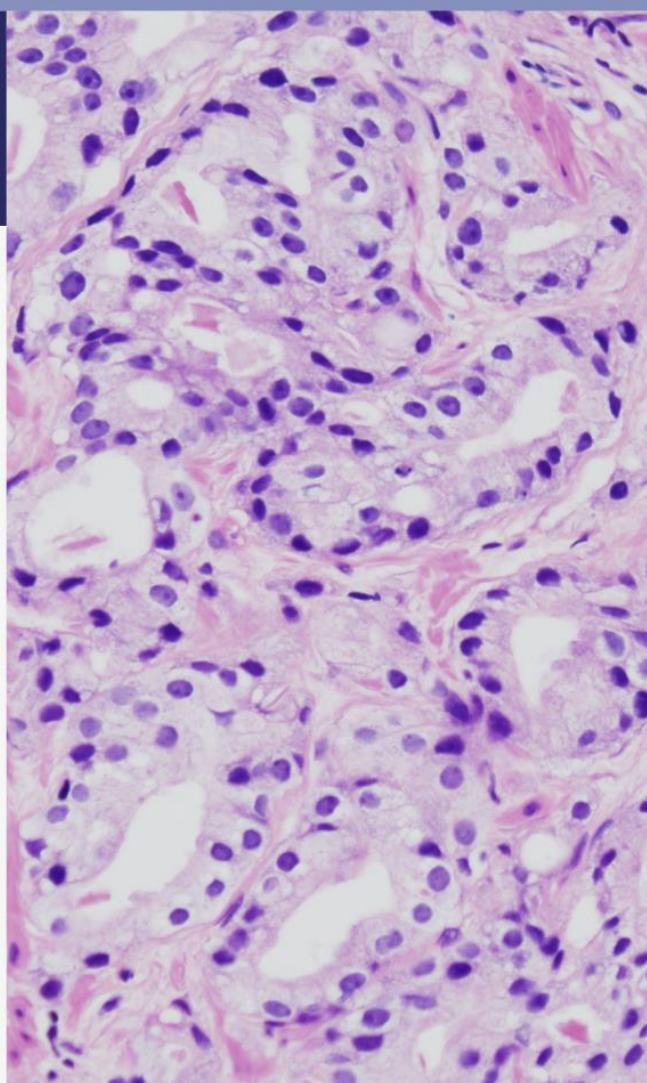
JUNIA ISEN

First Semester Project Report :

Classification of microscopic images of cancerous tissue using Artificial Intelligence

GROUP 18 - MASTER 1



STUDENTS :

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

PROJECT REFERENTS :

Bilel GUETARNI
Feryal WINDAL

2022-2023

Table of content

List of abbreviations

ABC: Activated B-Cell

AI: Artificial Intelligence

ANN: Artificial Neural Network

AUC: Area Under the Curve

BCa: Breast Cancer

CNN: Convolutional Neural Network

DL: Deep Learning

DLBCL: Diffuse Large B-Cell Lymphoma

ER+: Strogen Receptor Positive

FL: Follicular Lymphoma

GCB: Germinal Center B-cell

GLCM: Grey Level Co-occurrence Matrix

HES (or H&E): Hematoxylin and Eosin Staining

HOG: Histogram of Oriented Gradients

HSL: Hue, Saturation, Lightness

HSV: Hue, Saturation, Value

IHC: ImmunoHistoChemistry

LBP: Local Binary Pattern

NHL: Non-Hodgkin Lymphoma

RF: Random Forest

nRGB: normalized RGB

RGB: Red, Green, Blue

RNA-Seq: RNA sequencing

ROI: Region Of Interest

SEF: Supervised Encoder FusionNet

SEU: Supervised Encoder UNet

SVM: Support Vector Machine

WSI: Whole Slide Image

Introduction

Spreading across the body through the blood, the Non-Hodgkin Lymphoma (NHL) is an aggressive cancer impacting the lymphatic and immune systems. The Diffuse Large B-Cell Lymphoma (DLBCL), which represents 30% to 40% of NHL cases worldwide, is the most common subtype and can be lethal if not treated (Li *et al.*, 2018). Two main versions can be found for this subtype: ABC (Activated B-Cell) and GCB (Germinal Center B-cell), which have respectively a higher and a lower mortality rate. Performing an early diagnosis is essential as it can impact the patient's treatment and survival. Nowadays, diagnosis is done thanks to a biopsy and a genetic study, which can take a long time (two to four weeks). After the biopsy, hematoxylin and eosin staining (HES), and immunohistochemistry (IHC) are performed. Results are analyzed on Whole Slide Images (WSI), and diagnosis are made with high accuracy thanks to IHC which detects specific proteins. However, IHC is expensive and takes time. Therefore, the goal for this project is to be able to propose an alternative using artificial intelligence (AI) on WSI, which would meet the same accuracy as the IHC method. AI is an information processing tool that is particularly used for image classification. This project will therefore focus on the implementation of an AI that would be able to differentiate the two subtypes of DLBCL cancer. The language that will be used is Python, and its associated libraries and tools. The associated problematic is the following: How can AI technology be implemented in order to accurately classify the two subtypes of DLBCL cancer using WSI? We chose this project because we found that it corresponded well with our fields of study. Moreover, the research aspect associated with this project is also motivating us as it allows us to take a first step in this area of interest. In addition, the fact that the project is oriented around health adds a more concrete goal and relevance with AI technology than any other project could let appear. It is for all these reasons that we chose this project. In the first part of this report, a state of art on the matter of AI and image classification will be presented, in order to get a good knowledge of the actual studies and realization that can be inspiring for the project. Based on this research, in the second part, an approach to respond to this problem is then proposed. Finally, the last part describes the project management of the team, that will depict the overall organization of the future realization.

I. State of the art

1. Image processing

The first thing to think about before being able to use classification is the kind of data that will be used, and how it is possible to process it. For this project, WSI will be used as inputs of the model.

1.1. WSI patching

WSI are characterized by their size and weight, which can represent tens, or even hundreds of thousands of pixels and many gigabits. To tackle the issue, most approaches in the actual state of art consist of splitting the WSI into multiple patches beforehand.

An application of WSI splitting can be found in the article from Wei *et al.* "Pathologist-Level Classification of Histologic Patterns on Resected Lung Adenocarcinoma Slides with Deep Neural Networks" (2019). The aim of the study is to determine tumor grade of lung cancer by classifying histological patterns. For this purpose, 422 WSI are separated into different sets: 245 for the training set, 34 for the validation set and 143 for the testing set. The first step of their method is to divide the WSI into patches thanks to a sliding window approach. Window sliding consists of moving a rectangular-shaped frame across the image from side to side to split the image into patches. The code used for this

study is available on GitHub, and the patching has been done using Openslide. However, the dataset is not publicly available. Moreover, a limit that can be encountered using this program concerns the size of the WSI. When looking further into the code explanations, we can understand that the resolution used is of a small size of 3-5MB.

