POLITECNICO DI TORINO
Repository ISTITUZIONALE

Development and validation of an AI-based pathomics biomarker to predict response to first-line treatment in metastatic colorectal cancers

Original
Development and validation of an AI-based pathomics biomarker to predict response to first-line treatment in metastatic colorectal cancers / Nicoletti, G., Cafaro, D., Giannini, V., Mauri, G., Marchio, C., Lazzari, L., Sartore-Bianchi, A., Marmorino, F., Munoz, M.N., Gonzalez, N.S., Puccini, A., Di Como, M., Aquilano, M.C., Bonoldi, E., Siena, S., Marsoni, S., Regge, D.. - (2024). (10th International Conference on Bioengineering and Biotechnology (ICBB 2024) Barcelona (Spa) August 25-27 2024) [10.11159/icbb24.122].

Availability:
This version is available at: 11583/2997421 since: 2025-02-10T14:48:37Z

Publisher:
Avestia Publishing

Published
DOI:10.11159/icbb24.122

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

29 November 2026

Proceedings of the 10th World Congress on New Technologies (NewTech'24)
Barcelona, Spain - August 25-27, 2024
Paper No. ICBB 122
DOI: 10.11159/icbb24.122

Development and validation of an AI-based pathomics biomarker to predict response to first-line treatment in metastatic colorectal cancers

Giulia Nicoletti[1,2,3]*, Debora Cafaro[2]*, Valentina Giannini[2,3], Gianluca Mauri[4,5], Caterina Marchiò[6,7], Luca Lazzari[4], Andrea Sartore-Bianchi[8], Federica Marmorino[9,10], Maria Nieva Munoz[11], Nadia Saoudi González[12], Alberto Puccini[13], Martina Di Como[14], Maria Costanza Aquilano[14], Emanuela Bonoldi[14], Salvatore Siena[5], Silvia Marsoni[4], Daniele Regge[3]

[1]Dept. of Electronics and Telecommunications, Polytechnic University of Turin, Turin, Italy
giulia.nicoletti@polito.it
[2]Dept. of Surgical Sciences, University of Turin, Turin, Italy
debora.cafaro@unito.it, valentina.giannini@unito.it,
[3]Radiology Unit, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy
daniele.regge@ircc.it
[4]IFOM ETS, The AIRC Institute of Molecular Oncology, Milan, Italy
gianluca.mauri@ifom.eu, luca.lazzari@ifom.eu, silvia.marsoni@ifom.eu
[5] Dept. of Oncology and Hemato-Oncology, University of Milan, Milan, Italy
[6]Pathology Unit, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy
caterina.marchio@ircc.it
[7]Dept. of Medical Sciences, University of Turin, Turin, Italy
[8]Division of Clinical Research and Innovation, Grande Ospedale Metropolitano Niguarda, Milan, Italy
andrea.sartorebianchi@unimi.it
[9]Unit of Oncology, University Hospital of Pisa, Pisa, Italy
federica.marmorino@gmail.com
[10]Dept. of Translational Research, University of Pisa, Pisa, Italy
[11]Dept. of Medical Oncology, Hospital del Mar, Barcelona, Spain
maria.nieva.munoz@psmar.cat
[12]Dept. of Medical Oncology, Vall d'Hebron University Hospital, Barcelona, Spain
nsaoudi@vhio.net
[13]Medical Oncology Unit 1, IRCCS Ospedale Policlinico San Martino, Genoa, Italy
alberto.puccini@cancercenter.humanitas.it
[14]Unit of Surgical Pathology and Cytogenetics, Dept. of Hematology Oncology, and Molecular Medicine, Grande Ospedale Metropolitan Niguarda, Milan, Italy
martina.dicomo@ospedaleniguarda.it, mariacostanza.aquilano@ospedaleniguarda.it,
emanuela.bonoldi@ospedaleniguarda.it

* G.N. and D.C. are equal contributors to this work and designated as co-first authors.

