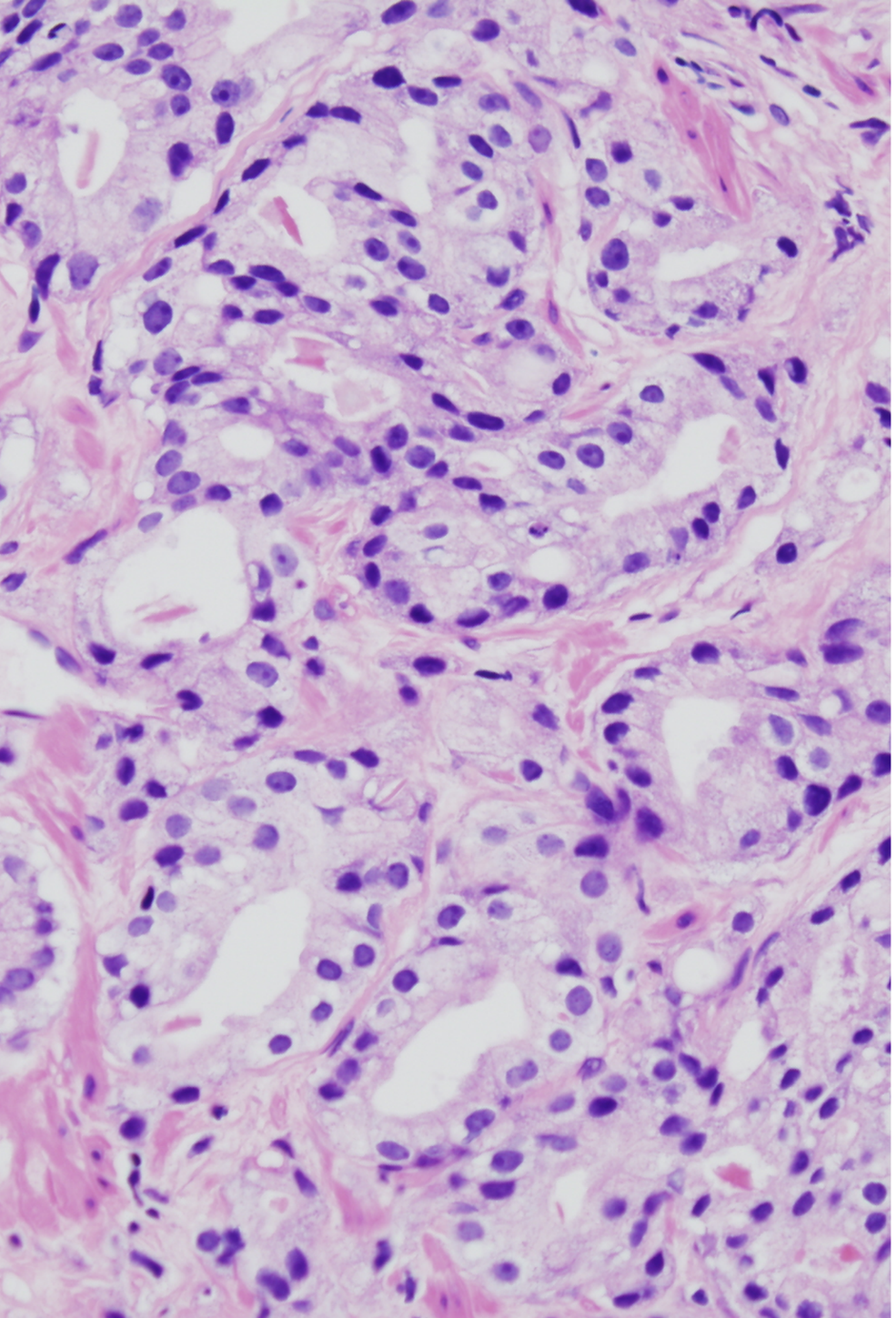


JUNIA Grande école d'ingénieurs

M1 Project - Group 18

Classification of microscopic image of cancerous tissue using Artificial Intelligence

