

Portfolio

Amine Hadj-Youcef. PhD
Senior Data Scientist & Machine Learning

15/01/2023



Hi 🙌

- Docteur Data Scientist, mon objectif est d'accompagner les clients sur leurs projets data, de la définition du besoin jusqu'à la mise en production, dans le but de résoudre des problématiques métiers concrets, d'apporter de la valeur ajoutée et d'augmenter le chiffre d'affaires.
- J'attache de l'importance au transfert de connaissances par le biais de l'encadrement et de la formation des juniors.



Compétence principales

1. Conseil et accompagnement des équipes métiers

- Définition du besoin avec les équipes métiers
- Traduction du besoin métier en une solution technique

2. Développement de solutions techniques en Data Science & Machine Learning

- Collecte, ingestion, filtrage, enrichissement, analyse et exploration des données
- Conception de dashboard analytiques pour mettre la data en valeur
- Modélisation et mise en production des solutions ML sur le cloud

3. Leadership, mentorat et management

- Suivie, orientation et gestion de projet
- Respect des deadlines et objectifs des projets

Compétence Techniques

	Outils
Data Science	Python, SQL, NumPy, pandas, Apache Spark
Machine Learning	Scikit-learn, AWS Sagemaker, HyperOpt, MLflow
Data Engineering	Airtable, postgres, Delta Lake
Data Visualisation	PowerBI, Tableau, Redash
Big Data	Apache Spark, Azure Databricks
DevOps	Git, GitHub, Docker, Flask, FastAPI, Postman
Cloud	Microsoft Azure, AWS, GCP
Computer Vision	OpenCV, Tensorflow
Gestion de Projet	Microsoft Teams, JIRA, Notion

Formations

- 🎓 Doctorat, Sciences et Technologies de l'Information
 - 2018: Université Paris-Saclay
- 🎓 Master, Traitement du Signal et de l'image
 - 2015: Université de Bordeaux
- 🎓 Ingénieur, Électronique
 - 2012: Ecole Nationale Polytechnique (ENP)

Certifications

- [Academy Accreditation - Databricks Lakehouse Fundamentals](#)
 - 2022 : Databricks
- [Hackmakers #BuildwithAI Global Hack: Mentor Certificate](#)
 - 2020: Hackmakers
- Deep Learning and Artificial Intelligence
 - 2018: École Polytechnique Palaiseau