Abstract - Microsatellite stable metastatic colorectal cancer (mCRC) patients are treated with a "one-fits-all" standard of care chemotherapy. However, responses occur in 20-30% of patients, while 15-20% are refractory. The latter are exposed to side effects that lower their quality of life. Therefore, the aim of this work is to develop a predictive biomarker, based on digital pathology images, that can help stratify patients according to their risk of resistance. Hematoxylin and eosin-stained (H&E) slides of mCRC resections were digitalized. Patches were extracted from the resulting whole slide images and automatically classified as belonging to one out of 9 classes, including the tumoral one, using a deep learning model. Based on texture features, clusters of patches were computed and were used to create the Bag of words (BoWs) that were subsequently used to train several machine learning classifiers. The best performances were obtained by a support vector machine, reaching a negative predictive value (NPV) of 90% (44/49; 95%CI=79-95%) and 82% (14/17;

95%CI=63-93%), in the training and validation sets, respectively. From a clinical perspective, NPV is the most relevant metric to ensure that sensitive patients are not wrongly prevented from receiving treatment. These preliminary findings should be further validated on a larger cohort of patients that we are collecting through a multi-institutional study.
***Keywords*:** Pathomics, digital pathology, artificial intelligence, machine learning, colorectal cancer, bag of words.

1. Introduction

Colorectal cancer (CRC) is the third most common and the third most lethal cancer worldwide, both in men and women [1]. Every year approximately 150.000 European patients die of CRC, and only 10-15% of CRC patients initially diagnosed with metastatic disease are still alive after 5-year from diagnosis [1], [2]. CRC is an immune-cold cancer driven in more than 50% of cases by mutations arising in the RAS gene, which is pharmacologically actionable only in a minority of cases [3]. Thus, contrary to other major cancer types such as melanoma or non-small cell lung cancer, the mainstay treatment of microsatellite stable (MSS) metastatic CRC (mCRC) remains a backbone of chemotherapy delivering the same three drugs in use since the last century [4]. However, response strongly varies across patients: 15%-20% of them have a chemo-refractory disease (i.e., no response to treatment), while up to 30% can reach a complete or long-lasting response.

To date, there are no available markers that can predict whether individual mCRC patients will respond or not to chemotherapy before treatment starts. This is clearly a major unmet clinical need, dramatically impacting both patients' survival and quality of life.

Beyond standard histopathology and molecular assessments, pathomics is emerging as a promising field that can enhance the diagnosis and management of patients, by providing valuable insights into the molecular and cellular characteristics of cancer. Pathomics refers to the use of cutting-edge algorithms and artificial intelligence tools to analyze digital images of tissue samples, such as those obtained from biopsies or surgical resections, and to extract complex patterns that are not easily visible to human experts and that might be correlated with tumor characteristics and prognosis.

Pathomics applied on whole slide images (WSI) of Hematoxylin and eosin stained (H&E) slides has been proven effective in diagnosing colorectal cancer [5], predicting mutations such as KRAS, NRAF, BRAF, PIK3CA, and microsatellite instability [6]. However, there is limited research on the application of AI in predicting CRC response to treatment. Most studies have assessed algorithms' ability to predict the response after therapy [7]. However, it would be of key importance to anticipate this prediction in order to save patients from unnecessary and toxic treatments.

The aim of this study is to develop an innovative AI framework applied to H&E WSI to predict the response outcome of first-line chemotherapy in mCRC patients before treatment.

2. Material and Methods

2.1. Patient population

A total of 196 mCRC patients who were treated with any first-line treatment were retrospectively enrolled from seven major oncology cancer centers across Italy (Grande Ospedale Metropolitano Niguarda, Milan; University Hospital of Pisa, Pisa; Istituto Nazionale dei Tumori, Milan; IRCCS Ospedale Policlinico San Martino, Genova) and Spain (Hospital del Mar, Barcelona; Institut Català d'Oncologia, Barcelona; Vall d'Hebron Institute of Oncology, Barcelona). Inclusion criteria were: a) information availability of the response to first-line treatment (see 2.2 for details), b)

availability of formalin-fixed paraffin-embedded (FFPE) samples. Exclusion criteria were: 1) absence of an FFPE sample of the resected tumor (i.e., only biopsy), 2) strong inconsistency between the tumor area and the automatic tumor mask, and 3) presence of artifacts in the digital histopathological image.