Marine GEORGES - Harith JADID - Carlos JIMÉNEZ GARCÍA - Michel SAUVAGE - Adrien Junior TCHUEM TCHUENTE

April 21st, 2023

Introduction

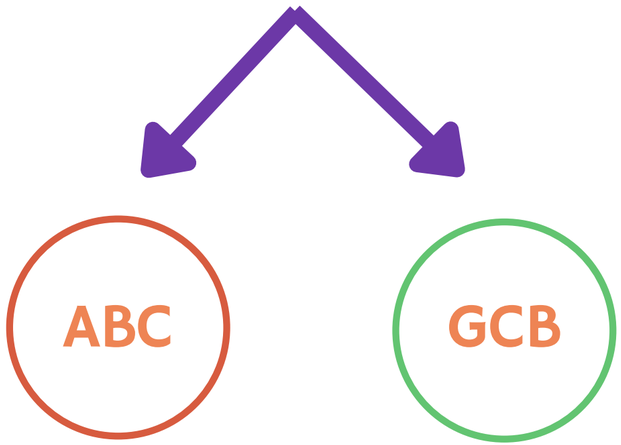
Non-Hodgkin's lymphoma (NHL)

DLBCL subtype

One of the most common

DLBCL subtype

VERSIONS



Steps

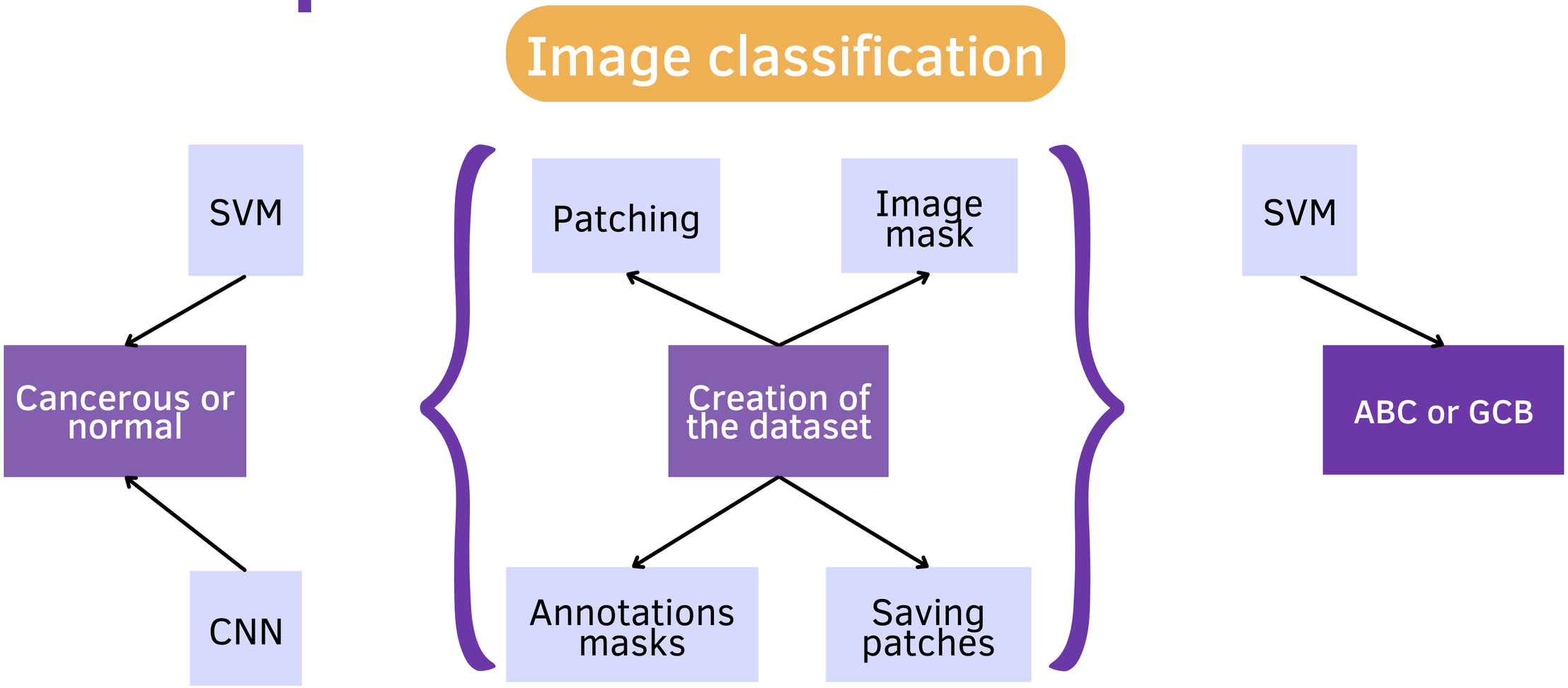


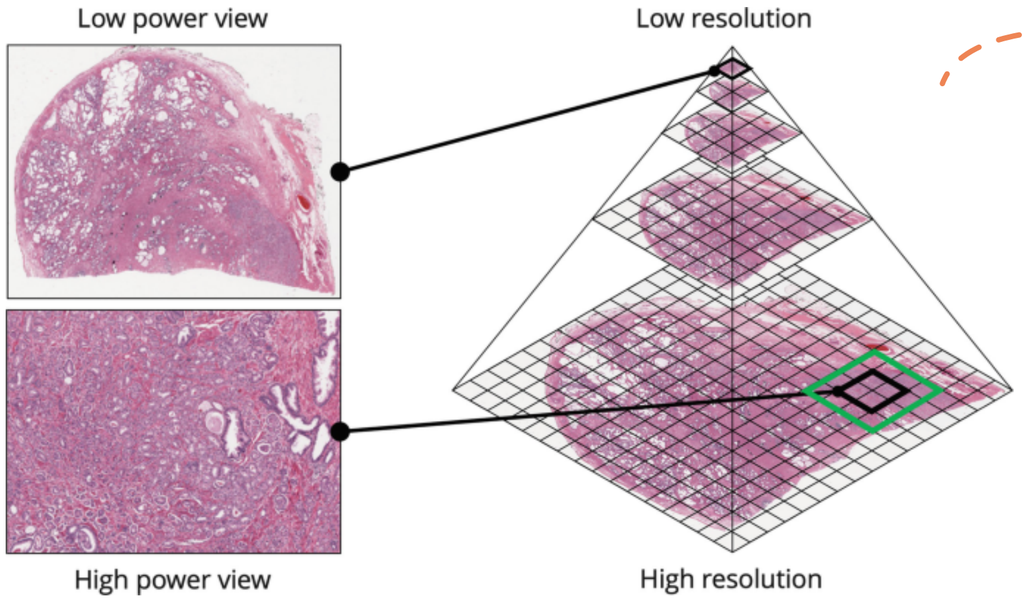
Image classification
SVM
Cancerous or normal
CNN
Patching
Image mask
Creation of the dataset
Annotations masks
Saving patches
SVM
ABC or GCB

Content of the presentation

- Creation of the dataset
- Option 1 : CNN implementation
- Option 2 : SVM implementation
- Final architecture
- Improvements & project feedback
- Conclusion

Creation of the dataset

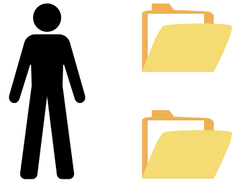
487 Whole Slide Images (WSIs) of patients' biopsies with annotations



from

to

A dataset patches in .png format organized by patient



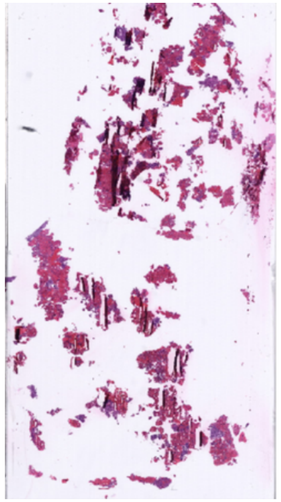
Manipulate the WSIs and their resolution levels with :

