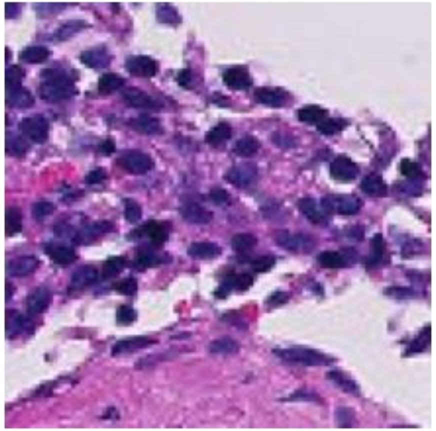
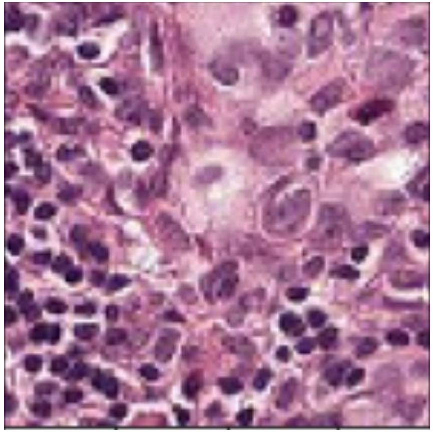
Pcam Challenge: Classification of Breast Cancer Metastasis

Keywords: Digital Histopathology, CNN, Gabor filters, XAI





MU4RBI07 Image Processing - Project

PICHON Fabien, DUSSARD Bastien, SIMON Louis, Master 1 ISI, Sorbonne Université

The PCamelyon challenge

Problematic:

- Perform fast and reliable classification of metastatic foci in sentinel lymph nodes

  ➥ Usually done by histopathologists with **high intra/inter-observer variability**

- Provide histopathologists with interpretable data for accurate breast cancer diagnosis

  ➥ Basic Deep-Learning algorithms provide **poor explanations**: less reliable for histopathologists

  ➥ Does not enrich the practice of histopathologists

Provided dataset [1]:

- Extended version of the Camelyon16 dataset with *already*-cropped patches
- Aiming at facilitating digital-histopathology using DL

Start of the art on metastasis classification

➨**Rotation Equivariant CNNs** [2]: G-CNN (Group Convolutional Neural Networks) for globally equivariant to rotation-metastatic cell detection

➨ **Deep Neural Network** [3]: Patch-level detection on WSI using Deep Neural Network (GoogLenet and VGG16)

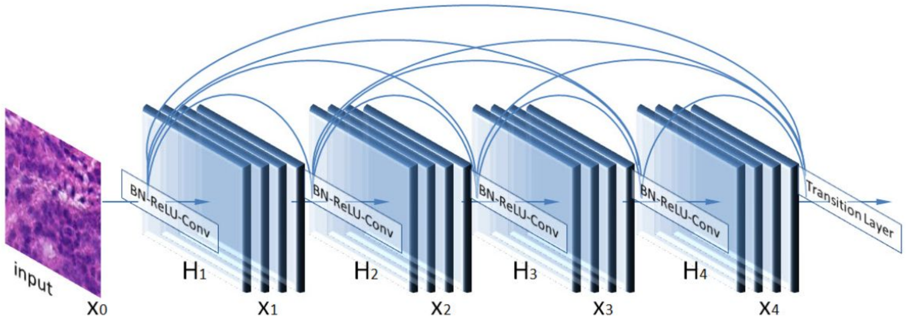
➨ **Semi-Supervised Learning** [4]: Dense-net combined with several training techniques, namely pseudo-labeling, Test Time Augmentation and ensemble learning

➨ **Single and multi-scale inception network** [5]: Patch-level detection using Inception V3 networks at different magnification

➨ **Interpretable CNN predictions** [6]: Local explanation on how CNN detects tumor tissues using LIME