Expériences professionnelles

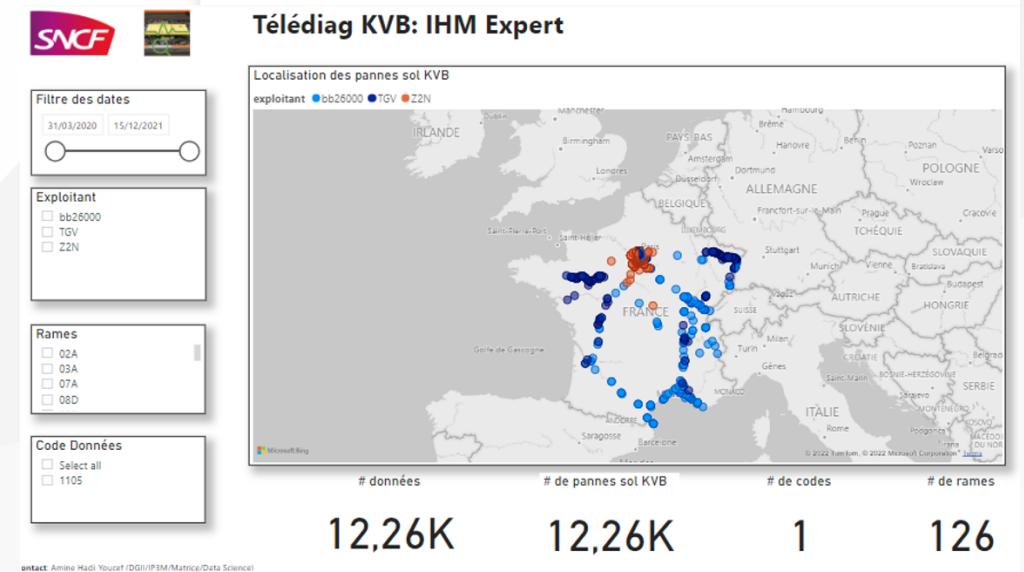
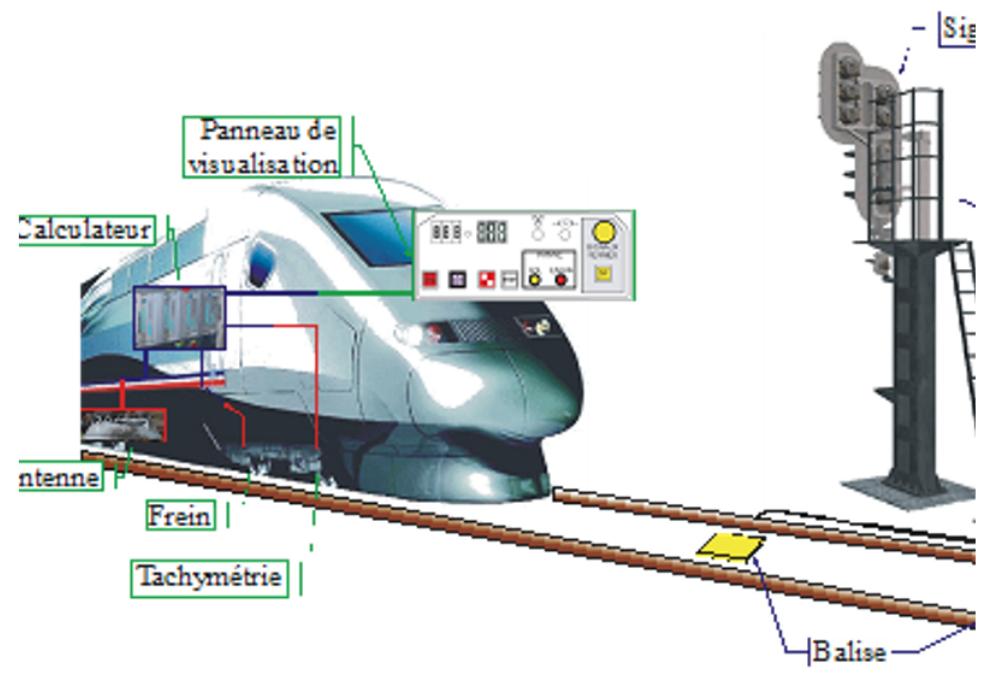


- Senior Data Scientist - Ingénieur Machine Learning
- Paris, France
- 2020-2022

- SNCF Réseau est une entreprise Française responsable de la maintenance et la sécurité du réseau ferroviaire dans toute la France. A travers le programme de supervision et de surveillance, la DGII assure la maintenance du réseau et la régénération des équipements. Des données massives sont collectées tous les jours par des capteurs connectés indiquant différentes mesures: vitesse des trains, températures des rails, état des appareils de voies,...
- Mon rôle de Data Scientist au sein de l'équipe *Data Science et Décision* est d'accompagner les différents équipes métier à définir le besoin, développer des solutions techniques adaptées au besoin, et d'assurer la communication sur le sujet avec les différents acteurs du projet.

- Parmi les tâches réalisées:
 - J'ai facilité l'accès aux données massives, nettoyées et enrichies, par la mise en place des pipelines d'ingestion sur le Lakehouse
 - Conception de dashboards interactives sur Power BI
 - Amélioration de la maintenance *des capteurs de la vitesse* des trains par la détection de pannes et l'envoi automatique des alarmes vers le centre de supervision
 - Automatisation de la détection d'anomalie des capteurs par l'industrialisation d'un algorithme de clustering des données
 - Traitement des données massives provenant des boîtes noires des trains

Tech: Apache Spark, Python, SQL, Databricks (Delta Live Table, Workflows), AWS, Azure, mlflow