The study was conducted in accordance with the Declaration of Helsinki and the International Conference on Harmonization and Good Clinical Practice guidelines. The study was approved by the Ethical Committee of IFOM IFOM-CPO003/2018/PO002 study (no. 617-122018 approved on 13 December 2018) and AlfaOmega-RETRO IFOM-CPO006/2019/PO005 (no. 145-07042020 approved on 7 April 2020). The study is embedded within the AlfaOmega Master Observational Trial (NCT04120935), therefore all patients signed dedicated informed consent to participate the study and allow FFPE samples.

2.2. Reference standard

Patients were dichotomized into two classes: 1) resistant, if they relapsed while on adjuvant oxaliplatin or they showed showed progressive disease (PD) to first-line or neoadjuvant chemotherapy, 2) sensitive if they reached complete response to first-line chemotherapy or partial response lasting more than 10 months with doublets or 12 months with triplets chemotherapy.

2.3. Model development

To develop the pathomics signature to predict response to treatment, we applied a method known as Bag of Words (BoW), that was initially used in Natural Language Processing where the frequency of occurrence of each word is used to perform document classification. Similarly, in computer vision, the BoW model can represent an image as a collection of local features that are counted based on their frequency and that can be grouped into clusters to form a *visual vocabulary*. The pipeline of our algorithm is depicted in figure 1 and consists of the following steps:

1) *Pre-processing and tumor mask creation*: 4 uM thin slides were stained with H&E and digitalized to obtain the WSI of the resected tumor. WSIs were subsequently tiled into patches of 224x224 pixels (0.5μm/pixel). All tiles were stain-normalized using the Macenko method [8], and classified, using a publicly available deep learning model [9], as belonging to one of the following 9 classes: adipose tissue, background, debris, lymphocytes, mucus, smooth muscle, normal colon mucosa, cancer-associated stroma, and colorectal adenocarcinoma epithelium. A mask containing only patches classified as belonging to colorectal adenocarcinoma epithelium was created and then cleaned by all tiles whose majority of pixels were white i.e., background pixels. The resulting mask was checked by an expert pathologist. In case of wrong tumor area prediction, i.e., the predicted area did not belong to the tumor or included more non-tumoral patches than tumoral ones, the corresponding patients were excluded from the subsequent analysis. Finally, tumoral masks were post-processed removing all connected areas smaller than 50 patches (figure 1.1). The pre-processing step and the creation of the tumor mask was carried out for all the WSIs of both training and validation sets.

