



Intelligence artificielle et cytologie

Thomas Boyer et Samy Dahmani

CHU Amiens

Congrès BiomedJ

Jeudi 9 mars 2023



ARTIFICIAL INTELLIGENCE

Programs with the ability to learn and reason like humans

MACHINE LEARNING

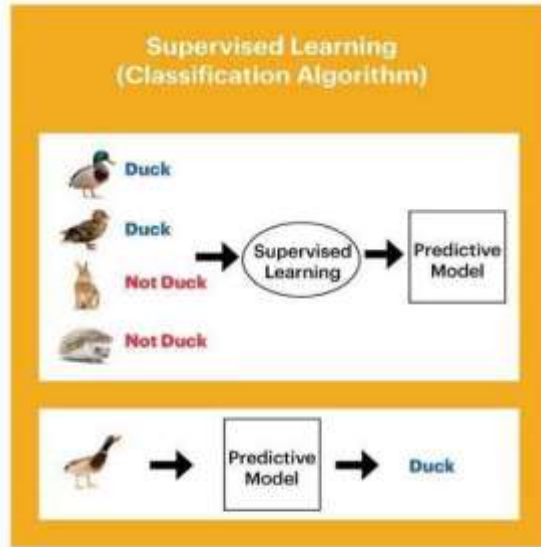
Algorithms with the ability to learn without being explicitly programmed

DEEP LEARNING

Subset of machine learning in which artificial neural networks adapt and learn from vast amounts of data



Apprentissage supervisé/non supervisé





Régression/Classification



Regression

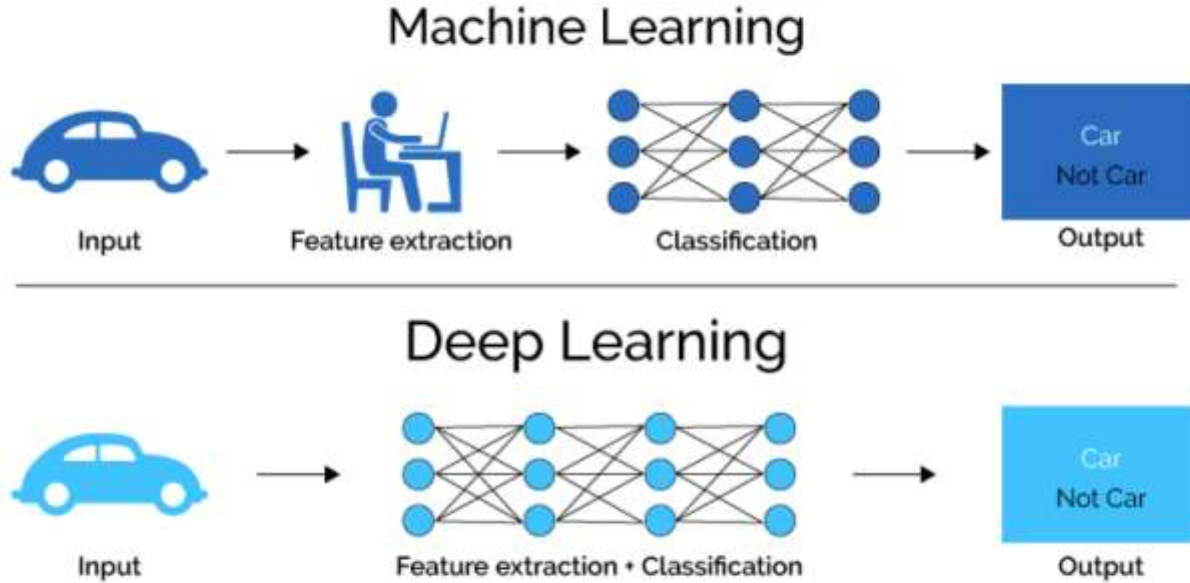
What is the temperature going to be tomorrow?



Classification

Will it be Cold or Hot tomorrow?