 TAG Heuer - LVMH

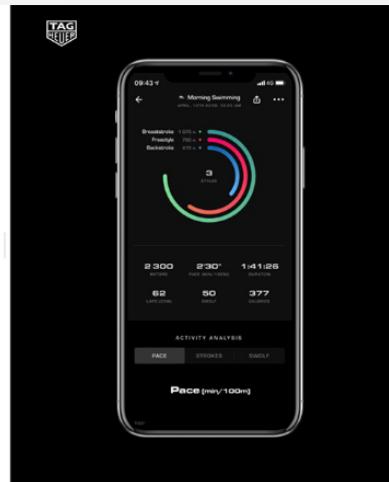
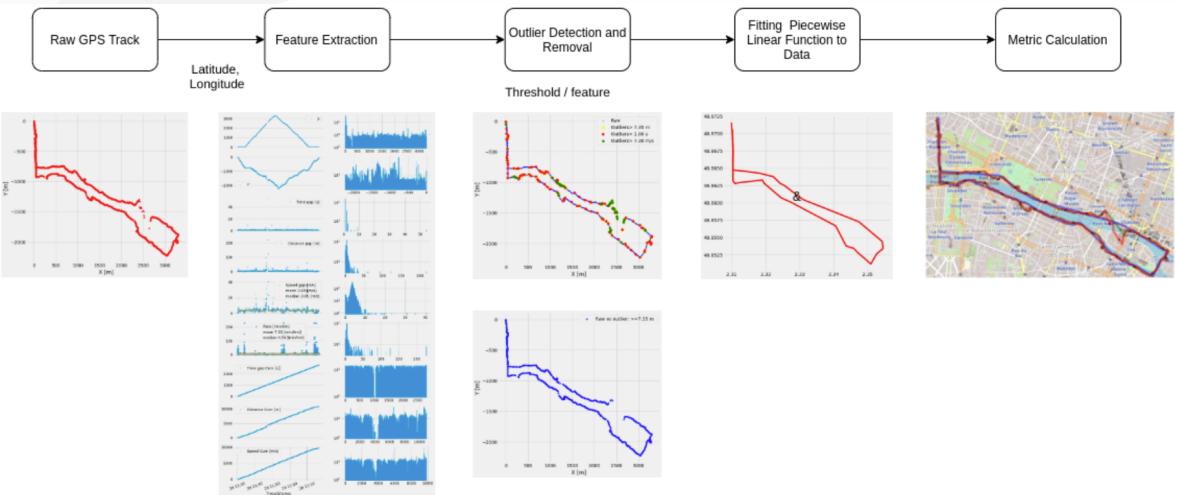


- Data Scientist - Data Engineer
- Paris, France
- 2020

- TAG Heuer est un horloger de luxe suisse qui conçoit, fabrique et commercialise des montres et des accessoires de mode. Dans le but d'élargir la gamme de ses produits, TAG Heuer a lancé au marché un nouveau modèle de montre intelligente de luxe, connectées, adaptés pour les activités sportives (running, swimming, golf). Dotée de différentes capteurs (WiFi, GPS, Podomètre, Cardiogramme), la montre embarque différents algorithmes pour traiter les données mesurées en temps réels, et calculer des métriques de sports.

- Mon rôle dans l'équipe Traitement du signal est de proposer des algorithmes de détection d'outlier causé par la perte du signal GPS, afin d'amélioration des KPI de l'application de la montre *TAG Heure Connected*
- En tant que Data Scientist, j'ai réalisé les tâches suivantes:
 - Traitement, nettoyage, préparation, analyse de données des capteurs (GPS...)
 - Développement d'algorithme machine learning pour détecter les anomalies
 - Conception d'une application web pour interagir avec les algorithmes développés
 - Communication des résultats et participation aux réunions

