

Université Numérique Cheikh Hamidou Kane (UN-CHK)

Master Big Data Analytics

Projet BIG DATA

Superviseurs :

Pr. Ousmane SALL

M. Abdourahmane BALDE

Réalisé par :

Moustapha Konteye

Amy Collé Barry

---

Domaine : Traitement d'Images

Thème : Développement d'un algorithme de segmentation pour les images microscopiques de frottis sanguins chez les patients atteints de drépanocytose.

---

PLAN

- I. Introduction
- II. Résumé
- III. Prétraitement des images
- IV. Choix d'un Algorithme de Segmentation
- V. Processus de Segmentation
- VI. Extraction et Caractérisation des cellules sanguines
- VII. Classification des cellules
- VIII. Présentation et Interprétation des résultats
- IX. Conclusion

**I. Introduction**

Un système de tri cellulaire consiste à trier les cellules présentes dans une image selon les classes de cellules qui nous ont été assignées. L'architecture d'un tel système de tri cellulaire repose sur trois parties s'enchaînent : la segmentation de l'image, l'analyse des caractéristiques cellulaires et la classification des cellules. La segmentation d'images microscopiques de frottis sanguins provenant de patients atteints de drépanocytose est essentielle pour analyser les caractéristiques des cellules sanguines et identifier les anomalies associées à cette maladie. Ce projet vise à développer une méthode de segmentation précise et efficace pour extraire les différentes composantes cellulaires des images de frottis sanguins.

Il existe trois principales approches en segmentation (Djemal & al. 2004, Djemal 2002). La segmentation fait souvent référence aux notions de similarité et de différence comme les perçoit le système visuel humain. On distingue trois approches couramment qualifiées d'approche classique

(méthode du gradient par seuillage), d'approche contour et d'approche région. Mais ce ne sont pas toutes ces méthodes qui conduisent directement à une segmentation de l'image.

Il sera donc nécessaire de procéder à une fermeture de contours si l'on souhaite obtenir une partition de l'image. L'approche région fait référence à des groupements de points ayant des propriétés communes. Contrairement à l'approche contour, l'approche région permet donc d'aboutir directement à une partition de l'image (chaque pixel étant affecté à une région unique).

Après la collecte d'images, l'image est la première segmentée afin d'isoler les parties intéressantes et supprimer le bruit et les composants indésirables. Ensuite, un processus d'extraction de caractéristiques est appliqué pour extraire les informations utiles provenant des cellules sanguines segmentées et enfin la classification peut être opérée selon la fonctionnalité extraite par l'étape précédente.

## II. Résumé

L'anémie est une maladie provoquée par un déficit en globules rouges. La forme des globules rouges prend la forme d'une faucille ou d'un croissant dans la drépanocytose. L'inspection manuelle d'images microscopiques est un processus très difficile et long. Dans cette recherche, des techniques de traitement d'image et d'apprentissage automatique sont utilisées pour automatiser le processus de détection des drépanocytoses dans des images microscopiques, puis classer les globules rouges en trois formes : circulaire, allongée (drépanocytose) et autre forme. L'image microscopique est prétraitée et la technique de seuillage Otsu est utilisée pour la segmentation. Ensuite, la segmentation des bassins versants est appliquée pour séparer les cellules superposées. Les caractéristiques géométriques, statistiques et texturales sont extraites des images. La machine à vecteurs de support de sable naïf de baye de régression logistique est utilisée

## III. Prétraitement des images

Le prétraitement d'image regroupe l'ensemble des processus visant à améliorer la qualité d'une image pour donner à l'utilisateur une image libérée de tous types de défauts. De ce fait, l'objectif principal est de mettre en relief l'information utile contenu dans l'image et par la suite atténuer, voire éliminer l'information inutile.

Dans cette partie, nous allons appliquer sur les images des algorithmes de filtrages tels que :

Le filtre gaussien, le filtre médian et le filtre bilatéral afin de supprimer le bruit.