OpenSlide + python™

https://doi-org.ezproxy.univ-catholille.fr/10.1007/978-3-030-83332-9_2

Creation of the dataset

Method used :



Original WSI
(low resolution)

Mask tissue

Tissue in white

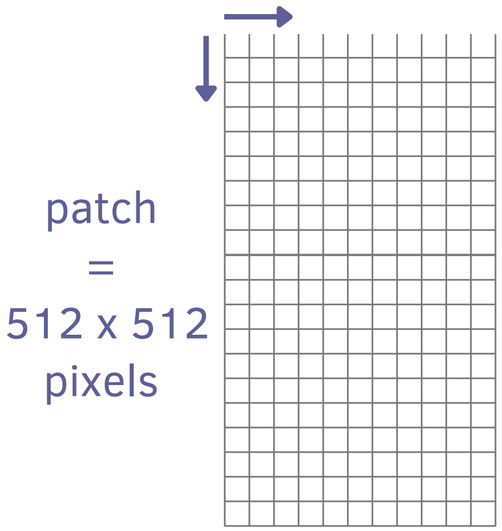
Mask annotations

Annotated area in white

- Colorimetric treshold on WSI at low-level resolution
- Library OpenCV

Creation of the dataset

Method used :



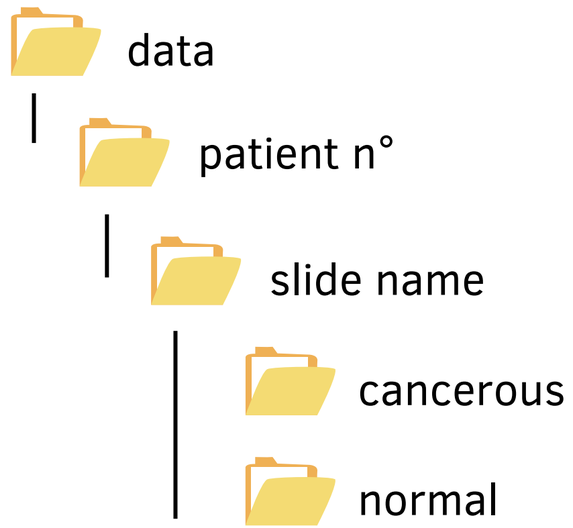
Move along coordinates patch by patch across the WSI

Mask tissue

Patch = tissue area ?

Mask annotations

Patch = cancerous or normal area ?

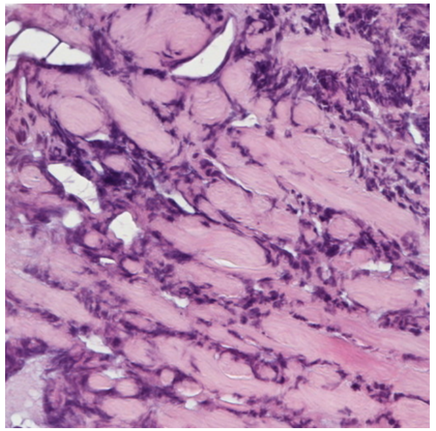


Saving patches in folders : cancerous / normal

CNN implementation

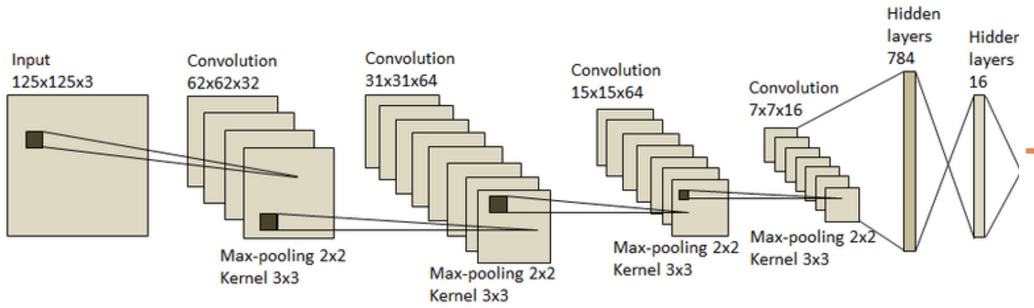
with TensorFlow

Input



Patch of 512 x 512 pixels

Convolutional Neural Network (CNN)