Tech: Python (pandas, Numpy, scikit-learn), Docker, Streamlit



 **Drone Volt**

DRONE VOLT®

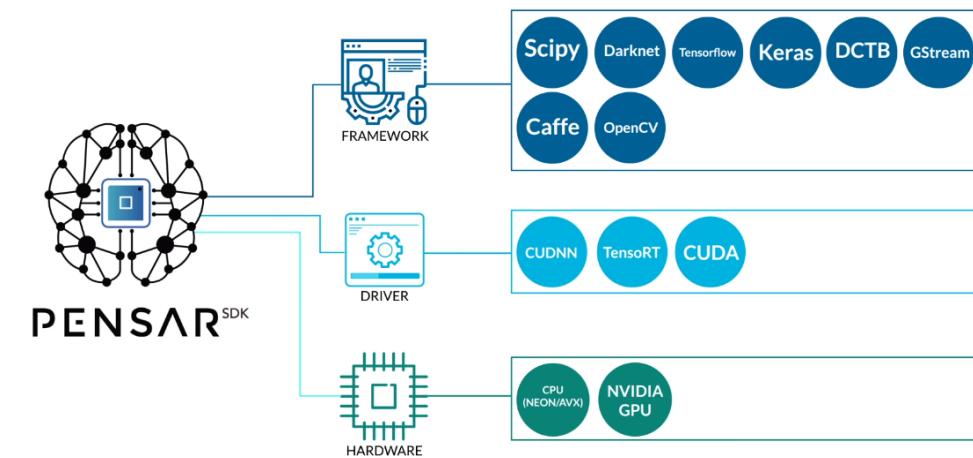
- Ingénieur IA - Computer Vision
- Villepinte, France
- 2019

- DRONE VOLT est un acteur international reconnu dans le domaine des drones civils professionnels et l'intelligence artificielle embarquée. Après l'acquisition d'Aérialtronics, Drone Volt a lancé une nouvelle caméra intelligente, **PENSAR**.
- Descriptif du produit : <https://pensarsdk.com/>

- En tant qu'ingénieur IA - Computer Vision, j'ai réalisé les tâches suivantes:
 - Collectes, préparation et annotation des images pour entraîner le modèle deep learning
 - Entraînement du modèle deep learning pour la détection du feu en temps réel
 - Travail d'équipe agile pour le développement du SDK de la caméra
 - Rédaction de la documentation technique de la solution IA
 - Formation de l'équipe commerciale du DRONE VOLT sur l'intelligence artificielle
 - Présentation de la solution aux clients et investisseurs du groupe



A NEW INTERFACE FOR AI APPLICATIONS DEVELOPMENT



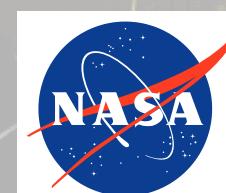


- Ingénieur Machine Learning - Traitement d'image

- Orsay, France

- 2015-2018

- Porteur du projet:





James-Webb Space Telescope (JWST) est un télescope spatial servant d'observatoire fonctionnant principalement dans l'infrarouge, développé par la NASA avec la participation de l'Agence spatiale européenne (ESA) et de l'Agence spatiale canadienne (ASC). Plus grand et plus onéreux télescope spatial à son lancement, le JWST est conçu pour poursuivre les travaux du télescope spatial Hubble,

Les principales objectifs de la mission **JWST** est d'étudier la formation et évolution des galaxies, de plus, la compréhension de la formation des étoiles et les systèmes planétaires.

$$\phi(\alpha, \beta, \lambda)$$

Light propagation direction

MIRI Imager
Optical Path
Simplified Scheme©ESA/STFC/CNRS/INAF/STScI/UMD
T. L. Smith et al.

Le JWST est doté d'un imageur infrarouge (MIRI) permettant de mesurer des observations dans un milieu infrarouge 5.6-7 microns. Divers problèmes sont rencontrés lors des mesures Limitation de la résolution spatiale par

A cause des limites physiques, la réponse optique des miroirs limite la résolution spatiale des images, et dépend de λ . De plus, le contenu spectrale de l'objet observé est intégré sur une large bande spectrale (filtre + détecteur)

- Mon objectif en tant qu'Ingénieur de traitement d'image / Machine Learning est de reconstruire objet spatio-spectral original en exploitant l'ensemble de données à différentes bandes spectrales
 - Proposition d'un modèle instrument de l'imageur infrarouge, prenant en compte la variation en λ de la PSF et de large intégration spectrale
 - Proposition d'un modèle direct par un traitement conjoint de données multi-filtre/instrument et choix du modèle linéaire par morceau
 - Augmentation de résolution spatial et de l'information spectrale