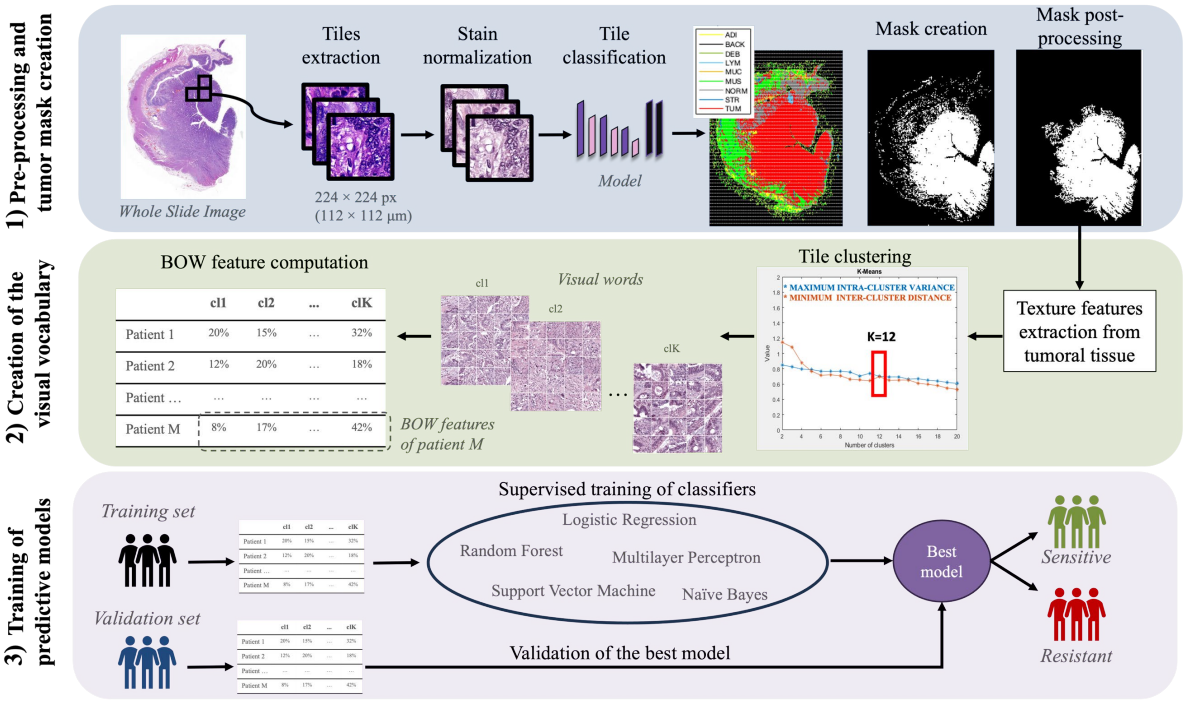

Fig. 1: Pipeline of the study.

2) *Creation of the visual vocabulary*: 24 texture features from the Gray Level Co-occurance Matrix were extracted from each of the patches belonging to the tumoral mask, using Pyradiomics v3.0.1. Then, all tumoral patches belonging to the training set were normalized using the p1p99 normalization and were clustered using the k-means algorithm based on the centroid Euclidean distance. The number of clusters (k) was ranged between 2 and 20 and for each value of k the inter-cluster distances and intra-cluster variance were computed, which were defined respectively as the distance between every couple of centroid, and the average distance between all elements of a certain cluster and the corresponding cluster centroid. Finally, the optimal k was chosen as the one that maximizes the minimum inter-cluster distance while minimizing the maximum intra-cluster variance to create homogeneous clusters with a high level of internal cohesion and a high distance between clusters. This unsupervised clustering allowed us to obtain the *visual words*, i.e., the patch clusters, that were used to characterize each patient according to the BoW features, i.e., a vector composed of the occurrences of the k visual words normalized for the total number of tumoral patches of the patient (figure 1.2).

3) *Training of predictive models*: BoW features of the patients of the training set were used to train different machine learning algorithms to perform a supervised classification. We trained several classifiers, including Random Forest, Naive Bayes, Support Vector Machine (SVM), and Multilayer Perceptron with different hidden layers and neurons (figure 1.3). During the training step, all hyperparameters of the classifiers were optimized. All algorithms were implemented using MatlabR2022b.

2.4. Model validation and statistical analysis

The classifier obtaining the best performance on the training set was finally validated on the validation set. BoW features of patients belonging to the validation set were created by computing the distances between each observation and and the centroids of the clusters found using the training set. For both the training and validation sets, the following metrics metrics were computed: 1) sensitivity: defined as the number of correctly classified sensitive patients over the total number number of sensitive patients; 2) specificity: defined as the number of correctly classified resistant patients over the total total number of resistant patients; 3) positive predictive value (PPV): defined as the number of correctly classified sensitive sensitive patients over the total number of patients classified as sensitive; 4) negative predictive value (NPV): defined as the as the number of correctly classified resistant patients over the total number of patients classified as resistant; 5) balanced balanced accuracy: defined as the average between sensitivity and specificity. To evaluate these metrics, the pathomics score score was dichotomized using the Youden Index, a commonly used measure of overall diagnostic effectiveness that optimizes optimizes accuracy in both groups.

3. Results and discussion

Figure 2 shows the composition of the training and internal validation sets. The final training set was composed of 94 patients (47 resistant and 47 sensitive), while the validation set was composed of 32 patients (20 resistant and 12 sensitive). The optimal number of clusters for the k-means was 12 (see tile clustering in figure 1.2), meaning that each patient was characterized by the normalized occurrence of the patches in each of the 12 clusters. The best results were obtained using an SVM with a polynomial kernel, in which the cutoff for the dichotomization of the posterior probabilities was set to 0.497

