Rapport IM05

Arthur Evain

19 avril 2026

Table des matières

1 Introduction **1**
- 1.1 Contexte du projet . . . 1
- 1.2 Vue d'ensemble de l'approche . . . 1
- 1.3 Problématiques identifiées et enjeux associés . . . 1
  - 1.3.1 Déséquilibre de classes . . . 1
  - 1.3.2 Qualité de l'acquisition et bruitage . . . 2
  - 1.3.3 Proximité morphologique et similitudes inter-classes . . . 2
  - 1.3.4 Enjeux de segmentation . . . 3

2 État de l'Art et Travaux Connexes **3**
- 2.1 Morphométrie classique : fondations et limites . . . 3
- 2.2 Complexité cytoplasmique et analyse texturale . . . 3
- 2.3 L'ère du deep learning : déséquilibre et maturation . . . 4

3 Méthodologie de Segmentation et Prétraitement **4**
- 3.1 Exploration du débruitage : algorithmes et limites . . . 4
  - 3.1.1 Débruitage via NL-Means et filtre bilatéral . . . 5
  - 3.1.2 Analyse des résultats et décision d'exclusion . . . 5
- 3.2 Stratégies de segmentation hybride : K-Means vs Chan-Vese . . . 5
  - 3.2.1 Clustering K-Means dans l'espace L*a*b* . . . 5
  - 3.2.2 Contours actifs (modèle de Chan-Vese) . . . 6
- 3.3 Validation géométrique et topologique . . . 6
- 3.4 Indices de confiance et score de fiabilité . . . 6
  - 3.4.1 Rapport signal-sur-bruit (SNR) de contraste . . . 6
- 3.5 Analyse exploratoire : séparabilité par PCA . . . 7

4 Apprentissage Profond et Entraînement du Modèle **8**
- 4.1 Architecture : ResNet50 et régularisation . . . 8
- 4.2 Prétraitement dynamique et augmentation de données . . . 9
  - 4.2.1 Intégration du CLAHE . . . 9
  - 4.2.2 Augmentations géométriques et colorimétriques . . . 9
- 4.3 Apprentissage progressif : Curriculum Learning . . . 9
- 4.4 Évaluation robuste : Stratified K-Fold . . . 9
- 4.5 Recherche d'hyperparamètres . . . 10

5 Résultats et Analyse des Performances **10**
- 5.1 Métriques globales . . . 10
- 5.2 Analyse de la matrice de confusion . . . 11
  - 5.2.1 Succès de classification . . . 11
  - 5.2.2 Confusions persistantes . . . 12
- 5.3 Analyse du compromis performance/rejet par seuil de confiance . . . 12
  - 5.3.1 Formalisation . . . 12
  - 5.3.2 Résultats . . . 13
- 5.4 Analyse de robustesse : impact du bruit d'acquisition . . . 13

6 Post-traitement Hybride : Correction Géométrique par PCA **14**
6.1 Détection des conflits et modélisation de la normalité 14
6.2 Rattrapage géométrique et inversion de label 14

7 Conclusion **15**

1 Introduction

1.1 Contexte du projet

L'analyse cytologique des frottis sanguins constitue le pilier du diagnostic hématologique moderne. L'examen consiste à identifier et à quantifier les différentes populations de leucocytes (globules blancs) afin de détecter d'éventuelles anomalies pathologiques. Historiquement, cette tâche repose sur l'expertise de biologistes qualifiés effectuant une lecture manuelle au microscope.

Cependant, la numérisation croissante des services de santé et l'augmentation du volume d'analyses quotidiennes imposent une transition vers l'automatisation. Ce projet s'inscrit dans cette dynamique en visant le développement d'un système de classification automatique capable de traiter des images de cellules isolées parmi 13 types leucocytaires distincts. L'enjeu est de garantir une précision diagnostique équivalente à celle d'un expert humain tout en réduisant les délais de traitement.

1.2 Vue d'ensemble de l'approche

La démarche adoptée est structurée en trois phases complémentaires, comme illustré en Figure 1.

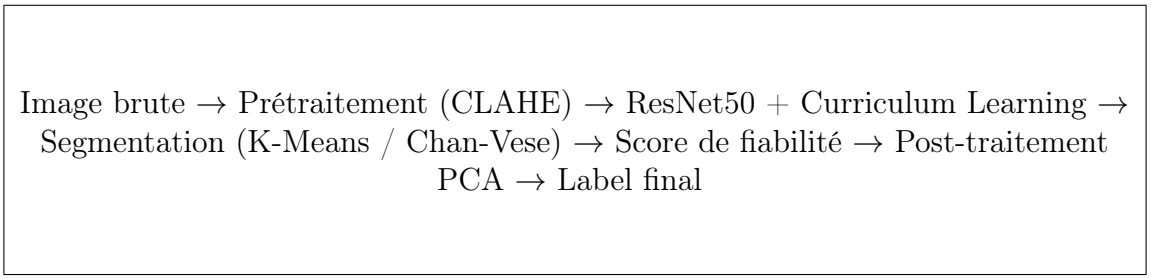

FIGURE 1 – Vue d'ensemble du pipeline hybride de classification.

Le code est organisé en 4 modules : le prétraitement et l'exploration (segmentation.ipynb), l'entraînement du modèle (train.py), le post-traitement (postprocess.py) et l'évaluation des résultats (eval.py).

1.3 Problématiques identifiées et enjeux associés

Le passage de l'image microscopique brute à une classification automatisée fiable se heurte à plusieurs difficultés structurelles propres au domaine de l'hématologie numérique.

1.3.1 Déséquilibre de classes

La distribution des leucocytes dans le sang humain est intrinsèquement asymétrique. Les classes majoritaires (Neutrophiles et Lymphocytes) surreprésentent le dataset, tandis que les classes pathologiques ou rares (comme les Prolymphocytes - PLY ou les Proplasmocytes - PC) ne disposent que de très peu d'exemples. Cette disparité crée un biais majeur : un modèle non contrôlé risque de négliger les classes minoritaires pour maximiser son taux de réussite global sur les classes fréquentes, ce qui est inacceptable dans un contexte médical où les cellules rares sont souvent les plus diagnostiquement significatives.

