

1. **Description of the assignment of the project**

Digital pathology: introduction to the problem

*In digital pathology a diagnosis is carried out by analysing histopathological samples which are pieces of tissues extracted via surgical operation. Specimens are typically stained with H&E (haematoxylin and eosin) so that different structures come in different shades between blue (haematoxylin binds to cell nuclei as they are negatively charged) and pink (eosin binds to extracellular matrix and cytoplasm as they are positively charged) for better distinguish between them (*Chan JK (2014). ["The wonderful colors of the hematoxylin-eosin stain in diagnostic surgical pathology"](https://www.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&tool=sumsearch.org/cite&retmode=ref&cmd=prlinks&id=24406626). Int J Surg Pathol. 22 (1): 12–32*). Nowadays digital slides are obtained by scanning specimens placed on conventional glass slides; such multi-resolution slides are called WSI (Whole Slide Images) and can be elaborated numerically, enabling different applications.*

# *It is well known that molecular expression of diseases tend to manifest in differences in the tissue architecture and morphology: the traditional approach consists in the visual examination of samples, carried out by a clinician, with the aim of detecting abnormalities related to a certain disease (i.e. if a tissue is cancerous or not). Visual examination is time consuming, prone to inter-reader and intra-reader variability, strongly depends on the skills of the operator and non-reproducible as the human eye is less adept to recognize changes in the tissues: these issues can be overcome making available to the pathologists a tool that supports them during the visual evaluation.(Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images, Pegah Khosravi, Ehsan Kazemi )*

# *Computed aided diagnosis systems (CAD) are thought to help clinicians in everyday tasks: the clinician is not put aside, but yet supported by tools that can, among all, improve the prediction of disease aggressiveness and of the patient outcome by suggesting details about, for example, medical images without substituting the clinician in the final decision. (Image analysis and machine learning in digital pathology: Challenges and opportunities, Anant Madabhushi, George Lee)*

*The focus of the project and clinical insights*

# *The focus of the project was developing a software able to produce an attention-map for cancer detection that drives pathologists’ attention to certain areas of the slice that might be pathological and might require further analysis. (Histopathological Image Analysis: A Review, Gurcan)*

*Adenocarcinoma and adenoma tissue samples were considered: adenocarcinoma is a cancerous tumour that interests epithelial tissue (i.e. tissue that interests inner and outer surface cavities in many organs and blood vessels) that has glandular origin and/or characteristics. On the other hand, adenoma is a benign tumour, but should be treated as pre-cancerous and requires attention because might turn into adenocarcinoma.*

*Such clinical aspects were kept into constant consideration during the development of the project: there is a huge variability between the appearance of tissues (cancers usually contain cells that are different grades) and intrinsic uncertainty that was modelled with an approximation of a Bayesian CNN trained with WSI images representing AD tissue (short for adenoma, a benign tumour of epithelial tissue with glandular origin or characteristics), AC tissue (short for adenocarcinoma, a malign tumour) and healthy tissue.. It can be seen that the tissue progressively loses coherence in gland patterns as it becomes* *carcinogenic. The morphology of the tissue or characteristics of nuclei are hallmarks for cancerous conditions: different metrics have been developed by clinicians for describe (after visual examination of samples) cancer basing on how abnormal the cells look and how quickly they grow; digital pathology makes possible a quantitative characterization of pathology imagery that is important not only for clinical purposes, but also for research, when providing reliable and innovative metrics of evaluation.* ***(****Deep Convolutional Neural Networks Enable Discrimination of**Heterogeneous Digital Pathology Images****,*** *Pegah Khosravi****)***

*(images)*

1. **Theory about CNNs focused on the application outline, providing references**

*CNNs*

*Neural networks are a machine learning approach that relies on several computational units, called neurons, differently interconnected via weights as NNs take inspiration from the way the brain is organised: weighted interconnections between neurons stand for biological synapses. The knowledge of the network is preserved in the weights and the behaviour of the network is related to its hierarchical architecture: the learning process consists in adjusting the weights extracting a mathematical model that fits training data giving as output a classification result or a prediction. (Deep Convolutional Neural Networks Enable Discrimination of  
Heterogeneous Digital Pathology Images, Pegah Khosravi). Every deep learning network begins with the assumption of random initialization of weights and, at each iteration, data is propagated through the network to compute the output.*