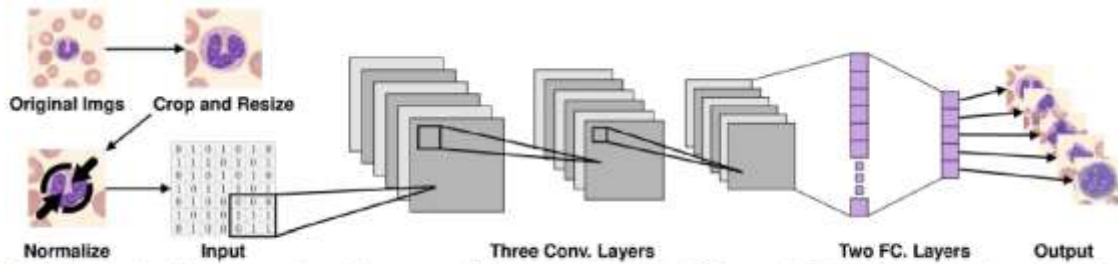
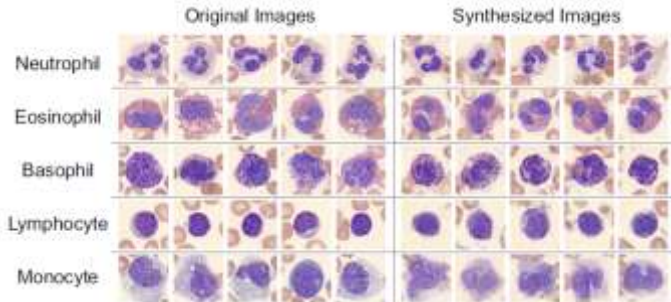


Fig. 2 An overview of the pre-processing and the proposed CNN-based architecture for WBC image classification. The pre-processing consists of cropping, re-sizing and normalizing. Three convolutional layers (including three pooling layers) are in charge of extracting and learning features, and two fully connected layers are in charge of classification



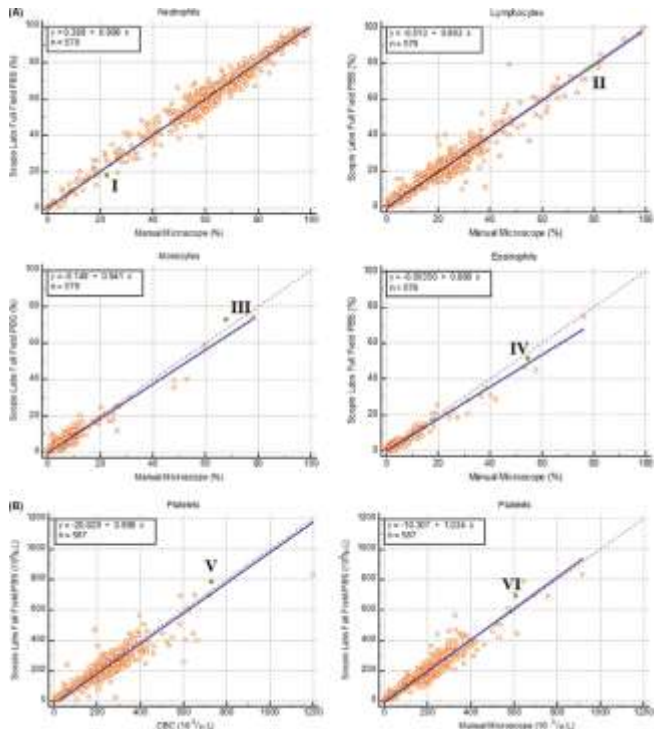
	NE. (%)	EO. (%)	BA. (%)	LY. (%)	MO. (%)
Fold-0	100	95	92	99	96
Fold-1	98	99	94	100	100
Fold-2	96	93	100	95	98
Fold-3	97	99	100	95	96
Fold-4	100	100	97	98	97
Fold-5	100	98	94	97	98
Fold-6	100	98	94	97	91
Fold-7	95	98	94	98	96
Fold-8	100	93	94	97	99
Fold-9	98	100	91	95	97
Avr. Acc.	98	97	95	97	97

The average accuracy for five classes is 97%

Utilisation CNN pour génération d'images pour leur algorithme d'IA



Evaluation Scpio Labs X100 (champs larges)



Distribution of abnormal samples

Acute inflammation/Bacterial infection	49	13	19	17
Chronic inflammation	32	13	8	11
Parasitic infection/Allergic reaction	44	18	12	14
Viral infection	41	13	14	14
Aplastic anemia/Chemotherapy	29	13	10	6
Lymphopenia	42	13	14	15
Acute leukemia	33	12	10	11
Severe anemia/Myeloproliferative disorders	40	13	17	10

