

Apercu de mon alternance au LIP6

Projets Cyclope | LigaPC

Imane EL Missaoui
M2 Data science en santé/alternante au LIP6 pour
la période Décembre 2023-Septembre 2024

13-09-2024



Cyclope

► **Description du projet :**

- Capsule endoscopique innovante pour le diagnostic du cancer colorectal.
- Capture, traitement d'images et reconstruction 3D.
- Transmission des images pertinentes pour faciliter le diagnostic.

► **Encadrant :** Andrea Pinna



► **Missions :**

- 1-Etude et développement de nouveaux protocoles pour la collecte des données médicales issues des future essais cliniques pour le test et la validation des premiers prototypes de la capsule.
- 2- Design et mise en place d'une base de donnée de capsules endoscopiques, organisation et structuration de la base.
- 3- Etude et conception d' un modèle d'inférence par apprentissage machine qui puisse être embarqué dans une capsule.

LigaPC

► **Description du projet :**

- Dispositif médical pour étudier la ligamentoplastie et l'instabilité de la cheville.
- Système électro-mécanique avec capteurs et Interface Homme-Machine (IHM).
- Tests in-vivo et développement d'algorithmes pour suivre et prédire la récupération.

► **Encadrant :** Sylvain Fregulio



► **Missions :**

- 1-Création et la structuration d'une nouvelle base de données multi-modale (accéléromètre, capteur de force, NIRS, électrodermie, température, etc.).
- 2- Réalisation d'un tableau de bord permettant de visualiser et traiter en temps réel différentes données émanant d'un ou plusieurs fichiers CSV se remplissant continuellement durant une acquisition afin d'afficher différents paramètres calculés (HR, BR, ...) sur les signaux acquis.

Cyclope

Mission 1: Protocoles de collecte de données :

1. Protocole d'acquisition de données pré-clinique avec des datasets existants

Choix des datasets pertinents qui contiennent des images d'endoscopie et des annotations (polypes, anomalies, etc.).
Déterminer les données d'intérêt à collecter

2. Protocole de confidentialité des données

Anonymisation/
Pseudonymisation : Data
Masking pour protéger la vie
privée des patients,
conformément aux
réglementations RGPD

3. Protocole de stockage et de gestion des données

Conception et mise en place
de la base de données
Sauvegarde et récupération :
Une copie de la base de
données a été créée pour la
récupération .

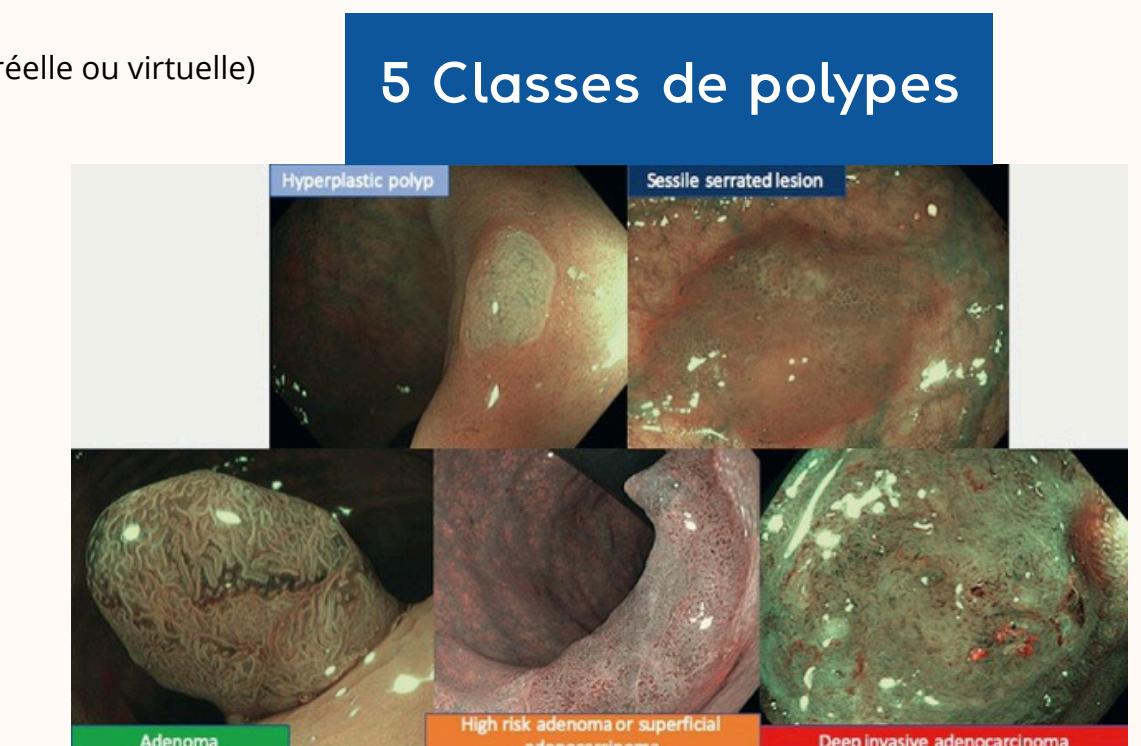
Cyclope

Protocoles de Collecte de données pour la Classification CONECCT :