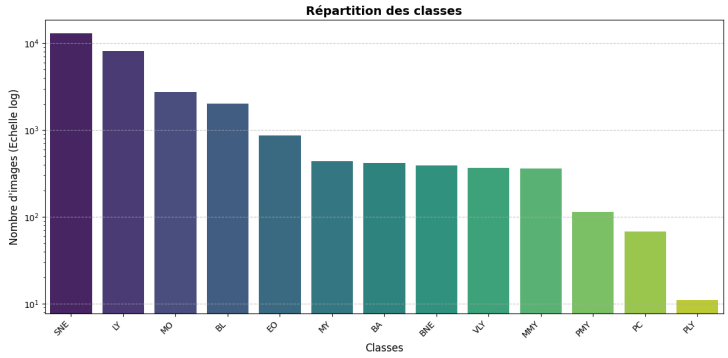

FIGURE 2 – Répartition des classes dans le dataset d'entraînement.

1.3.2 Qualité de l'acquisition et bruitage

Les images proviennent de conditions d'acquisition variées. Le bruit de capteur, les poussières optiques, le flou de mouvement et les variations d'éclairage introduisent des artefacts granuleux. Ces perturbations numériques peuvent être confondues avec des structures biologiques réelles, telles que la granulation du cytoplasme ou la texture de la chromatine, rendant complexe la séparation entre signal biologique et bruit électronique. La granularité étant déjà beaucoup critiquée dans les papiers, cela m'a convaincu de ne pas m'y intéresser.

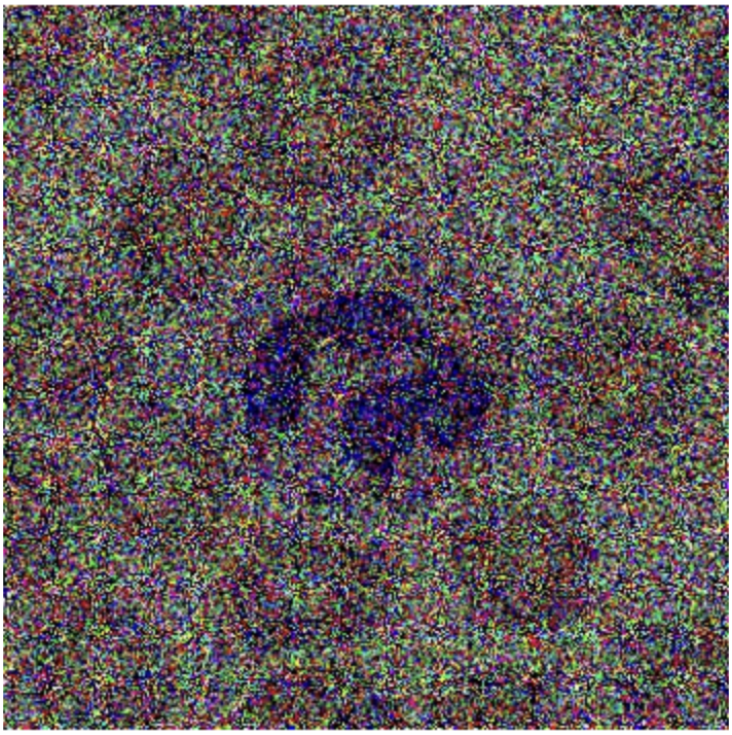

FIGURE 3 – Exemple d'image fortement bruitée dans le dataset.

1.3.3 Proximité morphologique et similitudes inter-classes

Contrairement à la classification d'objets du quotidien, les classes de leucocytes partagent un alphabet morphologique commun. La distinction entre un monocyte et un

grand lymphocyte, par exemple, repose sur des nuances de texture nucléaire ou des variations subtiles du rapport nucléo-cytoplasmique. Cette proximité augmente le risque de confusion, particulièrement pour les cellules en cours de maturation dont les frontières biologiques sont continues plutôt que discrètes.

1.3.4 Enjeux de segmentation

L'isolation précise de la cellule est un prérequis indispensable. La présence d'hématies (globules rouges) contiguës, de débris de coloration ou de cellules agrégées complique l'extraction d'un masque propre. Une segmentation imparfaite fausserait les caractéristiques géométriques et texturales transmises au classifieur final.

2 État de l'Art et Travaux Connexes

La classification automatisée des leucocytes est un domaine de recherche historiquement riche, marqué par une transition des méthodes morphométriques vers l'apprentissage profond. L'analyse de la littérature a été déterminante pour structurer ma démarche et m'a permis d'identifier les limites des approches isolées.

2.1 Morphométrie classique : fondations et limites

Historiquement, la reconnaissance des leucocytes reposait sur l'extraction de caractéristiques géométriques (« hand-crafted features »). Dès l'année 2000, des études ont montré qu'il était possible de discriminer les globules blancs en comptant le nombre de lobes du noyau (souvent 2,7 en moyenne pour les neutrophiles contre 1 pour les lymphocytes) et en extrayant des ratios d'aires stricts (noyau/cellule) [1]. Une décennie plus tard, une analyse statistique par test ANOVA a prouvé mathématiquement ($p < 0{,}001$) que des descripteurs tels que la compacité, l'excentricité et le ratio nucléo-cytoplasmique discriminaient significativement les classes [3].

Cependant, se baser exclusivement sur la géométrie du noyau présente des failles. Une étude de 2014 utilisant le seuillage d'Otsu a certes obtenu 100 % de précision sur les lymphocytes, mais s'est effondrée à 50 % sur les monocytes en raison de leur morphologie nucléaire extrêmement variable [2]. Ce résultat confirme qu'une approche purement géométrique est insuffisante pour les classes à forte variabilité — ce qui motive l'utilisation d'un réseau de neurones convolutif capable d'extraire des descripteurs de haut niveau.