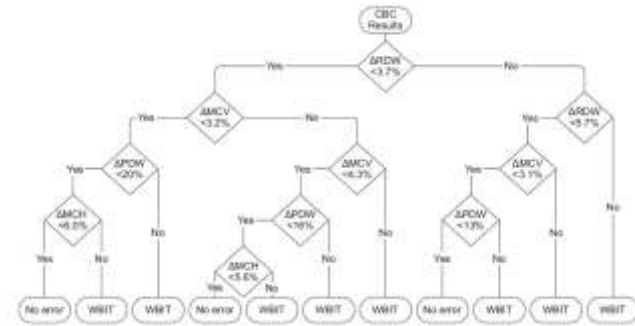
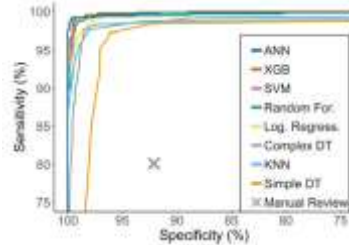
600 échantillons

3 hôpitaux différents

> 95% concordance avec technique manuelle



Algorithm	Accuracy (%) (95% CI)
Artificial neural network	99.1 (98.5-99.4)
Extreme gradient boosting	99.0 (98.5-99.4)
Support vector machine	99.0 (98.4-99.3)
Random forest	98.7 (98.1-99.1)
Logistic regression	98.3 (97.6-98.8)
Complex decision tree	98.0 (97.3-98.5)
K nearest neighbours	97.2 (96.4-97.9)
Simple decision tree	96.8 (95.9-97.5)
Manual review	85.7 (84.1-87.2)

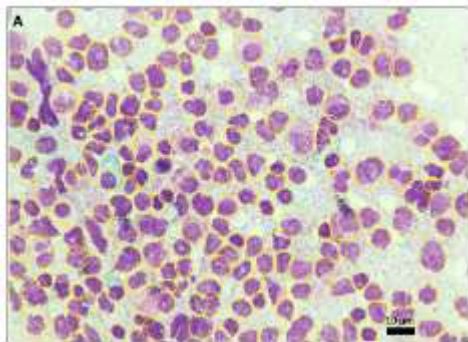


Arbre décisionnel proposé

Comparaison performances humaines/algorithmes de ML

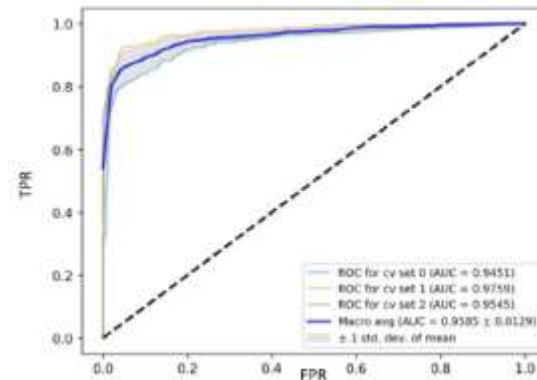


Deep Learning

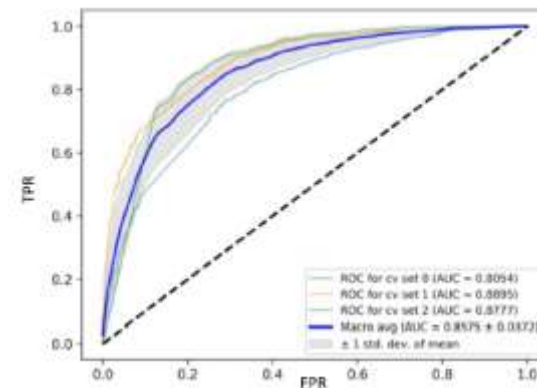


51 moelles LAP
1048 moelles LAM non LAP
236 moelles donneurs sains

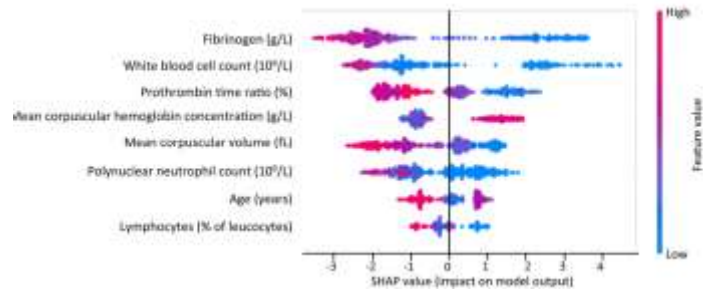
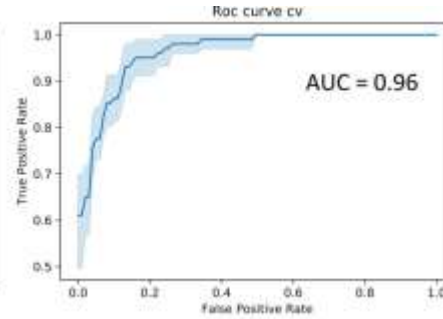
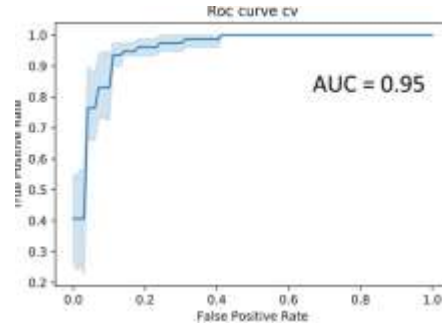
LAP vs MO normales



