



Projet 7 - Développez une preuve de concept

Aurélien Corroyer-Dulmont, PhD
Ingénieur imagerie médicale



Rappel de l'appel à projet

- **Contexte :**
 - Une méthode baseline est déjà en place pour répondre à une problématique, une veille technologique doit être faite afin de voir si un modèle innovant ne peut pas améliorer les performances
- **Objectif :**
 - Veille technologique de nouvelles méthodes de ML ou DL innovantes
 - Choix d'une nouvelle méthode à tester
 - Application de la nouvelle méthode et comparaison des performances avec la méthode baseline

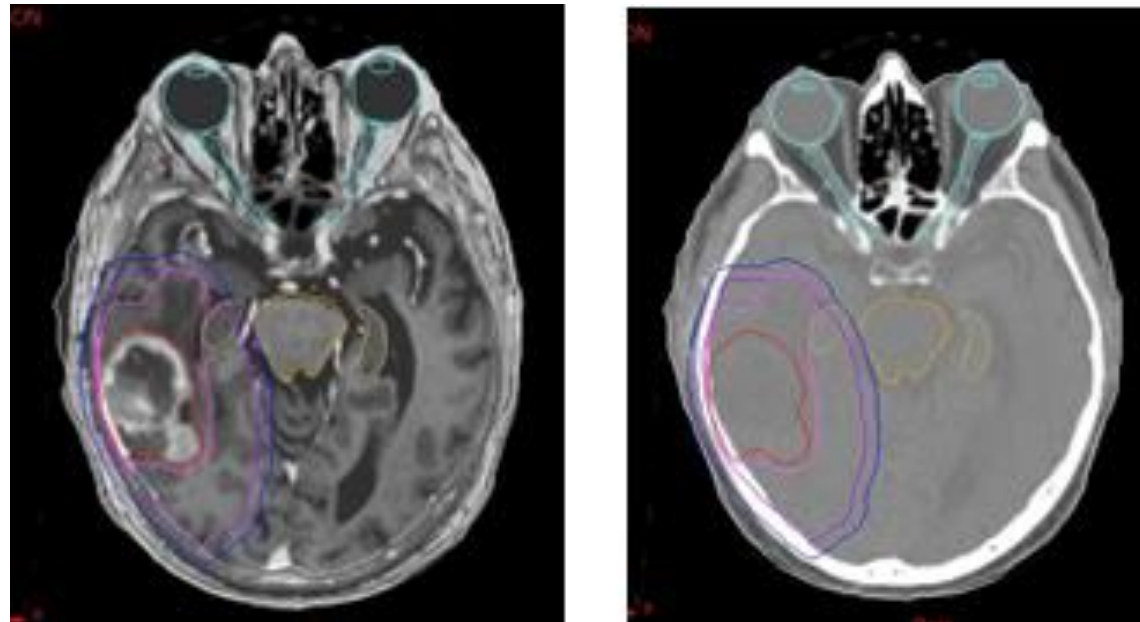
Projet choisi

“Optimisation par Machine Learning des contrôles qualités patients en radiothérapie au centre de lutte contre le cancer François Baclesse”

- **Thématique :**
 - Prise en charge d'un patient avec un cancer : étape de diagnostique



Diagnostic (imagerie IRM et CT)



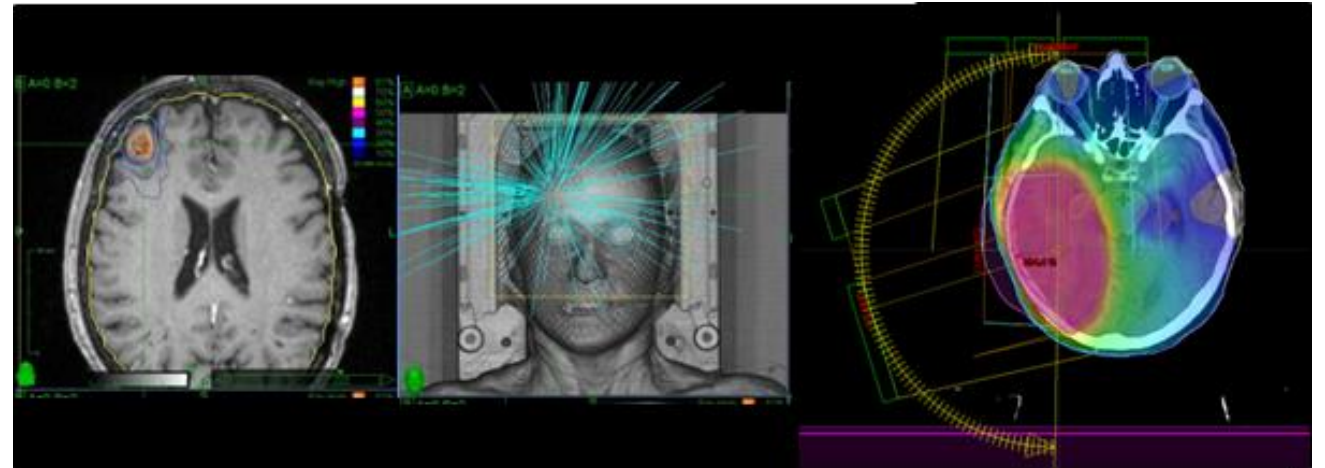
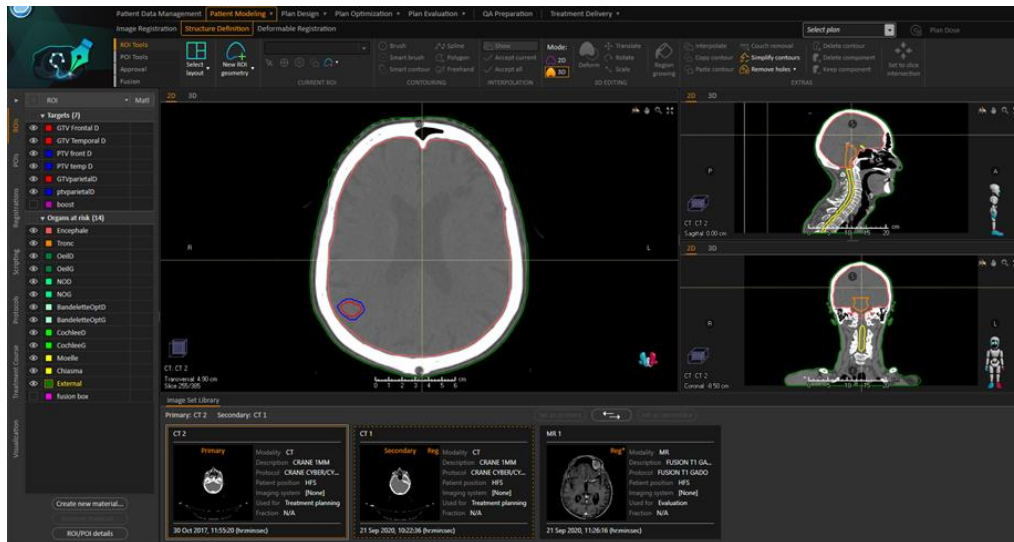
Projet choisi