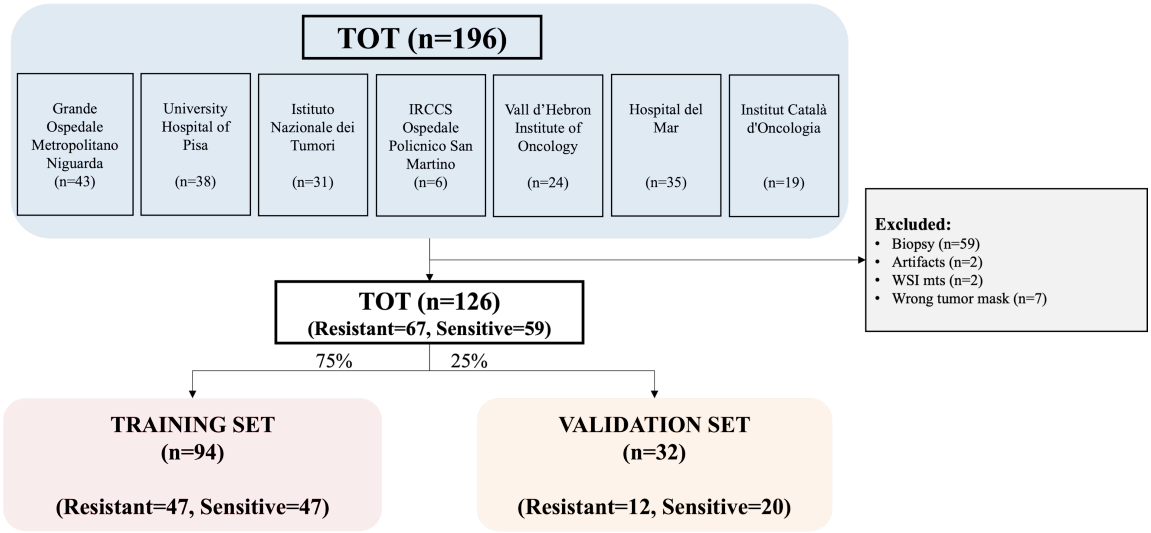

Fig. 2: Flowchart of the study population.

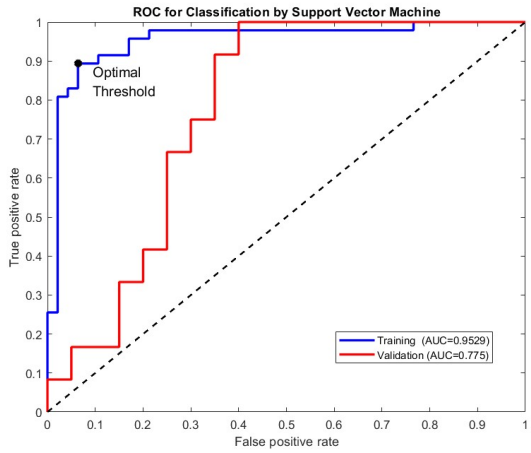

Fig. 3: Receiver Operating Characteristics (ROC) curve of the training (blue line) and validation (red line) sets. The blue dot is the threshold optimized on the training set by the Youden Index.

(see figure 3), based on the result of the Youden Index computed on the training set. Overall, we obtained reasonable high performances, with a specificity of 94% and 70% and a sensitivity of 89% and 75% in the training and validation sets, respectively (see table 1). From a clinical perspective, the NPV is the most important performance value since patients predicted as resistant would avoid an ineffective and toxic treatment that is commonly administered to all patients even if no improvement is obtained. In this regard, our model was demonstrated to be reliable when predicting a patient as resistant both in the training (90% NPV) and validation sets (82% NPV). From the waterfall plot of the predictions of training and validation sets (figure 4), it can be seen that most of the misclassified of both the training and validation sets are concentrated around the chosen cut-off and therefore it would be possible to further optimize the cut-off for the class of interest (i.e., resistant). However, given the high sensitivity of classification performance to the chosen cut-off, it is crucial to provide, alongside the dichotomized output prediction, the normalized pathomics score to clearly indicate the model's confidence level in the given prediction.

Table 1: Best model performances.