LAP vs LAM non LAP



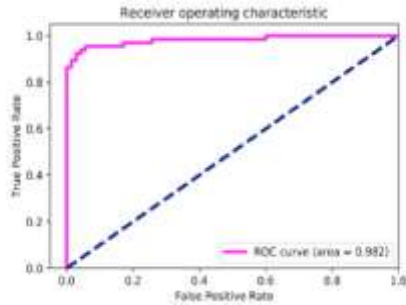
Algorithme XGBoost (ML)



https://share.streamlit.io/nico-facto/leukemia-apl-classification/main/Leucemie_app.py

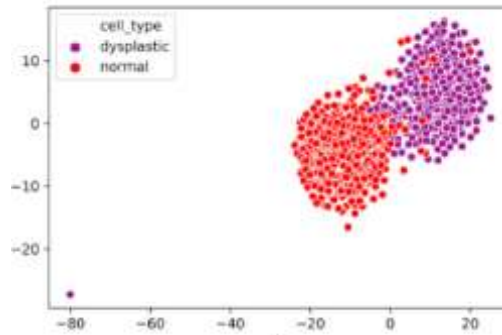


Algorithme DysplasiaNet (CNN)



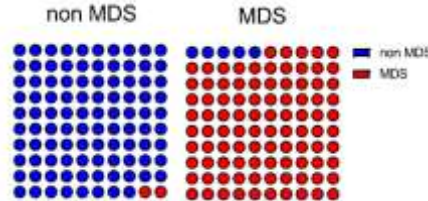
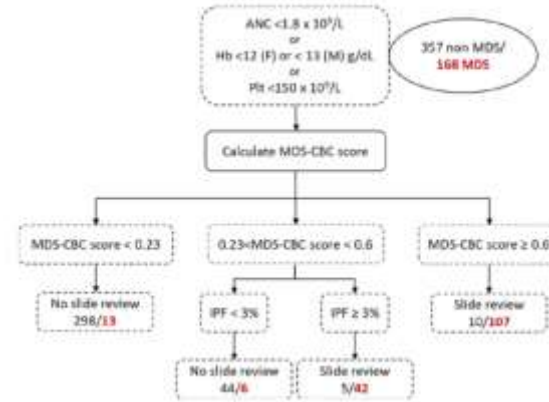
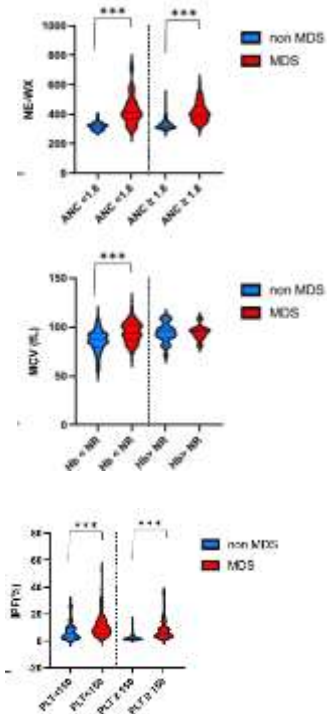
		Predicted	
		Dysplastic	Normal
True	Dysplastic	63 (95.5 %)	3 (4.5 %)
	Normal	4 (5.7 %)	66 (94.3 %)