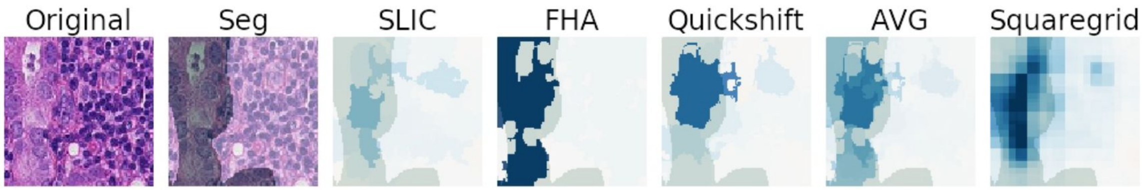
Review of methods used by other contestants

1) **Semi supervised** (Jaiswal et al.) → AUC = 0.98



- Use of DenseNet 201 network
- Learning on unlabelled data for entropy regularization
- Perform both classical learning and ensemble learning

2) **Interpretable CNN predictions** (Palatnik de Sousa et al.) → No performance given



- Use of VGG19 for classification
- Modified version of LIME
- Test of different segmentation algorithm for super-pixels extraction

Critical analysis of these methods

Method 1):

✔️ Reaches top AUC and accuracy,

✔️ Allows the use of unlabelled data in the training phase

❌ Does not produce any interpretable data (location, heatmap,...)

Method 2):

✔️ Provides readable data in the form of a heatmap

❌ Only provides location of tumors

Proposal of new methods

First method :

- CNN approach with LIME and Shap for interpretability
- Based on the article *Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases* [5]

✔️ Nice performances, uses different interpretability tools

❌ Complexity-depend model (more layers = more details but also more training time)

Second method :

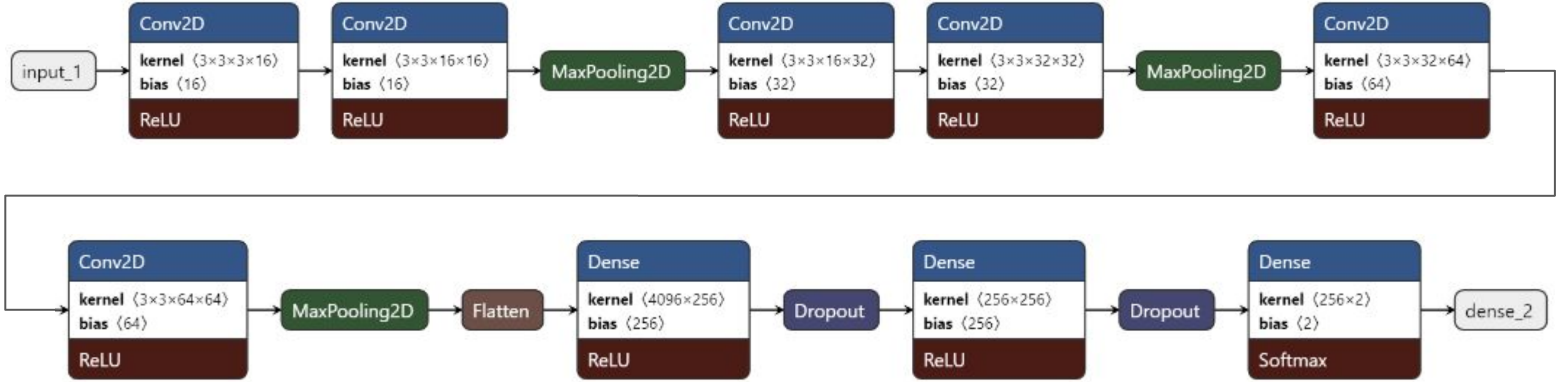
- Texture-based approach with Gabor filtering for feature extraction and a Random Forest classifier
- Fast and data-frugal training

✔️ Requires few data for training (~1000), provides information on the location of the tumor