- Etat de l'art sur la classification Connect
- Collecte des jeu de données publiques et privées qui pourront être exploitées pour la classification Conecct.
- Protocoles de Collecte de données pour la Classification:

- Les images doivent être prises par NBI ou équivalent
- Besoin d'avoir des données très bien et densément annotées avec les critères : Macro, Couleur(NBI), Vaisseaux(NBI ou équivalent), Fosses(endoscopie chromo réelle ou virtuelle)
- Besoin de concevoir un pipeline pour annotation

- **Label-I: Annotation de la Catégorie de Classification (I_H, I_{IA}, I_S, I_{IC}, III)**
- **Label-II: le masque/Bounding box**
- **Label-III: Délimitation/Bordure**
- **Label-IV: Description des attributs**
 - 4 attributs : Macro, Couleur, vaisseaux et Cryptes.



Cyclope

Protocoles de Collecte de données pour la Classification Conecct :

- Video/Image Annotation

VTT: Visual Object Tagging Tool: application pour annoter des objets dans des vidéos et des images + facilite le développement d'un pipeline d'apprentissage automatique de bout en bout englobant l'annotation/l'exportation/l'importation d'actifs. + Elle peut fonctionner en tant qu'application native ou via le web.

VIA: VGG Image Annotator. Web-app simple et autonome pour l'annotation manuelle d'images, d'audio et de vidéo.

Cyclope

Mission 2: Design et implementation de la base de donnée

- Design

Binary Large Object(stockage BLOB)	Stockage en Système de Fichiers	Codage en Base64
Images stockées sous forme de données binaires 	Image stockée sur le système de fichiers et la base de données contient une référence à l'emplacement du fichier. 	Image codée en Base64 et stockée dans une colonne de texte de la base de données. 

Meilleure option : stocker les images dans un système de fichiers et inclure le chemin d'accès et les métadonnées dans la base de données.

- Système de gestion de base de données



Pour :

- Niveaux de performance
- Vitesse et disponibilité élevées
- Documents de type JSON
- Syntaxe de requête simple
- Facilité d'utilisation

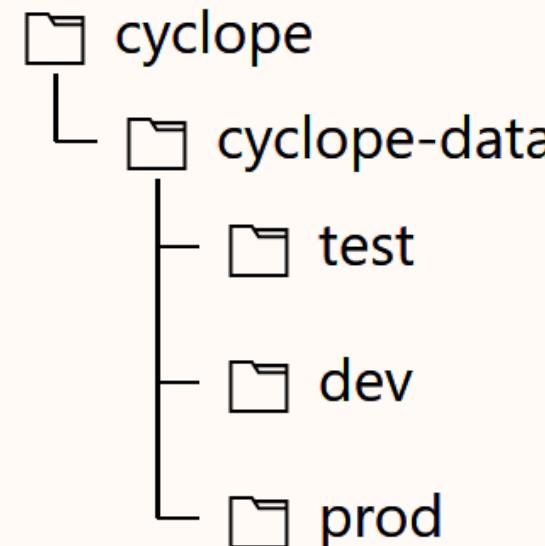
- Inconvénients :
- L'indexation peut être problématique
 - Utilisation élevée de la mémoire (RAM)

Cyclope

Mission 2: Design et implementation de la base de donnée

- 3 instances de la base de donnée MongoDB ont été créée: dev/test/prod sur le serveur nfs **users/outils/cyclope/mongodb-{dev}{test}{prod}/db**
- Une version de sauvegarde est créée automatiquement sur la machine “recover” au LIP6
- le système de fichiers a été créé dans le répertoire **/dsk/l1/misc/cyclope/cyclope-data/{dev}{test}{prod}** sur la machine “fox ”

Système de fichiers de la base de données



Métadonnées de la base de données

└── cyclope	
└── mongodb-dev	port 27017
└── mongodb-test	port 27018
└── mongodb-prod	port 27019

Cyclope

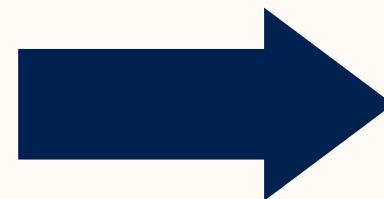
Mission 2: Design et implementation de la base de donnée

- Collecte de données d'imagerie, traitement de ces données avant de les fusionner pour alimenter notre nouvelle base de donnée dev.

```
└── Cyclope_Data
    ├── WCEcurated
    ├── CP-CHILD-A
    ├── CP-CHILD-B
    ├── CVC-300
    ├── CVC-ClinicDB
    ├── CVC-ColonDB
    ├── EndoScene
    ├── Kvasir
    ├── Kvasir-SEG
    ├── LDPolypVideo
    ├── PolypsSet
    └── Readme.txt
```

**11 datasets
9GB**

Data management



Filename standarization

Images: {datasetname}_{subset}_{class}_{id}.{extension}
Annotations: {datasetname}_{subset}_{class}_{id}_annotation.{extension}
Masks: {datasetname}_{subset}_{class}_{id}_mask.{extension}
Boundaries: {datasetname}_{subset}_{class}_{id}_boundary.{extension}

create_boundaries.py :Création des fichiers boundary
pour les images qui ont leurs masques correspondants