Another interesting approach of WSI patching can be found in the article from Miyoshi *et al*, "Deep Learning Shows the Capability of High-Level Computer-Aided Diagnosis in Malignant Lymphoma" (2020) which classifying technique is detailed above. Because of their large size, the WSI were cropped according to the annotations, and at different magnitudes (x5, x20 and x40), whereas other studies usually select only one magnification. To begin with, a first crop of 2048 x 2048 pixels is done in the annotated zone of the WSI. Then, cropping is performed again in the previously cropped image to obtain smaller patches. The patches are then randomly divided into sets in order to improve the efficiency of the prediction model with K-fold cross validation. For each magnitude, five tests are performed, using each one testing set of 100 patches and four training sets. This practice is advantageous as it uses the entire set as training, but it can also be a weakness if the final classifier is tested with a new set because it can become too specific to the known data. Using patches at different magnitudes allowed them to evaluate an average efficiency result with the ensemble model technique. It consists of calculating the average efficiency considering all the tests realized. Finally, each classifier diagnosis was compared to seven pathologists' diagnosis. As a result, the ensemble classifier shows its best accuracy at 97%, whereas the highest one for pathologists was 83.3%. One of the factors that can explain this performance can be the pixel-level scaling and the different magnitudes of patches, which allows looking at details that a human eye cannot see. Furthermore, this study is encouraging as the classifier does not require immunochemistry slides, which can be expensive and time-consuming to obtain. However, some limits can be highlighted. First of all, the classifier requires annotations on the WSI, which cannot be done without a professional. Then, this study only covers 3 of the 100 known subtypes, and therefore needs to be perfected. As a matter of fact, this study also considers only general subtypes. This method might not work as well for subtypes of DLBLC, which could be harder to differentiate.

1.2. Feature extraction

Feature extraction consists in converting the input data into a set of analyzable features. The main goal of feature extraction is to obtain the most relevant information from the original data and represent the information in a lower-dimensional space.

Management of patches

With the patching of WSI, one of the major areas to consider is the way to select the patches to be processed. Indeed, the selected patches largely determine the accuracy of the learning algorithm. A first common step in various studies is to be able to extract patches that contain important data. The method used in the article from Hsu *et al.* consists of discarding some patches according to the pixel intensities in order to improve the program's efficiency. In this article, WSI are used to detect glioma subtypes using Deep Learning. Therefore, areas which contain only a very low density of cells or no cells in the WSI must be removed. The selection is done by evaluating the mean of all pixel intensities in each patch. The mean must be between 50 and 150. Moreover, the difference between the extreme values of intensity should be smaller than 100. In the article from Fan *et al.*, another approach is used. For each WSI used, multiple patches are made and the blank patches where no cells are present are removed using the Otsu algorithm. Other than discarding unwanted patches, it can also be necessary to remove artifacts and staining invariability in the patches. Examples of such artifacts are shown in appendix 1. To facilitate the ignorance of artifacts in DL models, one approach consists of removing artifacts using images filters and the normalization (Magee *et al*, 2009). Finally, patches can also be categorized considering previous annotations on the WIS. This can allow to only select the patches of

the category of interest. In the article from Cruz *et al*, WSI from 162 patients are divided into 100x100 pixels using a grid sampling. Then, a comparison is made between the annotated regions of WSI and the patches. Of all the patches, the ones which contain at least 80% of the annotated cancerous region are considered as positives samples, while the others are considered as negatives samples.

Color

In the article "Color Spaces Advantages and Disadvantages in Image Color Clustering Segmentation" (2018), the effect of the color is studied from the point of view of the most popular color spaces available in image processing. Is also tested the effect of the selection of a given color space on one of the most common clustering Machine Learning algorithms (k-means).

RGB is the most widely used model in computer graphics for its ease of use and intuitiveness. However, it has some disadvantages when it is used in image processing as the model produces a discontinuous space, which makes the changes in color hue hard to follow. In addition, the color hue is easily affected by illumination changes which makes color monitoring and analysis a non-trivial task.

Normalized RGB (nRGB) is a variation from the RGB, and basically is created under the premise of protecting the color model from the illumination changes by using a proportion of three primary colors from the model, not a defined amount of each one.

HSL (Hue, Saturation, Lightness) can be computed from the RGB standard model by using a set of equations. The major advantage of using this model is the fact that it presents immunity to illumination changes, since the illumination is enclosed in the lightness component of the model. The other feature of this model can be found in the color hue changes; they are continuous and linear. However, this model also has some issues related to some incorrect color interpretations and undefined saturation. HSV (Hue, Saturation, Value) model is similar to HSL, and shares the same definition for the hue component, but differs in the way it interprets the color saturation. Sometimes, HSV is preferred due to the geometric representation, which is usually more natural than HSL and allows a better color hue manipulation.

Lab (also known as L * a * b*) is formed by a Lightness component and two chromatic or color components (a and b). This color model can represent colors that are not handled by other models. Unfortunately, its creation is a complex process, and the color space produced is not natural to humans.

To demonstrate the benefits and issues coming from selecting a given color space a series of tests were carried out with a set of three images with different characteristics (noise, shades, color definition...) and the following conclusions were reached. The results obtained from the test carried out with an image with some noise showed that normalized RGB was the best option. In addition, the following conclusions were drawn from the test performed with a natural image (appendix 2): RGB had errors in all the clusters. With HSV the background clusters presented some noise and HSL even more. Lab presented solid acceptable clusters for red and green shapes but for the blue color loses part of the lighter side shapes and Normalized RGB had a similar behavior to the Lab model. From all the performed scenarios, it is demonstrated that every color space presents its own issues that could affect the segmentation. But an important result observed from the behavior of each color space concerns the aspect of failure in different areas. It means that some color spaces were better for performing the segmentation of a given color, or a given illumination condition. So, an approach to follow is to use a given color space depending on the conditions of the scenario where it should be implemented.

Texture

The texture feature describes the surface properties of the object corresponding to the image or image area. It is an effective method when judging images with large differences in thickness and

density. However, it is difficult for the usual texture features to accurately reflect the differences between textures with different human visual perception. In the article "Texture Analysis of Breast Cancer via LBP, HOG, and GLCM techniques" (2020) the aim is to implement, compare and conclude which combinations between feature extraction methods (LVP. HOG, GLCM) and classification techniques (Logistic Regression, SVM, kNN, ANN) are the best in terms of accuracy.

The first technique used for describing the texture of an image is the Local Binary Pattern (LPB). The LBP operator describes the texture of an image by examining the signs of the differences between the central pixel and its neighbors. For each pixel in the image, a binary code is obtained by comparing (thresholding) the value of the center pixel with the values of its surroundings pixels. If the value of a neighbor pixel is greater than or equal to the threshold value, it becomes 1, and if it is less than the threshold value, it becomes 0 (appendix 3).

The second technique is called the GLCM (Grey Level Co-occurrence Matrix), which is a matrix that characterizes the texture of an image by calculating how often pairs of pixels with specific values and specific orientation occur in an image. By default, the spatial relationship is defined as the pixel of interest and pixel to its right, but there can be specified other spatial relationships between the two pixels (appendix 4) .

The last technique analyzed is HOG (Histogram of Oriented Gradients), based on the orientation of the gradient intensity in local areas of an image. The information obtained is projected into local histograms, calculated in small cell sizes, which are uniformly distributed over the entire image. These histograms provide information about the dominating contours in each of the image areas. The information obtained will allow us to distinguish the shape of the objects in an image and is a good basis for detecting and recognizing different objects.