❌ Long pre-processing of the image, depends on filter bank parameters

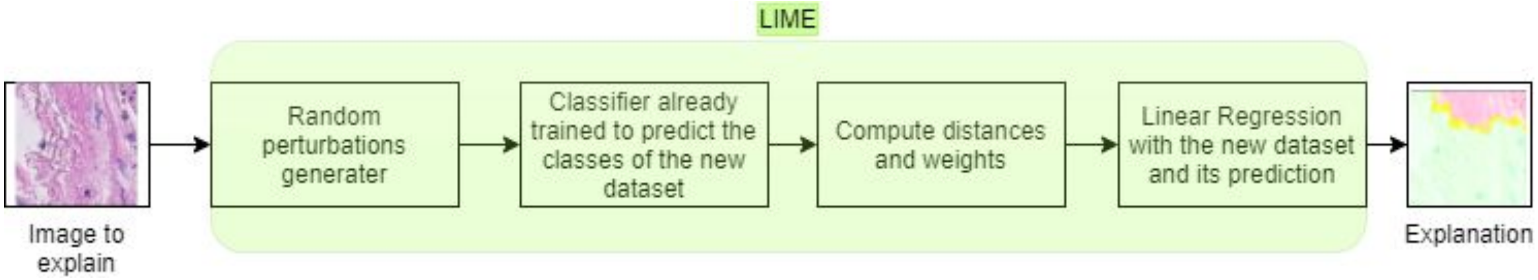
Development and Implementation of the algorithm: CNN

CNN :



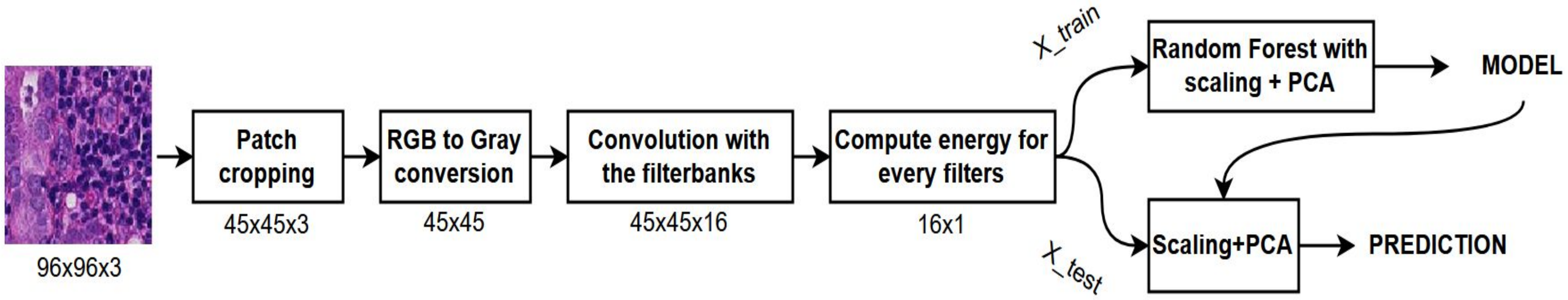
Total of trainable parameters = 1,187,218

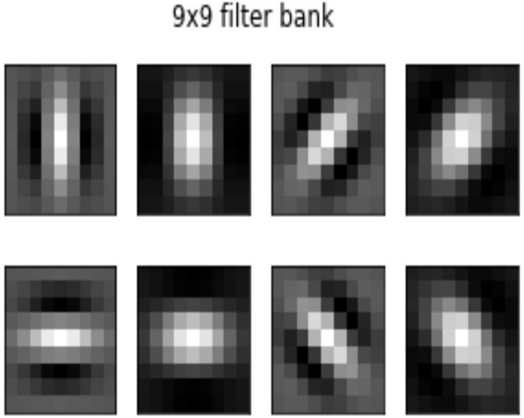
LIME:



Development and Implementation of the algorithm: Gabor

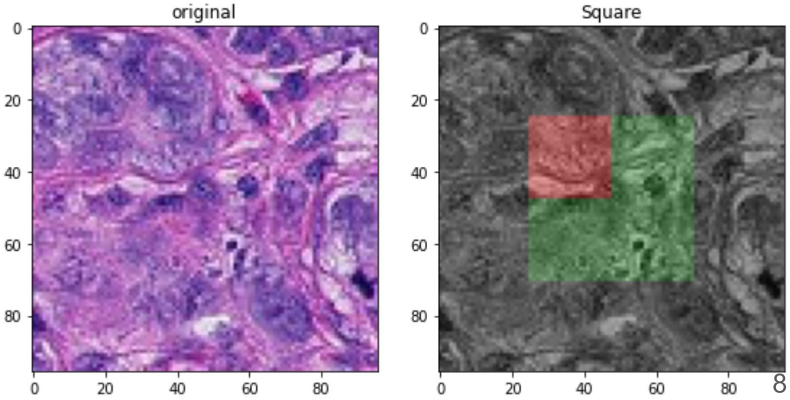
Used library: *Numpy, Matplotlib, Scikit-learn, Scipy*