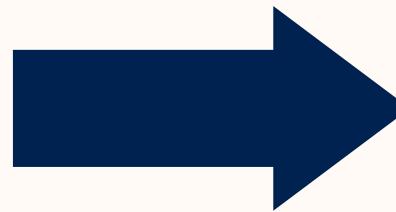
```
└── cyclope
    └── cyclope-data
        └── dev
            ├── Annotations
            │   └── {datasetname}_{subset}_{class}_{id}_annotation.{extension}
            ├── Boundaries
            │   └── {datasetname}_{subset}_{class}_{id}_boundary.{extension}
            ├── Images
            │   └── {datasetname}_{subset}_{class}_{id}.{extension}
            ├── Masks
            │   └── {datasetname}_{subset}_{class}_{id}_mask.{extension}
            └── Readme.txt
```

Système de fichiers de la base dev

Cyclope

Mission 2: Design et implementation de la base de donnée

Transfert des métadonnées à l'aide
de l'API MongoDB et la librairie
PyMongo



```
{  
    "_id": {  
        "$oid": "6661dd3d7f6d72a880399058"  
    },  
    "id": 18,  
    "dataset": "CP-CHILD-A",  
    "category": "non-polyp",  
    "unique_id": "114",  
    "image": {  
        "extension": "jpg",  
        "size": {  
            "width": 256,  
            "height": 256,  
            "path":  
        },  
        "resolution": {  
            "horizontal": 96,  
            "vertical": 96  
        },  
        "file_size": 30967  
    },  
    "mask": null,  
    "annotation": null,  
    "boundary": null,  
    "created_at": {  
        "$date": "2024-06-06T18:01:01.992Z"  
    },  
    "updated_at": {  
        "$date": "2024-06-06T18:01:01.992Z"  
    }  
}
```

MongoDB UI/ Instance Dev

Image metadata

The screenshot shows the MongoDB Compass interface with the database 'Fox_CyclopeDB' selected. The 'CyclopeDB' collection is open, displaying a single document. The document structure is highlighted with a red dashed box, focusing on the 'image' field which contains nested objects for 'extension', 'size', 'resolution', and 'file_size'. A large blue arrow points from the text 'Transfert des métadonnées à l'aide de l'API MongoDB et la librairie PyMongo' to this highlighted area. The text 'Image metadata' is written next to the arrow.

```
_id: ObjectId('6661dd3d7f6d72a880399047')  
id: 1  
dataset: "CP-CHILD-A"  
category: "non-polyp"  
unique_id: "114"  
image: Object  
    extension: "jpg"  
    size: Object  
    resolution: Object  
    file_size: 21431  
mask: null  
annotation: null  
boundary: null  
created_at: 2024-06-06T18:01:01.969+00:00  
updated_at: 2024-06-06T18:01:01.969+00:00  
  
_id: ObjectId('6661dd3d7f6d72a880399048')  
id: 2  
dataset: "CP-CHILD-A"  
category: "non-polyp"  
unique_id: "110"  
image: Object  
mask: null  
annotation: null  
boundary: null  
created_at: 2024-06-06T18:01:01.980+00:00  
updated_at: 2024-06-06T18:01:01.980+00:00
```

Cyclope

Mission 2: Design et implementation de la base de donnée

Organisation du système des fichiers pour les instances test et prod

La structure du système de fichiers doit suivre la logique de versioning et de prototypage :

- **Images** : Divisées par date ou par version du prototype, comme /V1/Images/2024-09-09/...
- **Annotations, Masques et Boundaries** : Même logique, séparées par type d'image ou version, comme /V1/Annotations/...
- **Logs et backups** : dossiers spécifiques pour conserver les logs techniques, avec des fichiers avec heure et data mentionés pour chaque version ou essai.

Cyclope

Mission 2: Design et implementation de la base de donnée

Base de donnée MongoDB Test et Prod

Collections MongoDB proposées :

1.Capsules (ou Devices)

- ID Capsule
- Date de création
- Type de prototype (V1, V2, etc.)
- Description technique (capteurs, sources lumineuses, dimensions, etc.)
- Version Firmware
- Statut de validation (ex. : en test, validé)

2.Patients (ou Subjects si les essais n'impliquent pas encore de patients réels)

- ID Patient
- Données anonymisées (âge, sexe, etc.)
- Protocole de test associé
- Date de début des tests
- Consentement éclairé (lien vers les documents)

3.Essais cliniques

- ID Essai
- Date de l'essai
- Prototype utilisé (référence à la capsule)
- Données de suivi (temps de fonctionnement, événements techniques, logs d'erreurs)
- Conditions de l'essai (p.ex. : environnement, réglages spécifiques)

