



DataScientest • com

Rapport Technique d'évaluation

Classification des cellules sanguines - leukopy

Promotion Mars 2021

Participants :

Mathieu Sarrat

Marie-Anne Mawhin

Laleh Ravanbod

Yahia Bouzoubaa

Contexte

Les globules blancs, plaquettes et globules rouges sont des composants cruciaux des réponses immunitaires. La présence de précurseurs des cellules sanguines dans la circulation est une indication importante d'infections ou de leucémies (Figure 1). Ainsi, plusieurs pathologies peuvent être détectées par une analyse hématologique, par exemple l'identification et la classification des leucocytes, des plaquettes et des érythrocytes est cruciale dans le diagnostic de l'anémie régénérative. Alors que plusieurs instruments de comptage automatique tels que le cytomètre en flux ou l'hémocytomètre fournissent des informations sur les comptes sanguins, l'évaluation qualitative de la cellule par imagerie est souvent nécessaire. Par exemple, en clinique, les frottis sanguins sont indiqués en cas de SIDA, de septicémie, de défaillance d'organe, de leucocytose inexpliquée, d'anémie ou de leucémie suspectée.

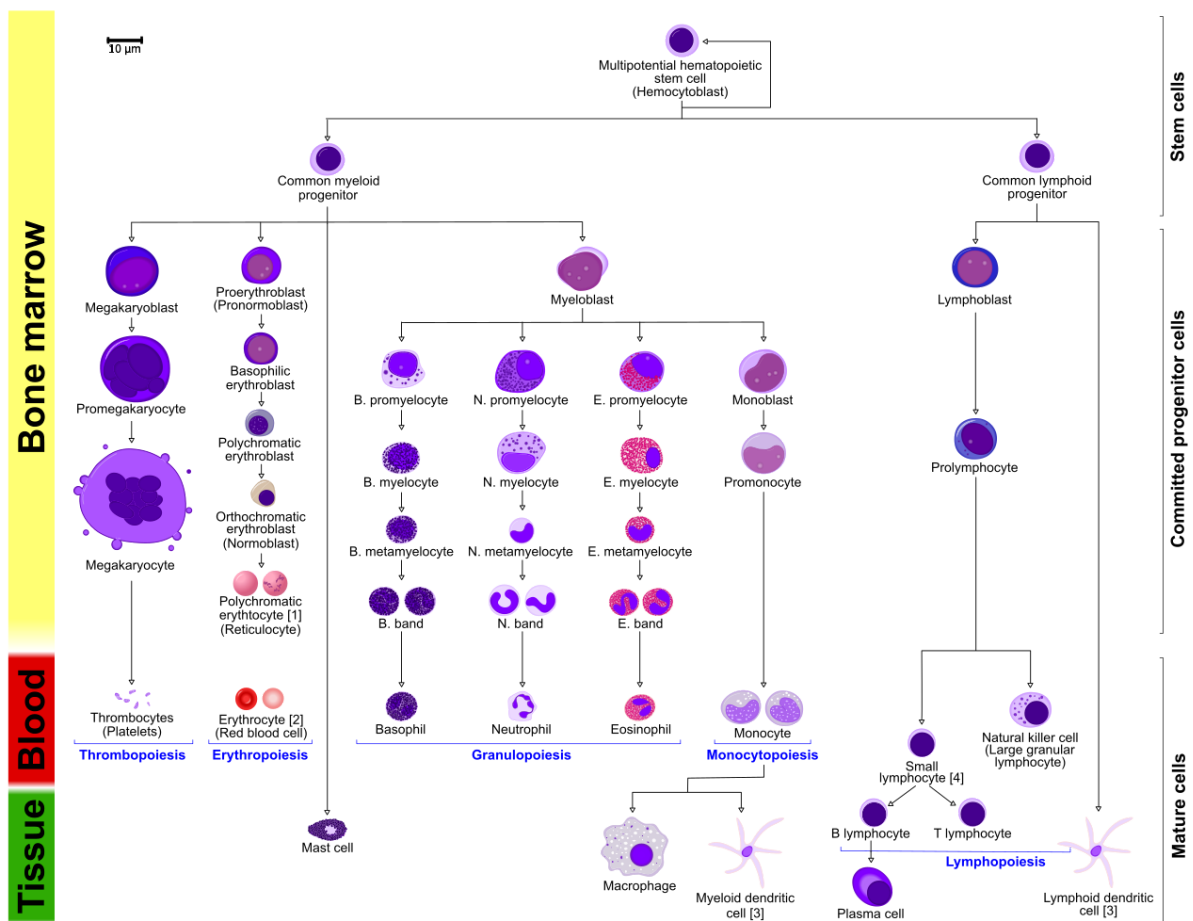


Figure 1 : Cellules sanguines périphériques et précurseurs.

L'identification manuelle des cellules sanguines périphériques est longue et coûteuse et nécessite la présence d'un spécialiste formé. De plus, la classification de certains précurseurs, comme les myélocytes et métamyélocytes, est sujette à de nombreuses controverses inter-spécialistes. Dans des contextes de recherche médicale, où les frottis sanguins sont souvent utilisés pour évaluer la qualité d'une préparation cellulaire, l'identification manuelle est un challenge. Enfin, l'évaluation manuelle est sujette à des résultats erronés. L'analyse assistée par ordinateur des frottis sanguins et l'identification des cellules anormales fourniraient une aide cruciale pour la recherche et les cliniciens, en termes de coût, de temps et de qualité.

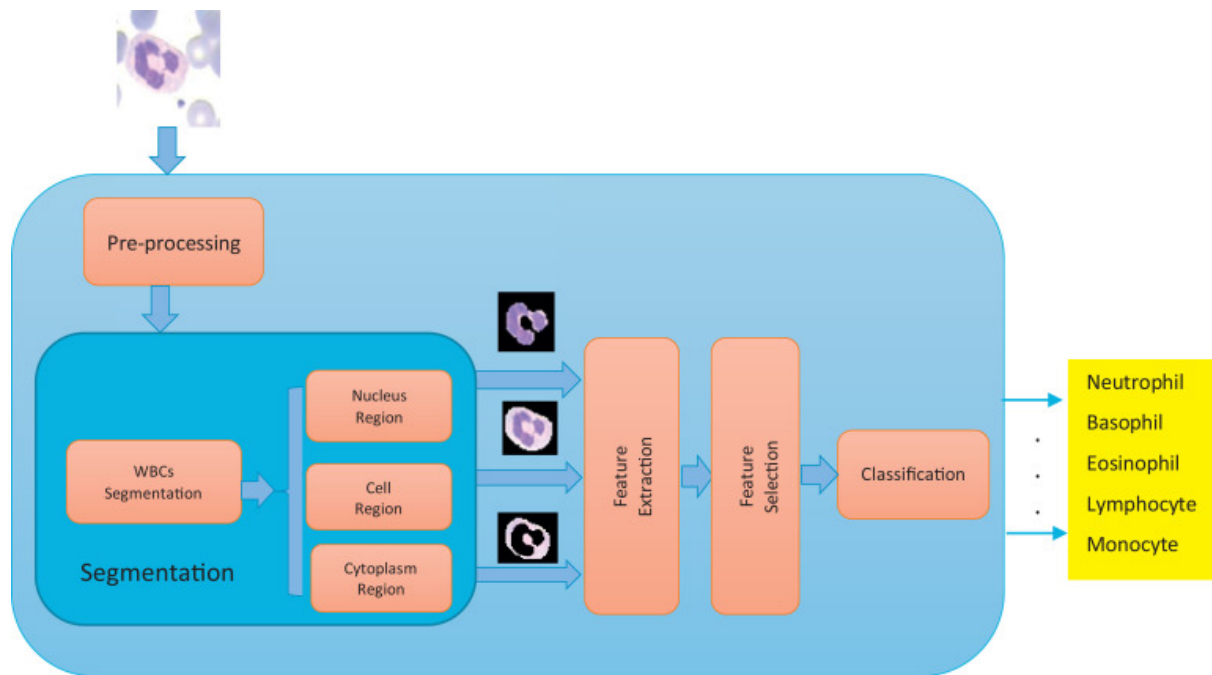


Figure 2: Identification des cellules sanguines selon un workflow classique. ¹⁾

Traditionnellement, les systèmes de reconnaissance des globules blancs reposent sur la segmentation, la séparation des composants, l'extraction de caractéristiques et la classification des globules blancs par des modèles de shallow machine learning (Figure 2). Ce type de méthode est difficile à généraliser, notamment, par la variabilité des colorations de Romanowsky et des systèmes d'acquisition, et par l'exigence non négligeable de sa conception en termes de pré-processing et d'extraction des features. Dans ce contexte, l'utilisation du deep learning prend tout son sens.

