1. **Description of the assignment of the project and the focus of your work**

Digital pathology: introduction to the problem

*In digital pathology a diagnosis is carried out by analysing histopathological samples which are pieces of human 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 [2].*

*The traditional approach consists in the visual examination of samples with the aim of detecting abnormalities, counting nuclei and so on, carrying out a certain diagnosis i.e. if a tissue is cancerous or not. Visual examination is time consuming, prone to inter-reader and intra-reader variability and non-reproducible: making available to the pathologists a tool that supports them during the evaluation can help overcoming these issues [3] as nowadays digital slides [4] 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 with many related advantages enabling, among all, telepathology that is a key requirement for second opinions on cases and remote consults. [1]*

*The focus of the project and clinical insights*

*The focus of the project is 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.*

*(image)*

*Uncertainty is a crucial aspect of such application and it is modelled with deep neural networks, more specifically with a Bayesian CNNs 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.*

*(images)*

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

*CNNs*

*Neural networks (NNs) are a machine learning approach that do not consist in one single computational unit but that relies on several units, called neurons, differently interconnected via weights: NNs take inspiration from the way the brain is organised. The knowledge of the network is preserved in the weights that stand for biological synapses and the behaviour 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 result or a prediction. Such model is usually too complex to be obtained by a human.*

*Among all the feed-forward artificial neural networks we have CNNs (short for Convolutional Neural Networks) where the local connectivity pattern between neurons is inspired by the organization of the animal visual cortex; 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: they consist in a input later, a variable number of hidden layers and a output layer. 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; such characteristics allow the network to be more flexible when extracting different combination of small patterns eventually combining them for the aim of the network.*

*(image)*

*Classification tasks and CNNs*

*There are many challenges in the automatic analysis of digital pathology images such as the variability of the morphology of the sample due to the pathology and to the preparation of slides or variations in staining and the variability between patients that have always made tedious to find handcrafted features that can make the system 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. [6]*

*The aim of CNNs is give a output that is a class or a prediction of the input and this is accomplished via several different layers that can be differently organized while information is processed similar to how the brain process information: the 3 characteristic layers are convolutional layers, non-linear layers and pooling layers. As hidden units are connected to local receptive fields and share weights, the input can have a high dimension without resulting in a large number of parameters, these parameters are learned during the training via backpropagation.[7]*

*The backpropagation algorithm is the most used method for training neural networks and consists in the update of the weights, initially random initialised, basing on an loss term that is computed with the output given by the network and the desired output.*

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

*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.*

*(details?)*

*The core of CNNs is the convolutional layers that is composed by a set of filters: each filter activates when a specific feature is detected in the input.*

*(details about the layers…?)*

***Bayesian CNNs***

*The risk of overfitting when the network is not trained on a large dataset and the falsely overconfidence [9] in the prediction related to the absence of a measure of uncertainty, are typical drawbacks of conventional deep learning methods.*

*Bayesian CNNs were thought to handle these problems: 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.*

*Bayesian CNNs are robust to outliers. It makes sense to use Bayesian inference in situations where it is very expensive to obtain a large amount of data.*

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 can be seen as 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.*

*Stochastic regularization techniques like dropout regularization can be used to approximate inference in Bayesian models without resulting in excessive computational costs. [8]*

Dropout

Dropout is a technique that prevents overfitting by randomly ‘dropping out’ (i.e. inactivating) units of a neural network with a certain probability.

Both training and testing?

Where

*RIFERIMENTO A NOSTRO PROGETTO*

*[9]*

Since MC Dropout is a type of VI (variational inference), it still gives an approximation to the posterior without any guarantee of convergence. But it is obvious that MC Dropout collects approximate inference even faster and scalable than VI. Furthermore, MC Dropout performs usually better in predictions than neural networks trained either by VI or MCMC (Gal & Ghahramani, 2016). However, we should keep in mind that Bayesian inference is not all about making better predictions. It seeks rather an understanding of the latent process that is supposed to have generated our observations. MC Dropout does not become more Bayesian than other methods if it performs better than other Bayesian methods.

Recent advances  
in variational inference introduced new techniques into  
the field such as sampling-based variational inference and  
stochastic variational inference

Given that good uncertainty estimates can be cheaply obtained from common  
dropout models, this might result in unnecessary additional  
computation.

*Dropout*

*Is a regularization technique used to prevent overfitting: for each layer of the networks a probability of removing units is set. The network is smaller: for each training sample different units are dropped out (= train using one of these reduced networks).*

*Addressing overfitting issues*

For **Monte Carlo dropout**, the dropout is applied at both training and test time. At test time, the prediction is no longer deterministic, but depending on which nodes/links you randomly choose to keep. Therefore, given a same datapoint, your model could predict different values each time.

So the primary goal of Monte Carlo dropout is to generate random predictions and interpret them as samples from a probabilistic distribution. In the authors' words, they call it **Bayesian interpretation**.

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)

*APPLICATION OUTLINE*

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

* Pipeline schematica
* Motivazione scelta dell’architettura della rete
* Processing delle immagini
* Validazione

Dropout method ([Srivastava et al., 2014](https://www-sciencedirect-com.ezproxy.biblio.polito.it/science/article/pii/S0895611117300502" \l "bib0255)) assists in reducing overfitting, especially when the available training data is limited such as the WSI data. During each iteration, individual nodes along with incoming and outgoing edges are removed from the network and are later returned along with their initial weights. In our approach, after each of the first two fully-connected layers, the dropout ratio i.e. the probability of dropping any input for both stages is set to 0.25. [7]

1. Detailed description of the results (provide graphs, tables, etc.);
2. Results discussion;
3. Future development

Possibilità di rendere disponibile il tool a diversi patologi per concordare sulla diagnosi in base alla mappa proposta, permettendo di annotare direttamente l'immagine prima di inviarla per un consulto 🡪 tele-istopatologia

**References**:

* **[1]** Image analysis and machine learning in digital pathology: Challenges and opportunities <https://www.sciencedirect.com/science/article/abs/pii/S1361841516301141>
* **[2]** 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. [*doi*](https://en.wikipedia.org/wiki/Digital_object_identifier):[*10.1177/1066896913517939*](https://doi.org/10.1177%2F1066896913517939). [*PMID*](https://en.wikipedia.org/wiki/PubMed_Identifier) [*24406626*](https://www.ncbi.nlm.nih.gov/pubmed/24406626)

# [3] Deep Convolutional Neural Networks Enable Discrimination of Heterogeneous Digital Pathology Images <https://www.sciencedirect.com/science/article/pii/S2352396417305078>

# [4] <https://www.sciencedirect.com/topics/medicine-and-dentistry/digital-pathology>

# [5] Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology [Harshita Sharmaa](https://www.sciencedirect.com/science/article/abs/pii/S0895611117300502#!) <https://www.sciencedirect.com/science/article/abs/pii/S0895611117300502>

# [6] 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)1 and [Anant Madabhushi](https://www.ncbi.nlm.nih.gov/pubmed/?term=Madabhushi%20A%5BAuthor%5D&cauthor=true&cauthor_uid=27563488)

# [7] Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology <https://www.sciencedirect.com/science/article/abs/pii/S0895611117300502>

# [8] Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning Yarin Gal

# [9] Kwon, Yongchan, et al. "Uncertainty quantification using bayesian neural networks in classification: Application to ischemic stroke lesion segmentation." (2018).

# 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>

* Deep Bayesian Active Learning with Image Data <https://arxiv.org/pdf/1703.02910.pdf>

# 