4.Images et données associées

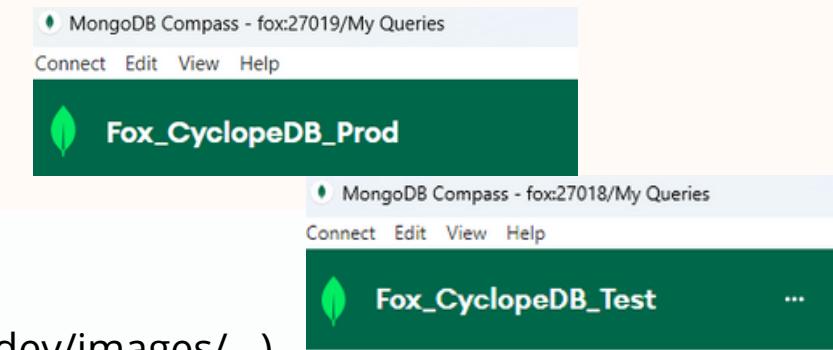
- ID Image
- Lien vers l'image dans le filesystem (ex : /dev/images/...)
- Type d'image (brute, annotée, segmentée)
- Résolution
- Date de capture
- ID Essai (référence à l'essai clinique)
- Type de capteur (pour chaque image)

5.Annotations et masques

- ID Annotation
- Lien vers le fichier d'annotation
- Type d'annotation (ex : segmentation, classification)
- Créeur de l'annotation (nom ou ID du médecin)
- Version de l'annotation
- ID Image (référence à l'image annotée)
- Date de création

6.Logs d'événements

- ID Log
- Date et heure
- Description de l'événement (erreur, action, ajustement)
- ID Capsule ou Essai (référence à l'objet concerné)
- Statut (succès, échec)



Cyclope

Mission 3: Etude et conception d' un modèle d'inférence par apprentissage machine qui puisse être embarqué dans une capsule.

- **Implémentation d'un modèle VGG16 + transfer learning**

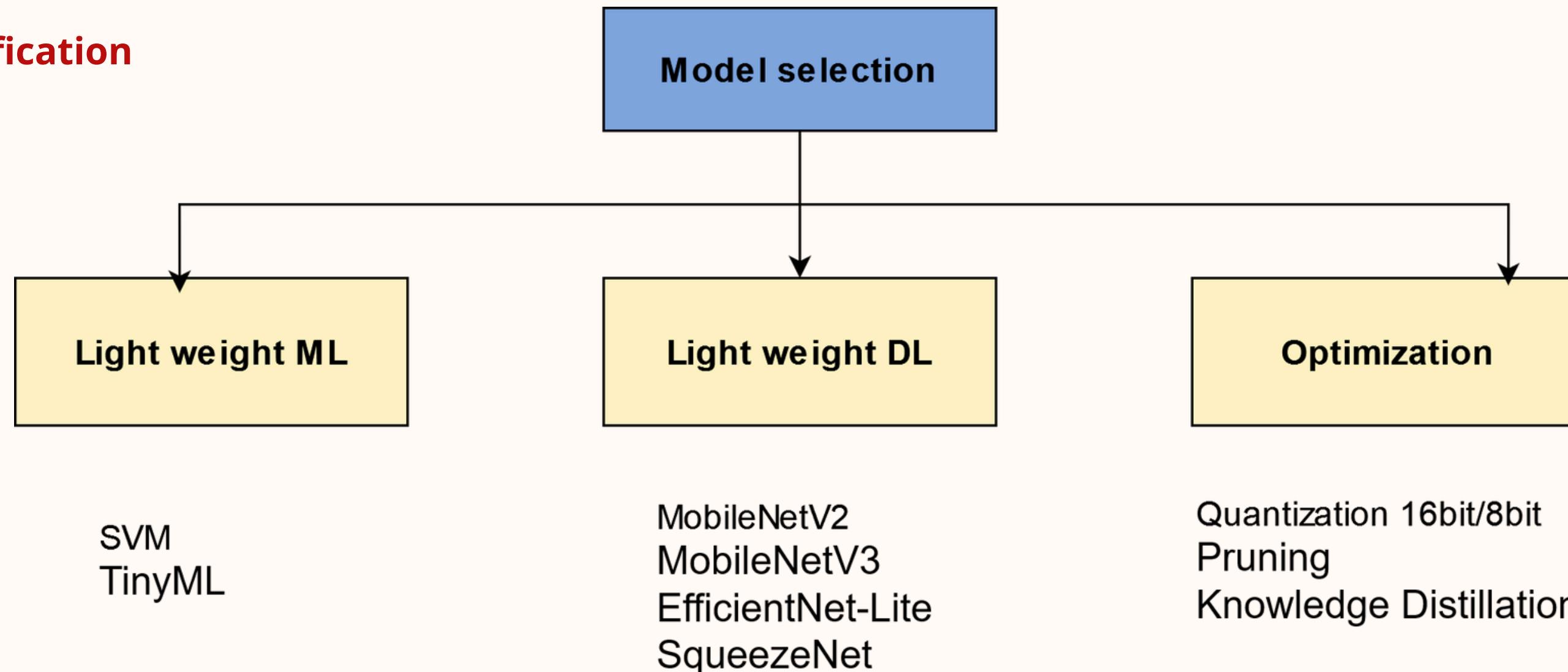
>>> Contraintes d'embarquabilité

>>> Équilibre précision/efficacité

Cyclope

Mission 3: Etude et conception d' un modèle d'inférence par apprentissage machine qui puisse être embarqué dans une capsule.