Ces travaux me conduisent à intégrer les caractéristiques géométriques (aire, périmètre, ratio N/C, circularité) non pas comme classifieur principal, mais comme « filet de sécurité » de post-traitement. Ils confirment également la nécessité d'exclure les leucocytes touchant les bords de l'image, qui faussent ces calculs d'aires.

2.2 Complexité cytoplasmique et analyse texturale

L'isolation précise du cytoplasme est un défi majeur, ses limites étant souvent floues ou incolores. Des travaux de 2011 ont proposé d'utiliser des algorithmes de contours actifs (*Snakes*) initiés par une dilatation du noyau [4]. Ces mêmes travaux ont mis en évidence la supériorité des descripteurs de texture *Local Binary Patterns* (LBP), plus rapides à calculer que les matrices de co-occurrence (GLCM) et invariants à la rotation. En 2022, il a été démontré que la combinaison de ces descripteurs avec des caractéristiques

chromatiques (RGB) crée un espace hybride séparant nettement mieux les classes que la géométrie seule [5].

Ces résultats motivent directement deux choix de ma pipeline : la segmentation noyau/cytoplasme comme étape centrale (Section 3), et l'intégration d'une étape de normalisation du contraste (CLAHE) pour stabiliser les caractéristiques chromatiques avant extraction.

2.3 L'ère du deep learning : déséquilibre et maturation

En 2020, une étude a prouvé la puissance d'une approche hybride : l'extraction de « Deep Features » via ResNet-50 et GoogLeNet couplées à un classifieur discriminant (QDA), atteignant des résultats quasi parfaits sur les lymphocytes et monocytes [6].

Néanmoins, l'utilisation de réseaux de neurones se heurte à deux problèmes majeurs :

- **Le déséquilibre des données :** La répartition naturelle est très asymétrique (jusqu'à 45 fois moins de basophiles que de lymphocytes). L'étude de 2022 a montré qu'en rééquilibrant les données via SMOTE, un modèle classique comme le *Random Forest* pouvait parfois surpasser un réseau de neurones victime de sur-apprentissage [5]. J'ai testé cette approche : sur ce dataset, SMOTE introduit des exemples synthétiques trop proches des frontières de décision pour des classes morphologiquement continues (notamment BNE/SNE), ce qui dégrade les performances. Le ResNet sans sur-échantillonnage s'est avéré supérieur, ce qui suggère que la pondération des classes dans la fonction de perte est une stratégie plus adaptée ici que la génération de données synthétiques.
- **Le problème de maturation :** Les leucocytes évoluent d'un stade « blaste » primitif à une forme mature. Cette continuité biologique rend la définition de frontières discrètes très complexe, induisant une confusion inévitable entre des stades contigus (par exemple, entre un neutrophile non segmenté « BNE » et un segmenté « SNE »).

Ces constats orientent ma stratégie vers trois axes : (1) transfer learning via ResNet avec pondération de classes, (2) suivi de performance par classe via le F1-score macro, et (3) un post-traitement géométrique pour corriger les confusions liées à la maturation.

3 Méthodologie de Segmentation et Prétraitement

La précision de la classification des leucocytes repose sur la qualité de l'isolation des structures cellulaires (noyau et cytoplasme). Cette section détaille le pipeline de segmentation hybride, qui sert à deux fins : me familiariser avec le dataset et extraire des caractéristiques morphologiques servant au post-traitement. Le tout peut être retrouvé dans le fichier segmentation.ipynb.

3.1 Exploration du débruitage : algorithmes et limites

L'acquisition d'images microscopiques est intrinsèquement liée à la présence de bruit électronique et de grain comme vu précédemment, qui peuvent potentiellement altérer la perception des textures biologiques fines. Dans l'optique d'optimiser les performances de ma segmentation et de ma classification, j'ai mené une phase d'expérimentation dédiée à la réduction du bruit.

3.1.1 Débruitage via NL-Means et filtre bilatéral

J'ai d'abord implémenté l'algorithme *Non-Local Means Denoising* (`fastNlMeansDenoisingColored`) Contrairement aux filtres de lissage classiques, cet algorithme recherche des patchs d'image similaires pour moyenner le bruit, ce qui est théoriquement plus respectueux des structures répétitives. Pour renforcer cette approche, j'ai couplé ce traitement à un filtre bilatéral (`bilateralFilter`). Ce dernier a pour propriété de lisser les zones de faible gradient tout en préservant les arêtes nettes, ce qui semblait idéal pour détacher la membrane cytoplasmique du fond de l'image.

3.1.2 Analyse des résultats et décision d'exclusion

Cependant, l'analyse comparative des performances a révélé que ce prétraitement était contre-productif pour mon application spécifique. Bien que les images résultantes paraissaient visuellement plus nettes, le lissage induit a provoqué une disparition des micro-textures haute fréquence.

En cytologie, ces détails granulaires (comme les granulations spécifiques des éosinophiles ou la finesse de la chromatine nucléaire) sont des descripteurs essentiels. En uniformisant ces zones, le débruitage a réduit la variance inter-classes et a dégradé la précision de mon modèle ResNet50 lors de mes tests préliminaires.

Constatant que cette étape n'apportait aucune amélioration statistique aux résultats finaux et qu'elle augmentait significativement le temps de calcul par image, j'ai décidé de ne pas intégrer le débruitage dans ma pipeline finale. J'ai préféré laisser au réseau de neurones la capacité d'apprendre à filtrer le bruit de manière endogène à travers ses propres couches de convolution.