Comment ces filtres fonctionnent ?

**Le filtre gaussien :** Le filtre gaussien est un type de filtre qui utilise une fonction gaussienne pour calculer la transformation à appliquer à chaque pixel de l'image. Il est utilisé pour réduire le bruit de l'image et atténuer les détails.

**Le filtre médian** est un filtre qui est couramment utilisé en traitement d'image. Il fonctionne en remplaçant la valeur chaque pixel par la médiane des valeurs de ses voisins dans l'image.

**Le filtre bilatéral** est un filtre qui réduit le bruit tout en préservant les bords de l'image. Il utilise deux paramètres de pondération : la similarité spatiale et la similarité des intensités des pixels.

La similarité spatiale est déterminée par la distance spatiale entre les pixels, tandis que la similarité des intensités de pixels est déterminée par la différence d'intensité entre les pixels.

Après l'application des algorithmes de filtres sur les images, des métriques (critère de performance) ont été utilisées pour choisir l'algorithme le plus performant.

### Critères de performance pour les différents filtres d'images utilisés :

1. **MSE (Mean Squared Error) :**
  - Plus la valeur de MSE est basse, meilleure est la performance du filtre.
  - MSE mesure la moyenne des carrés des différences entre les pixels de l'image originale et les pixels de l'image filtrée. Un MSE plus faible indique une meilleure correspondance entre les deux images.
2. **PSNR (Peak Signal-to-Noise Ratio) :**
  - PSNR mesure la qualité de l'image en termes de bruit.
  - Plus le PSNR est élevé, meilleure est la qualité de l'image.
  - Il est généralement utilisé pour comparer la qualité de l'image originale à l'image filtrée.
3. **MSSIM (Mean Structural Similarity Index) :**
  - MSSIM mesure à quel point les structures (contours, textures, etc.) sont préservées entre l'image originale et l'image filtrée.
  - Une valeur de MSSIM proche de 1 indique une bonne similarité structurelle entre les images.
  - MSSIM est souvent utilisé pour évaluer la qualité structurelle de l'image après le filtrage.
4. **MAE (Mean Absolute Error) :**
  - MAE mesure la moyenne des écarts absolus entre les valeurs des pixels de l'image originale et les valeurs des pixels de l'image filtrée.
  - Plus le MAE est bas, meilleure est la performance du filtre.
  - Il donne une mesure de l'erreur moyenne sans considérer la direction de l'erreur.

Moyennes des mesures pour chaque filtre :

MSE Gaussian	0.589276
PSNR Gaussian	50.419205
MSSIM Gaussian	0.991625
MSE Median	1.200443
MAE Gaussian	54.524289
PSNR Median	47.388256
MSSIM Median	0.982812
MAE Médian	70.385653

## IV. Choix d'un Algorithme de Segmentation

La segmentation d'images est une tâche fondamentale [de vision par ordinateur](#) qui consiste à diviser une image en régions significatives et sémantiquement homogènes. L'objectif est de simplifier la représentation d'une image ou de la rendre plus significative pour une analyse plus approfondie. Ces segments correspondent généralement à des objets ou à des régions d'intérêt au sein de l'image.

Dans le cadre de ce projet, nous utilisons l'algorithme de [watershed. Pourquoi ?](#)

Connu aussi sous le nom d'algorithme de bassin versant, l'algorithme de watershed est une technique classique de segmentation d'image basée sur le concept de transformation de bassin versant. Le processus de segmentation prendra la similitude avec les pixels adjacents de l'image comme référence importante pour connecter les pixels avec des positions spatiales et des valeurs de gris similaires.

### Fonctionnement de l'algorithme :

L'algorithme de bassin versant divise une image en segments à l'aide d'informations topographiques. Il traite l'image comme une surface topographique, identifiant les bassins versants en fonction de l'intensité des pixels. Les minima locaux sont marqués comme points de départ, et des inondations de couleurs remplissent les bassins versants jusqu'à ce que les limites des objets soient atteintes. La segmentation qui en résulte attribue des couleurs uniques aux régions, facilitant ainsi la reconnaissance des objets et l'analyse des images.