Once the tests combining the different feature extraction methods with each classification technique were carried out comparing them with previous works realized the following conclusions were reached. As is shown in the table of appendix 5, the best results in terms of accuracy were achieved by combining LBP-ANN (98,57% accuracy). Even so, the combinations between GLCM-KNN (92% accuracy), LBP-SVM (91% accuracy), HOG-SVM (90% accuracy) and LBP-Logistic regression (92,5% accuracy) also obtained optimal results. The limitation we can find in this article is that the data type used (MIAS) is different from the one to be used in our project (WSI). However, the information obtained is very useful when deciding on the techniques and methods to be developed. In addition, if necessary, the data type used could be modified to MIAS to improve the results.

1.3. Segmentation

Image segmentation is an image processing operation consisting of detecting and grouping the pixels according to criteria, in particular intensity or spatial criteria, the image thus appearing to be formed of uniform regions.

Thresholding

Image thresholding as used in "Whale Optimization Algorithm and Moth-Flame Optimization for Multilevel Thresholding Image Segmentation" (2017) is an image binarization technique, consisting in replacing the gray levels of an image by a set of pixels taking the value 0 or 255 (black or white) depending on whether its initial value is lower or higher than the value defined as threshold. According to this article, there are a lot of algorithms to find the optimal threshold, and the main two approaches are parametric and nonparametric. For the first one, some statistical parameters are computed for the classes in the image. Then, for the second one, the optimal threshold is found by maximizing some criteria such as the Otsu's criterion and Kapur's entropy.

Deep Learning

A part of the article “Deep Learning for Digital Pathology Image Analysis: A Comprehensive Tutorial with Selected Use Cases”*(2016)* develops nuclei segmentation in the case of breast cancer. As it is said in this article, there is a link between the organization of cancer nuclei and subtypes generally. This method uses the dataset of 141 WSI of 2000x2000 pixels patches of estrogen receptor positive (ER+) breast cancer (BCa), containing a subset of 12,000 an annotated nuclei (appendix 6) to be able to train the dataset to find the nuclei segmentation. Areas which are nonnuclear don't correctly absorb the coloration. For this reason, the Neural Network can use this difference to find the nuclei (appendix 7). However, the Neural Network alone is not efficient enough in certain cases and needs to be improved. That is why, in the study, they implanted a modulation patch to increase the precision of this AI by emphasizing the shapes. Results from this article demonstrate that the efficiency is better when using patches at a x40 magnification to detect segmentation nuclei, as the accuracy is 98%.

2. Image classification

Image classification is a fundamental step in computer vision because it is what makes it possible to describe what an image contains. For this project, the focus is especially on applications in the medical field. In this section, some general techniques and their applications in various studies are stated. Classifying techniques can be divided into two parts: Traditional Machine Learning, and Deep Learning.

2.1. Traditional Machine Learning

Machine Learning can be defined as an artificial intelligence technology that allows machines to learn without having been previously programmed specifically for this purpose.

Support Vector Machine (SVM)

The article “Survey on SVM and their application in image classification” (2021) explains how an SVM algorithm performs classification. To do so, the algorithm constructs a hyper-plane in higher dimensions. This hyper-plane is referred to as the decision plane. A decision plane is what distinguishes a group of data of one type from another. SVM searches vector points, referred as support vector, which define the decision boundary and give the large marginal separation between the classes. The margin is the smallest perpendicular distance to the data from the hyper-plane. The maximum marginal hyper-plane is the decision plane where the margin is the largest. SVM selects the maximum margin separating hyper-plane. Maximum marginal hyper-plane selection done by SVM increases the ability of accurate classification and reduces the chances of misclassification. SVM can be used with the scikit-learn library.

This method has been used to solve a lot of classification problems; we can take for example the one presented in the article “Non-small-cell lung cancer classification via RNA-Seq and histology imaging probability fusion" (2021), where an SVM algorithm using RNA-Seq data using 6 genes managed to classify lung cancer subtypes achieving a mean F1-Score, AUC and accuracy of 94.35%, 0.985 and 94.32% respectively. A CNN with WSIs as input was also used for this problem and obtained the following results: 83.39%, 0.947 and 86.03%. We can clearly see that in this case the SVM achieved much better results, thus might be more fit for these types of problems. Although it is used on RNA-Seq data in the study, SVM can also be used with images. In the context of our classification problem, an SVM could be used to identify which subtype a given sample belongs to, based on its features. The SVM can effectively handle high dimensional data and work well with non-linearly separable data which can make it a valuable tool in our case. But a problem that the SVM algorithm might face when using WSIs for classification is that the large number of pixels in each WSI could result in a too large number of features. Therefore, it could make it difficult for the model to learn useful decision boundaries.

Random Forest (RF)

A random forest classifier is an ensemble Machine Learning algorithm that uses a collection of decision trees to make predictions. A decision tree is a diagram in which every branch is associated with a decision according to the features of the data. All the predictions from the trees are selected in a random manner to make the final prediction. In "Metastasis Detection from Whole Slide Images Using Local Features and Random Forests" (2017), a RF classifier is used to detect the presence of metastasis in H&E stained WSI of breast cancer tissues. In the article, 270 WSI are used, some of them having annotations of ROI made by professionals. The RF was trained by using the features of the tissue samples. An ensemble model of 50 trees was implemented. A training and testing set were made in order to classify the images, and annotations were used to create the positive and negative training samples. As for the results, the study shows that the RF model is efficient to determine the metastasis in the WSI. The accuracy values presented for the AUC are 0.97–0.98 for tumor detection within whole image area, and 0.84–0.91 for tumor vs. normal tissue detection. A noticeable aspect that can be drawn from this study is its application possibility scope, as it could also be used with IHC images. However, as said in the article, the method could still need to be improved by adding Deep Learning to the model. Furthermore, WSI needs to be annotated for the training set to be made. Another limit mentioned concerns the similarity in color and shape between cancer cells and some other cells in the tissue. Due to this, false positives can be created. Finally, the quality of the images can impact the results obtained. Another paper where RF is detailed is "Context-based interpolation of coarse deep learning prediction maps for the segmentation of fine structures in immunofluorescence images", in which the segmentation of tumors is studied using immunofluorescent WSI. The results show that RF is accurate and fast, and that it can learn patterns. However, it used immunofluorescence and still need to be tested on more data.

Naïve Bayes classifier

Another method viewed in "Automatic Classification of Leukocytes Using Morphological Features and Naïve Bayes Classifier" (2016) that can be used for image classification is the Naïve Bayes classifier. This supervised learning algorithm is based on probabilities. When used for classification, this algorithm can make predictions rapidly and efficiently. As its name suggests, its functioning is related to the Bayes' Theorem, which gives a formula for the probability of an event A, with previous knowledge B. The term *naïve* means that all chosen classifying features are independent from each other. In this article from Gautam *et al.*, a concrete example is developed. The aim of their study is to classify leukocytes images automatically using a naïve bayes classifier. The dataset is made of 20 training images of microscopic blood smear and 68 testing images. Features like area, eccentricity, perimeter and circularity of leukocyte nucleus are extracted in a first step and then used to train the system. The predictions of the classifier are estimated thanks to the mean and variance of each feature. Therefore, according to the likelihood of each parameter with the training dataset, the algorithm can determine what kind of leukocyte is on the image. According to their study, an accuracy of approximately 80% was found using this classifying technique. An advantageous aspect of this method is that it only requires a small amount of data for the training set. Concerning the limits, as this method is based on probabilities, it might not be accurate enough for the project. Moreover, although the study seems to have a good performance with its very small dataset, a doubt remains concerning its use with a larger one. Finally, the images used in the datasets contain only one highlighted nucleus per image, and this algorithm might not work as well if many stained cells are present.