Left: 9x9 filter bank

Right: Interpretable output (green=1,red=0)

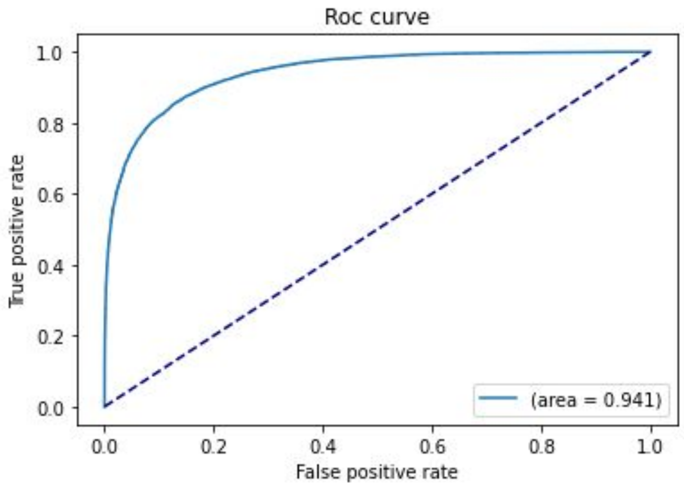


Testing the methods on the test dataset

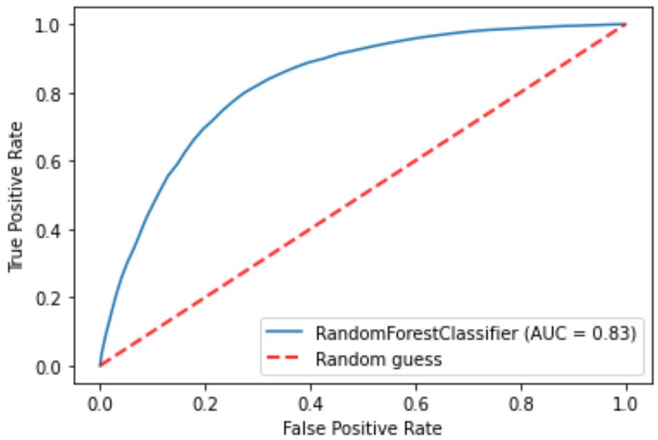
<u>Performances :</u> **CNN method:** AUC=0.94 & Accuracy=86%

Gabor method: AUC=0.83 & Accuracy=76% , (Sensitivity=0.82 & Specificity=0.71)

