

Medical Database for Detecting Neoplastic Lesions in Human Colorectal Cancer with Deep Learning



Rita Fioresi^{*1}, Francesco Faglioni² and Paola Sena³

¹Department of Mathematics, University of Bologna, Italy

²Department of Chemical and Geological Sciences, University of Modena and Reggio Emilia, Italy

³Department of Biomedical, University of Modena and Reggio Emilia, Italy

Received:  August 06, 2018; Published:  August 13, 2018

*Corresponding author: Rita Fioresi, Ufficio D9, Dipartimento di Matematica, Piazza Porta San Donato 5, 40126 Bologna, Italy, Tel: +39 051 2094477; Email: rita.fioresi@unibo.it

Abstract

Medical databases are fundamental for developing new techniques for early detection of neoplastic cells. They are however difficult to obtain, since the labelling of the images is often operator dependent, requires specialized skills and the written informed consent of the patient. The variability of structures in biological tissue poses a challenge to both manual and automated analysis of histopathology slides. Although some authors showed moderate to good agreement among expert pathologists, and satisfactory results on their intra-observer reliability, other studies found that even experienced pathologists frequently disagree on tissue classification, which may lead to the conclusion that solely using expert scoring as gold standard for histopathological assessment could be insufficient. Hence, there is a growing demand for robust computational methods in order to increase reproducibility of diagnoses. In this note we present a database containing images of preneoplastic and neoplastic colorectal tissues and in a forthcoming paper we will describe our proposed DL algorithm to classify them into the following categories: normal mucosa, early preneoplastic lesions, adenomas, cancer.

Aim and Scope

Colorectal cancer ranks among the three most common cancers in terms of both cancer incidence and cancer-related deaths in Western industrialized countries [1]. Every year in the world nearly 1.3 million new cases of CRC are reported and nearly 700.000 patients die [2]. Lifetime risk of colorectal cancer may reach 6% of the population living in developed countries. CRC is second in incidence in Europe only to lung cancer, and it causes around 204.000 deaths every year [3]. The age-specific incidence of colorectal cancer rises sharply after 35 years of age, with approximately 90% of cancers occurring in persons over 50 years old. As in other developed areas, in Italy CRC incidence ranks third for men (after prostate and lung cancers), and second for women (after breast cancer). The incidence for men had an upward trend until the mid of the first decade of the second millennium (+2.2% in the period 1999-2007) followed by a reduction (-6.8% per year after 2007), in part due to the activation of organized screening programs. The trend is similar in women: there was an increase (+2.1% /year in the period 1999-2006) and subsequently a reduction (-3.6% /year after 2006). The burden of the disease remains, however, serious in Italy as well as worldwide, because of the social impact, costs, and mortality. Accurate tumor grading is essential for patient survival and can be done most effectively in stained histopathological sections harvested via biopsy or during

surgery. Our goal is to develop an effective classification strategy through the construction of an effective database for a DL analysis of the above mentioned biological data (Figure 1).

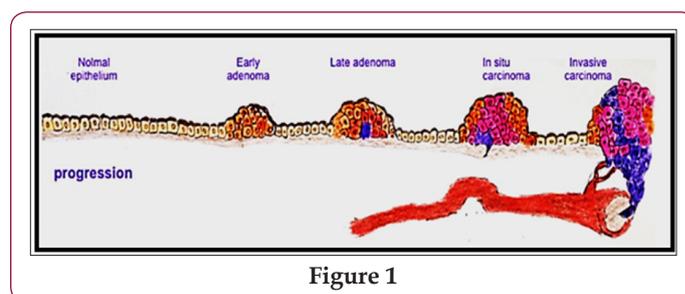


Figure 1

Results and Methodology

A typical histopathological image of colon glands contains four tissue components: lumen, cytoplasm, epithelial cells, and stroma (connective tissue, blood vessels, nervous tissue, etc.). The epithelial cells form the gland boundary, enclosing cytoplasm and lumen, whereas stroma is not considered part of the gland. If we just consider non-cancerous (benign) glands, DL algorithms must effectively be able to deal with significant variability in shape, size, location, texture and staining of glands. Moreover, in cancerous cases gland objects can significantly differ from benign glands, and the presence of corrupted areas (artifacts) further exacerbates

the problem. Therefore, machine learning approaches are predominantly used to develop robust models trained from labeled examples in order to cope with tissue variability. An effective algorithm for medical diagnoses needs large training datasets, in general extremely difficult to obtain, in order to correctly classify the different subtypes of benign and malignant gland types (see images below). Colorectal cancer presents heterogeneity during the adenoma carcinoma sequence [4].