ResNet

Output

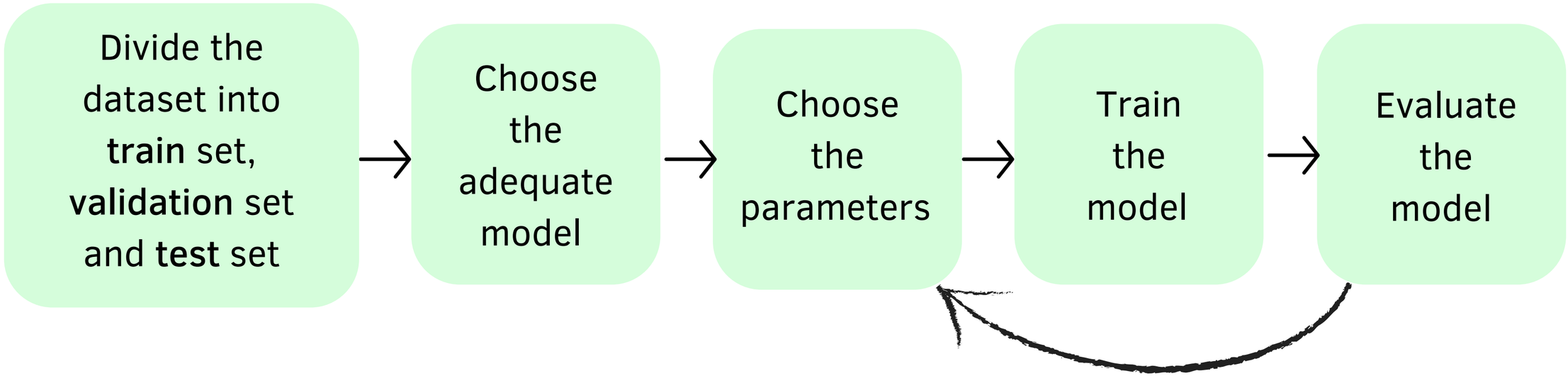
Class prediction

Predict if the patch is **cancerous** or **normal**

CNN implementation

with TensorFlow

Steps



CNN implementation

with TensorFlow

Results

1st try: ResNet50

(3 days of training)

74% of accuracy
Bad precision

2nd try: ResNet14

(3 days of training)

73% of accuracy
Bad precision

3rd try: ResNet14

with balanced classes

(2 days of training)

87% of accuracy
71% of precision

SVM implementation

Why Machine Learning ?

- save time on training
- more efficient on small data
- performant for image classification

What goes in entry ?

- Features extraction thanks to a Local Binary Pattern (LBP)