2.2. Deep learning

Deep Learning (DL) is a Machine Learning technique based on the model of neural networks: tens or even hundreds of layers of neurons are stacked to bring greater complexity to the establishment of rules.

Convolutional Neural Network

Convolutional Neural Network (CNN) is a type of Deep Learning artificial neural network that is typically used to analyze visual imagery. A CNN can have multiple layers (figure 1). Each of them learns to detect the different features of the input image. A filter or kernel is applied to each image to produce an output that gets progressively better and more detailed after each layer. At each successive layer, the filters increase in complexity to check and identify features that uniquely represent the input object. In our case, the input objects of the CNN are image patches from the WSI.

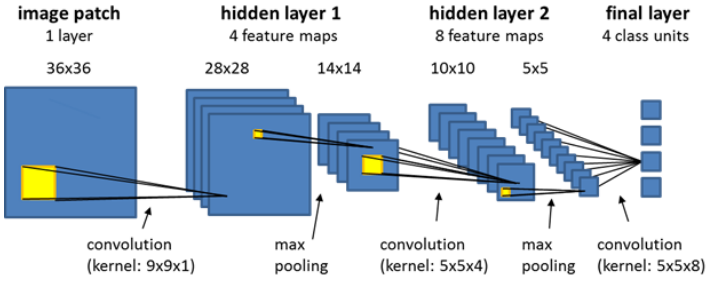


Figure 1 – Convolutional Neural Network schematic representation

In “Deep Learning Shows the Capability of High-Level Computer-Aided Diagnosis in Malignant Lymphoma” (2020), the problematic is to acknowledge the efficiency of AI technology in the diagnosis of three different subtypes: diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and reactive lymphoid hyperplasia (RH). Obtained thanks to biopsies and H&E staining, it is a total of 388 WSI that were considered and annotated by pathologists. As a comparation reference, each of these were previously diagnosed by immunohistochemistry for one of the three chosen subtypes. This method applies different magnifications on WSI and processes them with a CNN of 11 layers in order to classify the subtypes. The results demonstrate that the neural classifier is more effective than pathologists to differentiate the three subtypes, and that it can be used to support the diagnosis of malignant lymphoma. The best accuracy of this AI is 97% which shows that this method can be relevant. However, the problem with this method is that the HES slides need to be annotated which takes time and can be counterproductive. Moreover, there is no information concerning the processing time.

There are many existing architectures of CNN that can be used in order to classify WSI patches such as AlexNet, ResNet, and FusionNet. In the article “A Deep Learning Approach for Breast Invasive Ductal Carcinoma Detection and Lymphoma Multi-Classification in Histological Images” (2019), their approach in order to make a lymphoma multi-classification in H&E histological images (of 1388 × 1040 pixels splitted into patches of 128 × 128 pixels) is based on FusionNet. They propose two different scenarios: Classification by Reconstruction and Supervised Classification. The first scenario trains the FusionNet under a sparsity constraint in an unsupervised manner. Its accuracy is evaluated to 77,60%. For the second scenario, only the encoding part of the FusionNet is trained in a supervised manner. This scenario is referred as Supervised Encoded FusionNet (SEF). Its accuracy is evaluated to 97,67%. By comparing, the two scenarios, they showed that the SEF clearly outperformed approaches that used UNet or ResNet. The SEF method achieved an accuracy of 97.67% which is 20.07%, 32.57%, 4.87%, and 2.2% higher when compared with those of FusionNet, U-Net, SEU (Supervised Encoder UNet) and ResNet34, respectively (appendix 8).

The article from Janowczyk *et al.,* "Deep Learning for Digital Pathology Image Analysis: A Comprehensive Tutorial with Selected Use Cases"*(2016),* develops a study of breast cancer classification using the CIFAR-10 version of the CNN AlexNet. The strategy is developed step by step in the article. Moreover, a lymphocyte detection case and a lymphoma subtype classification case are especially detailed. For the detection on lymphocytes on the images, they obtain a F1 score of around 90% was obtained. For the Lymphoma Subtype Classification, the mean accuracy is 96.58% ±0.01% (on average 2.6 misclassified images in 75 tests). The strategy is developed step by step in the article (appendix 9).

II. Proposed strategy and feasibility of the project

1. Envisaged method

Our method for the realization of this project can be separated into multiple steps. First of all, it is necessary to preprocess the WSI, as they are too large to be processed raw. They have to be split into patches (by using the Openslide library) and sorted according to whether they can be considered cancerous or not with a CNN. From there, a database can be created. If this method doesn't work, we will take WSI in little resolution in order to use a segmentation between cell tissue and blanks. It would be composed only of cellular tissues, which will be divided into 3 parts (training set, validation set and test set). To deal with the hue coloration difference due to the origins of biopsies, we will apply a texture feature to normalized the shades of colors. We will use LPB if we choose a NN. However, if we choose to do a SVM, we will use HOG.

Then, our objective is to determine the cancerous areas in the cellular tissue. To do so, it is necessary to create the new dataset containing two disparate categories : cancerous and healthy tissues. These categories would be determined thanks to the annotations of the pathologist, using a deep learning algorithm AlexNet. If this architecture does not work, we will use Supervised Encoded FusionNet (SEF). If this solution still does not work, then, we will use a SVM to determine the cancerous zones on patches.

Finally, we create a dataset with ABC and GCB patches that were found, again thanks to the pathologist. For this part, we will use the to determine if an unannotated patch is of type ABC or GCB by using SEF which had 97,67% of accuracy in a study about Lymphoma Multi-Classification. If we have issue with this architecture, we can use ResNET-18. If the deep learning doesn't make the work, we can try to use a SVM. The program is expected to be able to tell which subtype detected is predominant, which should correspond to the type of lymphoma of each the patient.

2. Overall limitations and feasibility

Once the main strategy has been implemented, we must be cautious in analyzing its limitations and have more options available in case it does not work, so that we can continue developing the project. An exhaustive analysis of the strategy to be developed has been carried out and secondary plans have been established to deal with these possible issues.

If the Deep Learning method does not work for the classification process, the solution would be to analyze the arrangement of the cells with KNN (appendix 10). This new technique could tell us which type of cancer the biopsy corresponds to. We could also find that the accuracy of the classification is not the one we are looking for. To solve this issue, a good alternative would be to implement the IHC.

Another problem that we can find related to feature extraction is that the program can take too long to execute the instructions and, consequently, it can slow down the development of the project too

much. As already mentioned, there are numerous alternatives for feature extraction, so we could look for one that best fits the requirements and works in a more optimal way.

It should be also noted that the coloration of the cellular tissue may be different depending on the part of the body from which it comes. Therefore, we must be very careful with the normalization and make sure that we can work safely without making mistakes.

It is possible that once the tests have been developed, we do not achieve the desired accuracy with the LBP-SVM method-technique combination. If that happens, we can also try the HOG extraction technique as it has been demonstrated to have good results working with SVM. Even if it was decided in a previous step to modify the SVM selection method to another one, we could choose the feature extraction technique that best fits it, as is shown in the article "Texture Analysis of Breast Cancer via LBP, HOG, and GLCM techniques" (2020). The experiments on which we based our work use the MIAS data type that is different from the one to be used in our project (WSI), so we can also consider this a limitation. However, the information obtained is very useful when deciding on the techniques and methods to be developed. In addition, if necessary, the data type used could be modified to MIAS to improve the results.