¹ "White blood cells identification system based on convolutional deep"
<https://pubmed.ncbi.nlm.nih.gov/29173802/>. Accessed 27 Sep. 2021.

Objectifs

Consignes : Pour chacun des membres du groupe, préciser le niveau d'expertise autour de la problématique adressée ? Êtes vous entré en contact avec des experts métiers pour affiner la problématique et les modèles sous-jacents ? Si oui, détaillez l'apport de ces interactions.

L'objectif principal de cette étude est de développer un réseau de neurones profonds capables de classer les cellules saines du sang en 8 ou 11 classes principales de cellules sanguines directement à partir d'images :

- neutrophiles (segmentés) - SNE
- éosinophiles - EO
- basophiles - BA
- lymphocytes -LY
- monocytes -MO
- granulocytes immatures (métamyélocytes, myélocytes, promyélocytes) et "band" - IG ou séparés - MMY, MY, PMY, BNE
- plaquettes - PLT
- érythroblastes - ERB

Un objectif secondaire est de classer les cellules anormales, leucémiques, cependant par manque de ressources cet objectif a été écarté.

Une revue de la littérature a mis à jour plusieurs publications sur la classification des cellules sanguines par machine learning. La plupart des références utilisent des méthodes de segmentation, une phase d'extraction de features suivie de classifieur soit deep soit shallow². Cependant, plusieurs publications sont apparues récemment : les modèles présentés reposent sur le deep learning uniquement¹, avec des données pré-segmentées³ limitées à 5-6 classes^{4,5,6} ou un dataset unique^{7,8,9}. Ces articles scientifiques nous ont permis de mettre au point un workflow de base, notamment pour l'augmentation de données et le choix des modèles, plutôt de type vgg. Cela nous a permis aussi de déterminer des pistes d'amélioration, comme par exemple dans la généralisation limitée des modèles entraînés sur un seul jeu de données¹⁰.

² "Deep learning approach to peripheral leukocyte recognition - PLOS." 25 Jun. 2019, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0218808>. Accessed 26 Sep. 2021.

³ "Fine-grained leukocyte classification with deep residual learning for" <https://www.sciencedirect.com/science/article/pii/S0169260718300567>. Accessed 26 Sep. 2021.

⁴ "White blood cells detection and classification based on regional" <https://www.sciencedirect.com/science/article/pii/S0306987719310680>. Accessed 26 Sep. 2021.

⁵ "Fine-grained leukocyte classification with deep residual learning for" <https://www.sciencedirect.com/science/article/pii/S0169260718300567>. Accessed 26 Sep. 2021.

⁶ "Detection of subtype blood cells using deep learning - ScienceDirect." <https://www.sciencedirect.com/science/article/pii/S1389041718303760>. Accessed 26 Sep. 2021.

⁷ "White blood cell differential count of maturation stages in ... - PLOS." 11 Dec. 2017, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189259>. Accessed 26 Sep. 2021.

⁸ "Recognition of peripheral blood cell images using convolutional" <https://www.sciencedirect.com/science/article/pii/S0169260719303578>. Accessed 26 Sep. 2021.

⁹ "Human-level recognition of blast cells in acute myeloid leukaemia." 12 Nov. 2019, <https://www.nature.com/articles/s42256-019-0101-9>. Accessed 26 Sep. 2021.

¹⁰ "Generalizability in White Blood Cells' Classification Problem | bioRxiv." 28 Jul. 2021, <https://www.biorxiv.org/content/10.1101/2021.05.12.443717v3>. Accessed 26 Sep. 2021.

L'équipe a quelques bases en biologie et analyse d'images. Marie-Anne est à l'origine du projet, elle a un doctorat en immunologie et elle a donc des connaissances sur les cellules sanguines bien qu'elle ne soit pas experte en hématologie et cytologie. Mathieu est agrégé de physique et a un doctorat en physique des plasmas, cette problématique est donc nouvelle pour lui.

Data

Cadre

Cellules saines

1. Normal_Barcelone :

Nous avons commencé notre étude avec un jeu de données nettoyé par des experts pathologistes de l'Hôpital Clinique de Barcelone. Ce dataset constitue plus de 17000 images de cellules sanguines issues de frottis de donneurs sains et colorées par May-Grunwald Giemsa (MGG) (Table 1)¹¹.

2. Normal_CVB :

Nous avons aussi trouvé un petit dataset issu du blog de CellaVision et contenant 100 images de donneurs sains¹². Au vu de la taille de ce jeu de données, nous avons choisi de le garder comme validation et test de la **généralisabilité** de notre modèle (Table 1).

3. Normal_Raabin :

Nous avons ensuite eu accès à une base de données d'environ 15000 images, incluant les 5 principales classes de cellules sanguines¹³. Tous les échantillons sont sains mis à part quelques basophiles imagés d'un patient atteint de leucémie et viennent de trois laboratoires en Iran : Razi Hospital à Rasht, Gholhak Laboratory, Shahr-e-Qods Laboratory et Takht-e Tavous Laboratory à Téhéran.

Cellules leucémiques et saines

1. AML_Munich :

Nous avons utilisé le dataset AML Morphology Dataset qui contient 18,365 images de frottis sanguins labellisées par des experts et issues de 100 patients diagnostiqués avec une leucémie myéloïde aiguë et de 100 patients "sains" à l'Hôpital de Munich (Table 1)¹⁴.

2. APL_AML_Johns_Hopkins :

Nous avons aussi exploré un dataset de cellules sanguines cancéreuses de type leucémie myéloïde mais après une première exploration, beaucoup de labels se sont révélés erronés. Ce jeu de données a donc été écarté face à l'impossibilité de re-labelliser les données¹⁵.

¹¹ "A dataset of microscopic peripheral blood cell images for ... - NCBI." 8 Apr. 2020, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7182702/>. Accessed 18 Sep. 2021.

¹² "Fast and robust segmentation of white blood cell images by self" <https://www.sciencedirect.com/science/article/pii/S0968432817303037>. Accessed 18 Sep. 2021.

¹³ "Raabin-WBC: a large free access dataset of white blood ... - bioRxiv." 29 May. 2021, <https://www.biorxiv.org/content/10.1101/2021.05.02.442287v4>. Accessed 26 Sep. 2021.

¹⁴ "A Single-cell Morphological Dataset of Leukocytes from AML" 17 Feb. 2021, <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=61080958>. Accessed 26 Sep. 2021.

¹⁵ "Deep Learning for Distinguishing Morphological Features of Acute" 5 Nov. 2020, <https://ashpublications.org/blood/article/136/Supplement%201/10/472209/Deep-Learning-for-Distinguishing-Morphological>. Accessed 18 Sep. 2021.

	Normal Barcelone	Normal CVB	Normal_Raabin	AML_Munich
origin	Anonymised healthy blood smears stained with MGG (automatic stainer, QC)	Healthy donors	73 blood smears from healthy donors	100 patients diagnosed with Acute Myeloid Leukemia at Munich University Hospital between 2014 and 2017 + 100 healthy donors
staining	MGG	unknown (probably Giemsa)	Giemsa	Giemsa
size	17092 RGB images of individual cells	100 RGB images	14514 RGB images	18365 RGB images
extension	.jpg	.bmp	.jpg	.tiff
size	360 x 363 px	300 x 300 px	575 x 575 px	400 x 400 px
Acquisition system	CellaVision	CellaVision	Olympus microscope + Samsung camera	M8 digital microscope / scanner
annotation	Expert pathologist	NA	Two independent expert pathologists	Expert at Munich, re-annotated 2 times to account for variability inter-expert.
Cell type	8 Neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes (metamyelocytes, myelocytes and promyelocytes), erythroblasts, platelets	naming: in class_label_CVB.csv, labels (1- 5) correspond respectively to neutrophil, lymphocyte, monocyte, eosinophil and basophil.	Neutrophils, eosinophils, basophils, lymphocytes, monocytes,	BAS Basophil EBO Erythroblast EOS Eosinophil KSC Smudge cell LYA Lymphocyte (atypical) LYT Lymphocyte (typical) MMZ Metamyelocyte MOB Monoblast MON Monocyte MYB Myelocyte MYO Myeloblast NGB Neutrophil (band) NGS Neutrophil (segmented) PMB Promyelocyte (bilobed) PMO Promyelocyte
Ref	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7182702/	https://ashpublications.org/blood/article/136/Supplement%201/10/472209/Deep-Learning-for-Distinguishing-Morphological	https://www.biorxiv.org/content/10.1101/2021.05.02.442287v4?ijkey=7a2390aa1717177c60a3b67203f28fdcf23208d4&keytype=tf_ipsecsha	https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=61080958

Table 1: Description des datasets.