t-SNE avec réduction de dimensionnalité

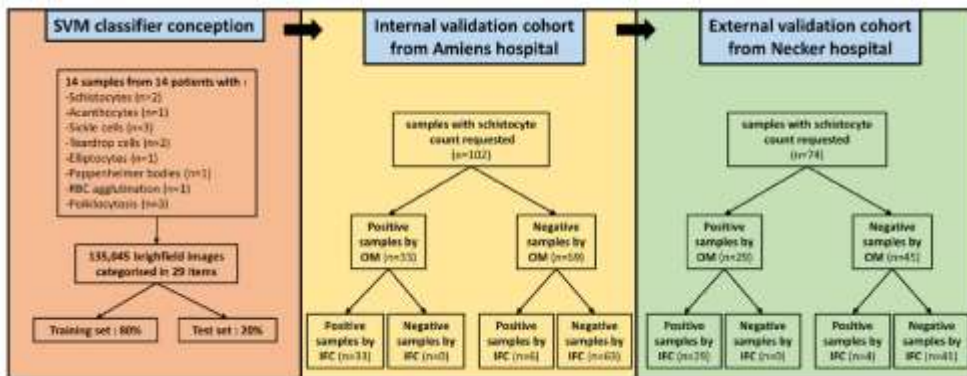




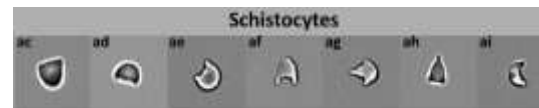
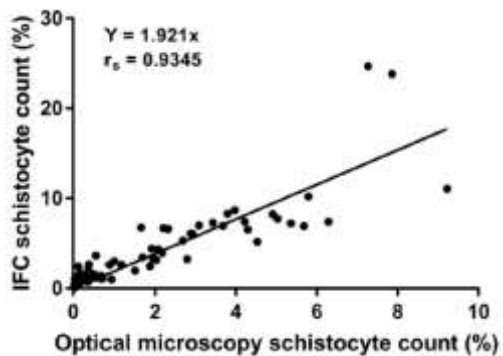
Random Forest + CART (ML)



98% des Non MDS et 95% des MDS correctement classés



SVM





Application pratique

- L'IA en cytologie est un sujet de recherche en constante évolution.
- Les études publiées en la matière ne partagent généralement pas de dataset et/ou de réseau de neurones, pas de retombée pratique
- Il y a de nombreuses équipes qui y travaillent mais pour avoir un impact, il faut mettre en place un dataset collaboratif de grande envergure et de référence
- La base de données ImageNet



ImageNet

- Base de données d'images pour la vision par ordinateur
- Plus de 14 millions d'images étiquetées dans plus de 20 000 catégories
- Créée en 2009 par une équipe de recherche de l'Université de Princeton
- A permis et a accompagné l'explosion du Deep Learning
- Devenu un dataset de référence





Dataset public

Dataset de Barcelone de 17092 images provenant du DM96 réparties en 8 classes (5 leucocytes normaux, granuleux immatures, érythroblastes et plaquettes géantes) (2020)

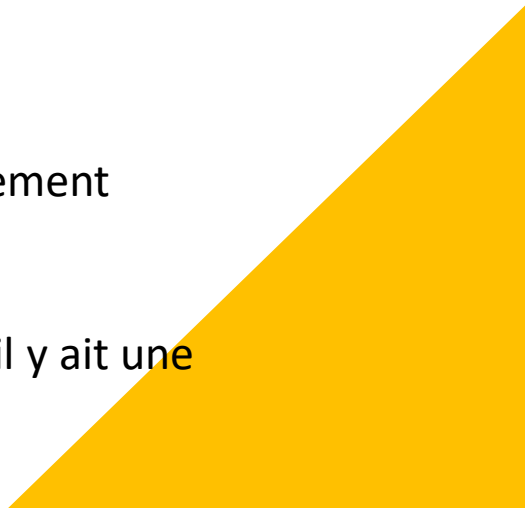
Dataset Rabbain-WBC de 40000 images, issues de 73 frottis. Ces images ont été capturées par des smartphones sur des microscopes

Dataset du Cancer Imaging Archive capturé sur un M8 scanner microscope comporte 18365 images, répartis en 16 classes issus 200 patients du même hôpital.