“Optimisation par Machine Learning des contrôles qualités patients en radiothérapie au centre de lutte contre le cancer François Baclesse”

- **Thématique :**
 - Traitement de radiothérapie : exposer les cellules cancéreuses d'une tumeur à des rayonnements de haute énergie



Préparation du plan de RT (logiciel TPS)

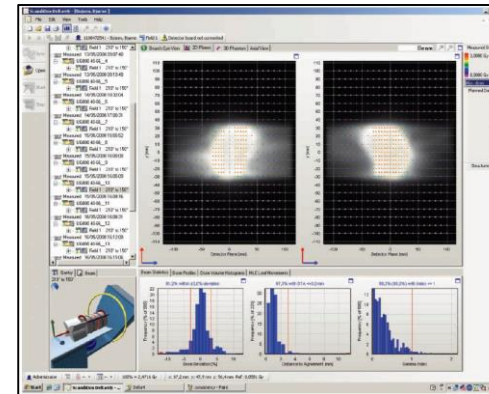
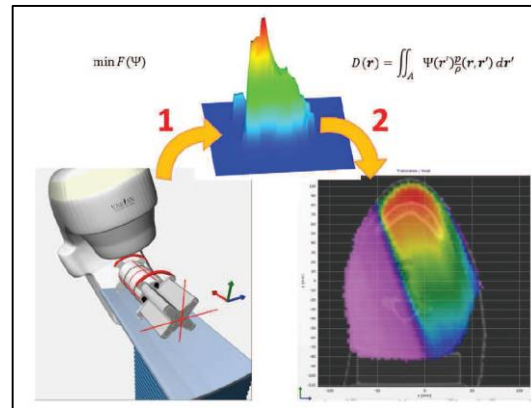


=> La dose planifiée est-elle réellement celle qui est prescrite ?

Projet choisi

“Optimisation par Machine Learning des contrôles qualités patients en radiothérapie au centre de lutte contre le cancer François Baclesse”

- Thématique :
 - Réglementation : contrôle qualité (CQ) obligatoire pour chaque patient !



=> Activité chronophage et indisponibilité de la machine de traitement pendant ce CQ

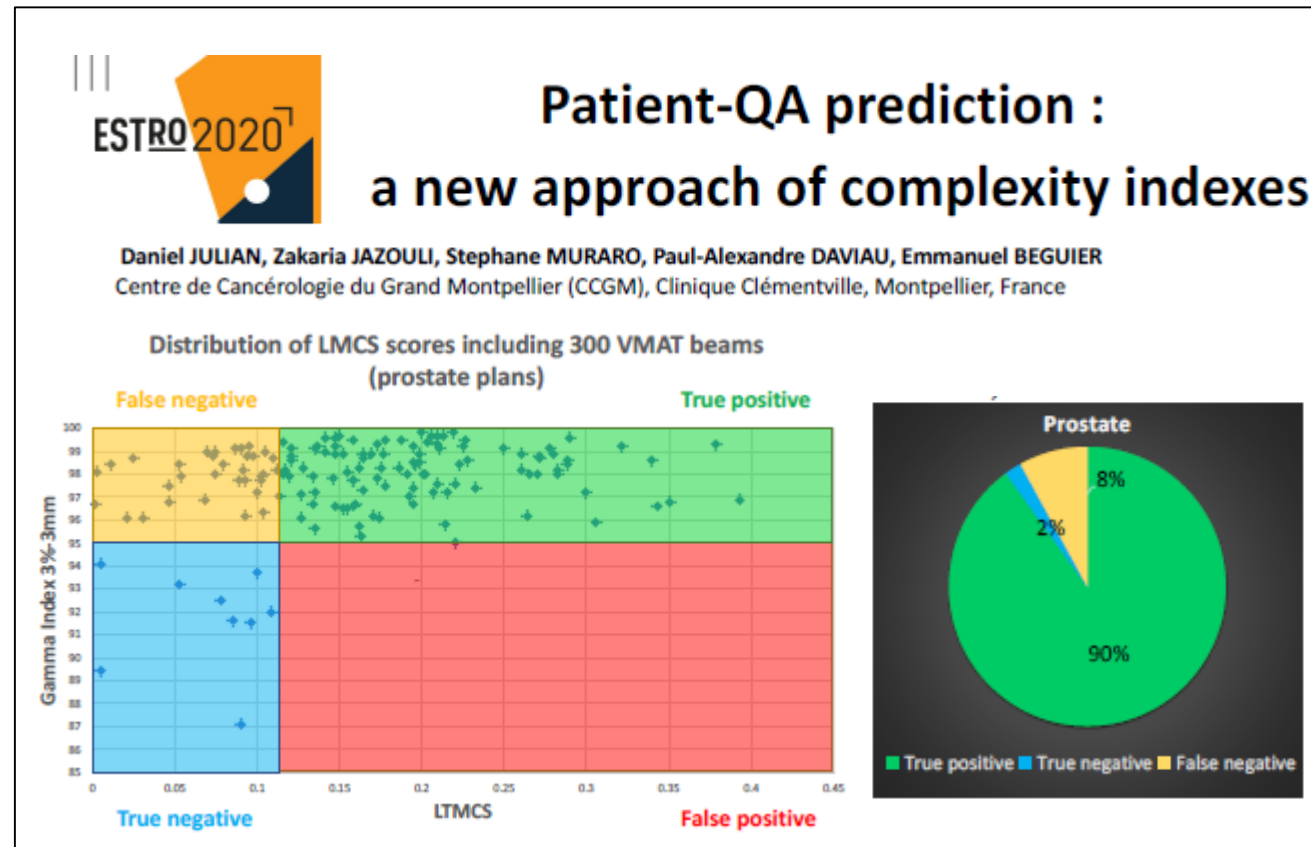
=> Méthode de prédiction des résultats ?