	Balanced accuracy	**Sensitivity**	**Specificity**	**Positive Predictive Value**	**Negative Predictive Value**
Training	91.5%	89% (42/47; 95%CI=77-96%)	94% (44/47; 95%CI=82-99%)	93% (42/45; 95%CI=82-98%)	90% (44/49; 95%CI=79-95%)
Validation	72.5%	75% (9/12; 95%CI=73-95%)	70% (14/20; 95%CI=76-88%)	60% (9/15; 95%CI=42-76%)	82% (14/17; 95%CI=63-93%)

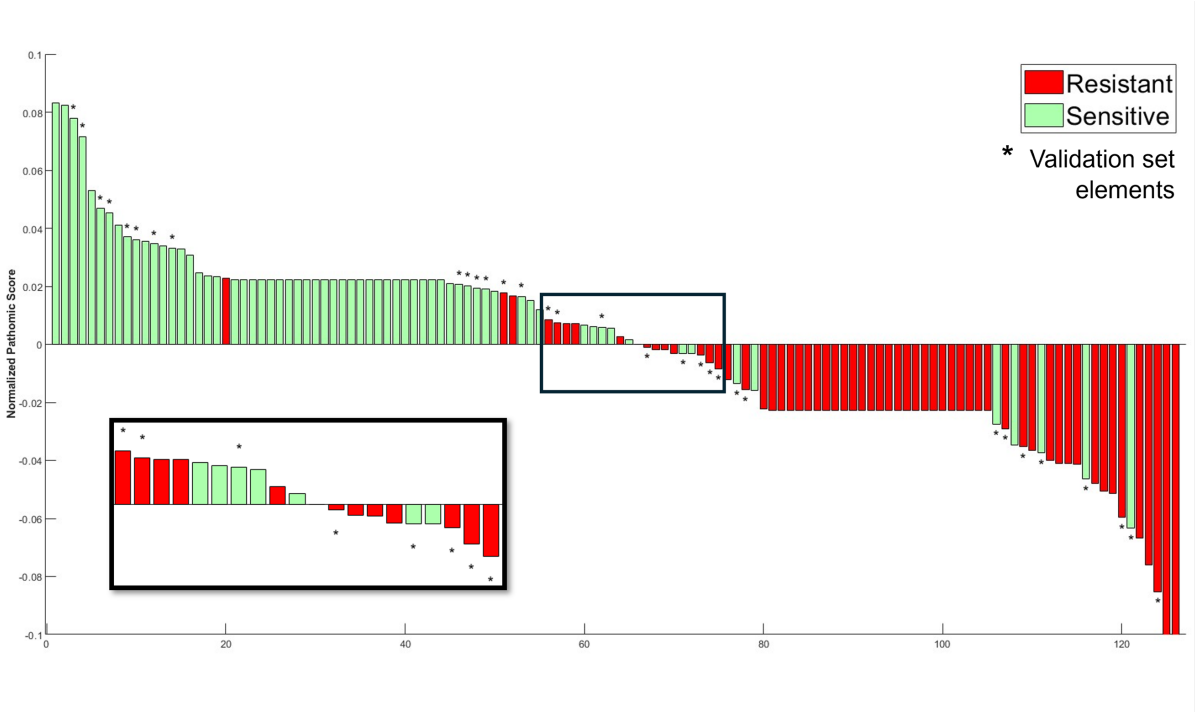

Fig. 4: Waterfall plot of the training and validation set predictions. The y-axis represents the Normalized Pathomic Score, i.e., SVM probability – classification cut-off (0.497). Negative Pathomic Scores represent the probability of being resistant while positive Pathomic Scores represent the probability of being sensitive. Therefore, red bars having a positive section represent patients incorrectly classified as sensitive, while green bars having a negative Pathomic Score are patients incorrectly classified as resistant. All bars marked with an asterisk (*) represent patients beloging to the validation set.

4. Conclusions

In this study, we developed an innovative AI framework for digital pathology images based on the method of the BoW. Using this system, we were able to develop and preliminary validate a pathomics signature to predict response to treatment of mCRC patients.

Further analysis should be performed to increase performances, e.g., by increasing the sample size of both the training and validation sets, and explainability, e.g., by including a visual assessment of patches performed by the expert pathologist to understand whether the patch types have real pathological meaning.