If during the image processing, we find some issues coming from the selected color space (HSL) we can also try to use other color spaces such as RGB, nRGB, HSV or Lab. As previously mentioned, each of these spaces has its advantages and disadvantages depending on the characteristics of the processed image (brightness, saturation...). So, considering the issues and the requirements found, we can choose from among the different options the most optimal one.

To conclude, it must be said that there is a high probability that during the development of the project we will realize new limitations as well as new solutions and alternatives to them. This is a continuous process that will be present throughout the entire project, so we must be always cautious and resolutive.

III. Project management

1. Objectives

As explained in the introduction, our goal for this project is to be able to achieve the classification of two subtypes of ABC and GCB lymphoma. To do this, we want to use an AI classified cancer thanks to WSI that undergo HES treatment. However, since this is a research project and the time allotted to us is quite short, what is expected here is mainly to be able to propose solutions and ideas, more than having a perfectly working model (although it would be great !).

2. Planification and division of tasks

Project planning is established using a Gantt chart that groups together the main tasks to be performed. These tasks were previously defined by a Work Breakdown Structure Diagram, available in the appendix 11. The designated team project leader is Michel Sauvage. The team will work with the XP method (eXtreme Programming), meaning that two people will work on the same computer to code. As not everyone has the same experience in the field of AI, the groups will be made in such a way that it is balanced, allowing each of the groups to work optimally. As we are five students, there will be a rotation in the groups at each stage. For each step, there will be a responsible person, who can guide the work by taking a step back on the progress. This responsible will ensure that the two teams work in a complementary way in order to optimize the work. Then, it is planned that at the beginning of each day there will be a daily scrum (Agile method), which is a 15-minute meeting, in a standing position, in

order to talk about what we did the day before and what remains to be done. To stimulate the team, the work environment will also be varied from time to time.

3. Risk analysis for the project realization

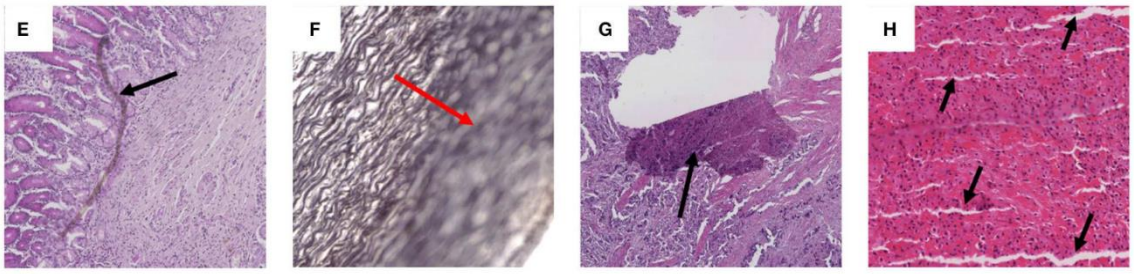
There are risks to consider regarding the progress of the project. First, regarding the achievement of objectives, several aspects must be considered. Indeed, we found few GitHub that could be relied upon, and this could slow down overall progress. In addition, absences of members, or even possible conflicts could also slow down the project. Then, another risk concerns the strategies and methods envisaged. It is possible that the working methods are not suitable or that technical unforeseen events jeopardize the overall strategy chosen. These risks must be anticipated as much as possible.

Conclusion

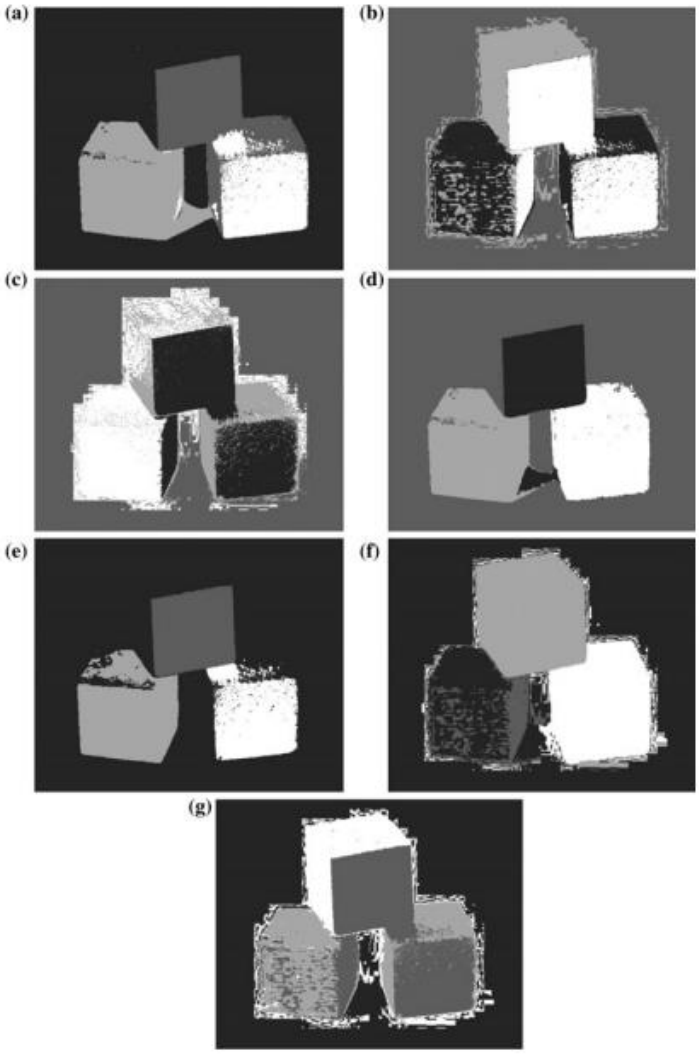
To conclude, the aim of this project is to be able to classify WSI according to DLBCL cancer subtypes in order to provide help in the diagnosis of cancers for affected patients. First of all, we carried out research on the state of the art related to this subject. This allowed us to have a global vision of the AI techniques used and their capabilities and limits in the context of applications related to the medical environment. Throughout these first weeks of research, we have learned a lot about the different applications and use of AI models for images, which was really instructive for us. Thanks to this research, we were able to establish a strategy for the realization of the project, which is as follows. First of all, we will process the WSI image by keeping only the cell tissue and cutting it into patches. After that, we will classify the cell tissue into two categories: cancerous tissue and non-cancerous. Finally, we will classify each patch of cancerous tissue according to their subtype. To determine the subtype of the biopsy, we will look at which subtype is predominant according to the classification made on the patches. The deliverable obtained through this strategy would be a program made in Python. Our objective is to be able to deliver a classification model approach in order to contribute to a research subject. Thus, the ideal would be to produce a functional program, which would come as close as possible to a satisfactory result. Regarding the management of the project, several points were discussed. To carry out this project in the second semester, we proposed a distribution of tasks, as well as working methods. In addition, a Gantt schedule was produced in order to be able to plan throughout the five weeks of work allocated. Some risks and apprehensions have been identified, particularly concerning the lack of experience in the environment and time management. Nevertheless, this project represents a stimulating challenge and motivates us to do our best for the conception part in the second semester.