*There are many challenges in the automatic analysis of digital pathology images, as said before, such as the variability of the morphology of the sample due to the pathology and to the preparation of slides and the variations in staining.*

*The variations between patients and clinical conditions have always made tedious to find handcrafted features that can be integrated in a system making it robust, efficient and reliable : deep learning methods overcome these issues deriving a feature space from the data itself and gaining the capability of generalization when unseen data is presented to the network. (Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases* [*Andrew Janowczyk*](https://www.ncbi.nlm.nih.gov/pubmed/?term=Janowczyk%20A%5BAuthor%5D&cauthor=true&cauthor_uid=27563488) *and* [*Anant Madabhushi*](https://www.ncbi.nlm.nih.gov/pubmed/?term=Madabhushi%20A%5BAuthor%5D&cauthor=true&cauthor_uid=27563488)*)*

*CNNs (Convolutional Neural Networks) are neural networks where the local connectivity pattern between neurons is inspired by the organization of the animal visual cortex and information is processed similarly to how the brain would do; cortical neurons respond to stimuli in a specific region of the space known as receptive field and this behaviour can be mathematically modelled via convolutions.*

*(image)*

*The 3 characteristic layers of CNN are the convolutional layers, the non-linear layers and pooling layers. The core of the network are the convolutional layers where, via a set of filters (kernels), feature maps are obtained from the input image and fed to the non-linear layer, characterized by an activation function: after this, the pooling layer reduces the number of features. As hidden units are connected to local receptive fields and share weights resulting in spatial invariance (i.e. a pattern can be recognized in different areas of the input image) and an optimization of the computation, the input can have a high dimension without resulting in many parameters: these parameters are learned during the training via the backpropagation algorithm.*

*Each hidden layer is dedicated to identifying a multiple feature of the input: low-level features are condensed in the deepest layers while problem-specific features belong to last layers (with no pre-existing assumptions about the particular tasks or dataset in form of encoded domain-specific information); such characteristics allow the network to be more flexible when extracting, during the training procedure, different combination of small patterns eventually combining them for the aim of the network. Regarding the training procedure, the backpropagation algorithm is the most used method and consists in the update of the weights, initially random initialised, basing on a loss term that is computed with the output given by the network and the desired output.*

*(Image?)*

* *Input features propagates in through the network in the forward direction computing the output and the*
* *The training loss is derivate with respect to the weights and computed back towards the input*

*This is an iterative procedure that is repeated until a certain stopping condition is reached: the tuning of the parameters of the backpropagation is proportional to the size of data.*

***Bayesian CNNs***

*The risk of overfitting when the network is not trained on a large dataset and the falsely overconfidence in the prediction related to the absence of a measure of uncertainty are typical drawbacks of conventional deep learning methods, based on point predictions.*

*Bayesian CNNs were thought to handle these problems and considered during this project: they are based on the Bayes’ theorem that is the fundamental of Bayesian inference, a way of quantifying model uncertainty. According to this, each observation is an opportunity to update the beliefs about a given deep learning model. Moreover, Bayesian CNNs are robust to outliers and are key solution when the lack of a large amount of data can result in unreliable networks. In theory, weights are not point estimated, but have a mean and a standard deviation, which are two hyperparameters that are updated during backpropagation. (Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning Yarin Gal)*

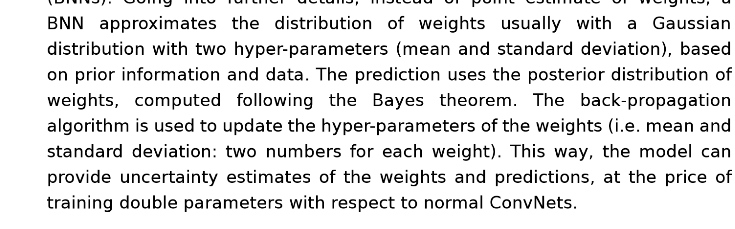
***Bayes’ theorem***

*The Bayes’ rule shows how the degree of belief in a model (posterior function, ) is related to the likelihood of the occurrence of the data , to the knowledge about the data (the prior ,) and to the evidence (marginal likelihood, ).*