- **Modèle de classification**



Cyclope

Mission 3: Etude et conception d' un modèle d'inférence par apprentissage machine qui puisse être embarqué dans une capsule.

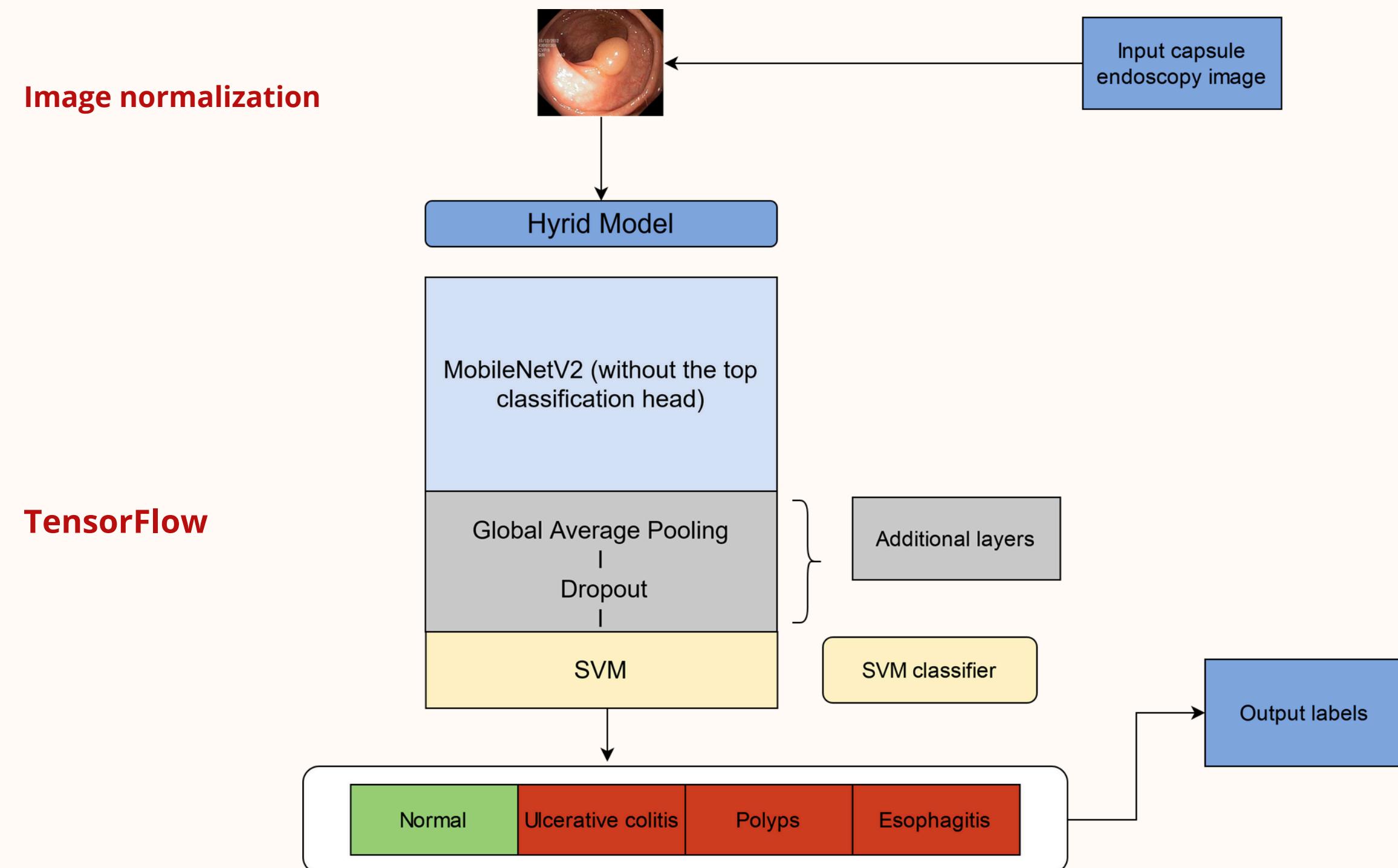
- Ensemble de données utilisé

-Wireless Capsule Endoscopy (WCE) curated

Class	Class name	Total images before augmentation	Total images after augmentation
0	Normal	1500	3000
1	Ulcerative colitis	1500	3000
2	Polyps	1500	3000
3	Esophagitis	1500	3000
	Total	6000	12000

Cyclope

- Diagramme d'architecture du modèle Hybride MobileNetV2+SVM



Cyclope

- Comparaison des métriques d'évaluation de diverses architectures CNN utilisées pour l'entraînement et la validation.