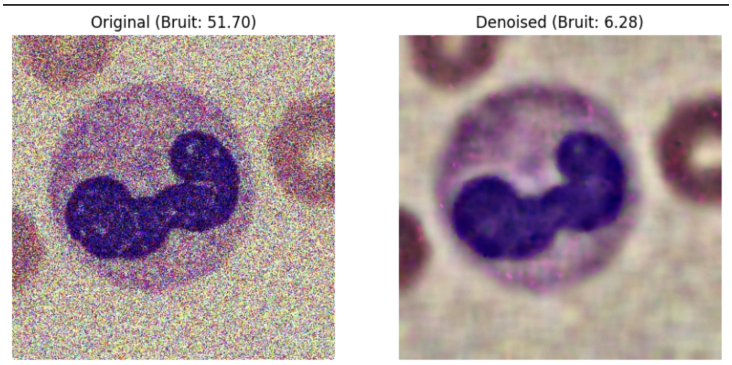

FIGURE 4 – Réduction du bruit sur une image

3.2 Stratégies de segmentation hybride : K-Means vs Chan-Vese

Pour garantir la robustesse face à la diversité morphologique des leucocytes, j'ai mis en place une stratégie de compétition entre deux approches fondamentales.

3.2.1 Clustering K-Means dans l'espace L*a*b*

Le K-Means est appliqué sur les composantes de couleur après conversion dans l'espace CIELAB. Cette décomposition permet d'isoler la luminance (L^*) de l'information chromatique (a^*, b^*), facilitant la distinction entre le noyau basophile (violet dense) et le

cytoplasme. L'algorithme segmente l'image en quatre classes triées par luminance, pour identifier de manière déterministe le compartiment nucléaire.

3.2.2 Contours actifs (modèle de Chan-Vese)

En parallèle, le modèle de Chan-Vese est utilisé pour les cellules dont les limites sont floues ou le contraste faible. Cette approche de segmentation sans contours, basée sur l'évolution de courbes de niveau (*level sets*), est particulièrement adaptée aux morphologies atypiques (Proplasmocytes, Prolymphocytes). Une itération sur une plage de paramètres (μ, radius) permet de capturer la géométrie la plus cohérente de la cellule.

3.3 Validation géométrique et topologique

Une fois les candidats extraits, une série de filtres biologiques est appliquée :

— **Inclusion stricte :** Vérification que les composantes du noyau sont intégralement contenues dans le cytoplasme via le test du polygone (*Point-in-Polygon*).

— **Circularité et convexité :** Calcul de la circularité sur l'enveloppe convexe pour rejeter les débris ou les agrégats cellulaires (hématies collées).

— **Contrainte homothétique :** Utilisation d'une « Safe Zone » générée par homothétie du noyau pour valider l'étendue du cytoplasme.

3.4 Indices de confiance et score de fiabilité

Pour garantir l'intégrité des données transmises au classifieur, j'ai implémenté un système de décision à trois niveaux basé sur des seuils de fiabilité (*Reliability Scores*).

Le score de fiabilité $\mathcal{R}$ croise la régularité géométrique (circularité) et la stabilité texturale (écarts-types locaux) :

— **Niveau 2 — Segmentation sûre :** Circularité élevée ($> 0{,}88$) et faible variance interne ($\sigma_{nuc} < 20$, $\sigma_{cyto} < 25$). Ces images constituent le « noyau dur » du jeu d'entraînement pour le post-traitement PCA.

— **Niveau 0 — Échec de segmentation :** Circularité dégradée ($< 0{,}8$) ou bruit aberrant ($\sigma_{nuc} > 35$ ou $\sigma_{cyto} > 45$). Ces masques sont rejetés car ils incluent des débris ou ont « fuité » sur le fond.

— **Niveau 1 — Acceptable :** Cas intermédiaires conservés mais signalés.

En pratique, environ 50 % des images obtiennent une segmentation de niveau 2. Cette limitation est importante : le post-traitement PCA (Section 6) ne peut s'appliquer qu'aux images pour lesquelles une segmentation fiable est disponible, ce qui constitue une borne supérieure sur son impact.

3.4.1 Rapport signal-sur-bruit (SNR) de contraste

Le SNR mesure la clarté de la différenciation entre chromatine nucléaire et cytoplasme :

$$SNR_{contrast} = \frac{|\mu_{nuc} - \mu_{cyto}|}{\sigma_{nuc} + \sigma_{cyto} + \epsilon} \tag{1}$$

Ce SNR est utilisé comme métrique de tri pour le Curriculum Learning : les images à SNR élevé correspondent aux exemples « faciles » présentés en premier au réseau.

3.5 Analyse exploratoire : séparabilité par PCA

Pour valider la pertinence des descripteurs morphologiques extraits, une Analyse en Composantes Principales (PCA) a été menée sur 100 images de niveau 2, confrontant les Lymphocytes (LY) et les Monocytes (MO) — les deux classes présentant les contrastes morphologiques les plus marqués.

Une standardisation Z-score préalable est indispensable pour que la PCA maximise la variance réelle des données et non celle induite par les unités de mesure :

$$z = \frac{x - \mu}{\sigma} \tag{2}$$

Les deux premières composantes principales expliquent environ 73 % de la variance totale et montrent une séparation linéairement séparable des deux populations :

- **PC1** (54 % de variance) est fortement corrélée à la taille cellulaire, isolant les Monocytes (grand diamètre) côté positif, et inversement proportionnelle au ratio noyau/cytoplasme, ce qui classe les Lymphocytes côté négatif.
- **PC2** capture les variations de densité chromatique, distinguant la chromatine dense des lymphocytes.

Variable morphologique	**Chargement sur PC1**
`cell_perimeter`	+0,500
`cell_area`	+0,496
`nucleus_area`	+0,485
`cell_mean_intensity`	+0,213
`cell_eccentricity`	+0,027
`nc_ratio`	−0,329
`nucleus_solidity`	−0,338

TABLE 1 – Coefficients de chargement des caractéristiques cellulaires sur la première composante principale (PC1).