Appendix

1 – Examples of artifacts on WSI patches

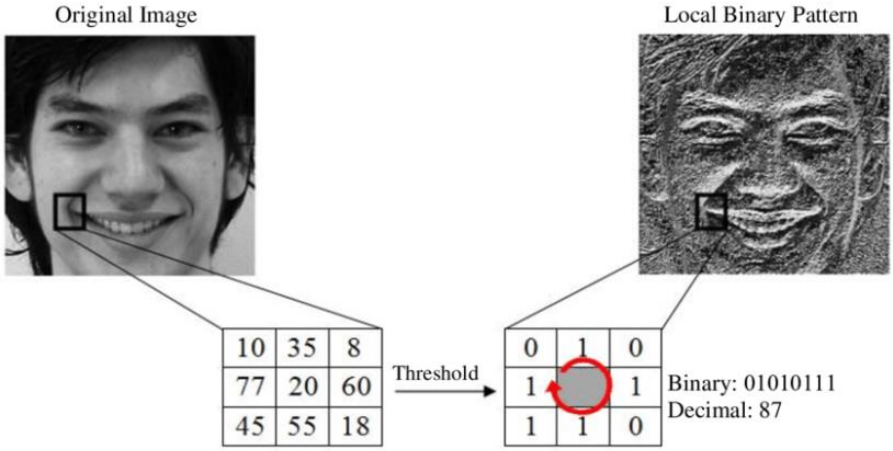


2 – Different clusters of color extraction1

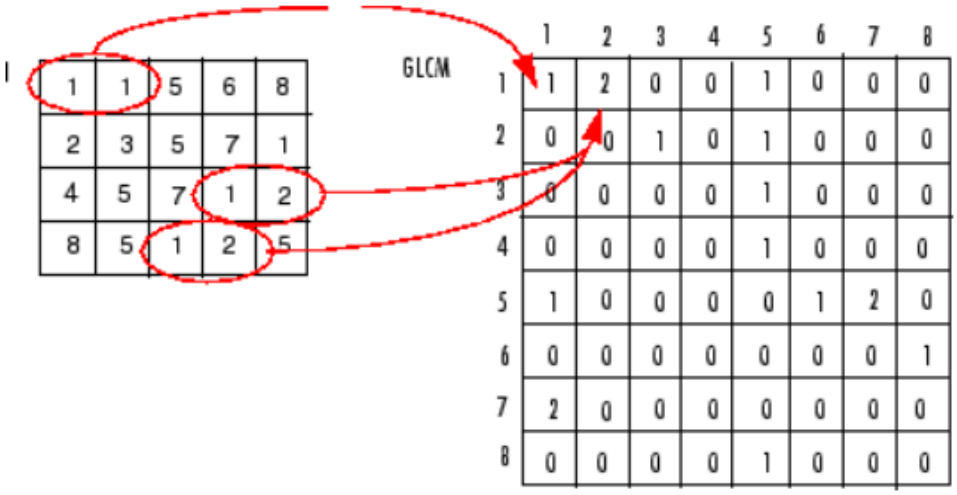


a Clusters resulting from RGB. b Clusters resulting from HSV. c Clusters coming from HSL. d Clusters coming from Lab. e Cluster coming from nRGB. f Clusters coming from H. g Clusters coming from HS

3 – Description of the LBP Process



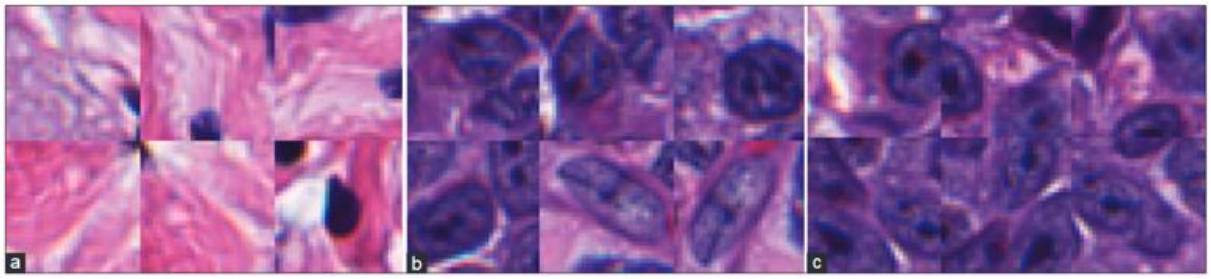
4 – Process used to create the GLMC



5 – Comparison approach with previous work

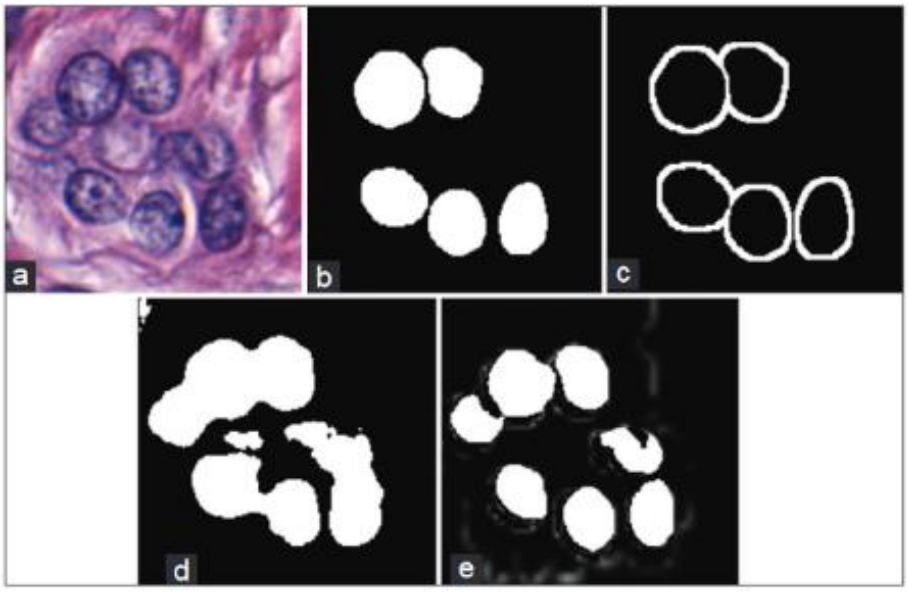
Title	Year	Feature extraction methods	Data	classifier	Accuracy %	Sensitivity %	Specificity %
A. Unni et al. [34]	2018	GLCM	MIAS	SVM	74.59	-	-
T. T. Htay et al. [35]	2018	GLCM	MIAS	KNN	92.0	-	-
N. Ponraj and M. Mercy [36]	2017	LBP	MIAS	SVM	91.0	90.0	92.0
M. M. Pawar et al.[37]	2018	LBP	MIAS	ANN	98.57	-	-
M. Abdel-Nasser et al.[38]	2016	HOG	MIAS	SVM	-	72.0	95.0
K. C. Tatikonda et al.[39]	2018	HOG	MIAS	SVM	90.0	76.66	98.89
My work [40]	-	LBP	MIAS	Logistic regression	92.5	88.0	97.0
My work	-	HOG	MIAS	SVM	90.0	79.0	100
My work	-	GLCM	MIAS	KNN	89.3	78.0	99.0

6 – Typical patches extracted for use a nuclear segmentation classifier



Six examples of (a) the negative class show large areas of stroma which are notably different than (b) the positive nuclei class and tend to be very easily classified. To compensate, we supplement the training set with (c) patches which are exactly on the edge of the nuclei, forcing the network to learn boundaries better

7 – The process of creation of training exemplars to enhance the result obtained *via* deep learning for nuclei segmentation