Pertinence

*Avez vous eu à nettoyer et à traiter les données ? Si oui, décrivez votre processus de traitement.
Quelles particularités de votre jeu de données pouvez-vous mettre en avant ?*

La variable cible dépend de la complexité choisie dans la classification des cellules, soit nous intégrons 8 classes :

- neutrophiles (segmentés) - SNE
- éosinophiles - EO
- basophiles - BA
- lymphocytes - LY
- monocytes - MO
- granulocytes immatures (métamyélocytes, myélocytes, promyélocytes, 'band')
- plaquettes - PLT
- érythroblastes - ERB

Soit nous intégrons 11 classes avec les granulocytes séparés : MMY, MY, PMY, BNE.

Dans un premier temps nous avons exploré le dataset "Normal_Barcelone". Les classes ne sont pas équilibrées avec une prédominance de neutrophiles et éosinophiles et un plus faible nombre de basophiles et lymphocytes (Fig. 3).

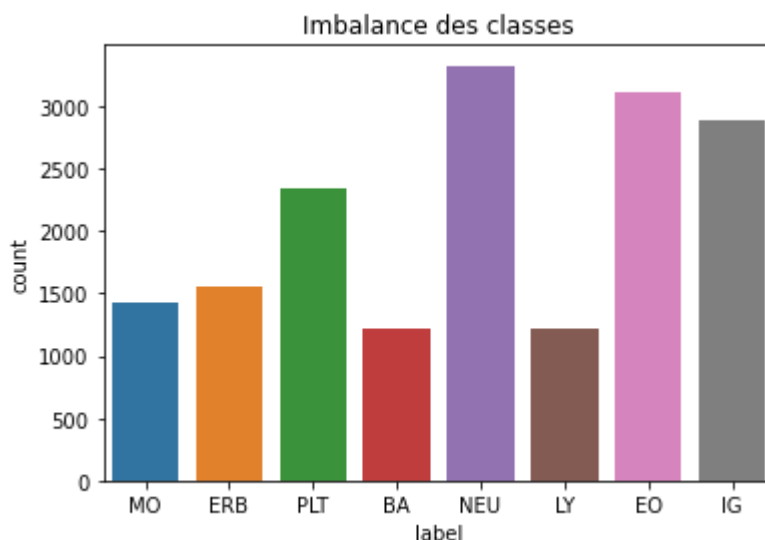


Figure 3: Distribution des classes dans le dataset de Barcelone.

En regardant les caractéristiques des cellules (Fig. 4), on peut remarquer des similitudes entre certains types de cellules : par exemple les monocytes et les granulocytes immatures ou les neutrophiles segmentés et bandes. On remarque aussi que certains éléments pourraient parasiter l'information importante, les globules rouges en fond et couleur de l'image.

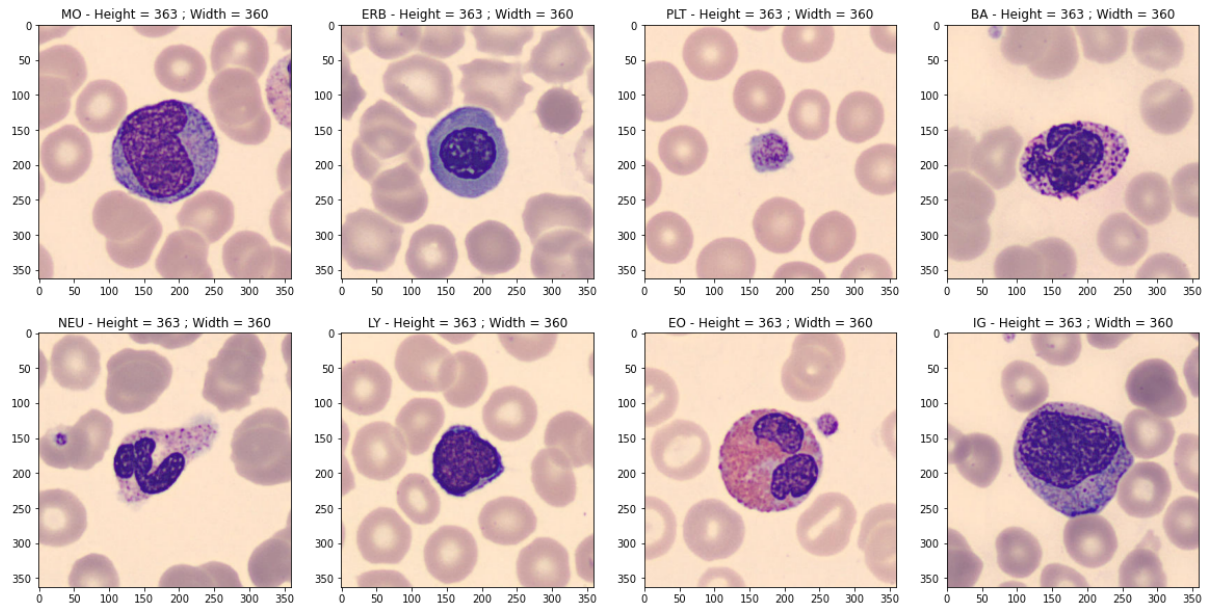


Figure 4 : Exemples des différents types cellulaires. La cellule d'intérêt présente généralement une forte coloration (bleue, violette ou rose selon le type de cellule) et se trouve au centre de l'image. Les petits objets de forme globalement arrondie situés autour sont des globules rouges.

Par une analyse de la luminosité de chaque image par classe de cellules, nous avons pu déterminer que toutes les classes présentent des outliers, essentiellement de faible luminosité, liés à la taille du noyau de la cellule d'intérêt, au nombre de globules rouges en fond, et à la présence de cellules multiples. De plus, les cellules de petites tailles comme les plaquettes et les érythroblastes ont une luminance plus haute. En utilisant des techniques de réduction de dimension telles que PCA, LDA ou UMAP (Fig. 5), nous avons confirmé que les plaquettes se distinguent bien du reste des cellules : cela pourrait s'expliquer par l'absence de noyau, qui conduit à une luminosité moyenne supérieure pour ces images.

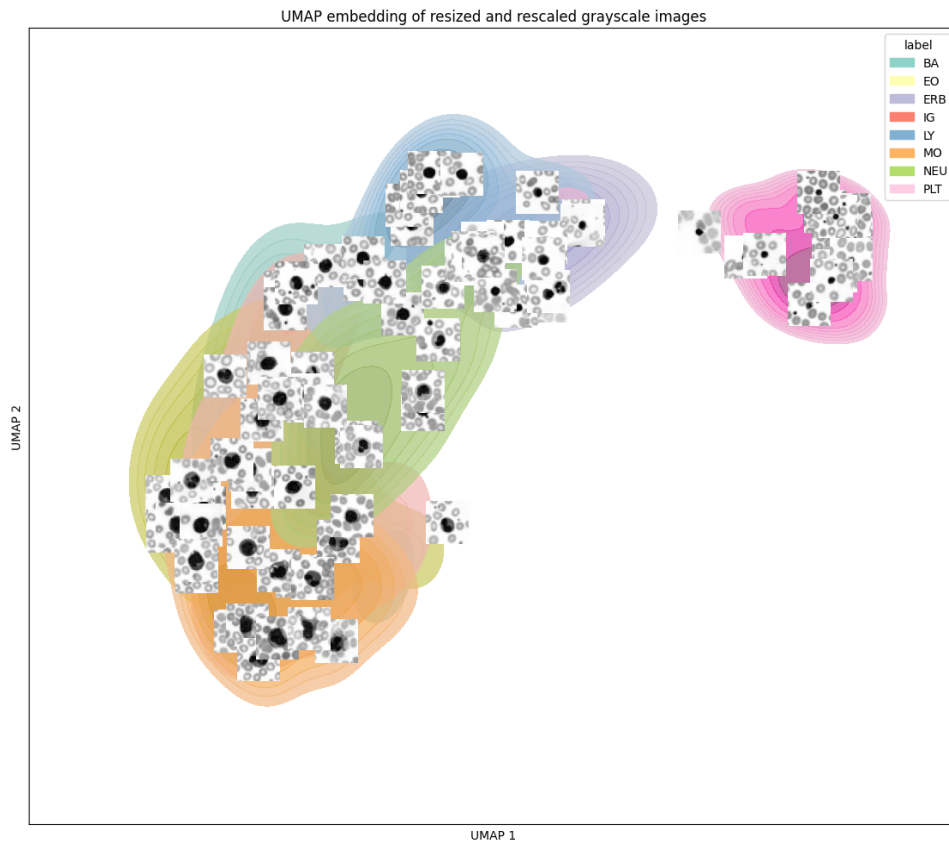


Figure 5: Réduction de dimension avec l'UMAP.