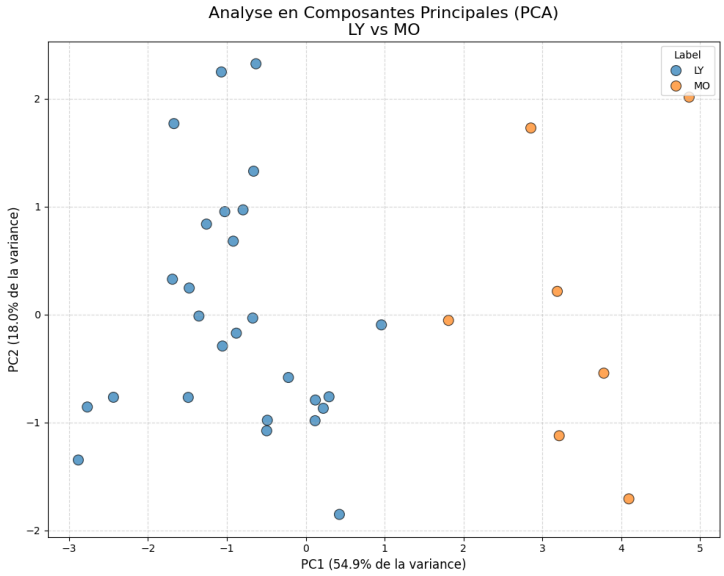

FIGURE 5 – Projection PCA (LY vs MO) sur les segmentations de niveau 2 : séparables linéairement

Ces résultats confirment la fiabilité du pipeline de segmentation pour les classes « pôles ». La persistance de chevauchements pour les formes immatures (notamment BNE/SNE) justifie le recours à un modèle ResNet50 : si la PCA suffit pour les cas simples, le réseau convolutif est nécessaire pour capturer les descripteurs de haut niveau requis par les classes morphologiquement proches.

4 Apprentissage Profond et Entraînement du Modèle

La section précédente a montré que la segmentation échoue dans environ la moitié des cas, et que les caractéristiques manuelles seules sont insuffisantes pour les classes proches. La classification finale est donc déléguée à un réseau de neurones convolutif (CNN), conformément aux approches les plus performantes de la littérature [6]. Le but est que celui-ci apprenne des représentations visuelles de haut niveau et maximise le F1-score macro sur les 13 classes.

4.1 Architecture : ResNet50 et régularisation

L'architecture **ResNet50** a été sélectionnée pour ses connexions résiduelles (*skip connections*), qui permettent d'entraîner des réseaux très profonds sans subir le phénomène de disparition du gradient (*vanishing gradient*). Ce mécanisme est crucial pour extraire des caractéristiques texturales complexes comme la condensation de la chromatine ou la granularité cytoplasmique.

Le modèle est initialisé avec des poids pré-entraînés sur ImageNet (*Transfer Learning*), conformément aux contraintes du projet. La couche de classification finale est remplacée pour correspondre aux 13 classes :

```
model.fc = nn.Linear(model.fc.in_features, num_classes)
```

L'ajout de couches de Dropout dans le classifieur final a été évalué lors de la recherche d'hyperparamètres, pour forcer le réseau à ne pas se reposer sur un sous-ensemble restreint de caractéristiques morphologiques.

Résultat de référence : Ce modèle seul (sans Curriculum Learning ni post-traitement) atteint un F1-score macro de **0,67** sur la validation croisée, ce qui constitue la ligne de base à partir de laquelle chaque amélioration est mesurée.

4.2 Prétraitement dynamique et augmentation de données

4.2.1 Intégration du CLAHE

L'amélioration locale du contraste (CLAHE) est encapsulée dans une classe `CLAHE_Transform` personnalisée, appliquée à chaque image au chargement par le *DataLoader* (espace LAB, `clip_limit=2.0`). Cela stabilise les gradients lors des premières itérations en normalisant les variations d'illumination entre images.

La normalisation de Macenko, également très citée dans la littérature, a été testée. Sur ce dataset, elle produit des artefacts de couleur sur les images dont la préparation de lame s'écarte significativement du frottis de référence. Le CLAHE, opérant localement sur la luminance sans supposer de frottis de référence stable, s'est révélé plus robuste à cette variabilité inter-lots.

4.2.2 Augmentations géométriques et colorimétriques

- **Invariances spatiales :** Flips horizontaux et verticaux, rotations aléatoires ($\pm 20°$), simulant l'orientation arbitraire des cellules sur le frottis.
- **Color Jitter :** Perturbation de la luminosité et du contraste ($\pm 10\,\%$), rendant le modèle résilient aux variations de coloration des réactifs (May-Grünwald-Giemsa).

4.3 Apprentissage progressif : Curriculum Learning

Le **Curriculum Learning** consiste à présenter initialement au modèle les exemples les plus simples, avant d'introduire progressivement la complexité [**?**]. En exploitant la métrique de bruit `noise_level`, le dataset est filtré dynamiquement :

1. **Époques 1 à 5 :** Cellules nettes uniquement (`noise_level` ≤ 20, $n = 16\,451$ images). Le modèle converge rapidement vers les formes géométriques fondamentales.
2. **Époques 6 à 10 :** Introduction des cellules modérément bruitées (`noise_level` ≤ 40).
3. **Époques 11+ :** Intégralité du dataset (sauf les images aberrantes avec `noise_level` > 100).

4.4 Évaluation robuste : Stratified K-Fold

Face au fort déséquilibre de classes, une simple division entraînement/validation risquerait de produire des ensembles de validation dépourvus de classes minoritaires — comme illustré ci-après par la classe PLY absente du fold 1 (Figure 6). La validation croisée à 5 plis stratifiée (**Stratified K-Fold**, $k = 5$) garantit que la proportion de chaque type de leucocyte est scrupuleusement respectée dans chaque pli.

La métrique retenue pour la sélection du meilleur modèle est le **F1-Score Macro**, qui accorde un poids équitable aux classes rares. Les résultats agrégés sur l'ensemble des 5 folds sont présentés dans la section suivante.

4.5 Recherche d'hyperparamètres

Le script d'entraînement est paramétrable en ligne de commande via `argparse`. Les expériences automatisées ont fait varier le taux d'apprentissage de l'optimiseur Adam (autour de 2×10^{-4}) et comparé les performances avec ou sans Curriculum Learning.