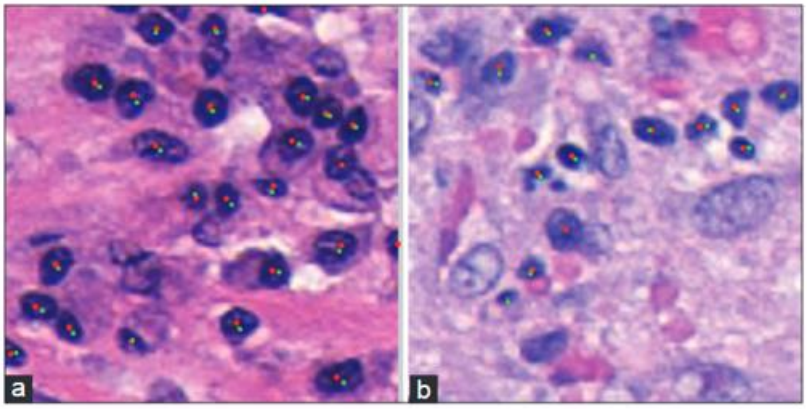


The original image (a) only has (b) a select dew of its nuclei annotated. This makes it difficult to find patches which represent a challenging negative class. Our approach involves augmenting a basic negative class, created by sampling from the thresholded color deconvoluted image. More challenging patches are supplied by (c) a dilated edge mask. Sampling locations from (c) allows us to create negative class samples which are of very high utility for the deep learning algorithm. As a result, our improved patch selection technique lead to (e) notably better-delineated nuclei boundaries as compared to the approach shown in (d)

8 – Results of the different approaches for the lymphoma multi-classification. The best performance is highlighted in bold.

Scenario	Network	Acc.	St. Dev.
Classification by reconstruction	FusionNet	77.60	5.4
	UNet	65.10	8.4
Supervised Classification	SEF	**97.67**	**1.3**
	SEU	92.80	3.2
	ResNet-34	95.47	2.0

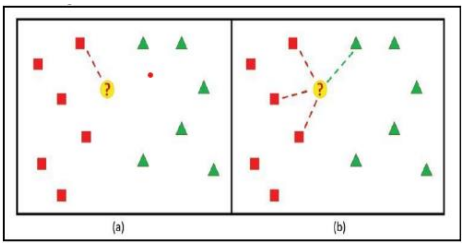
9 - Lymphocyte detection result where green dots are the ground truth, and red dots are the centers discovered by the algorithm



The image on the left (a) has 21 TP/2 FP/0 FN. The false positives are on the edges, about 1 o'clock and 3 o'clock. The image on the right (b) one has 11 TP/1 FP/2 FN. We can see the false négatives are quite small and not very clear making it hard to detect them without also encountering many false positives. The only false positives in the middle at around 7 o'clock though this structure does look lymphocyte-like.

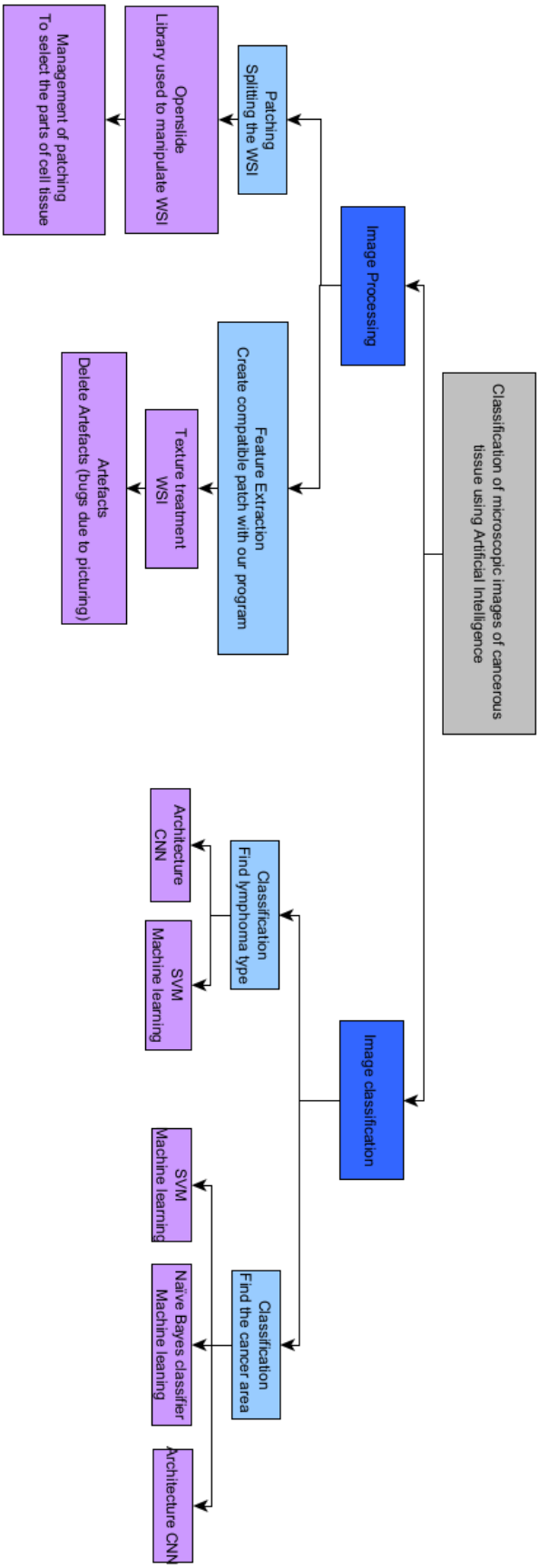
10 – Other classification approach : K-Nearest Neighbor (KNN)

In the article "Application of K-nearest neighbor (KNN) approach for predicting economic events theoretical background" (2013), KNN is used for some applications in the economy. Medicine applications are also mentioned. The article does not use precise data other than to say that it comes from data mining. The purpose of this article is to theoretically explain what is behind the KNN make it possible to classify data in order to make predictions. This algorithm can study points in a graphic and look around his K (a variable to set) neighbor(s) and calculate the distance between them. By doing so, an arena of points which are close to each other can be created. As stated in the article, in medical cases, this method is used to predict second heart attacks, calculate the amount of glucose in the blood, evaluate the risk of prostate cancer and analyze micro-array gene expression. One advantage of KNN is its robustness to the noise present in the data. However, with a large amount of data, the procedure to run the method takes a long time to finish. As for the limits of this method in our specifics case of lymphoma, KNN is great to visualize a concentration in an image. This could be interesting for the project as it could be applied on cell nuclei in our images.