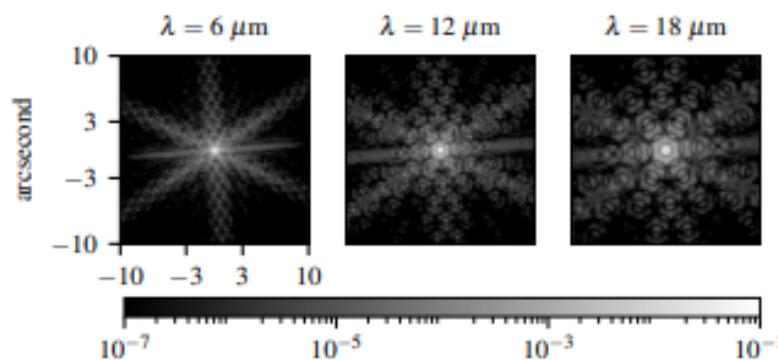


Fig. 1. Monochromatic PSF of the JWST/MIRI imager simulated at 6, 12, and $18 \mu\text{m}$ using WebbPSF [45] and displayed in logarithmic scale.

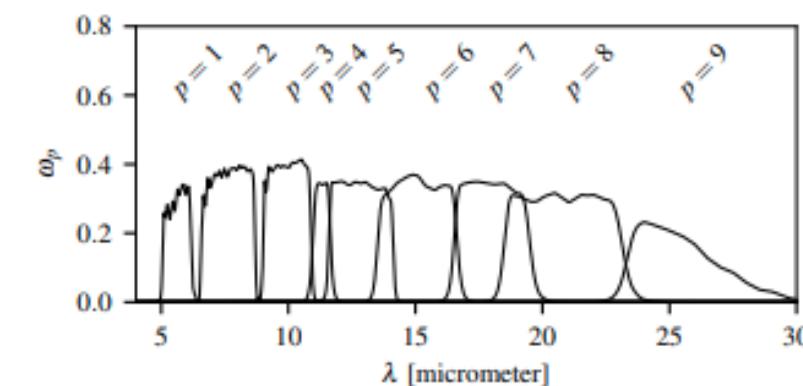
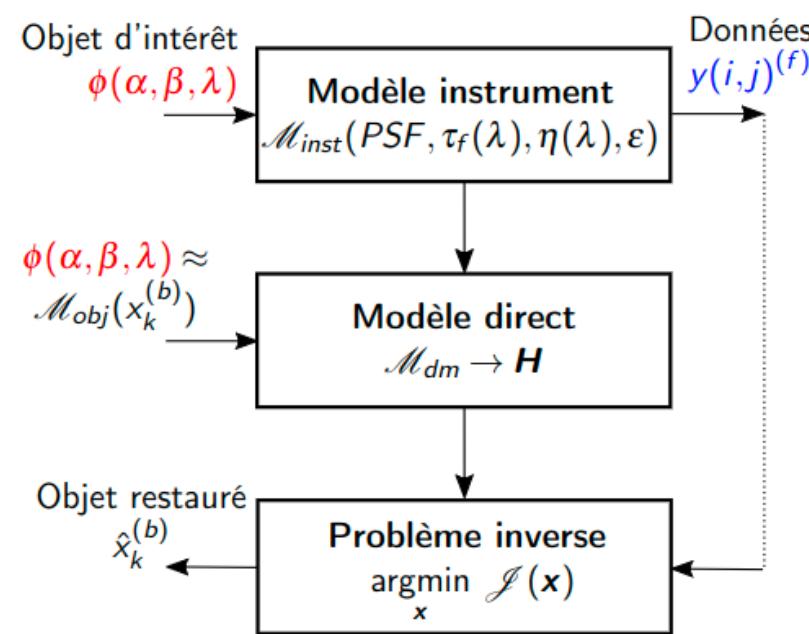


Fig. 2. Wide spectral bands ω_p of the JWST/MIRI Imager [47], from [48].