Pour pousser l'analyse, nous avons utilisé une technique de déconvolution des couleurs développée pour l'analyse histologique et la librairie employée utilise l'algorithme développé par G. Landini¹⁶. La déconvolution des couleurs est basée sur la loi de Beer-Lambert et peut séparer les images en trois canaux représentant l'absorbance de chaque coloration histologique : les globules rouges, les noyaux et le background. Des différences sont observées avec les plaquettes (PLT) et les érythroblastes (ERB) avec une intensité plus faible dans le canal noyau, plus haute dans le background et similaire dans le canal globules rouges, ce qui correspond bien à une taille plus petite (confirmée par une analyse de luminosité radiale des cellules moyennes, résultats non montrés dans ce rapport). Les basophiles (BA) présentent eux aussi un shift dans le canal background ce qui peut s'expliquer par leurs tailles et la présence d'agglomérats autour de ces cellules. Il n'y a donc probablement pas de biais de luminosité d'une classe de cellules à l'autre.

En termes de nettoyage, les images sont au format RGB et toutes n'ont pas la même taille. Nous avons donc dû redimensionner les images. Les données dont nous disposons pour les cellules saines dans le dataset de Barcelone sont de qualité et ne présentent pas de biais évident. Il faut toutefois ajouter qu'elles ont été enregistrées avec le même analyseur, et traitées dans le même laboratoire avec le même protocole de coloration : c'est une autre forme de biais, qui pourrait induire un overfitting de notre modèle et c'est pourquoi nous avons inclus d'autres datasets dans l'entraînement du modèle final. Pour débiter, nous avons comparé plusieurs modèles en travaillant à partir du dataset de Barcelone. Après avoir sélectionné un modèle "final", nous avons ensuite complété ce dataset avec les autres données que nous avons pu trouver.

¹⁶ "Colour deconvolution: stain unmixing in histological imaging - PubMed." 16 Jun. 2021, <https://pubmed.ncbi.nlm.nih.gov/32997742/>. Accessed 26 Sep. 2021.

Projet

Classification du problème

Avez vous utilisé d'autres métriques de performances qualitative ou quantitative) ? Si oui, détaillez.

Nous avons un **problème de classification d'images** à 8 ou 11 classes (en fonction de la complexité). Il s'agit de **computer vision**. La métrique choisie est le **F1-score pour chaque classe considérée individuellement** (le dataset étant déséquilibré, en plus de precision et recall) et l'**accuracy pour la performance globale** (toutes classes confondues, le déséquilibre n'étant toutefois pas énorme) . Nous avons aussi tenté de comprendre ce que faisaient nos modèles avec Grad-CAM, qui permet d'analyser les régions de l'image auxquelles le modèle accorde de l'importance.

Choix du modèle & Optimisation

Consignes :

- *Quels algorithmes avez vous essayé ?*
- *Décrivez celui / ceux que vous avez retenu et pourquoi ?*
- *Qu'est ce qui a engendré une amélioration significative de vos performances ?*
- *Avez-vous analysé les erreurs de votre modèle ?*
- *Cela a-t-il contribué à son amélioration ? Si oui, décrivez.*
- *Détaillez quelle a été votre contribution principale dans l'atteinte des objectifs du projet.*

Plusieurs modèles ont été essayés au cours des premières itérations du projet. La plupart de ces approches reposent sur les Réseaux Neuronaux Convolutifs, fréquemment utilisés dans le domaine de la classification d'images et connus pour obtenir de bonnes performances. Une approche plus originale, utilisant un modèle de VisionTransformer, a également été testée.

On soulignera que dans un premier temps nous avons tenté des modèles de Machine-Learning traditionnel (UMAP + KNN).

LeNet

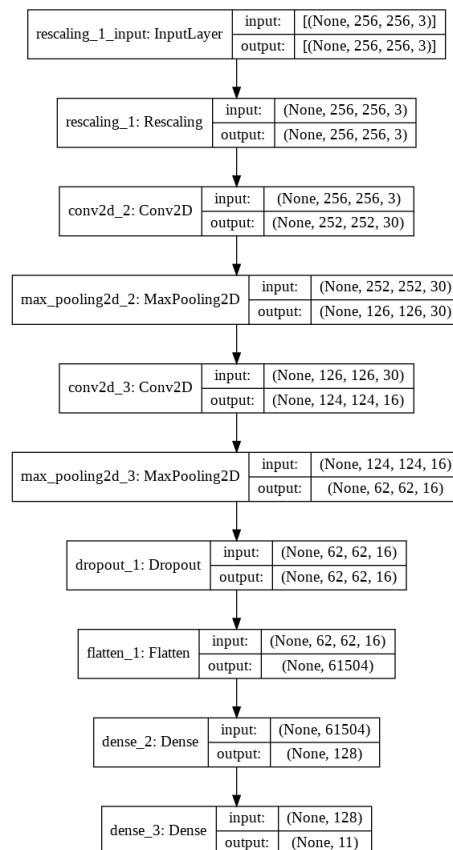
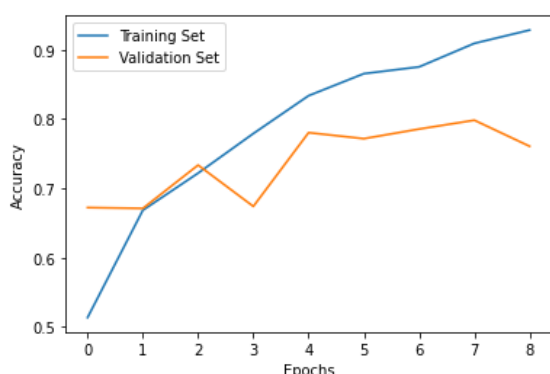
LeNet : baseline

Ce modèle, proposé par Yann LeCun en 1989, est l'un des tout premiers modèles de deep learning construit pour la classification d'images. Du fait de sa simplicité, il nous a servi à **établir une baseline** pour les données de Barcelone **mais aussi à nous familiariser avec les concepts de base** des CNN, Google Colab et le calcul sur GPU. L'**architecture du modèle LeNet utilisé est représentée sur la figure ci-contre, à droite.**

Le jeu de données a été divisé en trois ensembles : un jeu de test (2544 images), un jeu de validation (1730 images) et un jeu d'entraînement (12677 images). Les images ont été redimensionnées en 256*256*3 (RGB) et traitées par batchs de 32, sans augmentation de données. Ci-dessous, **la distribution des classes des images pour le jeu d'entraînement** :

EO	PLT	SNE	BNE	ERB	MO	LY	BA	MY	MMY	PMY
2352	1743	1263	1212	1162	1068	946	903	833	734	461

On compare sur le graphique ci-dessous l'accuracy globale mesurée sur le jeu d'entraînement (bleu) et sur le jeu de validation (orange) durant l'entraînement. On voit que le modèle over-fitte de manière importante les données, l'écart d'accuracy mesurée sur les deux jeux de données étant de plus en plus important au fur et à mesure de l'entraînement. **L'accuracy sur le jeu de validation plafonne aux alentours de 79%** (79,31% sur le jeu de validation, 79,83% sur le jeu de test).



Classification Report :				
	precision	recall	f1-score	support
BA	0.860825	0.883598	0.872063	189
BNE	0.610224	0.712687	0.657487	268
EO	0.983213	0.949074	0.965842	432
ERB	0.889362	0.881857	0.885593	237
LY	0.963415	0.940476	0.951807	168
MMY	0.537815	0.405063	0.462094	158
MO	0.810256	0.76699	0.78803	206
MY	0.453039	0.445652	0.449315	184
PLT	0.991667	0.983471	0.987552	363
PMY	0.481203	0.831169	0.609524	77
SNE	0.733906	0.652672	0.690909	262

Figure 7 : Gauche : accuracy globale du modèle LeNet sur les jeux d'entraînement et de validation. Droite : rapport de classification pour le jeu de test : précision, rappel, f1-score de chaque classe. La colonne support indique le nombre d'images dans chaque classe (labels réels).

La **lecture du rapport de classification** nous permet de mettre en évidence de très bons F1-Scores pour les éosinophiles (cellules roses, EO), sur les lymphocytes (LY) ainsi que sur les plaquettes (PLT). En revanche, le modèle est **médiocre pour les MMY, PMY et MY, ainsi que pour les SNE/BNE**.

Regardons maintenant la matrice de confusion obtenue pour le jeu de test afin de déterminer quelles sont les confusions les plus fréquentes commises par le modèle : sur l'axe vertical nous représentons les labels réels des images ; les labels prédits se trouvent quant à eux sur l'axe horizontal. Chaque case de la matrice contient le pourcentage d'images du label X classées dans la catégorie Y : la somme des valeurs d'une ligne vaut 1. Ainsi, on peut lire sur la première ligne que 88,4% des images de BA ont été classées BA par le modèle. La couleur de chaque case est représentative du pourcentage d'images qui s'y trouve : **un modèle qui classe bien les images implique une matrice de confusion dont seule la diagonale est sombre**.