CNN architectures	Accuracy	AUC	Precision	Recall
Training				
VGG19	0.967	0.998	0.969	0.966
ResNet50	0.983	0.999	0.983	0.983
ResNet101	0.980	0.999	0.981	0.980
EfficientNetV2B1	0.984	0.999	0.984	0.983
GastroNetV1	0.987	0.999	0.987	0.987
MobilNetV2+SVM	0.999	0.999	0.999	0.998
Validation				
VGG19	0.963	0.999	0.964	0.961
ResNet50	0.980	0.999	0.981	0.980
ResNet101	0.976	0.999	0.978	0.976
EfficientNetV2B1	0.985	0.999	0.985	0.985
GastroNetV1	0.992	0.991	0.993	0.993
MobilNetV2+SVM	0.977	0.997	0.978	0.977

Spécificité pour la classe polyp: 0.977

Sensibilité pour la classe polyp: 0.989

Cyclope

- **Création d'une application web de test avec le modèle MobileNetV2-SVM lite intégré**

-Streamlit

-Outil qui pourrait aider les médecins dans l'analyse, génération de rapports,...

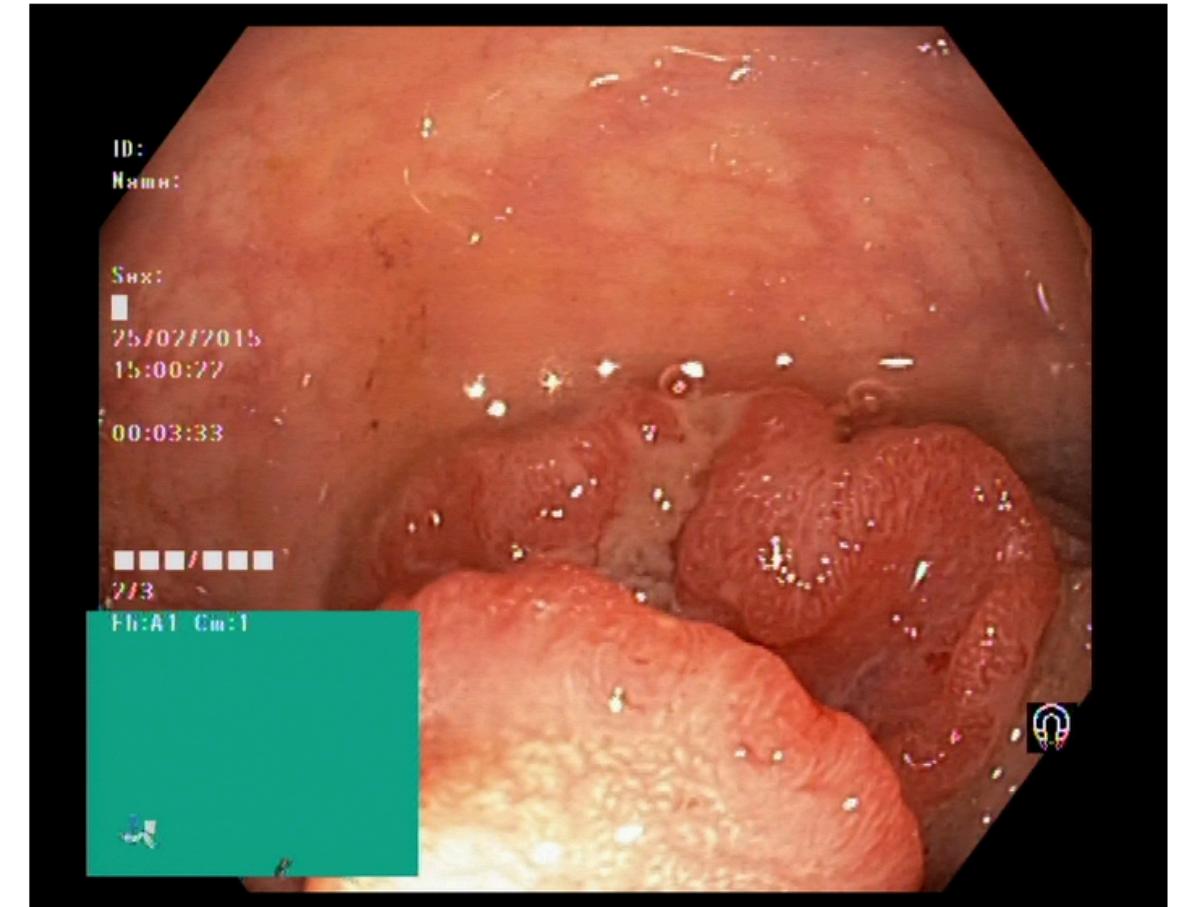
Medical Image Classification

Upload an endoscopic image and classify it using the Hybrid MobileNetV2-SVM model.

Choose an image...

Drag and drop file here
Limit 200MB per file • JPG, JPEG, PNG

 test_polyps_(47).jpg 187.1KB 



Uploaded Image.

Predicted Class: Polyps

Real Class: Polyps

The model predicted correctly!

LigaPC

Mission 1: Création et la structuration d'une nouvelle base de données multi-modale (accéléromètre, capteur de force, NIRS, électrodermie, température, etc.).