2 possibilities

SGD

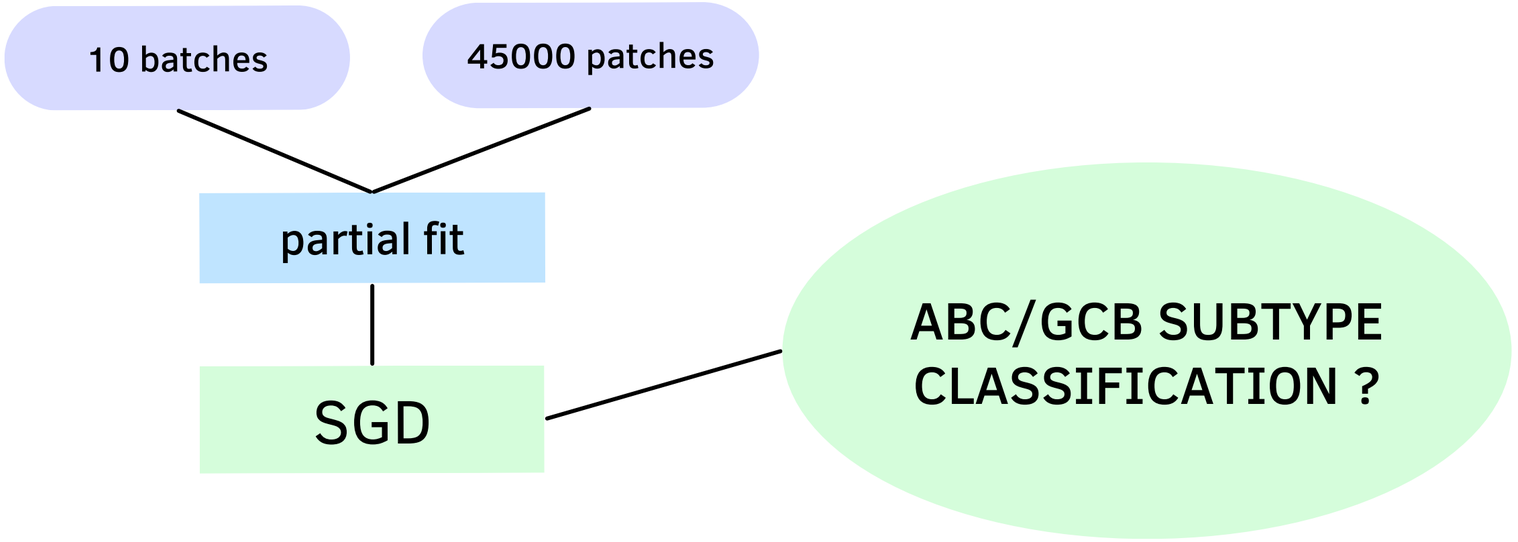
Stochastic Gradient Descent: optimization method used to minimize the loss function in machine learning models.

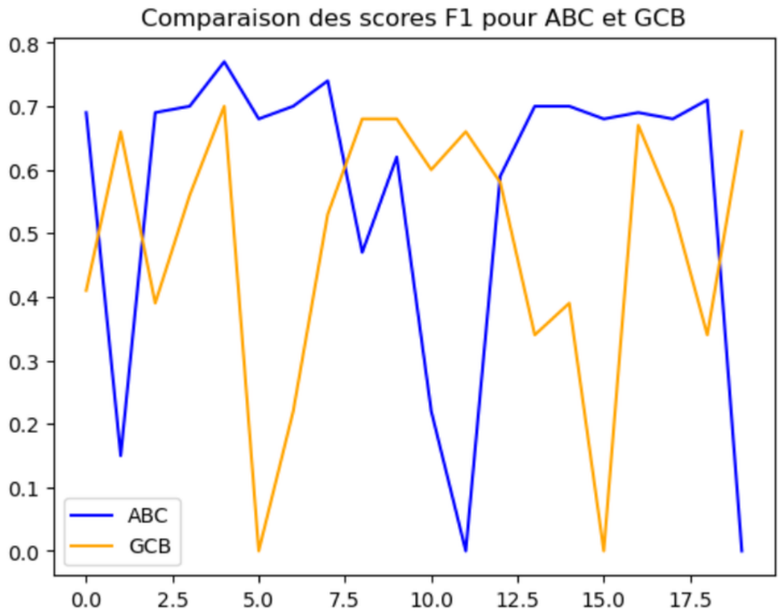
SVC

Support Vector Classification: supervised learning algorithm that aims to find the optimal hyperplane to separate the data into different classes.