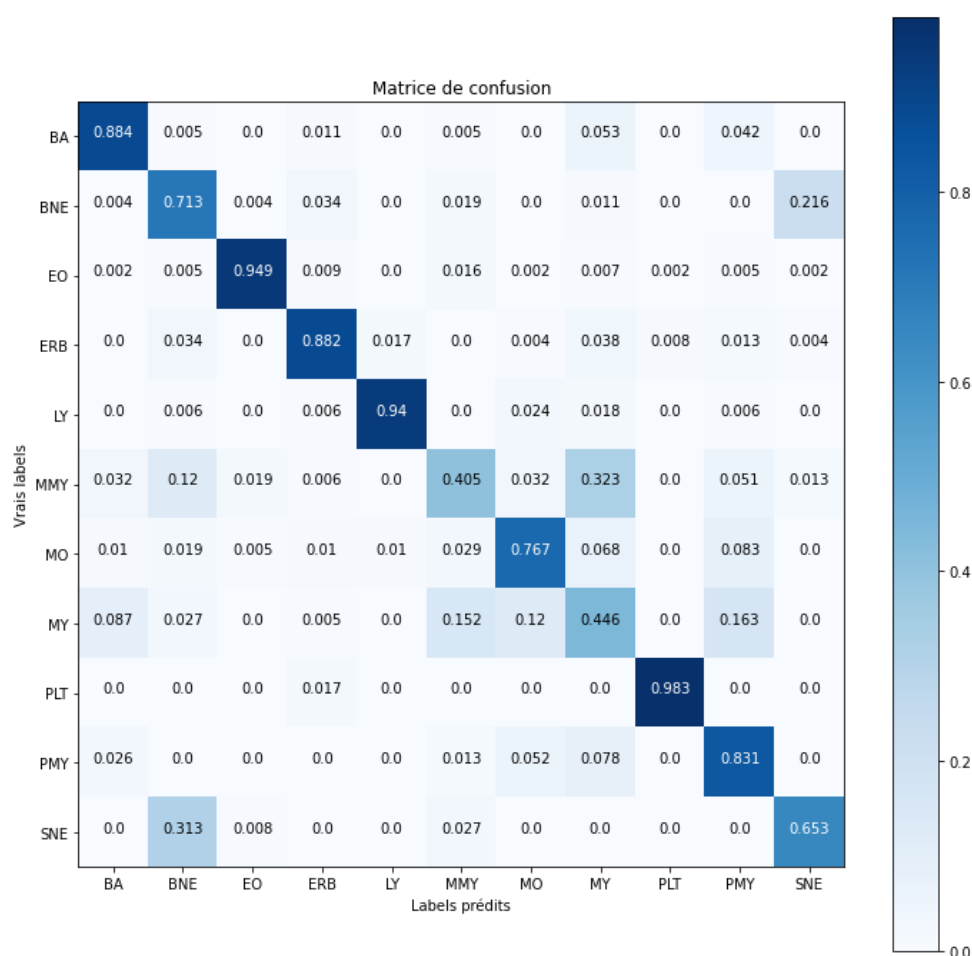


Figure 8 : Matrice de confusion sur le jeu de test. Modèle LeNet, 11 classes.

Ce n'est clairement pas le cas avec le modèle LeNet. Sans entrer dans les détails (nous le ferons plus loin), soulignons les principales erreurs du modèle :

- **confusion entre BNE et SNE** (31,3% des SNE sont classés comme BNE, 21,6% des BNE sont classés comme SNE);
- **confusion entre MY et MMY**, et dans une **moindre mesure avec les PMY (bien mieux détectés) et les monocytes (MO)**.

Il apparaît que **les MY sont les plus difficiles à classer**, dans le sens où on trouve un pourcentage important de ces images dans pas moins de 5 classes.

LeNet : considérations sur le label des cellules

À titre d'expérience et de comparaison, **entraînons maintenant le modèle sur 8 classes**, comme c'est le cas dans l'article accompagnant le dataset Barcelone Normal : pour cela, nous allons regrouper PMY, MY et MMY dans une même classe (IG pour Immature Granulocyte) et SNE et BNE dans la classe NEU (Neutrophil).

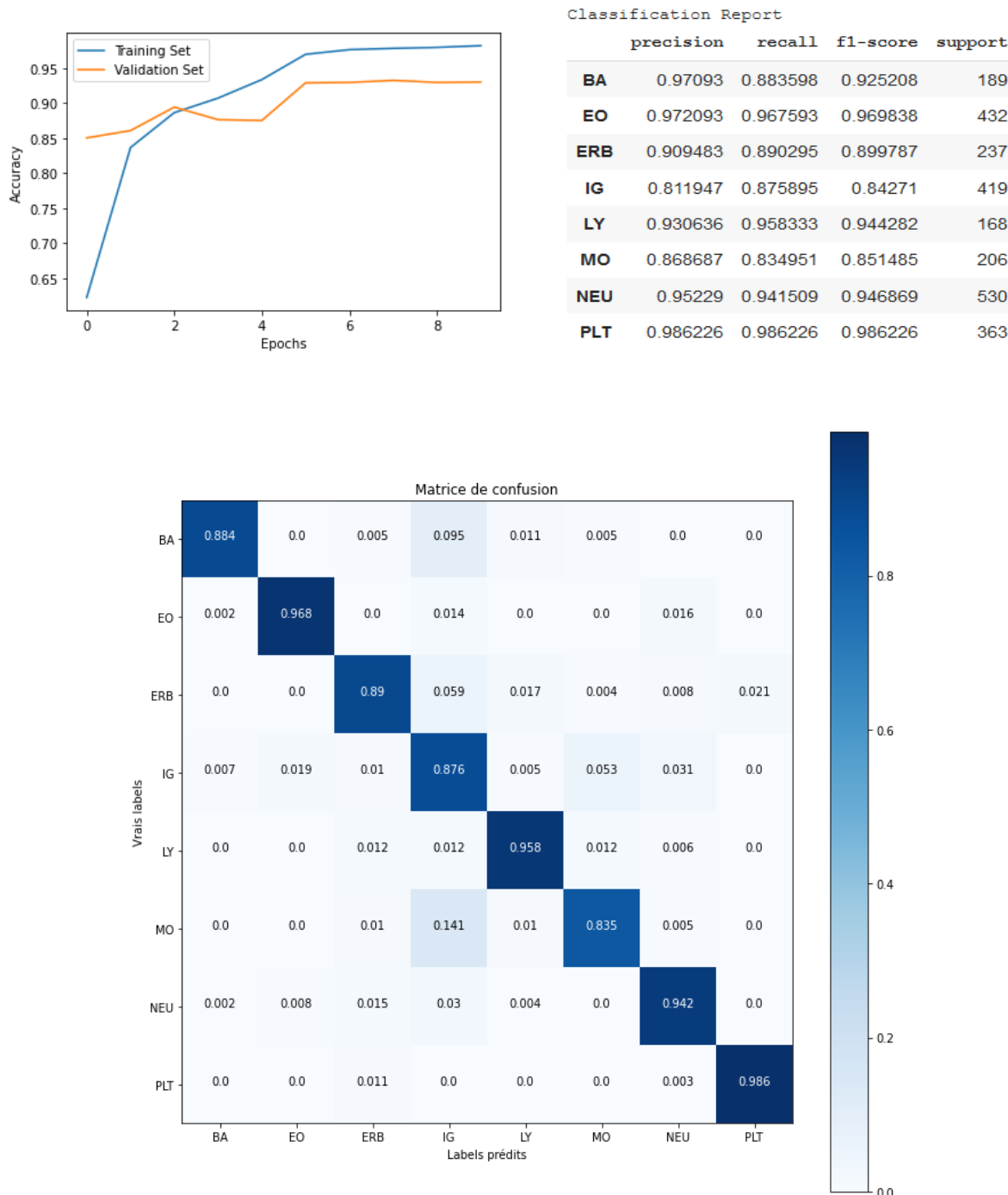


Figure 9 : Modèle LeNet, 8 classes. Résultats pour le jeu de test.

Le résultat de cette expérience est une **amélioration drastique des performances apparentes du modèle LeNet**, puisque nous obtenons une accuracy globale de 92 ~ 93% sur les jeux de test et de validation (soit plus de 10% d'amélioration par rapport au cas à 11 classes, **et pourtant avec exactement les mêmes données** dans chacun des trois jeux). Le résultat est légèrement moins

bon que dans l'article Le surapprentissage apparaît nettement réduit et le modèle classe légèrement mieux d'autres types de cellules (BA et MO).