Algorithm 1 The FJMR algorithm

```

1: procedure FJMR( $H, D_h, D_v, g, \mu$ )
2:   Initialization:  $b_h = b_v = \mathbf{0}$ 
3:    $\bar{F} = \text{diag}\{\mathbf{F}, \dots, \mathbf{F}\}$ 
4:    $\bar{D}_h = \text{diag}\{D_h, \dots, D_h\}$ 
5:    $\bar{D}_v = \text{diag}\{D_v, \dots, D_v\}$ 
6:   Compute the Hessian matrix
7:    $Q \leftarrow H^T H + \mu (\bar{D}_h^T \bar{D}_h + \bar{D}_v^T \bar{D}_v)$ 
8:   Compute the block diagonal matrix  $\Upsilon$ 
9:    $\Upsilon \leftarrow \bar{F}^\dagger Q^{-1} \bar{F}$ 
10:  while criterion is not reached do
11:    1 — Compute the solution (mixing coefficients)
12:     $q \leftarrow H^T g + \mu (\bar{D}_h^T \hat{b}_h + \bar{D}_v^T \hat{b}_v)$ 
13:     $\hat{f} \leftarrow Q^{-1} q = \bar{F}^\dagger \Upsilon \bar{F} q$ 
14:    2 — Update the auxiliary variables in parallel
15:     $\hat{b}_h \leftarrow \bar{D}_h \hat{f} - \frac{1}{2} \varphi' (\bar{D}_h \hat{f})$ 
16:     $\hat{b}_v \leftarrow \bar{D}_v \hat{f} - \frac{1}{2} \varphi' (\bar{D}_v \hat{f})$ 
17:  end while
18:  return  $\hat{f}$ 
19: end procedure

```



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

- Ingénieur Machine Learning
- Alger, Algérie
- 2013

EEG

L'objectif du projet est de développer une méthode de détection des épilepsies (en état de repos), qui soit assimilable à un diagnostic précoce de la maladie. Cette méthode peut être utilisée comme un outil de diagnostic assisté par ordinateur par les praticiens de la neurophysiologie. Cela permettra d'anticiper le traitement de la pathologie et de protéger le patient en cas de crises d'épilepsie.

En tant qu'ingénieur en machine learning, j'ai travaillé sur un sujet de classification des signaux Electro-encéphalogramme (EEG) pour la détection d'épilepsie.

- Etude de l'état de l'art et comparaison entre les méthodes dans la littérature
- Préparation et analyse de données
- Concevoir une chaîne de traitement des séries temporelles basé sur le machine learning et la transformation en ondelettes sous Matlab
- Obtention de très bon résultat de classification par rapport à l'état de l'art (précision >90%)
- Pour plus de détail :
 - <https://ieeexplore.ieee.org/abstract/document/6602363>