Projet choisi

“Optimisation par Machine Learning des contrôles qualités patients en radiothérapie au centre de lutte contre le cancer François Baclesse”

- **Etat de l’art :**
 - Recherche de méthode de prédiction des résultats
 - Utiliser des métriques issues des plan de RT ?
 - Utilisation des indices de complexité
 - Application d’un seuil sur une de ces métriques ?

=> **Simpliste, FN trop important et flexibilité pour localisations tumorales médiocres**



Jazouli, Z et al., (2021). Patient-QA prediction: a new approach of complexity indexes. *Physica Medica*

Projet choisi

“Optimisation par Machine Learning des contrôles qualités patients en radiothérapie au centre de lutte contre le cancer François Baclesse”

- Etat de l’art :
 - Méthodologie plus complexes avec du machine learning/deep learning

=> Relativement peu de littérature sur la RT VMAT

L. Simon et al. Cancer/Radiothérapie 25 (2021) 623–626

Table 1
Artificial intelligence for quality assurance in radiotherapy: different uses of machine learning for patient-specific quality assurance.

Reference	Machine learning type	Nature of training data	Data quantity	Expected outcome
Kimura et al. [3]	Convolutional neural network	Three-dimensional patient-specific quality assurance results	161 prostate plans	Detection of errors in patient-specific quality assurance results
Valdes et al. [4]	Poisson regression	Bidimensional patient-specific quality assurance results	498 intensity-modulated radiotherapy plans	Virtual patient-specific quality assurance (prediction of gamma index pass rate results)
Interian et al. [6]	Convolutional neural network	Plan complexity metrics Same data as [4]	Same data as [4]	Virtual patient-specific quality assurance (prediction of gamma index pass rate results)
Tomori et al. [7]	Convolutional neural network	Bidimensional patient-specific quality assurance results Plan complexity metrics Planning target volume and rectum volume	60 prostate plans	Virtual patient-specific quality assurance (prediction of gamma index pass rate results)
Granville et al. [8]	Support vector classifier	Three-dimensional patient-specific quality assurance results Plan complexity metrics Linear accelerator performance metrics	1620 volumetric arctherapy plans	Virtual patient-specific quality assurance (prediction of gamma index pass rate results)
Lam et al. [9]	Random forest	Bidimensional patient-specific quality assurance results Plan complexity metrics Linear accelerator performance metrics	182 intensity-modulated radiotherapy plans	Virtual patient-specific quality assurance (prediction of gamma index pass rate results)
Carlson et al. [10]	Random forest	Planned multileaf collimator positions and velocities Delivered multileaf collimator positions and velocities (dynalog)	74 volumetric arctherapy plans	Incorporating predicted multileaf collimator errors into treatment planning system computations
Chuang et al. [11]	Several regression models	Planned multileaf collimator positions and velocities Delivered multileaf collimator positions and velocities (dynalog) Gravity vector and gantry velocity	142 intensity-modulated radiotherapy, 125 volumetric arctherapy	Incorporating predicted multileaf collimator errors into treatment planning system computations

Simon, L et al., (2021). Artificial intelligence for quality assurance in radiotherapy. *Cancer/Radiothérapie*

Projet choisi

“Optimisation par Machine Learning des contrôles qualités patients en radiothérapie au centre de lutte contre le cancer François Baclesse”

- Etat de l’art :
 - Méthodologie plus complexes avec du machine learning/deep learning
 - SVM classifieur

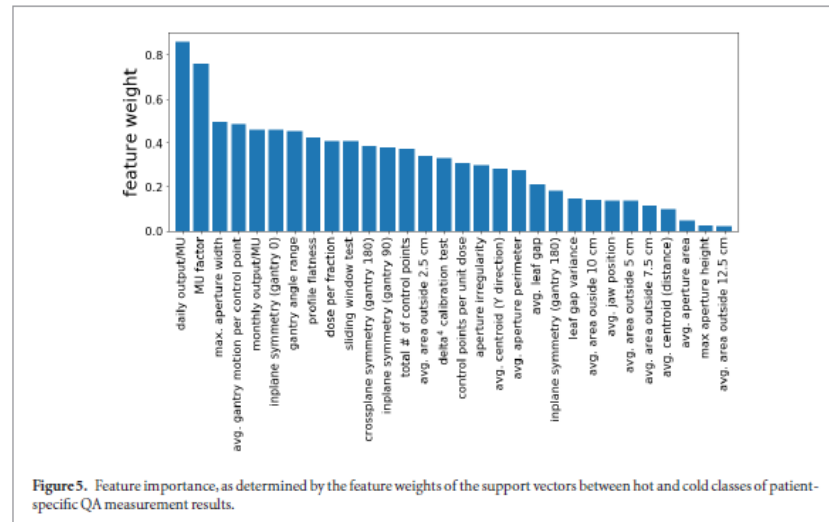
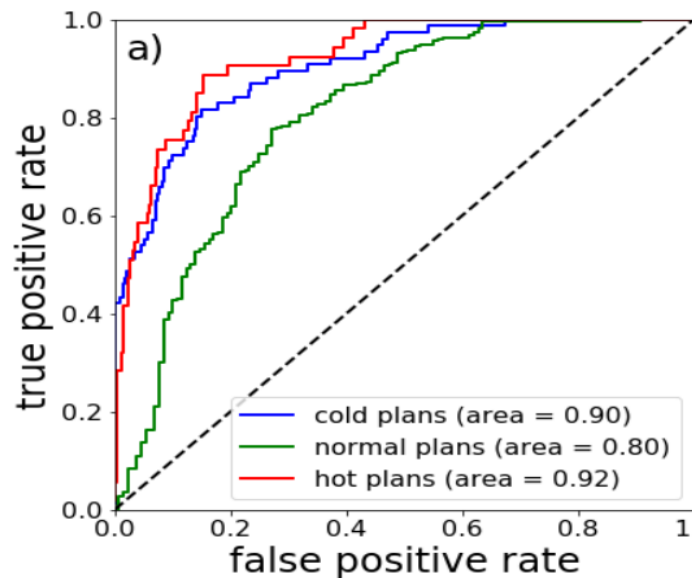