5 Résultats et Analyse des Performances

5.1 Métriques globales

Configuration	**F1-macro (val. croisée)**
ResNet50 (baseline)	0,67
+ CLAHE	0.70
+ Curriculum Learning	0.71
+ Post-traitement PCA	0.75
Modèle final	0.76

TABLE 2 – Table d'ablation : contribution de chaque composant du pipeline au F1-score macro. Valeurs de validation croisée ($k = 5$, moyenne sur les 5 folds).

Ces résultats correspondent globalement à ce qui était attendu : le transfer learning depuis ImageNet apporte une base solide (0,67), et chaque composant apporte un gain marginal positif. La principale surprise a été l'inefficacité de SMOTE et de la normalisation Macenko, détaillée dans les sections précédentes, qui aurait pu être prédite *a priori* en considérant la variabilité inter-lots du dataset.

La validation croisée à 5 folds donne un F1-macro moyen de 0.73 ± 0.03 (écart-type), confirmant la stabilité du modèle.

5.2 Analyse de la matrice de confusion

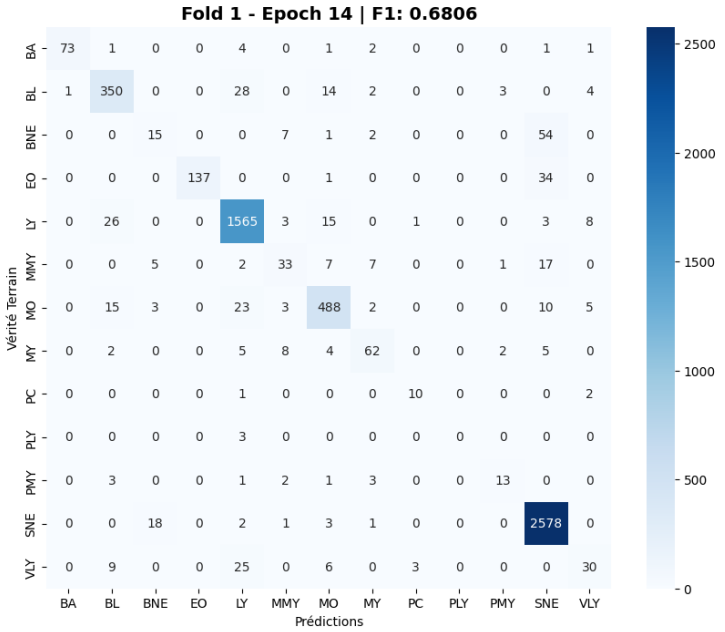

FIGURE 6 – Matrice de confusion (fold 1, ensemble de validation).

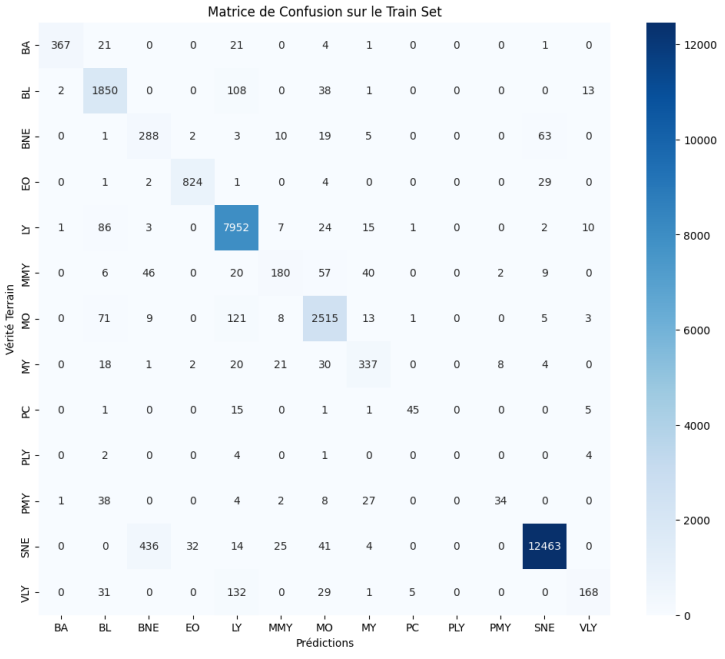

FIGURE 7 – Matrice de confusion du même fold sur l'ensemble d'entraînement complet.

5.2.1 Succès de classification

Les classes présentant les caractéristiques les plus distinctives sont très bien classées :

— **Neutrophiles (SNE) et Éosinophiles (EO) :** Grande précision, leurs caractéristiques granulaires uniques étant parfaitement capturées par les filtres de convolution.

— **Lymphocytes (LY) :** Classifiés avec une grande fiabilité grâce à la densité caractéristique de leur chromatine.

Ces résultats correspondent aux attentes issues de l'état de l'art : les classes à morphologie distinctive sont les plus aisément séparables.

5.2.2 Confusions persistantes

Confusion LY vs PLY. La classe PLY est absente du fold 1 de validation (Figure 6), ce qui illustre l'utilité d'une validation croisée : l'absence dans un seul fold n'impacte pas l'évaluation globale. Cette confusion était attendue : la distinction entre un lymphocyte mature et son précurseur prolymphocytaire repose sur la présence d'un nucléole parfois masqué par le bruit résiduel, une subtilité que même des experts humains confondent parfois.

Confusion BNE vs SNE. La classe BNE est classée presque quatre fois plus souvent en SNE qu'en BNE — un résultat surprenant, qui dépasse la confusion attendue pour des stades contigus de maturation. L'explication probable réside dans le déséquilibre de classes : SNE étant bien plus fréquente, le modèle lui accorde un biais a priori fort. La Figure 8 montre la projection PCA de ces deux classes sur les images segmentées (niveau 2), révélant un chevauchement partiel qui ouvre la voie au post-traitement géométrique.