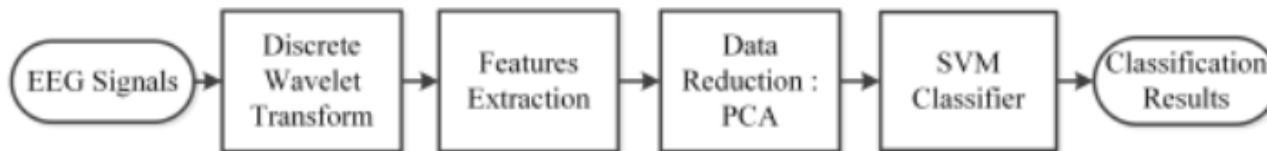


Fig. 1. Block diagram of the proposed method

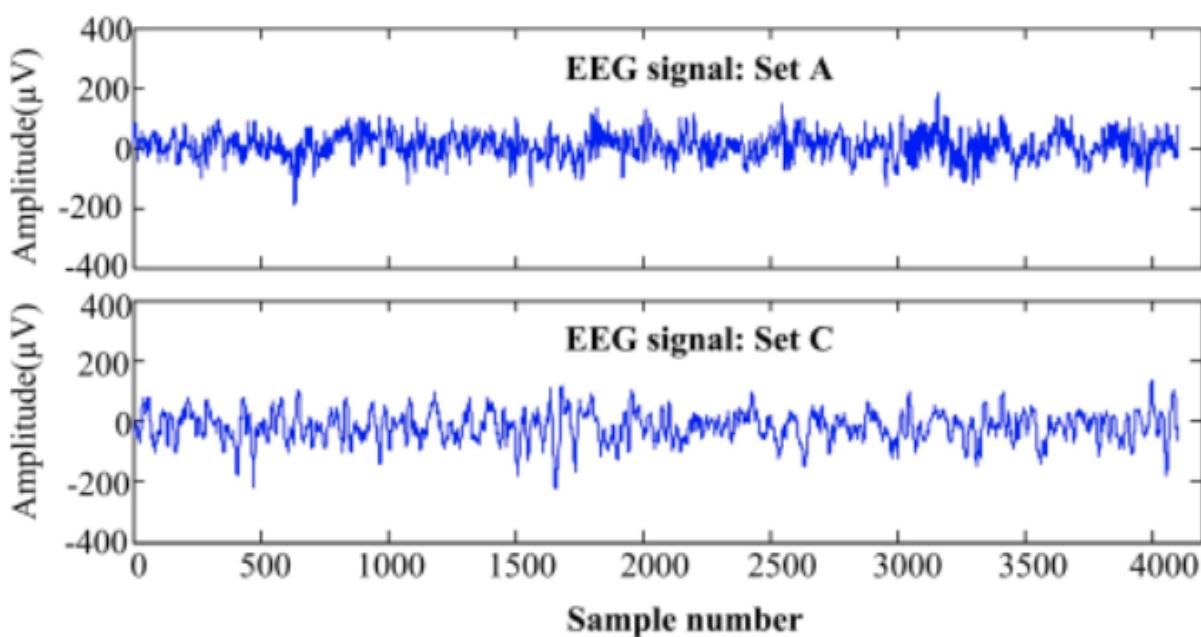


Fig. 2. Example of EEG signal from set A (Healthy) and set C (Epileptic).

