De Pietro, G., Gallo, L., Howlett, R.J., Jain, L.C. (eds.) *KES-IIMSS 2017*. SIST, vol. 76, pp. 21–30. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-59480-4_3
39. Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., Liu, C.: A survey on deep transfer learning. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. (eds.) *ICANN 2018*. LNCS, vol. 11141, pp. 270–279. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01424-7_27
40. Komura, D., Ishikawa, S.: Machine learning methods for histopathological image analysis. *Comput. Struct. Biotechnol. J.* **16**, 34–42 (2018). <https://doi.org/10.1016/j.csbj.2018.01.001>
41. der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, (2008)
42. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vision* **128**(2), 336–359 (2019). <https://doi.org/10.1007/s11263-019-01228-7>
43. Deng, J., Li, K., Do, M., Su, H., Fei-Fei, L.: Construction and analysis of a large scale image ontology. Presented at the (2009)