Example of using KNN with (a) K=, and (b) K=4

11 – Work Breakdown Structure of the project and Gantt diagram

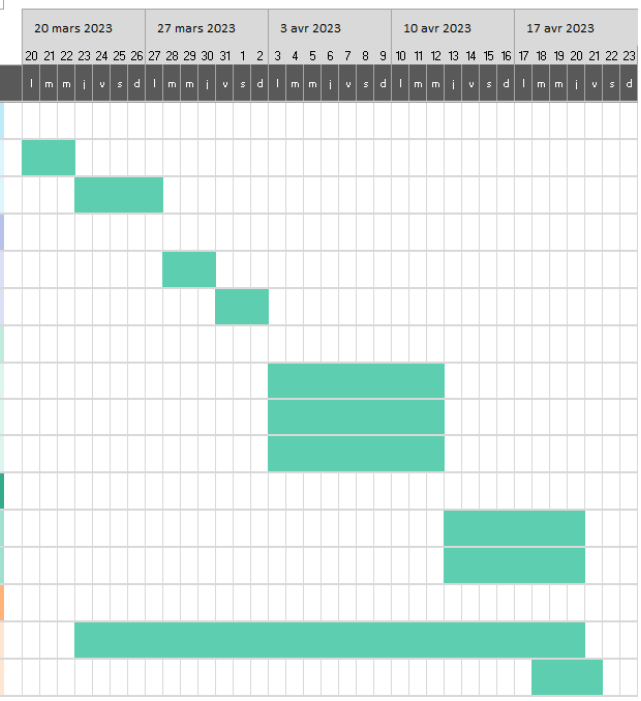


Project M1 : Classification Lymphoma Subtypes

Junia ISEN
Group 18

lun, 3/20/2023
1

TÂCHE	AVANCEMENT	DÉBUT	FIN
Patching : Splitting the WSI			
Openslide Library used to manipulate WSI	0%	20/3/23	22/3/23
Management of patching To select the parts of cell tissue	0%	23/3/23	27/3/23
Feature Extraction :Create compatible patch with our program			
Texture treatment WSI	0%	28/3/23	30/3/23
Delete Artefacts (bugs due to picturing)	0%	31/3/23	2/4/23
Classification : Find the cancer area			
SVM machine learning	0%	3/4/23	12/4/23
Architecture CNN	0%	3/4/23	12/4/23
Naive Bayes Classifier	0%	3/4/23	12/4/23
Classification : Find lymphoma type			
SEF CNN	0%	13/4/23	20/4/23
SVM machine learning	0%	13/4/23	20/4/23
Présentation finale			
Rédaction du rapport	0%	23/3/23	20/4/23
Presentation Orale	0%	18/4/23	21/4/23



Bibliography

Aziz, Mohamed Abd El, Ahmed A. Ewees, and Aboul Ella Hassanien. 'Whale Optimization Algorithm and Moth-Flame Optimization for Multilevel Thresholding Image Segmentation'. Expert Systems with Applications 83 (October 2017): 242–56. https://doi.org/10.1016/j.eswa.2017.04.023.

Brieu N, Caie DP, Gavriel GC, Schmidt G, Harrison JD. 'Context-based interpolation of coarse deep learning prediction maps for the segmentation of fine structures in immunofluorescence images.' In: Medical Imaging 2018: Digital Pathology. Houston, TX (2018).

Chandra, Mayank Arya; Bedi, S. S. «Survey on SVM and their application in image classification » "International Journal of Information Technology", Volume 13 (5): 11 – Oct 1, 2021

Chavolla, Edgar, Daniel Zaldivar, Erik Cuevas, and Marco Cisneros. 'Color Spaces Advantages and Disadvantages in Image Color Clustering Segmentation'. In Studies in Computational Intelligence, 3–22, 2018. https://doi.org/10.1007/978-3-319-63754-9_1.

Cruz-Roa, Angel, Ajay Basavanhally, Fabio González, Hannah Gilmore, Michael Feldman, Shridar Ganesan, Natalie Shih, John Tomaszewski, and Anant Madabhushi. "Automatic Detection of Invasive Ductal Carcinoma in Whole Slide Images with Convolutional Neural Networks." edited by Metin N. Gurcan and Anant Madabhushi, 904103. San Diego, California, USA, 2014. https://doi.org/10.1117/12.2043872.

Dimitriou, Neofytos, Ognjen Arandjelović, and Peter D. Caie. “Deep Learning for Whole Slide Image Analysis: An Overview.” Frontiers in Medicine 6 (2019). https://www.frontiersin.org/articles/10.3389/fmed.2019.00264.

Fan, Kun, Shibo Wen, and Zhuofu Deng. “Deep Learning for Detecting Breast Cancer Metastases on WSI,” 137–45, 2019. https://doi.org/10.1007/978-981-13-8566-7_13.

Gautam, Anjali, Priyanka Singh, Balasubramanian Raman, and Harvendra Bhadauria. ‘Automatic Classification of Leukocytes Using Morphological Features and Naïve Bayes Classifier’. In 2016 IEEE Region 10 Conference (TENCON), 1023–27, 2016. https://doi.org/10.1109/TENCON.2016.7848161.

Hsu, Wei-Wen, Jing-Ming Guo, Linmin Pei, Ling-An Chiang, Yao-Feng Li, Jui-Chien Hsiao, Rivka Colen, and Peizhong Liu. “A Weakly Supervised Deep Learning-Based Method for Glioma Subtype Classification Using WSI and MpMRIs.” Scientific Reports 12, no. 1 (April 12, 2022): 6111. https://doi.org/10.1038/s41598-022-09985-1.

Li, Shaoying, Ken H. Young, and L. Jeffrey Medeiros. “Diffuse Large B-Cell Lymphoma.” Pathology, 50th anniversary review issue, 50, no. 1 (January 1, 2018): 74–87. https://doi.org/10.1016/j.pathol.2017.09.006.

Imandoust, S.B., and Mohammad Bolandraftar. ‘Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events Theoretical Background’. Int J Eng Res Appl 3 (1 January 2013): 605–10.

Janowczyk, Andrew, and Anant Madabhushi. ‘Deep Learning for Digital Pathology Image Analysis: A Comprehensive Tutorial with Selected Use Cases’. Journal of Pathology Informatics 7 (26 July 2016): 29. https://doi.org/10.4103/2153-3539.186902.

Magee RD, Treanor D, Crellin D, Shires M, Smith EK, Mohee K, et al. Colour normalisation in digital histopathology images. In: Proc. Opt. Tissue Image Anal. Microsc. Histopathol. Endosc. London (2009). p. 100–11.

Miyoshi, Hiroaki, Kensaku Sato, Yoshinori Kabeya, Sho Yonezawa, Hiroki Nakano, Yusuke Takeuchi, Issei Ozawa, et al. ‘Deep Learning Shows the Capability of High-Level Computer-Aided Diagnosis in Malignant Lymphoma’. Laboratory Investigation 100, no. 10 (October 2020): 1300–1310. https://doi.org/10.1038/s41374-020-0442-3.

Simon, Olivier, Rabi Yacoub, Sanjay Jain, John E. Tomaszewski, and Pinaki Sarder. ‘Multi-Radial LBP Features as a Tool for Rapid Glomerular Detection and Assessment in Whole Slide Histopathology Images’. Scientific Reports 8, no. 1 (1 February 2018): 2032. https://doi.org/10.1038/s41598-018-20453-7.

Valkonen, Mira, Kimmo Kartasalo, Kaisa Liimatainen, Matti Nykter, Leena Latonen, and Pekka Ruusuvuori. “Metastasis Detection from Whole Slide Images Using Local Features and Random Forests.” *Cytometry Part A* 91, no. 6 (2017): 555–65. https://doi.org/10.1002/cyto.a.23089.

Wei, Jason W., Laura J. Tafe, Yevgeniy A. Linnik, Louis J. Vaickus, Naofumi Tomita, and Saeed Hassanpour. “Pathologist-Level Classification of Histologic Patterns on Resected Lung Adenocarcinoma Slides with Deep Neural Networks.” Scientific Reports 9, no. 1 (March 4, 2019): 3358. https://doi.org/10.1038/s41598-019-40041-7.