Dataset public

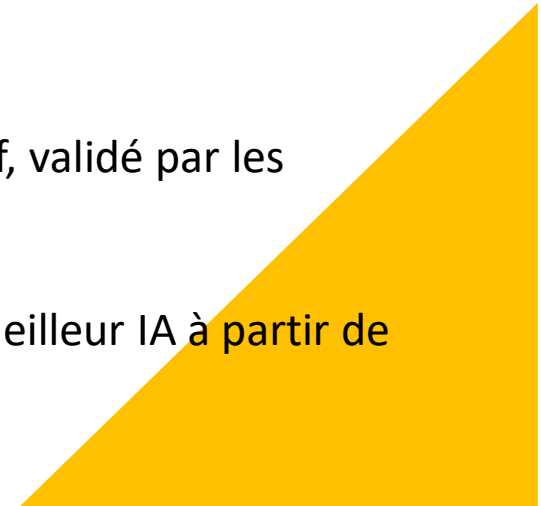
- Très bons datasets
- Ne sont pas encore des "ImageNet"
- Pas assez de classes, d'images ou de variabilité (généralement monocentriques et sur un faible nombre de patients)
- Il faut que tous les efforts soient mis en commun, et qu'il y ait une collaboration des différents centres





Datachallenge

- Appel à projet du Health Data Hub
- Réalisation de datachallenge en santé
- Mise en place d'un dataset de qualité, varié et collaboratif, validé par les experts
- Compétition d'équipes internationales pour produire la meilleur IA à partir de ce dataset





Projet Cytologia

- **Réaliser une intelligence artificielle permettant de classer les leucocytes normaux et pathologiques**
- **1ère étape** : Constituer une base de données d'image de leucocytes normaux et pathologiques de bonne qualité et variée
- **2ème étape** : Datachallenge afin de sélectionner le meilleur algorithme



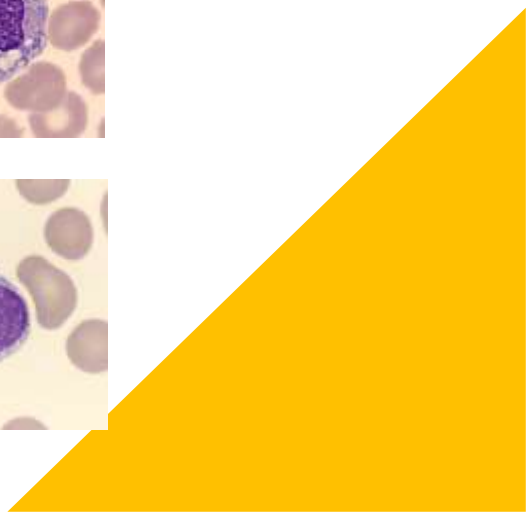
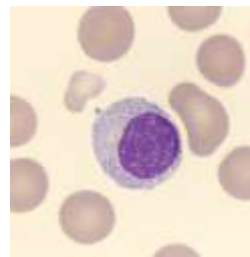
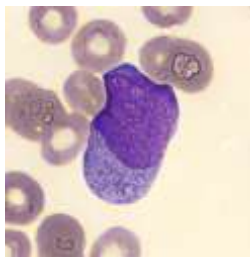
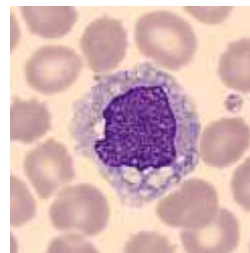
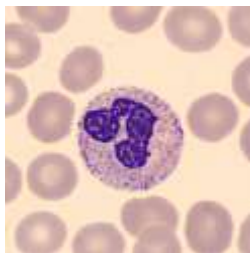
Dataset

15 classes utilisées en routine

Minimum 1000 images par classe

Classes équilibrées

Multicentrique et collaboratif
en mettant à contribution le
réseau du GFHC





Datachallenge

Datase d'entrainement et dataset de test

Compétition public internationale

Dernier trimestre de 2023

Récompense prévue pour les 3 premiers équipes et partage de leurs algorithmes





Perspectives

- Dataset **riche**, **transversale** et **varié**
- IA et dataset partagés
- Dataset sera utile à d'**autres projets de recherche** et pour la formation
- **Valorisation** du dataset via des publications
- Emergences de projets similaires (cytologie = domaine large).





Conclusion

- Collaboration multicentrique essentielle pour maximiser l'efficacité de l'IA .
- Appel à contribution pour la constitution de la base de données
- Envoyez un mail à Boyer.Thomas@chu-amiens.fr





Merci





Liste des classes

- Polynucléaire neutrophile
- Polynucléaire basophile
- Polynucléaire éosinophile
- Lymphocyte
- Monocyte
- Myéloblaste
- Lymphoblaste
- Monoblaste
- Lymphocyte réactionnel
- Cellules lymphomateuses
- Granuleux immatures
- Grandes plaquettes
- Erythroblastes
- Cellules lysées
- Artéfacts