The normal epithelium becomes a hyperproliferative mucosa and subsequently gives rise to a benign adenoma, which can then evolve into a carcinoma and metastases. Currently, but we are in the process of expanding our database, we have collected over 2000 images, divided into the four categories detailed above. We used an optical microscope Leica IRBE with CCD Rising Tech. Sony CCD

Sensor (USB2.0 5.0MP CCD ICX452AQ) camera and Software Image Pro Plus 4.5 with a 20X magnification and 600 dpi of resolution. Then, each slide was scanned without any data that could be linked to patient ID. The ethical committee of the University of Modena has approved the procedure. Three pathologists of the Unit of Pathology of Modena independently reviewed the images in order to correctly identify the stage of colorectal lesions [5-7]. Each image was divided into 9 equal parts and then, after relabelling by experts, rotated and reflected, so to obtain a database of over 10000 images, thus suitable for our machine learning algorithm. This database was furtherly subdivided into training (8000 images), validation (1000 images) and test (1000 images) sets. We obtained with our DL algorithm over 95% accuracy and we describe our procedure in detail in the forthcoming paper [7] (Figure 2).

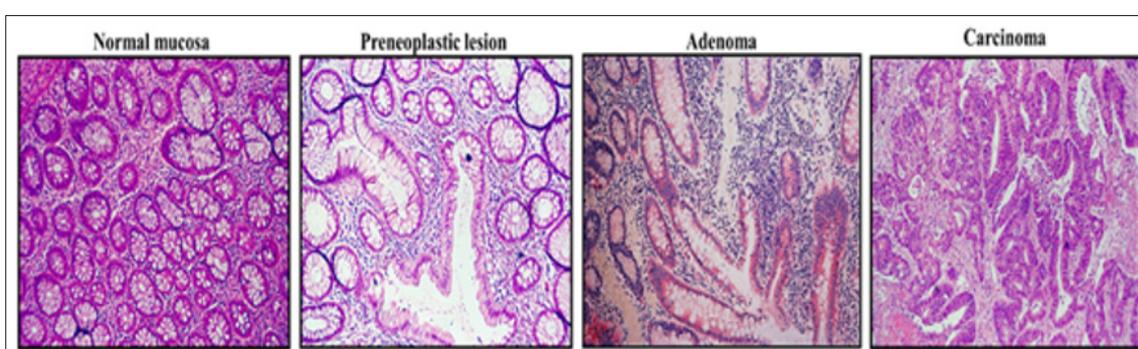


Figure 2

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Jemal A, et al. (2015) Global Cancer Statistics 2012. *CA a cancer J Clin* 65(2): 87-108.
2. Das V, Kalita J, Pal M (2017) Predictive and prognostic biomarkers in colorectal cancer: A systematic review of recent advances and challenges. *Biomed Pharma cother* 87: 8-19.
3. Ferlay J, Steliarova Foucher E, Lortet Tieulent J, Rossos S, Coebergh JW, et al. (2013) Cancer incidence and mortality patterns in Europe: Estimates for 40 countries in 2012. *Eur J Cancer* 49(6): 1374-1403.
4. Di Gregorio C, Losi L, Fante R, Modica S, Ghidoni M et al. (1997) Histology of aberrant crypt foci in the human colon. *Histopathology* 30(4): 328-334.
5. Roncucci L, Stamp D, Medline AA, Cullen JB, Bruce WR (1991) Identification and quantification of aberrant crypt foci and microadenomas in the human colon. *Human Pathology* 22: 287-294.
6. Gupta AK, Pinsky P, Rall C, Mutch M, Dry S et al. (2009) Reliability and accuracy of the endoscopic appearance in the identification of aberrant crypt foci. *Gastrointest Endosc* 70(2): 322-330.
7. Sena P, Fioresi R, Faglioni F, Losi L, Faglioni G, Roncucci L (2018) detecting preneoplastic and neoplastic lesions in human colorectal histological images. Preprint.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2018.07.001572

Rita Fioresi. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>