De ceci, il y a plusieurs commentaires à faire :

- nous passons d'un F1-Score inférieur à 70% à un score supérieur à 94% pour l'ensemble des neutrophiles. L'augmentation du F1-Score pour les IG (MY, MMY et PMY regroupés) est tout aussi spectaculaire, et s'accompagne d'une amélioration importante pour les MO. **Cela semble aller dans le sens d'une confusion entre d'une part les sous-classes de neutrophiles : SNE/BNE et d'autre part entre les sous-classes d'immatures (MY, MMY et PMY).** Cette confusion peut provenir de **deux causes : le modèle qui peine à distinguer les cellules ou des erreurs lors du labelling des images.** Nous avons en tout cas deux **“super-classes” (IG et NEU) qu'il est difficile de subdiviser ;**
- le choix des classes à décrire correctement et le labelling en amont du modèle peuvent donc avoir un impact très important sur le résultat final et **conduire à une augmentation / à une diminution artificielle des performances d'un modèle.** Il est ainsi très simple de “mettre la poussière sous le tapis” sans s'en rendre compte : il convient donc d'être prudent et de **réfléchir au rôle de chaque sous-classe de cellule** pour savoir si oui ou non il est pertinent de considérer deux classes ou bien une seule.

En réalité, les neutrophiles segmentés (SNE), les éosinophiles et les basophiles font partie d'une catégorie de globules blancs appelés “granulocytes”, selon la capacité qu'elles ont à absorber certains types de colorants (neutres, acides ou basiques). **Chacune de ces sous-catégories a sa fonction spécifique dans l'organisme :** les éosinophiles sont par exemple impliqués dans la lutte contre les parasites et dans le mécanisme des allergies. Il est donc **pertinent de chercher à les décrire correctement.**

Ces cellules sont des cellules arrivées à maturité, qui circulent dans le sang et dont la concentration change lorsqu'il y a une pathologie. Avant d'arriver à cette maturité, elles passent par plusieurs stades d'évolution, comme on peut le voir sur la Figure 1 : promyélocyte (PMY), myélocyte (MY) puis métamyélocyte (MMY) et enfin granulocyte de bande (dont les BNE, qui sont des neutrophiles de bande). Cette évolution se produisant de façon continue, il est possible de trouver des cellules “entre deux stades”, ce qui complique la tâche de labelling et de classification.

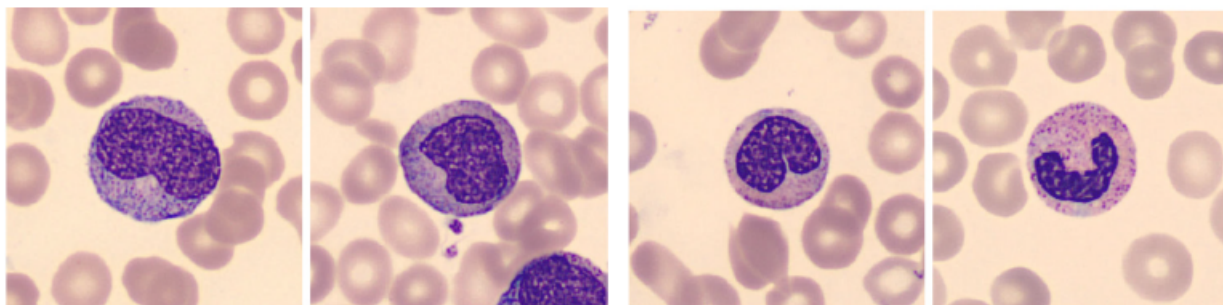


Figure 10 - De gauche à droite, un promyélocyte, deux métamyélocytes et un BNE.

Leur présence dans le sang est révélatrice d'une inflammation ou d'une infection, comme si l'organisme ayant besoin de défenseurs supplémentaires envoyait les "bleus" en cours de formation aider leurs homologues matures. Il est donc **important de distinguer les cellules immatures des cellules matures** (et donc d'avoir une classe "IG") mais il est peut-être possible de se passer de distinguer MY/MMY/PMY/BNE. Ajoutons qu'aucun des datasets à notre disposition ne distingue les éosinophiles et les basophiles immatures des éosinophiles et basophiles matures (peut-être parce que la présence de ces cellules matures dans le sang est déjà rare en comparaison d'autres classes).

Nous avons cependant choisi de distinguer toutes les sous-classes (donc 11 classes au total) : les modèles fondés sur les CNN sont capables de classer correctement dans un nombre bien plus important de classes (cf. le dataset ImageNet), nous préférons donc garder du détail dans notre modélisation. Nous avons aussi décidé d'**afficher les probabilités d'appartenance aux classes** plutôt que la seule classe dominante afin de **fournir une prédiction moins "exclusive"** et de permettre au spécialiste utilisant le modèle de trancher.

LeNet : effet d'une augmentation de données

VGG-16

EfficientNet

VGG-19

Vision Transformer

Nous avons aussi essayé un modèle de vision transformer (ViT)¹⁷. Ce type modèle est apparu comme prometteur du fait de la couche d'attention qui **permet au modèle de se concentrer sur un zone spécifique** de l'image (Fig. 6 et Fig. 7).



Figure 6: La carte d'attention des ViT. (source Keras implementation of ViT (Vision Transformer) - GitHub)

De plus, ces modèles requièrent moins de temps de calcul. Comme les couches d'attention utilisent une opération quadratique (pairwise distance), appliquer cette couche sur une image de 300x300 se révèle coûteux. Une solution est de découper l'image en patches (16x16 ou 32x32, flatten) connectés par un embedding positionnel. De plus, les processus d'attention (self attention) sont dits 'multi-head' et sont exécutés en parallèle. A l'inverse des transformeurs utilisés sur des données textuelles, les ViT n'ont pas de décodeur mais une tête MLP (multilayer perceptron). Nous avons utilisé une librairie disponible sur fond Keras; **ViT-Keras**¹⁸. Les ViT sont recommandés uniquement en transfert learning, puisqu'ils n'ont pas l'équivariance translationnelle des CNNs¹⁹.

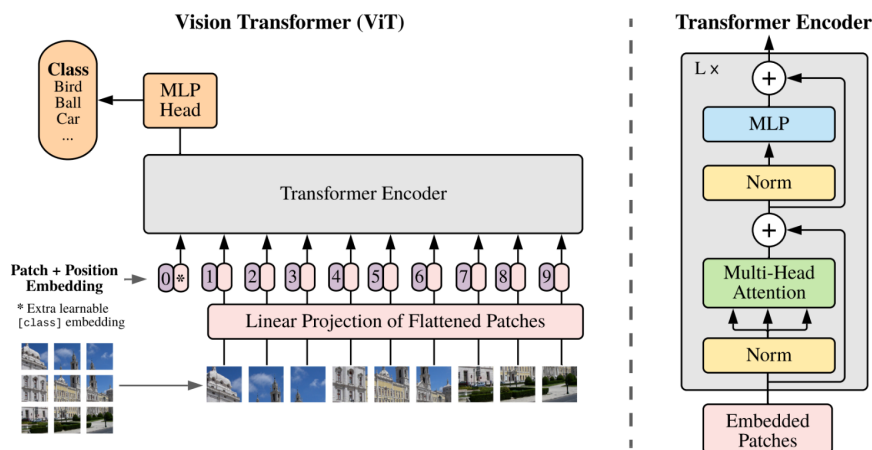


Figure 7: Vision Transformer: architecture. (source An Image is Worth 16x16 Words: Transformers)

¹⁷ "[2010.11929] An Image is Worth 16x16 Words: Transformers - arXiv." <https://arxiv.org/abs/2010.11929>. Accessed 27 Sep. 2021.

¹⁸ "Keras implementation of ViT (Vision Transformer) - GitHub." <https://github.com/faustomorales/vit-keras>. Accessed 27 Sep. 2021.

¹⁹ "How to train your ViT? Data, Augmentation, and Regularization in" 18 Jun. 2021, <https://arxiv.org/pdf/2106.10270>. Accessed 28 Sep. 2021.

Pour nos images, nous avons utilisé une résolution de 352 x 352 (patch 32, ViT-b32) ou 224 x 224 (patch de 16x16, ViT-b16). La librairie ViT-Keras offre la possibilité de fine-tuning avec les code suivant:

```
image_size = 224
model = vit.vit_b32(
    image_size=image_size,
    pretrained=True,
    include_top=False,
    pretrained_top=False)
```

En utilisant ce code avec une couche Dense avec une activation softmax nous avons obtenu une accuracy de 86-88%. Cependant, en inspectant le nombre de paramètres et les couches du ViT nous avons constaté que le modèle était entièrement ré-entraîné sur nos données. Nous avons donc essayé de bloquer les couches du ViT mais l'accuracy a diminué à 60%. En regardant le code source sur GitHub il est difficile d'identifier comment améliorer le code ou quelles sont les raisons de cette baisse.