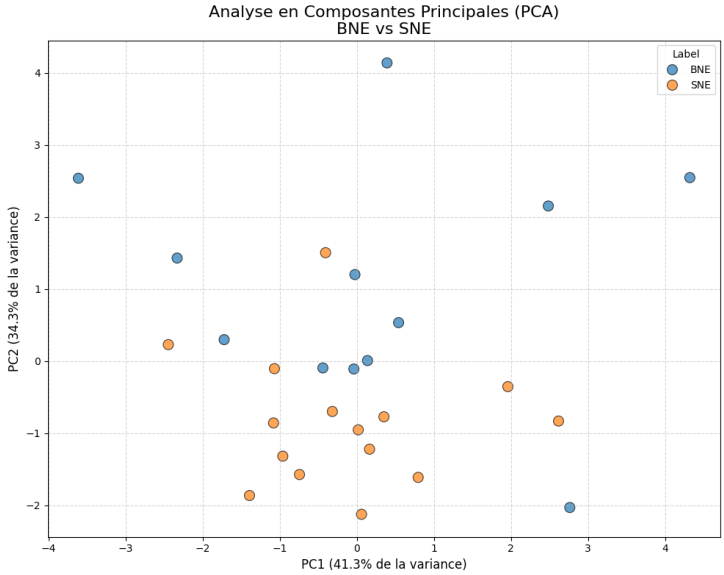

FIGURE 8 – Projection PCA des classes BNE et SNE sur un échantillon de segmentations de niveau 2.

5.3 Analyse du compromis performance/rejet par seuil de confiance

5.3.1 Formalisation

ResNet50 produit un vecteur de logits z. La fonction Softmax transforme ces scores en probabilités :

$$P(y = i \mid \mathbf{x}) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}, \quad K = 13 \tag{3}$$

Le taux de confiance est défini comme la probabilité maximale attribuée par le réseau :

$$C = \max_{i \in \{1 \dots K\}} P(y = i \mid \mathbf{x}) \tag{4}$$

5.3.2 Résultats

— **Phase de purification (0,0 à 0,65) :** Le F1-Score Macro croît continûment. Le point optimal est atteint à un seuil de $C = 0{,}65$, en conservant plus de 93 % du dataset.

— **Effondrement (C > 0,80) :** À des seuils élevés, les classes minoritaires (PLY, PC) sont totalement exclues du calcul, car le modèle n'atteint pas une certitude de 95 % sur ces morphologies complexes — ce qui fait mécaniquement s'effondrer le F1 macro.

Ce comportement était partiellement attendu : il est connu que les architectures profondes souffrent de mauvaise calibration. En revanche, l'amplitude de l'effondrement au-delà de 0,80 est plus marquée qu'anticipé, illustrant à quel point les classes rares sont morphologiquement ambiguës même pour le réseau. Pour une mise en production, un seuil de rejet à $C = 0{,}60$ constitue le meilleur compromis.

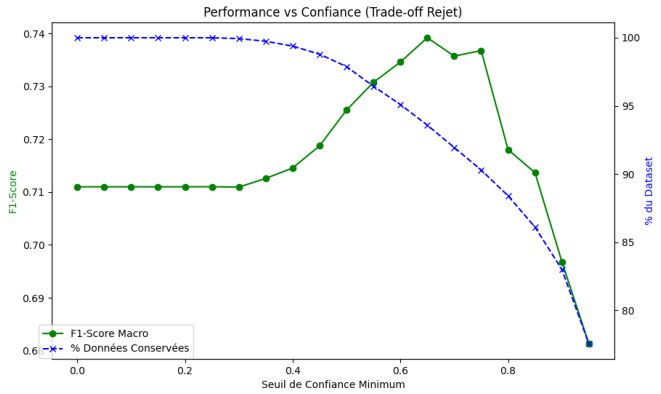

FIGURE 9 – Évolution du F1-Score Macro en fonction du seuil de confiance minimal (rejet des prédictions sous seuil).

5.4 Analyse de robustesse : impact du bruit d'acquisition

— **Zone optimale (bruit 0–20) :** $F_1 \approx 0{,}74$, correspondant à $n = 16\,451$ images. Les textures fines sont préservées, et le modèle peut exploiter la chromatine et les granulations.

— **Palier de sensibilité (20–60) :** Chute de $\approx -10\%$. Le bruit masque le signal biologique, dégradant particulièrement les classes à faible contraste (monocytes).

— **Seuil de rupture (60–80) :** $F_1 \approx 0{,}58$, valeur minimale.

— **Anomalie statistique (80–100) :** Rebond apparent à $F_1 \approx 0{,}65$. Cet intervalle ne contient que 157 échantillons ($< 1\,\%$ du dataset) : une performance correcte sur les Neutrophiles suffit à biaiser la moyenne macro sur un si petit échantillon.

Ces résultats valident *a posteriori* le Curriculum Learning : en entraînant d'abord sur les images nettes (bruit 0–20), le modèle converge vers des représentations idéales, et l'introduction progressive des données bruitées agit comme une régularisation. Ce mécanisme était anticipé, et la Figure 10 le confirme quantitativement.

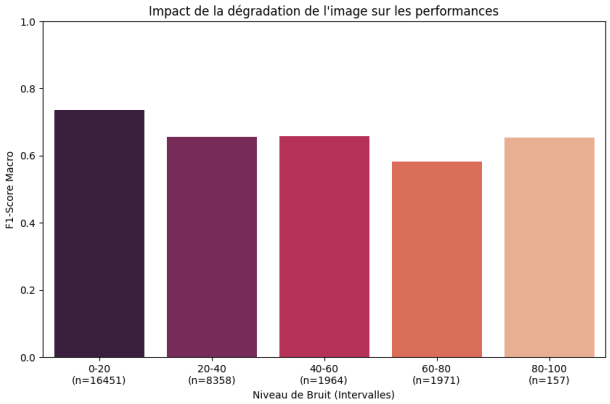

FIGURE 10 – F1-Score Macro en fonction du niveau de bruit de l'image.