*The posterior function is the probability distribution of interest that summarizes the knowledge about the model parameters given data and needs to be estimated given that the aim is obtaining the parameters of the model in order to get the correct output for a given input. The prediction of new observations is made through model update on the posterior predictive distribution, the neural network of interest being a conditional model parameterized by the weights.*

*The exact Bayesian inference is intractable, and Bayesian CNNs come with a high computational cost: the estimation can only be approximated via several method. (Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. Yongchan Kwon)*

*Stochastic regularization techniques like dropout regularization can be used to approximate inference in Bayesian models without resulting in excessive computational costs. (Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning Yarin Gal)*



***Dropout***

*Dropout is a regularization technique that prevents overfitting and improve generalizability by randomly ‘dropping out’ (i.e. inactivating) units of a neural network with a certain probability: for each training sample different units are dropped out, resulting in a training procedure on reduced networks. (Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopatholog**y* [*Harshita Sharmaa*](Harshita%20Sharmaa)*)*

*When the dropout is applied at both training and test time, we have the Monte Carlo dropout: setting the dropout rate and the number of iterations, the same element of the dataset is presented to the network different times and, for each presentation, a different result is obtained. At test time the prediction is no longer deterministic but depends on which nodes is randomly choose to be kept: given a same datapoint, the model can predict different values each time. The primary goal of Monte Carlo dropout is to generate random predictions and interpret them as samples from a probabilistic distribution. (Kwon, Yongchan, et al. "Uncertainty quantification using bayesian neural networks in classification: Application to ischemic stroke lesion segmentation."  
(add details)*

1. **Detailed description of the method you used (describe most salient aspects of your code too)**

Dataset creation

The dataset consists in digital histopathological images of different dimensions, belonging to different patients and representing different classes: adenoma, adenocarcinoma and healthy tissue.

(grafico dataset – numero immagini e dimensioni)

As the dataset only included 70 images, crops had to be generated with the purpose of data augmentation.

1. Images were cropped with squared crops of 1344, 2240 and 3136 pixels: all crops dimensions are multiple of 224 (crops are then resized to 224x224 in order to feed them to the input layer of the network). Crops dimensions were chosen for incorporating information at multiple resolution and at different level of detail.

The dimension of crops kept into consideration domain knowledge and required a visual analysis of images at different resolutions using Aperio ImageScope, investigating the appearance of the tissues isolated with different crop sizes, considering the contextual information and neighbourhood: some areas might be difficult to differentiate without neighbourhood information if the view field is small.

(immagine con sovrapposti vari crop)

It is relevant to say that no information was given to the network about crops that included blood vessels and other part of tissues not of interest and no crops were removed in such sense, but only crops with a percentage of white pixels greater than 70% (where such a high quantity of white pixels belonged to the slide and not to the tissue) where excluded.

(esempi crop vari e percentuali di bianco)

The overlap between crops was distributed between crops …. (distribuzione overlap: per ridurre il rischio di tagliare strutture e raggruppamenti (la morfologia è importante dal punto di vista clinico, distribuzione di nuclei))

1. No pre-processing steps where applied to the images (such as brightness, contrast and intensity adjustments or affine transformations) in order to preserve the salient texture, colour and morphological properties of the original stained images.
   1. The dataset was divided into training and testing: crop belonging to the same patient were considered entirely or in the training set or in the dataset. This led to an unbalanced dataset that was balanced using 1120 pixels squared crops randomly obtained from the original image. The integration with such crops instead of the deletion of crops was considered a better approach for the balancing of the dataset: in this way no information is deleted and the dataset is kept representative, taking into account the fact that given original dataset was small.

(grafico numero crop per ogni classe per visualizzare lo sbilanciamento)

1. Crop normalization?
2. Formato immagini fornite alla rete

The network architecture

(grafico rappresentazione rete + dimensioni layers)

This is the approximation of a Bayesian neural network: it consists in….