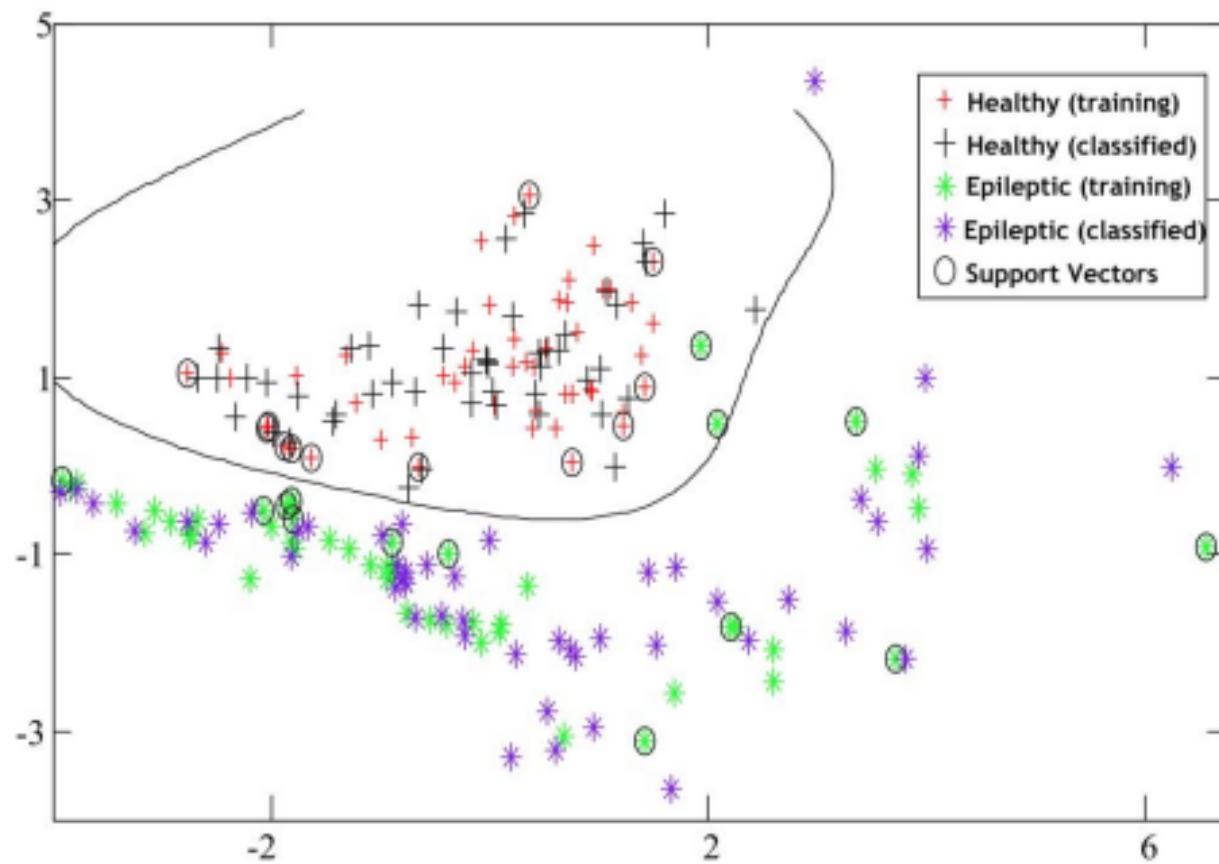


Fig. 7. Example of classification using SVM classifier and standard deviation feature (std) with 10-fold cross-validation

Publication



Publication dans des journaux scientifiques

- Hadj-Youcef, M. A., Orieux, F., Abergel, A., & Fraysse, A. (2020). Fast Joint Multiband Reconstruction From Wideband Images Based on Low-Rank Approximation. *IEEE Transactions on Computational Imaging*, 6, 922–933.
- Boucaud, A., Bocchio, M., Abergel, A., Orieux, F., Dole, H., & Hadj-Youcef, M. A. (2016). *Convolution kernels for multi-wavelength imaging*. October, 7.
<https://doi.org/10.1051/0004-6361/201629080>



Publication dans des conférence internationaux

- Hadj-Youcef, M. A., Bousbia-Salah, A., & Adnane, M. (2018). Feature selection applied to wavelet packet transform for an efficient EEG signal classification. 2018 International Conference on Electrical Sciences and Technologies in Maghreb.
- Hadj-youcef, A. (2014). Reconstruction des Images par des méthodes d'optimisation convexes et non lisses.
- Hadj-youcef, M. A., & Abergel, A. (2017). Restauration d'objets astrophysiques à partir de données multispectrales floues et une réponse instrument non stationnaire. Gdr Isis, 123(908), 2017.
- Orieux, F., Fraysse, A., & Abergel, A. (2018). Restoration from Multispectral Blurred Data with Non-Stationary Instrument Response. ii, 1–5.

- Hadj-Youcef, M. E. (2018). Spatio spectral reconstruction from low-resolution multispectral data: Application to the Mid-Infrared instrument of the James Webb Space Telescope. Université Paris-Saclay (ComUE).
- Hadj-Youcef, M. A., Orieux, F., Fraysse, A., & Abergel, A. (2018). Spatio-spectral multichannel reconstruction from few low-resolution multispectral data. 2018 26th European Signal Processing Conference (EUSIPCO), 1980–1984.
- Hadj-Youcef, M. A., Adnane, M., & Bousbia-Salah, A. (2013). Detection of epileptics during seizure-free periods. 2013 8th International Workshop on Systems, Signal Processing and Their Applications, WoSSPA 2013, May, 209–213. <https://doi.org/10.1109/WoSSPA.2013.6602363>

Publication dans Medium

- Convolutional neural network for image classification with implementation on python using pytorch
- Have you optimized your deep learning model before deployment
- Pensar-sdk
- How to write and publish a research paper
- Simplify your data science project with this tool
- How to package a python application using anaconda and docker
- Model centric vs data centric view in the age of ai
- Data visualization using redash on the cloud

Poster de conférence

- Poster GRETSI 2017

Contactez moi  / 