6 Post-traitement Hybride : Correction Géométrique par PCA

Le réseau ResNet50 démontre une forte capacité à extraire des descripteurs texturaux, mais l'analyse post-entraînement révèle qu'il ne tient pas compte de contraintes géométriques strictes — notamment pour les paires de classes liées par la maturation (BNE/SNE). Pour corriger ces erreurs systématiques sans dégrader les cas bien classés, j'ai développé une méthode d'ensemble qui délègue la décision finale à un espace de caractéristiques manuelles via PCA, lorsque la prédiction du réseau est géométriquement incohérente.

6.1 Détection des conflits et modélisation de la normalité

Le pipeline se déroule en trois phases :

1. **Identification des confusions :** La matrice de confusion est calculée sur les prédictions d'entraînement. Si le taux d'erreur orienté d'une classe vers une autre dépasse 30 % (`CONFUSION_THRESHOLD = 0.3`), le couple est enregistré comme conflit majeur.
2. **Entraînement de la PCA par paire :** Pour chaque paire conflictuelle, une PCA est entraînée exclusivement sur des vecteurs de caractéristiques (aire, circularité, ratio N/C) issus de segmentations de niveau 2. Cela garantit que la PCA apprend une « vérité géométrique » exempte de biais d'acquisition.
3. **Calcul des centroïdes :** Le centroïde μ_C de chaque classe dans l'espace PCA 2D représente la morphologie typique de la classe, au sens géométrique.

6.2 Rattrapage géométrique et inversion de label

Lors de l'inférence, chaque prédiction impliquée dans un conflit surveillé est soumise à un test de cohérence. Les descripteurs de la cellule sont projetés dans l'espace PCA du conflit, et les distances euclidiennes aux centroïdes sont calculées :

$$d_{pred} = \|x_{pca} - \mu_{pred}\|_2 \quad \text{et} \quad d_{alt} = \|x_{pca} - \mu_{alt}\|_2 \tag{5}$$

Le label est inversé si la cellule est géométriquement beaucoup plus proche de la classe alternative que de la classe prédite :

$$\text{Correction si } d_{alt} < 0{,}75 \times d_{pred} \tag{6}$$

Le seuil de 0,75 a été déterminé par une recherche sur grille sur l'ensemble de validation grâce à la librairie optuna : des valeurs inférieures (ex. 0,50) corrigent trop peu de cas ; des valeurs supérieures (ex. 0,90) introduisent des inversions erronées sur des cas ambigus.

Cette hybridation conserve la puissance du Deep Learning pour la grande majorité des cas, tout en appliquant un garde-fou déterministe et interprétable sur les confusions critiques. Il faut noter que cette correction ne peut s'appliquer qu'aux 50 % d'images pour lesquelles une segmentation de niveau 2 est disponible, ce qui limite son impact global.

7 Conclusion

Ce projet avait pour objectif la conception d'un système automatisé de classification des leucocytes parmi 13 types cellulaires. Les défis principaux — déséquilibre de classes, variabilité colorimétrique, bruit d'acquisition et proximité morphologique — ont été adressés par un pipeline hybride en trois phases.

1. **Prétraitement et extraction de caractéristiques :** Un algorithme de segmentation K-Means dans l'espace L*a*b*, complété par le modèle Chan-Vese, permet d'isoler le noyau et le cytoplasme. Un score de fiabilité géométrique et textural filtre automatiquement les segmentations défaillantes. La normalisation CLAHE stabilise les gradients lors de l'entraînement.
2. **Apprentissage profond et Curriculum Learning :** ResNet50 (transfer learning depuis ImageNet) constitue le cœur du classifieur, atteignant une ligne de base de F1 = 0,67. Le Curriculum Learning, en présentant d'abord les images nettes, améliore la stabilité de la convergence et les performances finales. La validation croisée stratifiée garantit une évaluation robuste malgré le déséquilibre de classes.
3. **Post-traitement géométrique (PCA) :** Pour les confusions systématiques entre classes morphologiquement contiguës (notamment BNE/SNE), un rattrapage par PCA réévalue les prédictions du réseau à l'aune de règles géométriques apprises sur les segmentations de confiance.

Résultats et bilan. La principale limite de cette approche est la couverture partielle du post-traitement PCA, conditionnée à la disponibilité d'une segmentation fiable. À l'avenir, le remplacement de la segmentation maison par CellPose — entraînée sur un dataset beaucoup plus large — permettrait d'obtenir des masques robustes pour une proportion bien plus élevée des images, et donc d'étendre l'impact du post-traitement géométrique.

Références

[1] Sawsan F. Bikhet, Ahmed M. Darwish, Hany A. Tolba, and Samir I. Shaheen. Segmentation and classification of white blood cells. In *2000 IEEE International Conference on Systems, Man, and Cybernetics*, volume 3, pages 2259–2261. IEEE, 2000.

[2] Anjali Gautam and Harvindra Bhadauria. Classification of white blood cells based on morphological features. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 2363–2368. IEEE, 2014.

[3] Madhumala Ghosh, Devkumar Das, Subhodip Mandal, Chandan Chakraborty, Mallika Pal, Ashok K. Maity, Surjya K. Pal, and Ajoy K. Ray. Statistical pattern analysis of white blood cell nuclei morphometry. In *Proceedings of the 2010 IEEE Students' Technology Symposium*, pages 59–66. IEEE, 2010.

[4] Seyed Hamid Rezatofighi and Hamid Soltanian-Zadeh. Automatic recognition of five types of white blood cells in peripheral blood. *Computerized Medical Imaging and Graphics*, 35(4) :333–343, 2011.

[5] Furqan Rustam, Naila Aslam, Isabel De La Torre Díez, Yaser Daanial Khan, Juan Luis Vidal Mazón, Carmen Lili Rodríguez, and Imran Ashraf. White blood cell classification using texture and RGB features of oversampled microscopic images. *Healthcare*, 10(11) :2230, 2022.

[6] Mesut Toğaçar, Burhan Ergen, and Zafer Cömert. Classification of white blood cells using deep features obtained from Convolutional Neural Network models based on the combination of feature selection methods. *Applied Soft Computing*, 97 :106810, 2020.