* Input layer…
* Other layers details
* Output layer: we consider the maximum calculated probability between the three classes taken into consideration (H, AD, AC) in order to assign the class label to a crop. As we are considering an approximation of a Bayesian NN, a crop is presented to the network for a certain number of iterations: at each iteration different units are activated, and the output varies. Mean and variance of the prediction are computed taking into consideration all the outputs (add details?). Dropout rate and iterations are further discussed.
* The result of the network is used so that an attention map can be visualized superimposed to the original image: (details about how we choose the color and intensity, colormap etc)

**Motivazione scelta classi: il numero di immagini dei pazienti sani era troppo basso (aggiungere** dettagli)

(grafico sui crop – immagini sane sono più piccole!!)

Casting – definition of the network

**Dire perché abbiamo scelto questa rete piuttosto che un’altra.**

Training the network

Parametri e grafici (batch size etc)

Dropout: le varie iterazioni! Dalla letteratura

* Crop dimensions
* Dropout – iterazioni direttamente proporzionali

The implementation

* Librerie python (Keras, TensorFlow, numpy)
* Colab etc
* Colori su immagine: probabilità e varian

Python code

* **+ DA SCRIVERE parte sul codice!**

(Pipeline schematica)

1. **Detailed description of the results (provide graphs, tables, etc.);**

* Tempo – crop size – training time & accuracy
* Drop-out e iterazioni
* Test time
* Manual segmentation – varianza e accuratezza integrate

(Grafici)

GUI and visualization

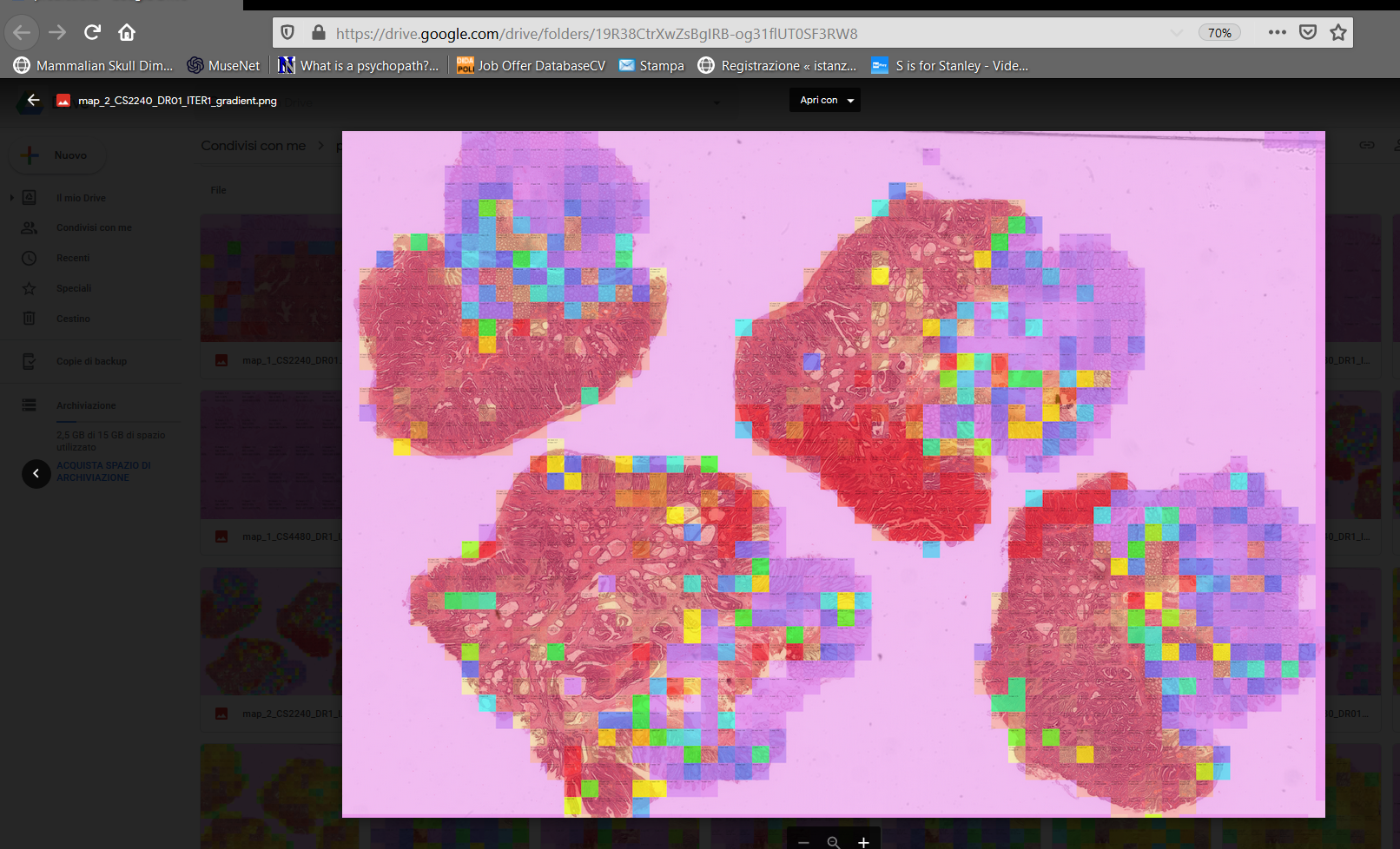
The GUI was developed basing on considerations about the user experience: the clinician needs a ready to use and intuitive interface that can display the attention map related to the image taken into consideration.