- Une base de donnée SQL sur PostgreSQL
- La base de donnée contiendra les fichiers csv contenant les données cliniques enregistrées par le dispositif électro-mécanique + données des patients
- La base de données sera hébergée sur un serveur lip6

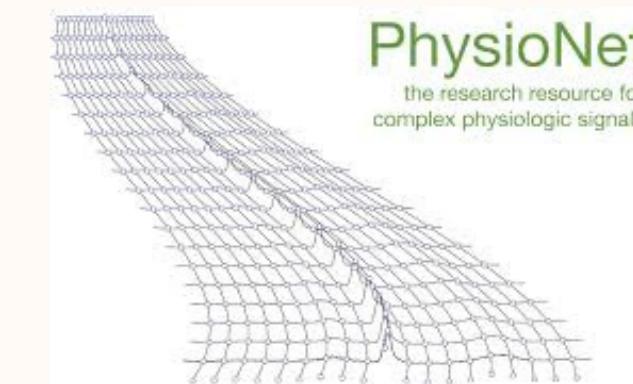


LigaPC

Mission 2: Réalisation d'un tableau de bord permettant de visualiser et traiter en temps réel différentes données émanant d'un ou plusieurs fichiers CSV se remplissant continuellement durant une acquisition afin d'afficher différents paramètres calculés (HR, BR, ...) sur les signaux acquis.

- **Dataset utilisé**

- Open access, ensemble de données haute résolution et synchronisées dans le temps.
- Plusieurs capteurs portés à différents emplacements sur le corps.
- Photopléthysmogramme (PPG), données inertielles, pression et ECG.
- 22 sujets sains effectuant 3 activités physiques(assis,en marche, en course).
- 66 enregistrements d'ondes provenant de PPG multi-sites et multi-longueurs d'onde, pressions d'attachement des capteurs, températures des capteurs ECG + pressions artérielles systolique et diastolique et niveaux de saturation en oxygène du sang (SpO₂).



LigaPC

Mission 2: Réalisation d'un tableau de bord permettant de visualiser et traiter en temps réel différentes données émanant d'un ou plusieurs fichiers CSV se remplissant continuellement durant une acquisition afin d'afficher différents paramètres calculés (HR, BR, ...) sur les signaux acquis.

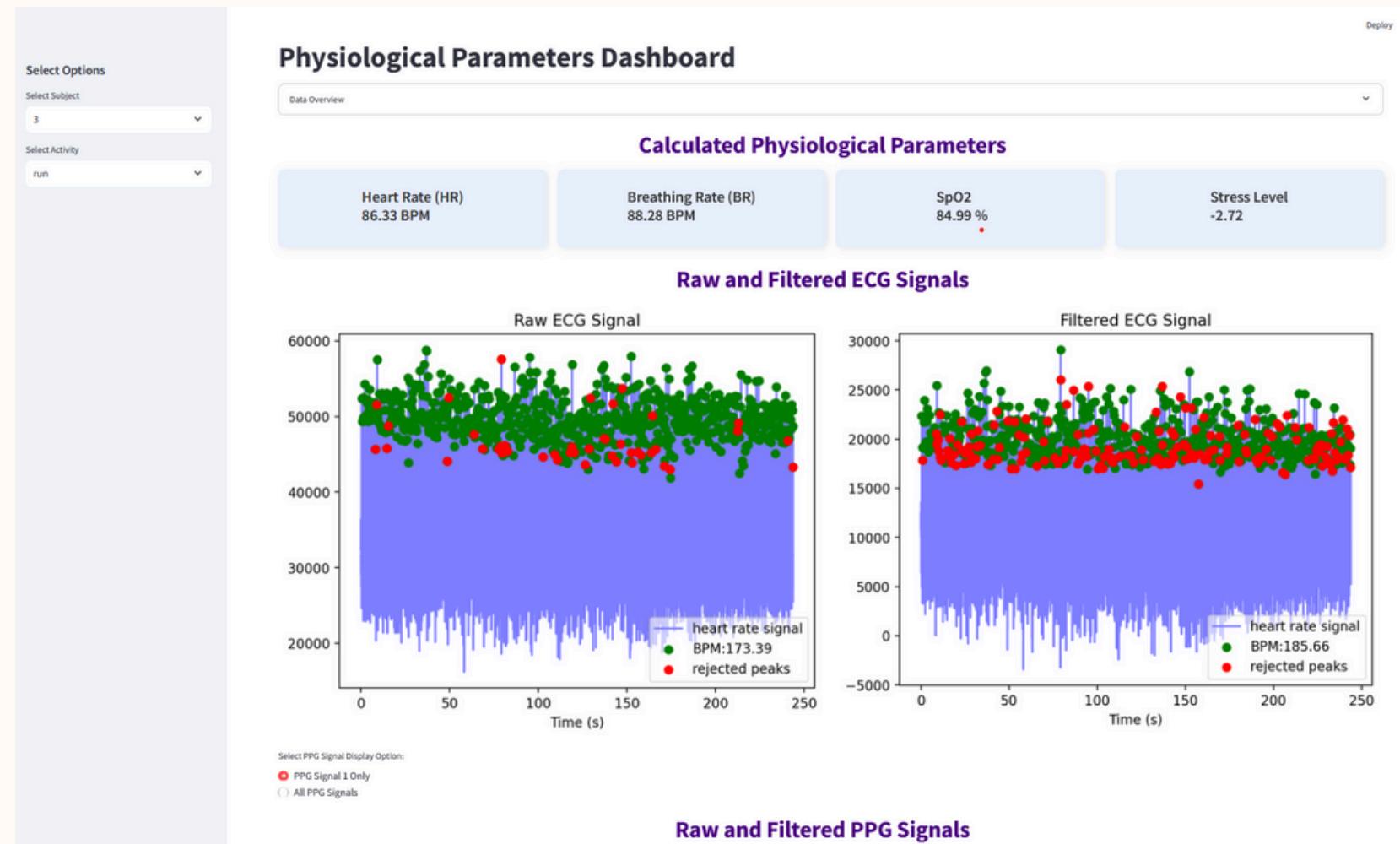
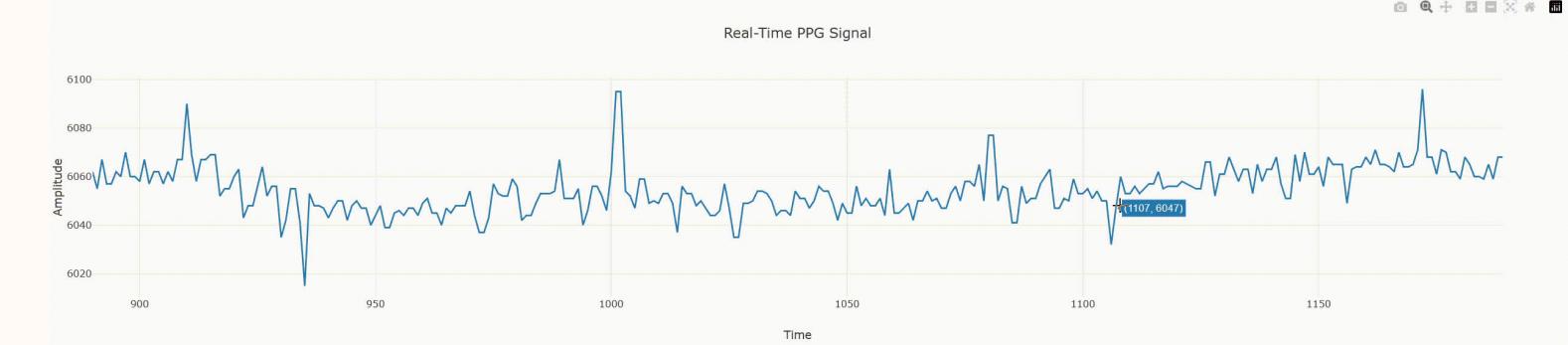