Figure 5. Feature importance, as determined by the feature weights of the support vectors between hot and cold classes of patient-specific QA measurement results.

- ⇒ **Spécificité/sensibilité moyenne**
- ⇒ **Nécessité nombre important de features potentiellement compliqué à obtenir en routine clinique**
- ⇒ **~Propre à chaque centre**

Projet choisi

“Optimisation par Machine Learning des contrôles qualités patients en radiothérapie au centre de lutte contre le cancer François Baclesse”

- Objectif :
 - Développer une solution baseline utilisant le machine learning pour la prédiction du résultat des CQ des plans de RT VMAT au centre François Baclesse
 - Comparer les performances de cette approche à des méthodes innovantes de la littérature

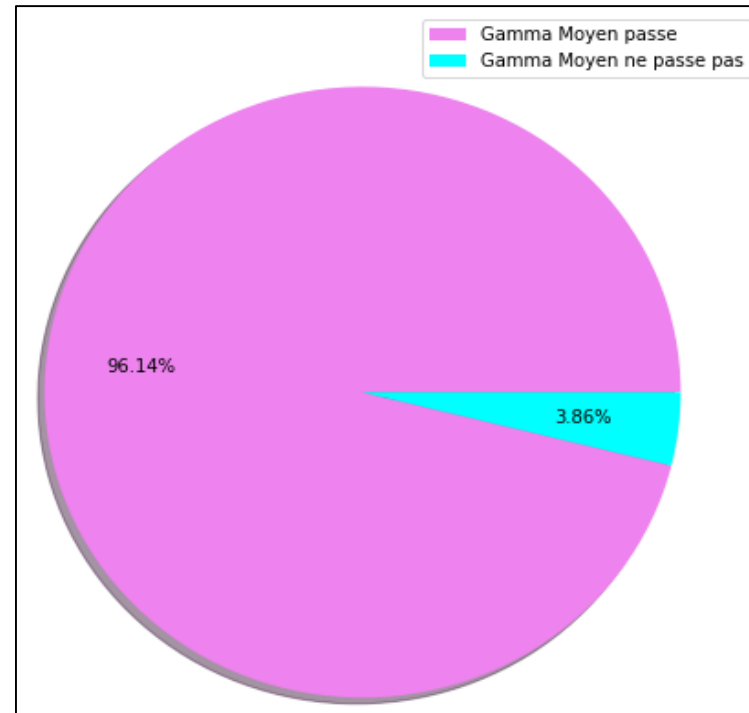
Jeu de données

- 1632 plan de RT de patients différents
 - 6 indices de complexité pour chaque plan
 - Résultat du CQ patient sur ces plans (conforme = 1, non-conforme = 0)

SAS10	MCSv	LT	LTMCS	AAV	LSV	Class_GM_10%
0.6864	0.0606	27.0829	0.0524	0.0764	0.7822	0
0.612	0.1045	26.2473	0.0908	0.1481	0.7034	0
0.4311	0.1344	32.1237	0.1128	0.1777	0.7517	0
0.544	0.1207	27.6456	0.104	0.1844	0.6491	0
0.8101	0.0773	20.0849	0.0695	0.1069	0.7245	1
0.5887	0.1187	28.0969	0.1021	0.1547	0.7653	1
0.6799	0.093	27.7451	0.0801	0.1227	0.7559	1
0.4956	0.116	34.4167	0.096	0.1636	0.7159	1
0.4156	0.1459	30.955	0.1233	0.1907	0.7642	1
0.309	0.1399	65.666	0.094	0.1831	0.763	1
0.4616	0.1411	27.8564	0.1214	0.1955	0.7155	1
0.8793	0.1033	19.8011	0.0931	0.1753	0.5935	1
0.4726	0.1427	25.4205	0.1245	0.1909	0.7421	1
0.4403	0.1234	30.1972	0.1047	0.1673	0.7367	1
0.6297	0.1184	22.1768	0.1053	0.2084	0.5649	1
0.5275	0.134	22.6211	0.1188	0.183	0.7314	1
0.4636	0.1427	24.0118	0.1256	0.2007	0.7128	1