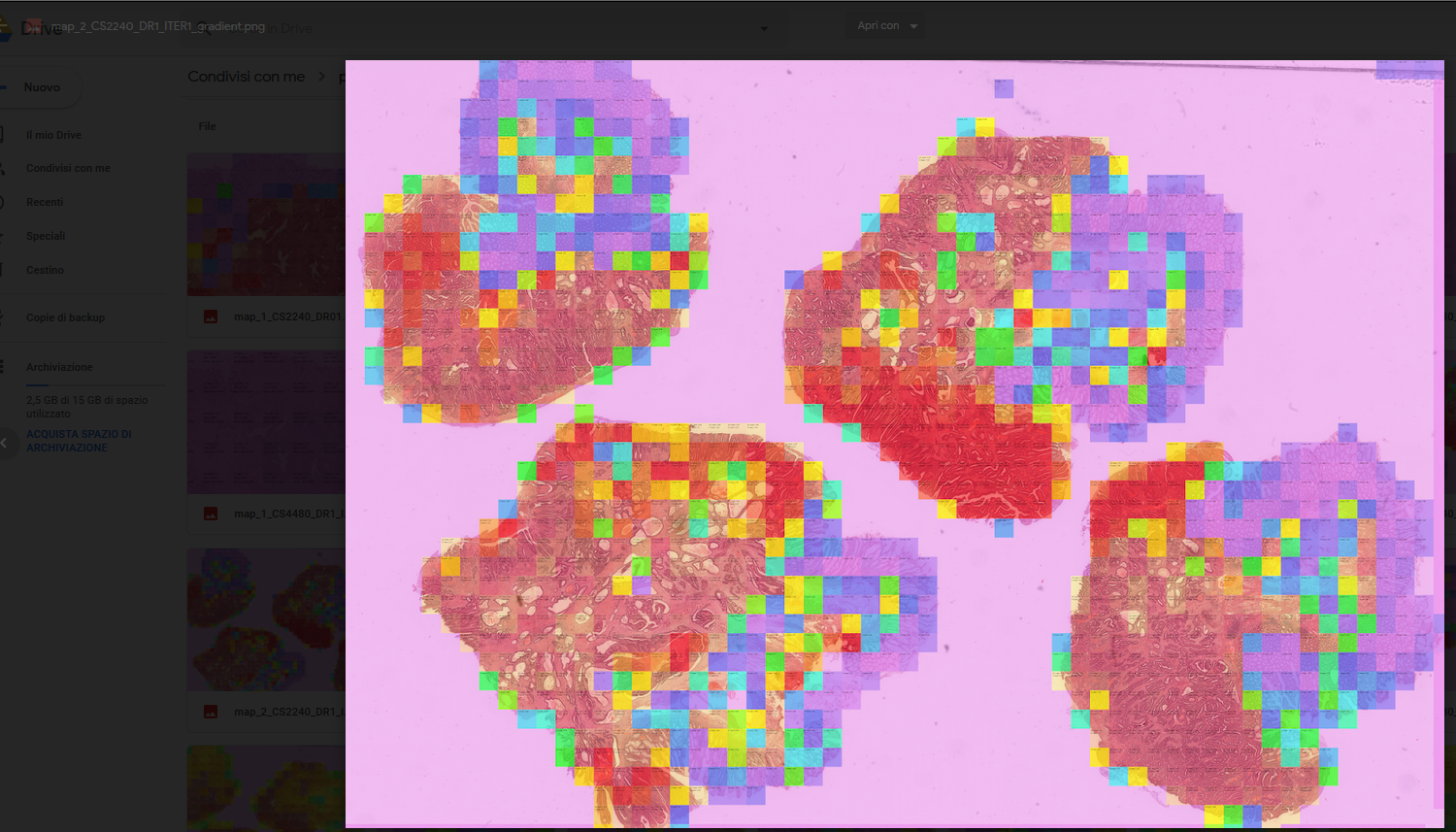
Features of the GUI:

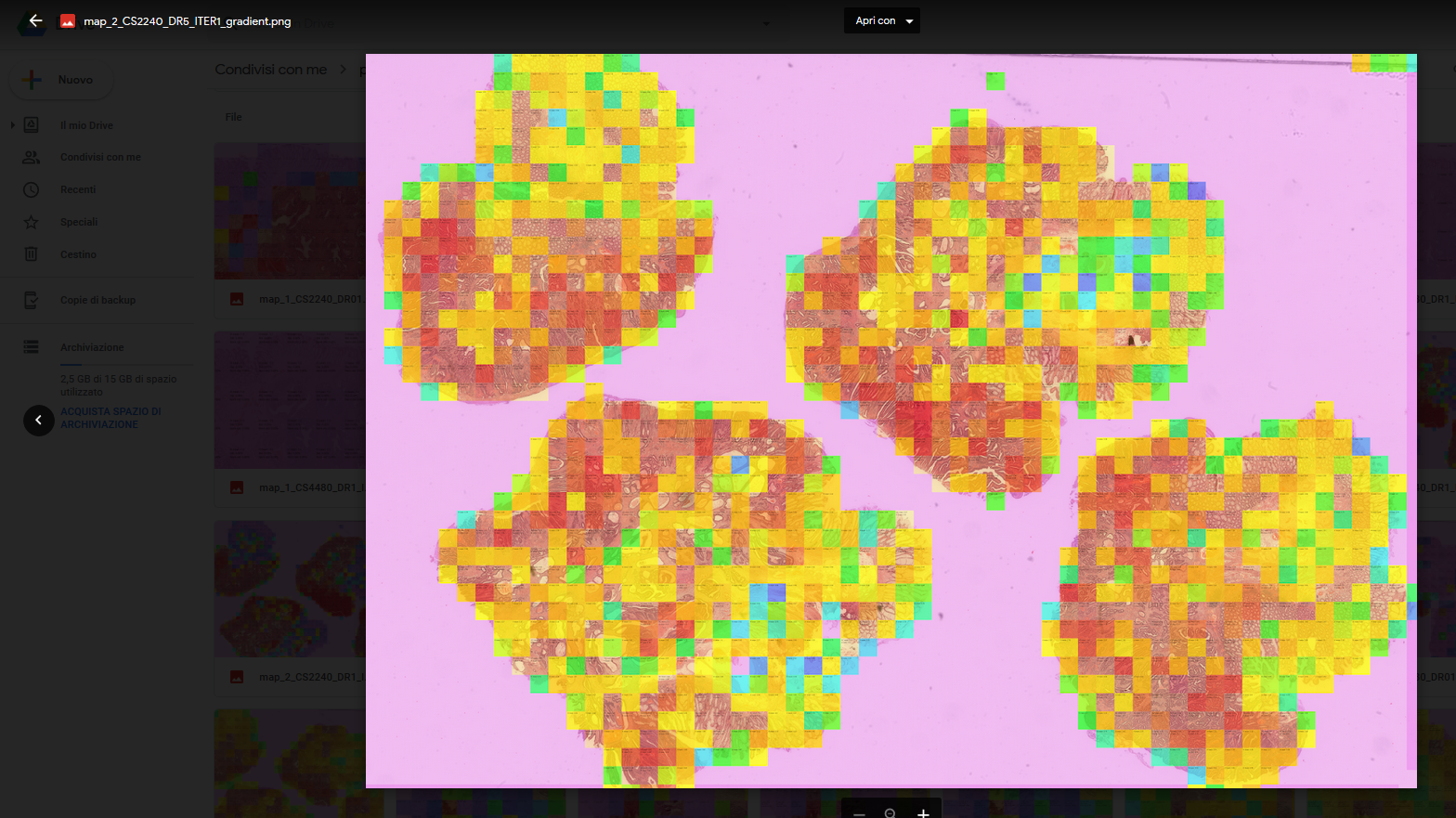
* Batch mode
* Single image visualization
* Selection of folders
* Saving?
* Color map per il medico