In conclusion, we presented preliminary results of a non-invasive biomarker that can potentially improve the use of personalized medicine in mCRC patients, being a feasible alternative to the "one size fits all" approach currently used. This

biomarker could promote a paradigm shift to personalized routine care of mCRC patients delivering better outcomes such as increased survival and quality of life.

Acknowledgements

This work was supported by 5 per Mille 2018-ID. 21091 program/Italian Association for Cancer Research.

References

[1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," CA. Cancer J. Clin., vol. 73, no. 1, pp. 17–48, Jan. 2023, doi: 10.3322/caac.21763.

[2] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, and A. Jemal, F. Bray, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," CA. Cancer J. Clin., vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.

[3] M. G. Fakih Marwan, L. Salvatore, T. Esaki, D. P. Modest, D. P. Lopez-Bravo, J. Taieb, M. V. Karamouzis, E. Ruiz-Garcia, T. Kim, Y. Kuboki, F. Meriggi, D. Cunningham, K. Yeh, E. Chan, J. Chao, Y. Saportas, Q. Tran, C. Cremolini, and F. Pietrantonio, "Sotorasib plus Panitumumab in Refractory Colorectal Cancer with Mutated KRAS G12C.," N. Engl. J. Med., Oct. 2023, doi: 10.1056/NEJMoa2308795.

[4] A. Cervantes, R. Adam, S. Roselló, D. Arnold, N. Normanno, J. Taïeb, J. Seligmann, T. De Baere, P. Osterlund, T. Yoshino, and E. Martinelli, "Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up," Ann. Oncol., vol. 34, no. 1, pp. 10–32, Jan. 2023, doi: 10.1016/j.annonc.2022.10.003.

[5] K. S. Wang, G. Yu, C. Xu, X. H. Meng, J. Zhou, C. Zheng, Z. Deng, L. Shang, R. Liu, S. Su, X. Zhou, Q. Li, J. Li, J. Wang, K. Ma, J. Qi, Z. Hu, P. Tang, J. Deng, X. Qiu, B. Y. Li, W. D. Shen, R. P. Quan, J. T. Yang, L. Y. Huang, Y. Xiao, Z. C. Yang, Z. Li, S. C. Wang, H. Ren, C. Liang, W. Guo, Y. Li, H. Xiao, Y. Gu, J. P. Yun, D. Huang, Z. Song, X. Fan, L. Chen, X. Yan, Z. Li, Z. C. Huang, J. Huang, J. Luttrell, C. Y. Zhang, W. Zhou, K. Zhang, C. Yi, C. Wu, H. Shen, Y. P. Wang, H. M. Xiao, and H. W. Deng, "Accurate diagnosis of colorectal cancer based on histopathology images using artificial intelligence," BMC Med., vol. 19, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/s12916-021-01942-5.

[6] A. Echle, N. Ghaffari Laleh, P. L. Schrammen, N. P. West, C. Trautwein, T. J. Brinker, S. B. Gruber, R. D. Buelow, P. Boor, H. I. Grabsch, P. Quirke, and J. N. Kather, "Deep learning for the detection of microsatellite instability from histology images in colorectal cancer: A systematic literature review," ImmunoInformatics, vol. 3–4, p. 100008, Dec. 2021, doi: 10.1016/J.IMMUNO.2021.100008.

[7] H. Qiu, S. Ding, J. Liu, L. Wang, and X. Wang, "Applications of Artificial Intelligence in Screening, Diagnosis, Treatment, and Prognosis of Colorectal Cancer," Curr. Oncol., vol. 29, no. 3, pp. 1773–1795, Mar. 2022, doi: 10.3390/curroncol29030146.

[8] M. Macenko, M. Niethammer, J. S. Marron; D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009, 2009, pp. 1107–1110, doi: 10.1109/ISBI.2009.5193250.

[9] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C. Weis, T. Gaiser, A. Marx, N. A Valous, D. Ferber, L. Jansen, C. C. Reyes-Aldasoro, I. Zörnig, D. Jäger, H. Brenner, J. Chang-Claude, M. Hoffmeister, and N. Halama, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," PLoS Med., vol. 16, no. 1, p. e1002730, 2019, doi: 10.1371/journal.pmed.1002730.