N'ayant pas encore reçu le training sur PyTorch et du fait de la nouveauté du modèle (essentiellement Kaggle notebooks), nous n'avons pas poursuivi plus en profondeur le tuning du modèle à ce jour. Une piste que nous souhaitons encore explorer est l'augmentation plus intense des données (plus de transformations) comme suggéré dans la référence sur le training des ViT.

De plus, nous avons écarté ce modèle par manque de temps mais une exploration avec des modèles hybrides CNN-ViT ou des architectures plus récentes en utilisant PyTorch pourrait mettre ce type de modèle en avant.

Description des travaux réalisés

Répartition de l'effort sur la durée et dans l'équipe

Le diagramme de Gantt en annexe offre une description détaillée de la répartition des tâches et de l'implication des membres de l'équipe.

Bibliographie

Modèles:

"[2010.11929] An Image is Worth 16x16 Words: Transformers - arXiv."
<https://arxiv.org/abs/2010.11929>. Accessed 27 Sep. 2021.

"Keras implementation of ViT (Vision Transformer) - GitHub."
<https://github.com/faustomorales/vit-keras>. Accessed 27 Sep. 2021.

"How to train your ViT? Data, Augmentation, and Regularization in" 18 Jun. 2021,
<https://arxiv.org/pdf/2106.10270>. Accessed 28 Sep. 2021.

Classification de cellules sanguines :

"White blood cells identification system based on convolutional deep"
<https://www.sciencedirect.com/science/article/pii/S016926071730411X>. Accessed 26 Sep. 2021.

"Deep learning approach to peripheral leukocyte recognition - PLOS." 25 Jun. 2019,
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0218808>. Accessed 26 Sep. 2021.

"Fine-grained leukocyte classification with deep residual learning for"
<https://www.sciencedirect.com/science/article/pii/S0169260718300567>. Accessed 26 Sep. 2021.

"White blood cells detection and classification based on regional"
<https://www.sciencedirect.com/science/article/pii/S0306987719310680>. Accessed 26 Sep. 2021.

"Detection of subtype blood cells using deep learning - ScienceDirect."
<https://www.sciencedirect.com/science/article/pii/S1389041718303760>. Accessed 26 Sep. 2021.

"White blood cell differential count of maturation stages in ... - PLOS." 11 Dec. 2017,
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0189259>. Accessed 26 Sep. 2021.

"Recognition of peripheral blood cell images using convolutional"
<https://www.sciencedirect.com/science/article/pii/S0169260719303578>. Accessed 26 Sep. 2021.

"Human-level recognition of blast cells in acute myeloid leukaemia." 12 Nov. 2019,
<https://www.nature.com/articles/s42256-019-0101-9>. Accessed 26 Sep. 2021.

"Generalizability in White Blood Cells' Classification Problem | bioRxiv." 28 Jul. 2021,
<https://www.biorxiv.org/content/10.1101/2021.05.12.443717v3>. Accessed 26 Sep. 2021.

"Colour deconvolution: stain unmixing in histological imaging - PubMed." 16 Jun. 2021, <https://pubmed.ncbi.nlm.nih.gov/32997742/>. Accessed 26 Sep. 2021.

Datasets:

"A dataset of microscopic peripheral blood cell images for ... - NCBI." 8 Apr. 2020, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7182702/>.

"Fast and robust segmentation of white blood cell images by self" <https://www.sciencedirect.com/science/article/pii/S0968432817303037>. Accessed 18 Sep. 2021.

"Raabin-WBC: a large free access dataset of white blood ... - bioRxiv." 29 May. 2021, <https://www.biorxiv.org/content/10.1101/2021.05.02.442287v4>. Accessed 26 Sep. 2021.

"A Single-cell Morphological Dataset of Leukocytes from AML" 17 Feb. 2021, <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=61080958>. Accessed 26 Sep. 2021.

"Deep Learning for Distinguishing Morphological Features of Acute" 5 Nov. 2020, <https://ashpublications.org/blood/article/136/Supplement%201/10/472209/Deep-Learning-for-Distinguishing-Morphological>.

Difficultés rencontrées lors du projet

Le principal verrou scientifique rencontré lors du projet reste la qualité et la disponibilité des **données** :

- **les datasets sont déséquilibrés** : certains types de cellules comme les éosinophiles (EO), les neutrophiles (surtout SNE) sont sur-représentés par rapport à d'autres catégories plus difficiles à trouver²⁰ et que le modèle a du mal à classer (PMY, MMY, MY, en faible nombre et difficiles à reconnaître à l'œil nu, pour un observateur peu expérimenté ; il existe de plus une controverse scientifique sur la pertinence de distinguer ou non ces catégories). **Ce déséquilibre s'explique cependant par des raisons biologiques : certains types de cellules, comme par exemple les SNE, sont présentes en quantités plus importantes que d'autres dans le sang. La présence accrue de cellules immatures (comme les BNE ou les PMY/MMY/MY) peut être considérée comme révélatrice d'infections ou d'inflammations, ce qui explique la faible quantité de ces cellules chez des patients sains ;**
- **le problème des labels** : nous avons des doutes quant aux labels de certaines images (notamment lorsqu'il s'agit des SNE des BNE, qui sont en fait la même cellule mais observée à deux instants différents de sa vie). Ces doutes sont nourris par les résultats de nos modèles, mais aussi par ce que nous avons pu lire dans la littérature scientifique. On citera notamment la Table n°4 de l'article de recherche présentant le dataset Raabin, où les auteurs montrent les résultats de l'identification des images par deux experts indépendants ne s'étant pas consultés²¹ : à titre d'exemple, près de 30% des images classées comme BNE par un expert ont été

²⁰ On peut citer les PMY, MMY, MY qui ne sont pas présents dans tous les datasets, ainsi que les BNE, eux aussi en faible nombre et parfois regroupés avec les SNE sous l'appellation générale "Neutrophile").

²¹ "Raabin-WBC: a large free access dataset of white blood ... - bioRxiv." 29 May. 2021, <https://www.biorxiv.org/content/10.1101/2021.05.02.442287v4>. Accessed 29 Sep. 2021.

classées comme SNE (le terme exact est Neutrophil dans l'article) par l'autre. Plusieurs milliers d'images ont été classées comme "Not labelled" ou "Not recognized", attestant de la difficulté de l'exercice pour un observateur humain, même chevronné ;

- **finalement, assez peu de sources de données différentes**, ce qui pose de gros **problèmes en termes de généralisation**. Beaucoup de travaux dans le domaine de la reconnaissance des cellules sanguines par du deep-learning mettent en avant des modèles obtenant des performances excellentes, mais utilisent des données provenant d'une seule origine. Lorsque nous avons testé la première version du modèle VGG19, entraîné sur le dataset Barcelona Normal, nous obtenions une accuracy de plus de 95% pour la plupart des classes de cellules (notamment 99,8% sur les EO). Lorsque ce même modèle a été testé sur les données du dataset Raabin, l'accuracy globale du modèle a littéralement plongé (passage de 92,6% à 47,4%, en tombant à 37,7% pour les seuls EO de Raabin).

Par souci d'éviter de construire un modèle trop spécialisé, nous avons cherché à compléter notre dataset. Cela nous a conduit à affronter des difficultés d'ordre prévisionnel :

- Le dataset AML_APL_John-Hopkins a été exploré mais n'a pas été utilisé à cause d'un **mauvais labelling**.
- **D'autres datasets ne proposaient que des images ayant subi une procédure de segmentation** : l'image avait été découpée autour des cellules, pour en éliminer le fond contenant les globules rouges. Il nous a paru peu pertinent de mélanger ces images avec celles dont nous disposions et pour lesquelles les globules rouges étaient présents. De plus et comme nous l'avons souligné dans l'introduction de ce document, bien que les systèmes de reconnaissance existants des globules blancs utilisent la segmentation, **ce type de méthode est difficile à généraliser** : variabilité des colorations utilisées pour traiter les échantillons de sang, variabilité des systèmes d'acquisition. L'utilisation d'un modèle reposant sur des CNN permet potentiellement de contourner ce problème, pourvu que les données disponibles soient suffisamment diverses. Nous avons cependant **essayé de mesurer l'impact de la segmentation sur les performances d'un de nos modèles**.
- Le **téléchargement des datasets** d'une part, puis leur **transfert sur Google Drive** d'autre part, a également posé plusieurs problèmes : des problèmes de temps, les durées de téléchargement ayant été parfois très longues (jusqu'à plusieurs jours pour les transferts sur Drive), des problèmes de connexion internet, des problèmes de capacité de stockage sur Drive.