SVM implementation

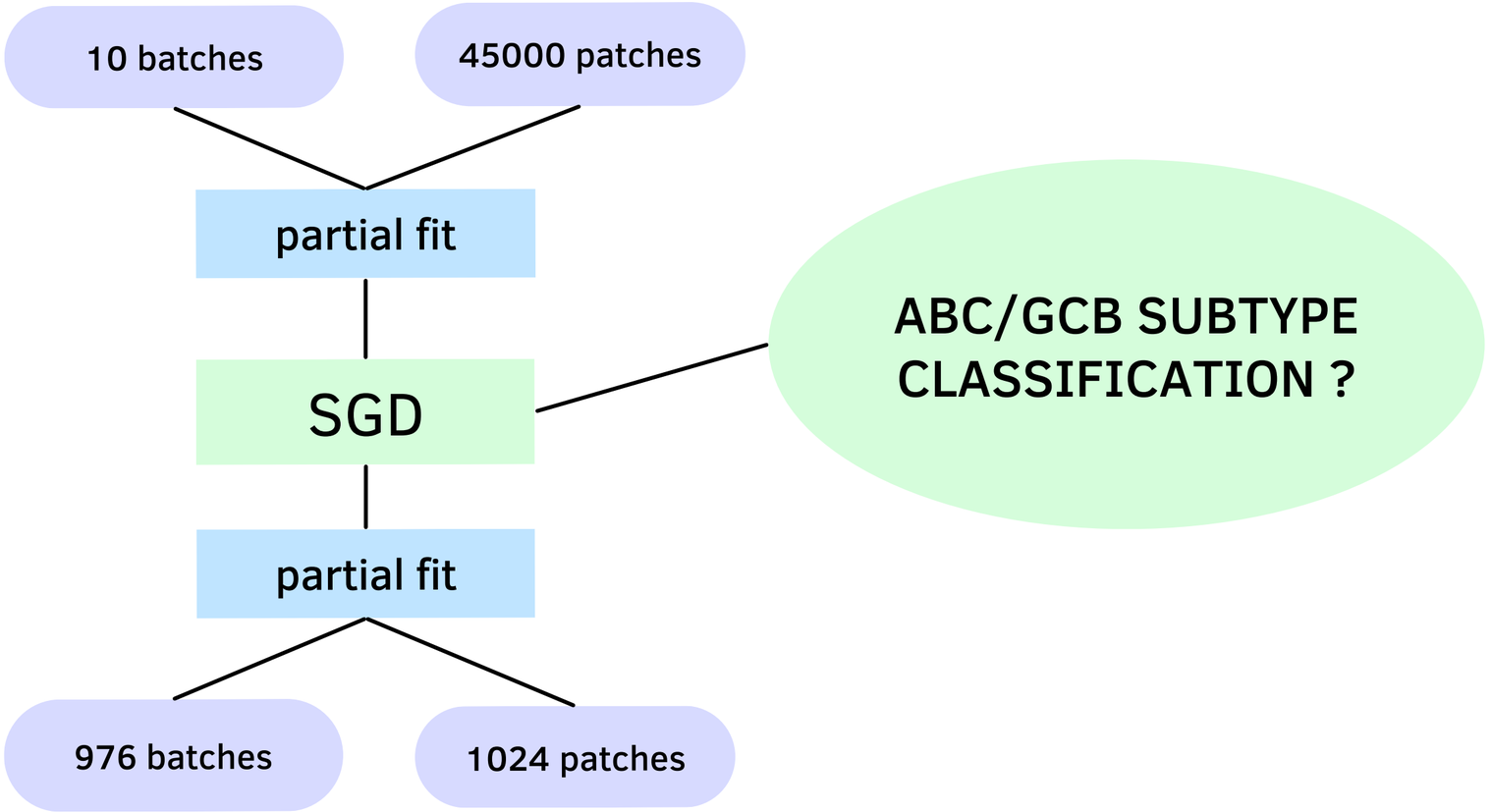
with scikit learn





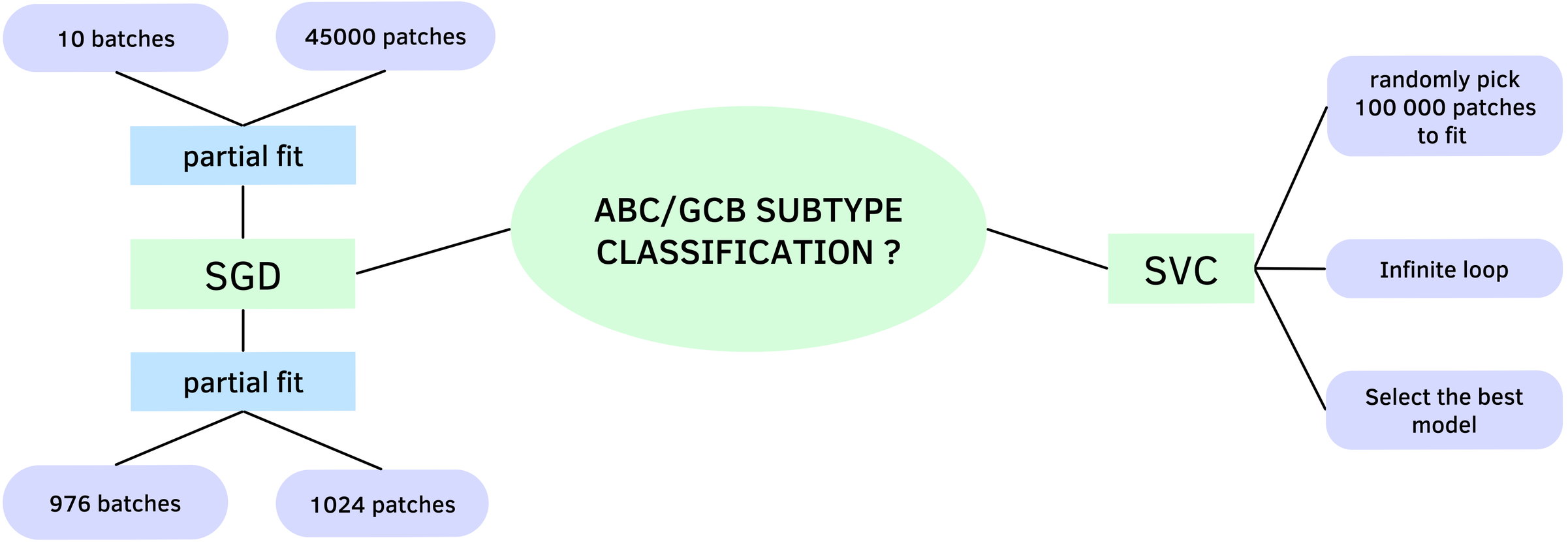
SVM implementation

with scikit learn



SVM implementation

with scikit learn



SVM implementation

with scikit learn

ML	ABC	GCB
SGD 45000 patches	f1-score : 0.73	f1_score : 0.46
SGD 1024 patches	f1-score : 0.58	f1-score : 0.08
SVC	?	?

Final architecture

Best results at the end of the project

TISSUE CLASSIFICATION		BEST MODEL
CNN results: 86.7% accuracy	SGD results: 83% accuracy	→ CNN

SUBTYPE CLASSIFICATION		
SGD results: 70% accuracy	SVC results: still waiting for results	→ SGD

Final architecture

A first "prototype" pipeline using SGD

SELECT PATIENT

WSIs into patches

TISSUE PREDICTION

only cancerous-predicted patches go to next step

SUBTYPE PREDICTION

one label predicted for each patch

FINAL SUBTYPE PREDICTION

Majority voting

ABC ?

GCB ?

(with the condition of a gap of 20% minimum)

non-significant ?

DATASET

training & validation sets

TRAINED SGD MODEL

Cancerous - Normal tissue classification

TRAINED SGD MODEL

ABC - GCB subtype classification

Improvements & tests

Get the accuracy of the model on each patient individually

- Better understand the shortcomings of the model
- Adapt the dataset accordingly

Test new architectures for the model

- Less training time
- Better accuracy

Use other libraries such as PyTorch

- Use the original algorithms for the architectures
- More room for personalization of the neural network

Project feedback

Proficiency in research

Acquisition of knowledge in ML/DL and image processing

Planning, meetings and distribution of tasks

Communication skills

Conclusion

CNN	SGD
Cancerous-normal	ABC-GCB
87%	70%

- Not optimal results and improvements to add but efficient overall system
- Development of a solid and strong project base
- Encouraging for the future of Non-Hodgkin's Lymphoma diagnosis

Thank you for your attention

Questions ?