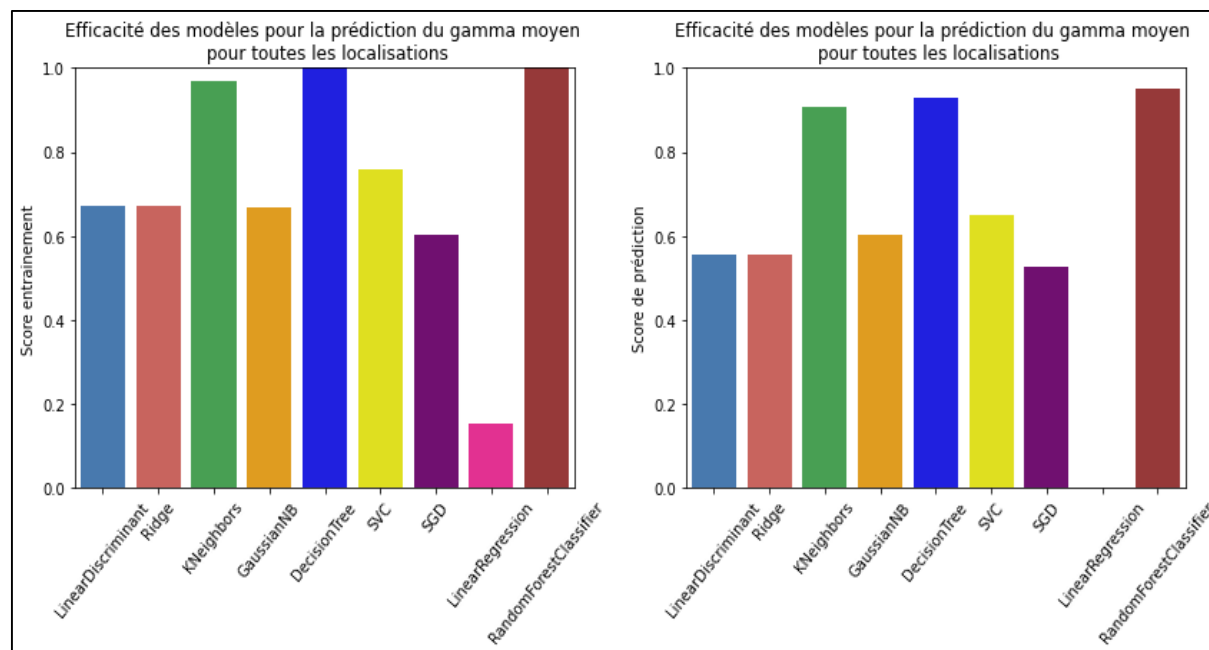
Jeu de données

- Proportion différentes dans les deux classes à prédire
 - La non-conformité des plans est beaucoup moins courantes
 - Nécessité de prendre cette information en compte pour l'efficacité du modèle



Modélisation Baseline - Machine Learning

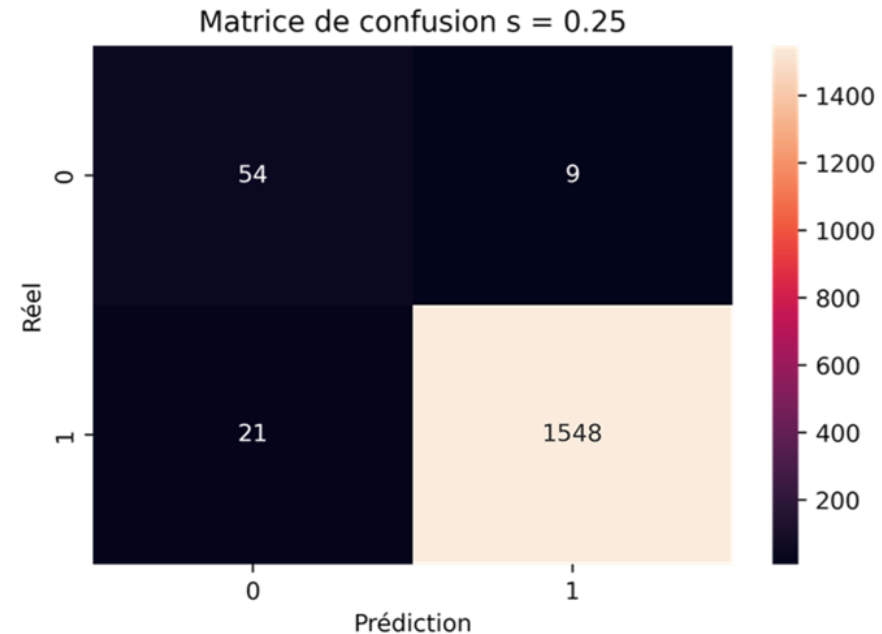
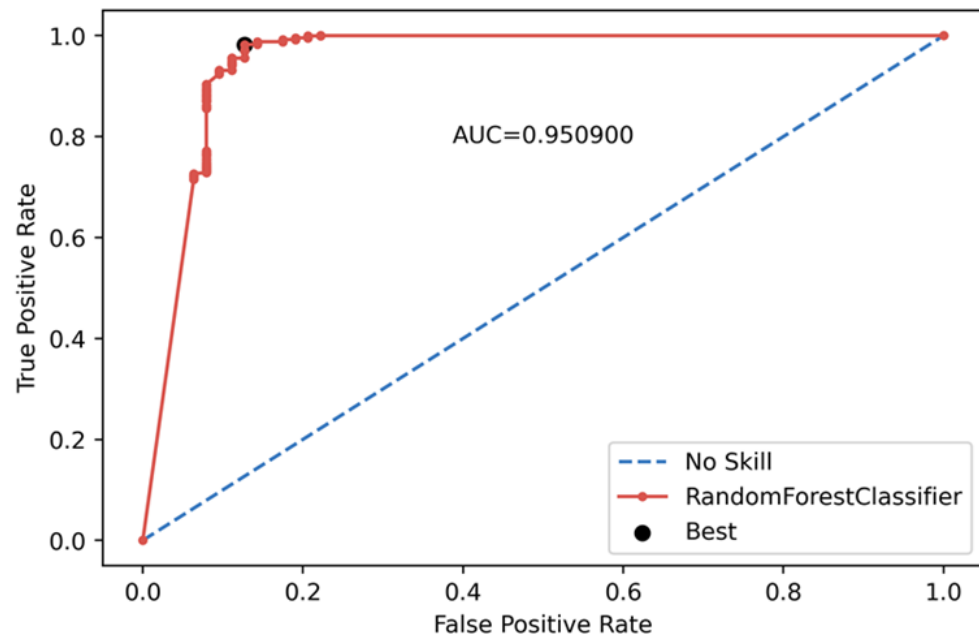
- Modèles testés :
 - LinearDiscriminant, Ridge, KNeighbors, GaussianNB, DecisionTree, SVC, SGD, LinearRegression et RandomForestClassifier
- Optimisation des hyperparamètres par Gridsearch et validation croisée
- Evaluation des performances par études du score d'entraînement et de prédiction



Le modèle RandomForestClassifier donne les meilleurs résultats

Modélisation Baseline - Machine Learning

- Performance du modèle de RandomForestClassifier :
 - Etude de la spécificité/sensibilité : Courbe ROC (Receiver Operating Characteristic) et valeur AUC (Area Under the Curve)