Notre volonté d'explorer des sentiers parfois non balisés par la formation nous a conduit à prendre du temps pour acquérir un certain nombre de compétences, via la documentation des librairies utilisées et grâce à des cours et tutoriaux trouvés sur internet. On citera notamment :

- **la familiarisation avec Google Colab**, pour le transfert des données depuis le Drive et l'utilisation du calcul sur GPU. Il aurait bien d'avoir une machine virtuelle à partager pour le cloud computing;
- **la découverte et l'implémentation de CAM / Grad-CAM** pour analyser le comportement du modèle, élément absent de la formation;
- l'entraînement du **ViT** a été plus difficile qu'envisagé au premier abord et a demandé beaucoup de recherches bibliographiques. De plus, le manque de connaissances sur les **transformeurs et PyTorch** a ralenti le développement de ce modèle. Avec le recul, nous pensons qu'**il aurait été pratique d'avoir eu accès au module 152 sur les CNN** avec PyTorch, en plus des 151 et 155 qui nous ont été proposés.

Bilan & Suite du projet

En utilisant une approche combinatoire avec des sources variées de données, nous avons pu obtenir un modèle capable de classer plus de 11 classes de leucocytes circulants ou immatures. Ce type de modèle pourrait être utilisé comme base pour un fine-tuning et une classification d'autres types cellulaires ou de cellules leucémiques.

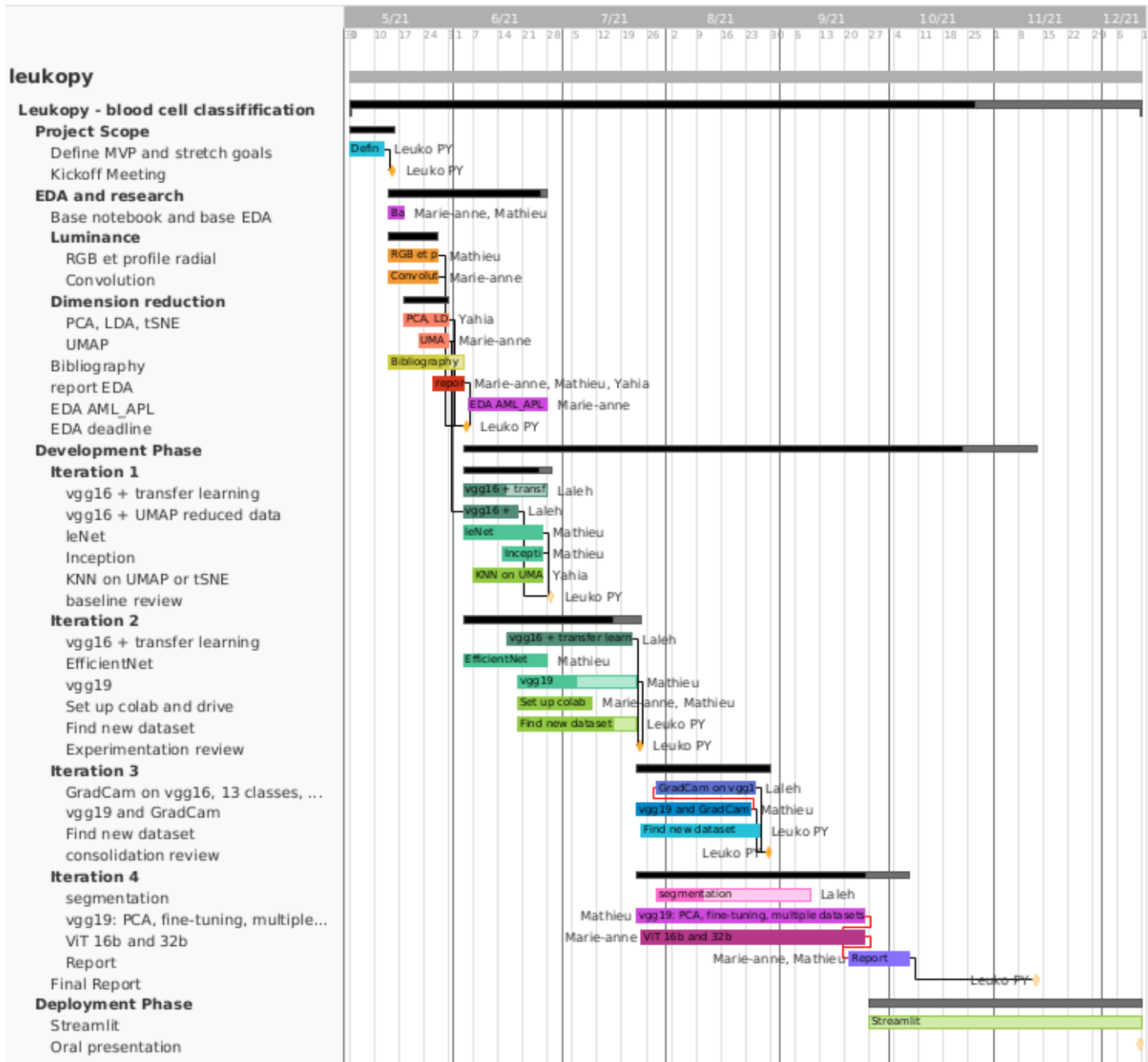
Notre objectif principal de classer les cellules saines circulantes a été atteint et nous avons pu développer une analyse poussée de l'interprétabilité de nos modèles. Nous sommes confiants quant à notre capacité à construire un outil utile et accessible pour les experts pathologistes, leur permettant d'identifier les cellules sur des frottis sanguins plus efficacement et avec plus de confiance.

De nombreuses améliorations sont cependant encore possibles et souhaitables. Voici plusieurs pistes qu'il reste à explorer ou à approfondir :

- **Diversifier davantage la provenance des images du dataset**, dans l'optique de construire un modèle généraliste, susceptible de fournir des prédictions correctes dans des hôpitaux et laboratoires différents. **Développer l'augmentation de données** peut aussi contribuer à cet objectif.
- L'utilisation de **GANs** serait un moyen de générer de nouvelles images, notamment pour les classes de cellules saines en faible effectif.
- **S'assurer de la qualité des labels des images utilisées**, par exemple à travers un partenariat avec des spécialistes qui évalueraient indépendamment chaque image, dans le but de ne conserver que celles faisant consensus.
- Avec davantage de temps et de moyens, il serait possible d'abandonner le transfer-learning et de construire un modèle depuis le départ. On pourrait avoir recours à des méthodes d'ensemble à partir des features extraites par le réseau neuronal pour améliorer les performances.

Annexes

Diagramme de Gantt



Description des fichiers de code

notebooks

- └─ leukopy_lib.py: helper functions
- └─ cleaning
 - └─ generate_dataframes.ipynb: notebook pour generer les datasets de training, validation et test.
 - └─ Importer_PBC_AML_APL.ipynb: nettoyage et EDA rapide du dataset AML_APL (exclu)
 - └─ Importer_PBC_normal.ipynb: nettoyage et EDA rapide du dataset de barcelone
- └─ EDA
 - └─ First_EDA_Barcelona.ipynb: premières explorations de dataset de barcelone
 - └─ Dimensionality reduction (PCA, umap...).ipynb:
 - └─ EDA_deconv_hist.ipynb; deconvolution
 - └─ EDA_mean_color.ipynb: exploration des niveaux de RGB et profil radial
 - └─ EDA_mean_gray.ipynb: exploration des niveaux de gris
 - └─ EDA_mean_gray_UMAP.ipynb: to clean?
 - └─ EDA_PCA_LDA_TSNE_V3-3.ipynb: to clean?
 - └─ EDA_PCA_LDA_TSNE_V3.ipynb: to clean?
 - └─ EDA_PCA_V1.ipynb: to clean?
 - └─ EDA_PCA_V2.ipynb: to clean?
- └─ model
 - └─ Dimension_red_KNN.ipynb: test de dimension de reduction et KNN en classifieur
 - └─ efficientNet
 - └─ EfficientNet_subclasses.ipynb
 - └─ EfficientNet_subclasses_test_tf_dataset.ipynb
 - └─ EffNetB0_model.ipynb
 - └─ EffNetB3_GPU.ipynb
 - └─ lenet
 - └─ leuko_lenet_gc.ipynb
 - └─ vgg16
 - └─ Models_151_Transfer_Learning_class_rep.ipynb
 - └─ Models_151_Transfer_Learning_LR.ipynb
 - └─ Models2_151_Transfer_Learning_LR.ipynb
 - └─ Model_Transfer_Learning_13classes.ipynb
 - └─ Model_Transfer_Learning_9classes.ipynb
 - └─ Model_Transfer_Learning_9classes_Suit2.ipynb
 - └─ Model_Transfer_Learning_9classes_Suite.ipynb
 - └─ vgg19
 - └─ VGG19_TL_11.ipynb
 - └─ vgg19_tl_11.py
 - └─ ViT
 - └─ Vit_b16.ipynb
 - └─ ViT_b32.ipynb