Reference [4]: AUC=0.98

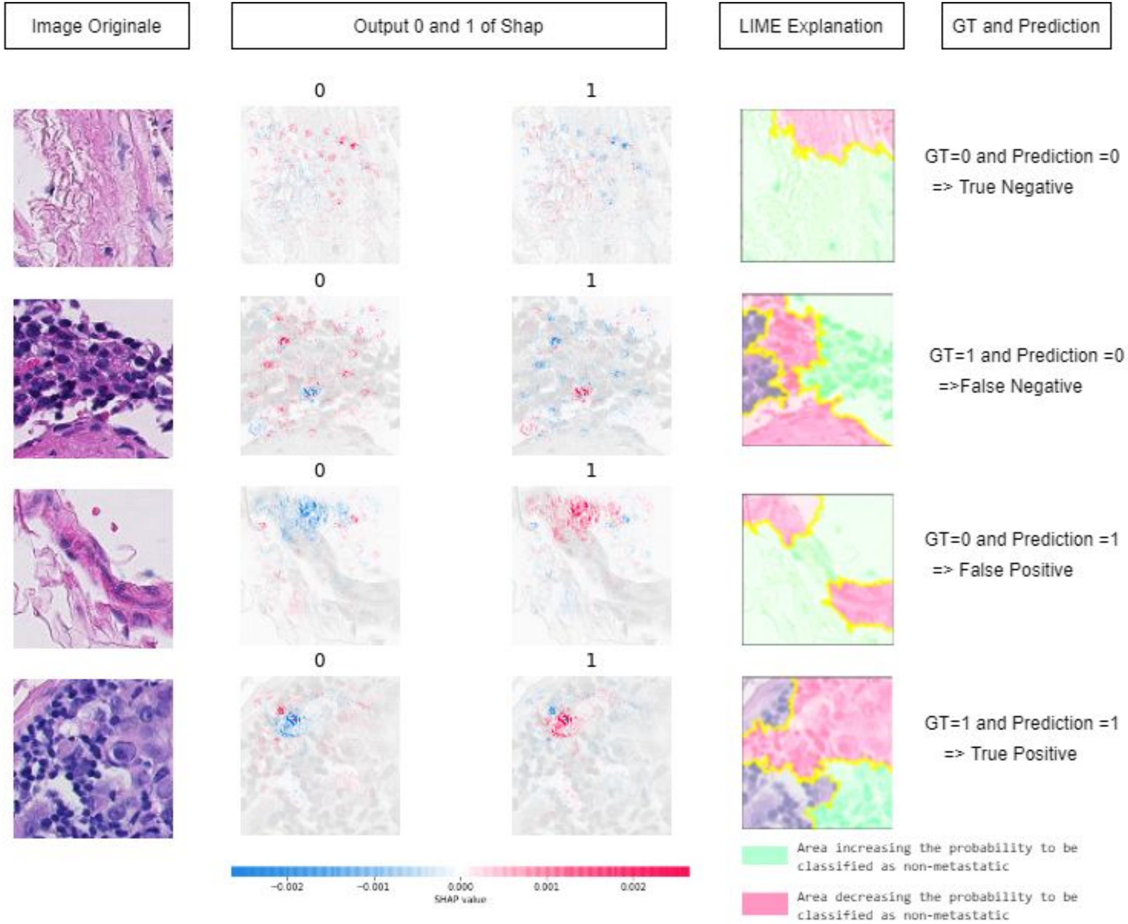


ROC curve CNN



ROC curve Gabor

Examples of generated explanations



Conclusion and perspectives

- Proposed methods perform classification of WSI with a decent accuracy
- Clear explanations on the location as well as the texture of the cancerous cells are provided

- Perform multiscale texture analysis to produce better explanations (wavelets)
- Data augmentation could lead to higher accuracy
- Test other segmentation algorithm for LIME
- Merge both methods (CNN architecture with texture images as inputs ?)

Thank you for listening, any questions?

Histopathological T-Rex

References

[1] « PatchCamelyon - Grand Challenge », *grand-challenge.org*. https://patchcamelyon.grand-challenge.org/Introduction/ (consulté le mai 10, 2021).

[2] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, et M. Welling, « Rotation Equivariant CNNs for Digital Pathology », *arXiv:1806.03962 [cs, stat]*, juin 2018, Consulté le: mai 12, 2021. [En ligne]. Disponible sur: http://arxiv.org/abs/1806.03962

[3] D. Wang, A. Khosla, R. Gargeya, H. Irshad, et A. H. Beck, « Deep Learning for Identifying Metastatic Breast Cancer », *arXiv:1606.05718 [cs, q-bio]*, juin 2016, Consulté le: mai 12, 2021. [En ligne]. Disponible sur: http://arxiv.org/abs/1606.05718

[4] A. K. Jaiswal, I. Panshin, D. Shulkin, N. Aneja, et S. Abramov, « Semi-Supervised Learning for Cancer Detection of Lymph Node Metastases », *arXiv:1906.09587 [cs]*, juin 2019, Consulté le: mai 12, 2021. [En ligne]. Disponible sur: http://arxiv.org/abs/1906.09587

[5] Y. Liu *et al.*, « Detecting Cancer Metastases on Gigapixel Pathology Images », *arXiv:1703.02442 [cs]*, mars 2017, Consulté le: mai 12, 2021. [En ligne]. Disponible sur: http://arxiv.org/abs/1703.02442

[6] I. Palatnik de Sousa, M. Maria Bernardes Rebuzzi Vellasco, et E. Costa da Silva, « Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases », *Sensors (Basel)*, vol. 19, nº 13, juill. 2019, doi: 10.3390/s19132969.