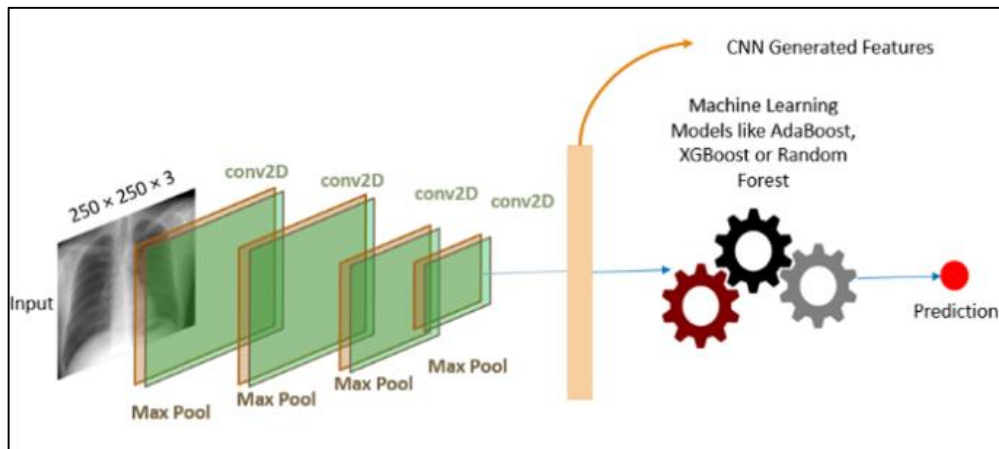


Preuve de concept : étude bibliographique

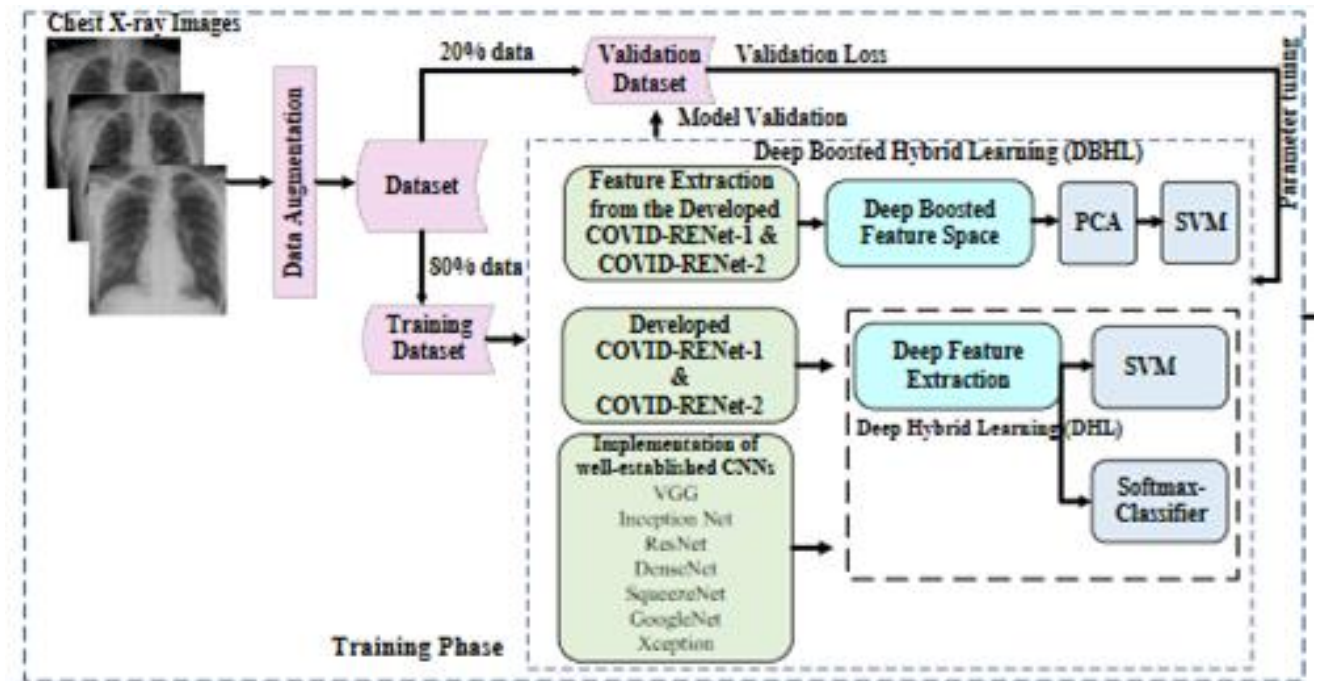
- Source bibliographique :
 - Pubmed
 - Arxiv
 - Paper with code
- Articles sélectionnés :
 - **Tushar Sarkar.** XBNet: An Extremely Boosted Neural Network. **Arxiv. (2022)**
 - <https://arxiv.org/pdf/2106.05239v3.pdf>
 - **Zhuoning Yuan *et al.*,** Large-scale Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification **Arxiv. (2021)**
 - <https://arxiv.org/pdf/2012.03173.pdf>

Modèle innovant n°1 - XBNet

- Classification - adapté à notre problématique
 - Deep Hybrid Learning : combinaison d'une architecture de deep learning avec ML classifieur



<https://towardsdatascience.com/deep-hybrid-learning-a-fusion-of-conventional-ml-with-state-of-the-art-dl-cb43887fe14>