1. **Results discussion**

* **Metrica di confronto segmentazioni**
* **Identification for structures in the edges**



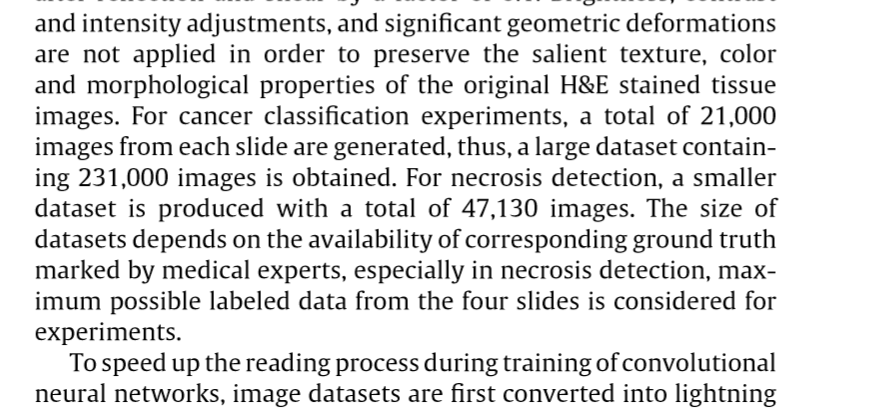




1. **Future development**

The importance of digital pathology and the integration of tools that can help and support decision in the clinical workflow have a huge impact on the treatment and study of pathologies: having the samples electronically scanned makes, for example, easier to collect second opinions and share knowledge; the tool developed in this project might integrate the possibility of annotating images basing on the attention map and share such results between clinicians via tele-histopathology (si parla anche di milioni di patches).

As future development, an integration of the dataset should be taken into consideration in order to make it more consistent and capable of generalization. Regarding this, is known that differences in staining in different samples can make a huge difference: it could be interesting integrating images acquired in different laboratories and make the system resistant to variations by normalizations of images and inhomogeneity corrections.

 dall’articolo Shaarma ([2] latex)

Moreover, integrating in the network crops belonging to tissues not of interest such as blood vessels, might improve performances.

Consider 2 or 3 classes?

GPU implementation?

**References**:

# [5] <https://www.sciencedirect.com/science/article/abs/pii/S0895611117300502>

# https://ieeexplore-ieee-org.ezproxy.biblio.polito.it/document/5299287

# Receptive fields, binocular interaction and functional architecture in the cat's visual cortex [D. H. Hubel](https://www.ncbi.nlm.nih.gov/pubmed/?term=Hubel%20DH%5BAuthor%5D&cauthor=true&cauthor_uid=14449617) and [T. N. Wiesel](https://www.ncbi.nlm.nih.gov/pubmed/?term=Wiesel%20TN%5BAuthor%5D&cauthor=true&cauthor_uid=14449617)

# Bayesian Neural Network Series Post 1: Need for Bayesian Neural Networks <https://medium.com/neuralspace/bayesian-neural-network-series-post-1-need-for-bayesian-networks-e209e66b70b2>

# CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more… <https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5>

* https://medium.com/neuralspace/bayesian-convolutional-neural-networks-with-bayes-by-backprop-c84dcaaf086e
* [*https://medium.com/@shridhar743/a-beginners-guide-to-deep-learning-5ee814cf7706*](https://medium.com/@shridhar743/a-beginners-guide-to-deep-learning-5ee814cf7706)