L'ensemble du processus de l'algorithme de bassin versant peut-être résumé dans les étapes suivantes :

- ✓ Placement des marqueurs : la première étape consiste à placer des marqueurs sur les minima locaux, ou sur les points les plus bas de l'image. Ces marqueurs servent de points de départ au processus d'inondation.
- ✓ Flooding : L'algorithme inonde ensuite l'image de différentes couleurs, en commençant par les marqueurs. Au fur et à mesure que la couleur se propage, elle remplit les bassins versants jusqu'à atteindre les limites des objets ou des régions de l'image.
- ✓ Formation des bassins versants : Au fur et à mesure que la couleur s'étale, les bassins versants se remplissent progressivement, créant une segmentation de l'image. Les segments ou régions résultants se voient attribuer des couleurs uniques, qui peuvent ensuite être utilisées pour identifier différents objets ou caractéristiques de l'image.
- ✓ Identification des limites : L'algorithme de partage des eaux utilise les limites entre les différentes régions colorées pour identifier les objets ou les régions de l'image. La segmentation résultante peut être utilisée pour des tâches de reconnaissance d'objets, d'analyse d'images et d'extraction de caractéristiques.

## V. Processus de Segmentation

### **PROCESSUS WATERSHED:**

Le dataset est constitué de 106 d'images microscopiques de sang frottis prélevés sur des patients drépanocytaires.

- **Dans le prétraitement des données**, étape où l'image microscopique est convertie en gris image, le filtre gaussien est utilisé pour éjecter le bruit du gris image à l'échelle. La dernière étape du prétraitement des données est l'amélioration de l'image dans laquelle les petits objets sont supprimés et la netteté de l'image est effectuée pour obtenir une meilleure qualité image.
- **Segmentation en tant que telle** : La technique de seuillage D'otsu est utilisée pour séparer les trois classes, cette méthode est l'une des meilleures méthodes de seuillage pour détacher l'objet de l'image. Otsu sélectionne le seuil qui réduit la variance interclasse du seuil noir et des pixels blancs.
- **Segmentation des bassins versants** : Watershed est basé sur la visualisation d'une image en trois dimensions, cela fournit un cadre simple pour intégrer des contraintes basées sur les connaissances dans le processus segmentation et utilise la morphologie de l'image. La cellule superposée est séparée en utilisant cette méthode, après la séparation, la cellule est étiquetée et la boîte englobante est formée sur la drépanocytose. Cette segmentation est effectuée sur une image en échelle de gris.
- **Opération morphologique** : L'opération morphologique consiste à rejeter les objets indésirables de l'image. En cela, les trous de l'objet sont couverts ; les bordures de l'image sont effacées et petites, les objets sont éliminés de l'image.

## VI. Extraction et Caractérisation des cellules sanguines

La caractérisation évalue les caractères extraits à la segmentation. Cela consiste à fournir les données représentatives de chaque cellule : taille, surface, texture, etc.

Une fois l'image initiale  $I$  contenant les cellules segmentées, on obtient une image  $I'$  qui contient les différents objets à analyser. Chaque objet de l'image a une étiquette ou référence et pour pouvoir classer chaque cellule (pour connaître son type), il nous faut des critères mesurables.

Westerman et Bacus ont classé les cellules dans le sang de patients drépanocytaires en 14 classes différentes, basé sur six caractéristiques : taille (surface), teneur en hémoglobine (densité optique intégrée), pâleur centrale, circularité, allongement et sécularité. Cette étude rapporte les caractéristiques statistiques dans la classification des cellules du sang comme normales et anormales.

Dans notre cas, nous allons procéder de la manière suivante pour extraire et caractériser les cellules sanguines :

- **Caractéristique géométrique**

La circularité, le rapport hauteur/largeur et l'excentricité sont appliquées.