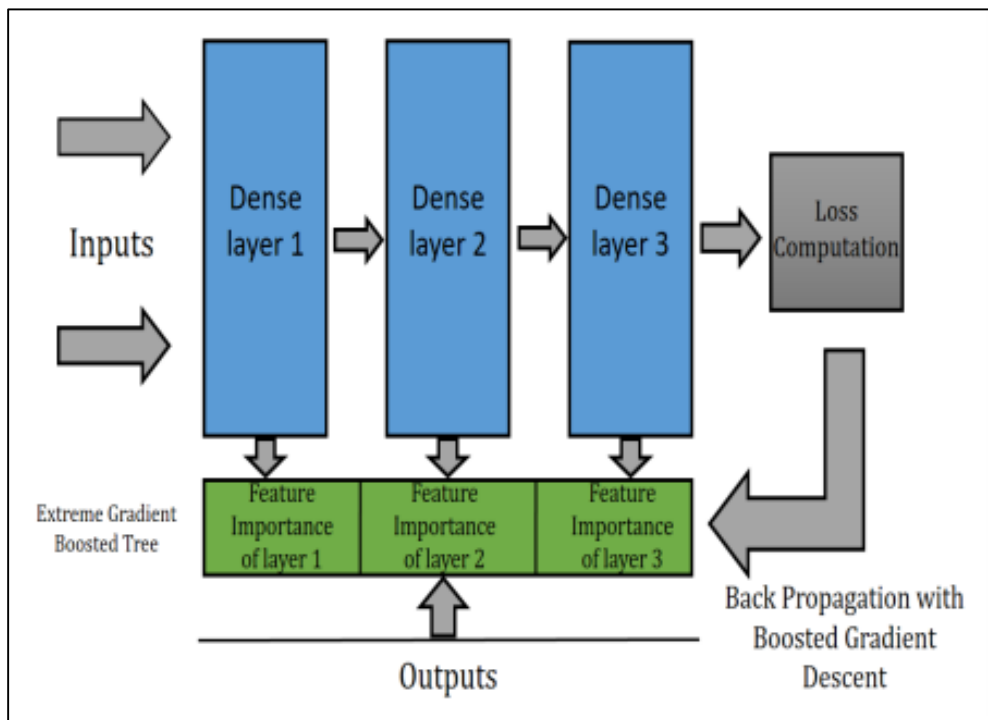


Hussain Khan et al., (2021). COVID-19 Detection in Chest X-ray Images using Deep Boosted Hybrid Learning. Computers in Biology and Medicine

Modèle innovant n°1 - XBNet

- Classification - adapté à notre problématique
 - Tushar Sarkar. XBNet: An Extremely Boosted Neural Network. *Arxiv*. (2022)

<https://github.com/tusharsarkar3/XBNet>



Our architecture creates a Sequential structure of layers with the first and last being the input and output layers respectively. The weights of the first layer are not initialized randomly but it is the feature importance of a gradient boosted tree which is trained at the time of initialization of the model.

$$w^{[1]} = \text{tree.train}(X, y).importance$$

weights of all the layers get updated according to gradient descent once and then instead of going to the next epoch of training it goes through all the layers again and updates its weights again based on the feature importance of the gradient boosted tree that is trained on the layers respectively.

$$f^{[l]} = \text{tree.train}(A^{[l]}, y^{(i)}).importance$$

where $A^{[l]}$ is the output of the l^{th} layer that is activated with an activation function $g(x)$, $y^{(i)}$ is the outputs for the mini-batch that is fed to the system.

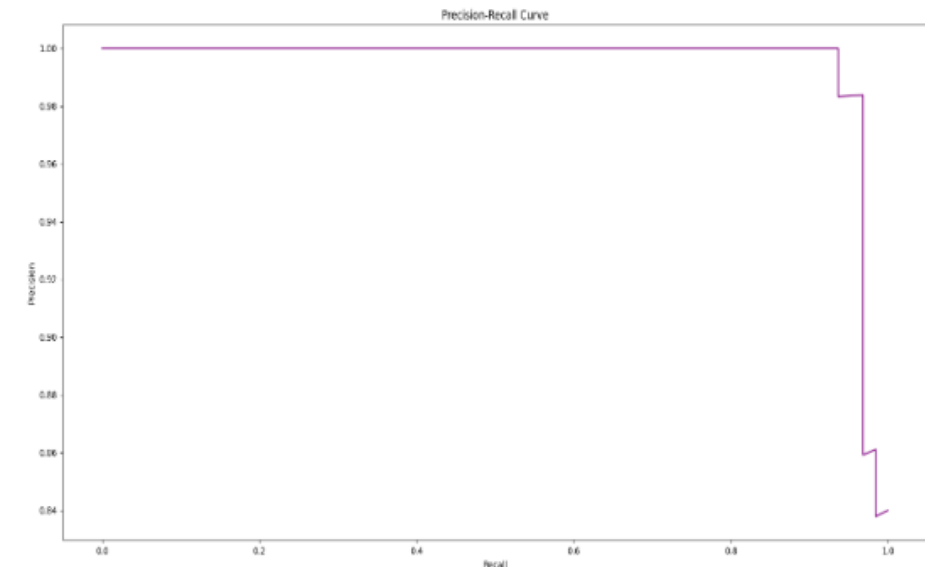
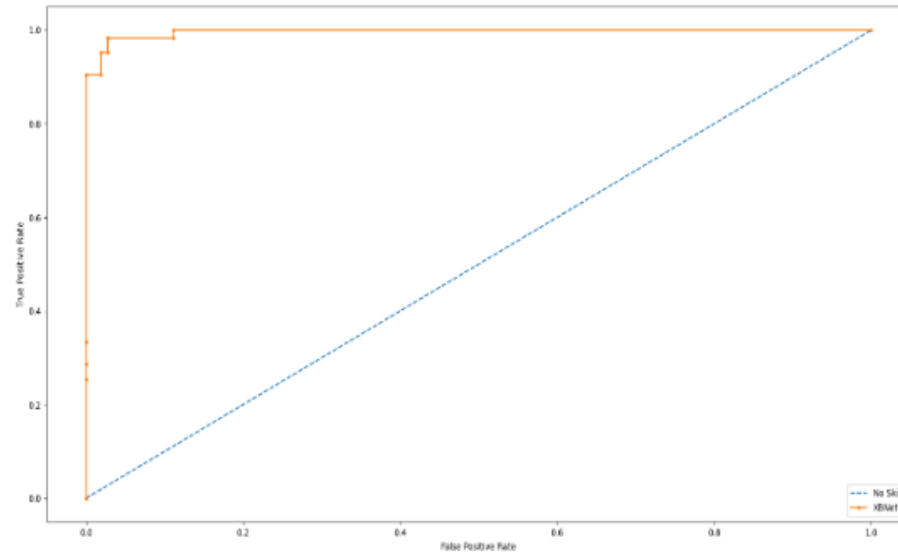
Modèle innovant n°1 - XBNet