Tableau de bord streamlit



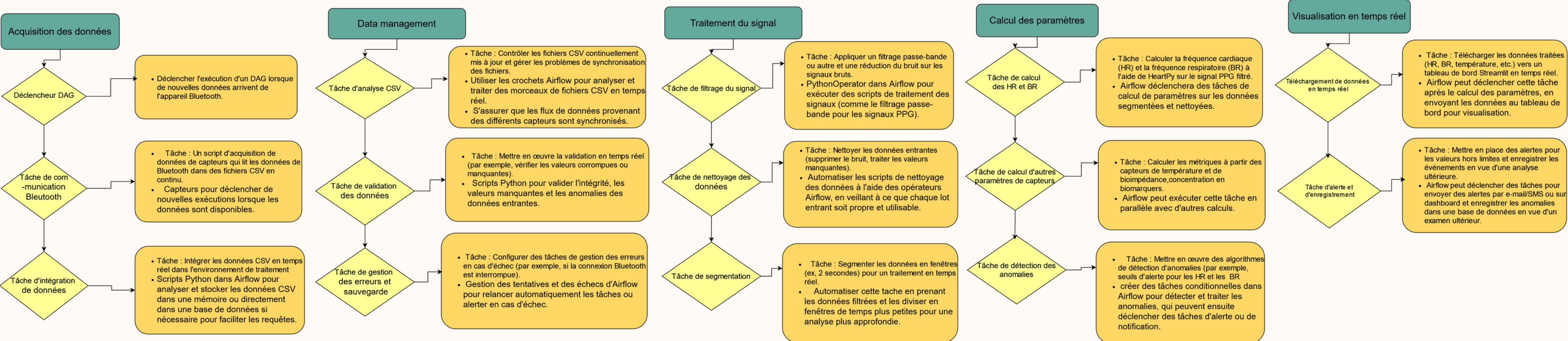
Signal PPG en temps réel

LigaPC

Mission 2: Réalisation d'un tableau de bord permettant de visualiser et traiter en temps réel différentes données émanant d'un ou plusieurs fichiers CSV se remplissant continuellement durant une acquisition afin d'afficher différents paramètres calculés (HR, BR, ...) sur les signaux acquis.



- **Design du pipeline d'automatisation sur Airflow**



Suite

- **Estimation de l'embarquabilité du modèle hybride**
- **Développement du tableau de bord, et automatisation des taches de traitement de signal et de calculs d'indicateurs en temps réel,déploiement du tableau de bord.**

Merci pour 
votre attention 22