Circularité - Les valeurs de la circularité sont par calculer par  $(4 \cdot \text{Zone} \cdot \pi) / (\text{Périmètre})^2$ .

Rapport d'aspect – Le rapport d'aspect est le rapport des principaux axe à axe mineur qui est utilisé pour distinguer cellule circulaire, autre et allongée.

Excentricité – L'excentricité d'une image est le rapport de  $\sqrt{1 - (\text{axe mineur}/2)^2} / (\text{axe majeur}/2)$ .

- **Fonctionnalité statique et de texture**

Dans les caractéristiques statiques et de texture, variance, entropie, l'écart type et l'énergie sont utilisés.

## VII. Classification des cellules

Après avoir extrait les caractéristiques de chaque cellule de l'image, ces derniers sont enregistrés dans un dataframe afin de pouvoir utiliser un algorithme de classification. Une fois chaque cellule caractérisée par différentes valeurs mesurables, le système informatique de cytologie se sert de ces valeurs pour classer la cellule.

Dans ce cas-ci, nous avons utilisé la régression logistique et le Vecteur Support Machine SVM comme algorithme de classification. C'est à ce stade qu'intervient donc une décision. Le système de décision doit intégrer les résultats de mesures quantitatives avec éventuellement des données cliniques.

## VIII. Présentation des résultats

Résultat de la caractérisation :

	Cellule	Aspect Ratio	Eccentricity	Circularity	Standard Deviation	Variance	Entropy	Energy	category
0	Cellule #1	1.887194	0.848068	0.345474	51.494960	2651.730867	7.504007	222346	Circulaire
1	Cellule #2	2.001169	0.866194	0.296758	53.417827	2853.464214	7.520746	206056	Circulaire
2	Cellule #3	1.995141	0.865321	0.318503	51.032987	2604.365770	7.972556	405302	Circulaire
3	Cellule #4	1.774280	0.826042	0.472714	57.109053	3261.443903	7.898656	353402	Circulaire
4	Cellule #5	1.890928	0.848721	0.415250	55.997996	3135.775571	7.998933	401346	Circulaire
5	Cellule #6	2.031237	0.870419	0.246096	50.336988	2533.812409	8.435628	599104	Autres
6	Cellule #7	1.965093	0.860836	0.331816	54.645772	2986.160366	8.121542	415517	Circulaire
7	Cellule #8	1.910344	0.852047	0.358119	54.030381	2919.282069	7.898962	327230	Circulaire
8	Cellule #9	2.091731	0.878320	0.252167	53.343025	2845.478316	8.436682	380090	Circulaire
9	Cellule #10	1.274278	0.619802	0.342807	48.529453	2355.107810	9.668347	1808570	Allongée
10	Cellule #11	1.924635	0.854422	0.315292	54.007660	2916.827305	8.389001	509528	Circulaire
11	Cellule #12	1.942417	0.857297	0.365861	53.049117	2814.208863	7.735946	274487	Circulaire

## IX. Conclusion

La drépanocytose est une situation dans laquelle les globules rouges ne sont pas produits comme ils devraient l'être. Dans ce projet, nous avons utilisé la méthode de segmentation de watershed qui est une méthode de segmentation d'image rapide, stable, avec une grande précision sur les frontières. obtenues et offre une distinction parfaite des régions collées. Néanmoins, elle consomme de la mémoire, elle est sensible aux bruits, et peut présenter une sur-segmentation. Pour contrer ces contraintes il suffit de bien débruiter l'image et supprimer tous les minima à faible contraste. Les caractéristiques telles que les caractéristiques géométriques et statistiques sont extraites des cellules.

Nous avons utilisé la régression logistique et le Vecteur Support Machine SVM comme algorithme d'apprentissage pour classer les cellules circulaires, allongés et autres selon notre ensemble de données.

Parce qu'il n'existe pas de méthode de segmentation universelle, on devra adapter le processus de segmentation à chaque problématique.