- Classification - adapté à notre problématique
 - **Tushar Sarkar**. XBNet: An Extremely Boosted Neural Network. *Arxiv*. (2022)

Comparison with XGBOOST

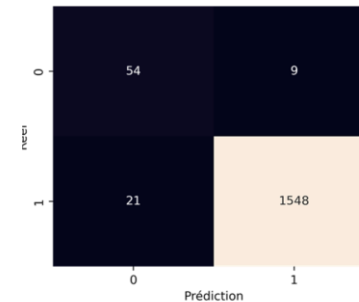
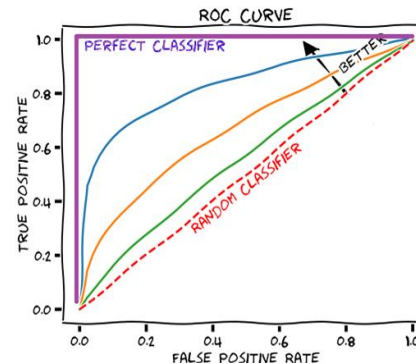
XBNET VS XGBOOST testing accuracy on different datasets

Dataset	XBNET	XGBOOST
Iris	100	97.7
Breast Cancer	96.49	96.47
Wine	97.22	97.22
Diabetes	78.78	77.48
Titanic	79.85	80.5
German Credit	71.33	77.66
Digit Completion	86.11 85.9	77.66



Modèle innovant n°2 - LibAUC

- Application médicale – adapté à notre problématique
- Entrainement/optimisation du modèle avec comme métrique principale l'AUC
 - AUC : importance dans la classification
 - AUC : importance pour disproportion
- AUC margin loss vs AUC square loss



Decomposition of Square loss:

$$A(\mathbf{w}) = E[(h_{\mathbf{w}}(\mathbf{x}) - a(\mathbf{w}))^2 | y = 1] + E[(h_{\mathbf{w}}(\mathbf{x}') - b(\mathbf{w}))^2 | y' = 1] + (1 + b(\mathbf{w}) - \bar{a}(\mathbf{w}))^2$$

$a(\mathbf{w})$ ($b(\mathbf{w})$): average score of positive data (negative data)

Margin-based Loss: (under prepration)

$$A_1(\mathbf{w}) = E[(h_{\mathbf{w}}(\mathbf{x}) - a(\mathbf{w}))^2 | y = 1] + E[(h_{\mathbf{w}}(\mathbf{x}') - b(\mathbf{w}))^2 | y' = 1] + (m + b(\mathbf{w}) - a(\mathbf{w}))_+^2$$

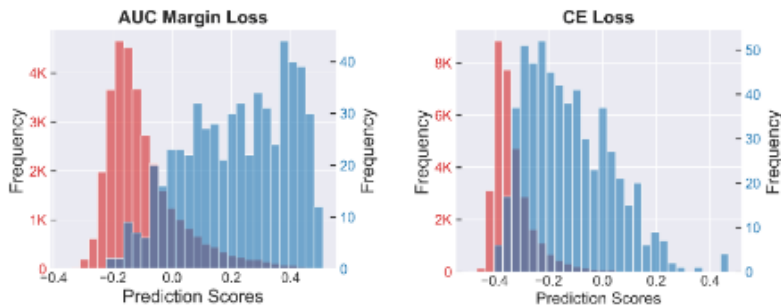
where $[s]_+ = \max(0, s)$, m is a margin parameter.

data (imratio)	CE	Focal	AUC-S	AUC-M
Cat&Dog (1%)	0.718±0.018	0.713±0.009	0.803±0.018	0.809±0.016
CIFAR10 (1%)	0.698±0.017	0.700±0.007	0.745±0.010	0.760±0.006
STL10 (1%)	0.641±0.032	0.660±0.027	0.669±0.070	0.703±0.030
CIFAR100 (1%)	0.588±0.011	0.591±0.017	0.607±0.010	0.614±0.016

Modèle innovant n°2 - LibAUC

- Application médicale – adapté à notre problématique
- Classification d'image de mélanome

<https://github.com/Optimization-AI/LibAUC>



w/ TTA(30)		
Loss	Public	Private
CE	0.9447	0.9345
Focal	0.9424	0.9303
AUC-S	0.9502	0.9364
AUC-M	0.9503	0.9393
AUC-S (Meta)	0.9501	0.9409
AUC-M (Meta)	0.9520	0.9423
Our Submission	0.9685	0.9438

Table 2: Summary of Medical Classification Tasks.

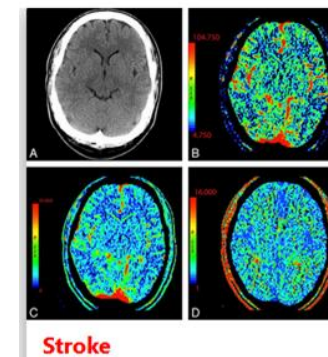
Dataset	Image Domain	Imratio	# Training
CheXpert	Chest X-ray	20.21%	224,316
Melanoma	Skin Lesion	7.1%	46,131
DDSM+	Mammogram	13%	55,000
PatchCamelyon	Microscopic	1%	148,960

- Détection de lésions pulmonaires :



CheXpert is a large dataset of chest X-rays and competition for automated chest x-ray interpretation. Our Deep AUC (ROC) Maximization method has achieved the **1st place** on Stanford CheXpert Competition organized by Andrew Ng's ML group on August 2020. This competition is for automatically detecting related diseases based on Chest X-ray images.

Model	AUC	NRBC	Rank
Stanford Baseline [22]	0.9065	1.8	85
YWW [40]	0.9289	2.8	5
Hierarchical Learning [31]	0.9299	2.6	2
DAM (Ours)	0.9305	2.8	1



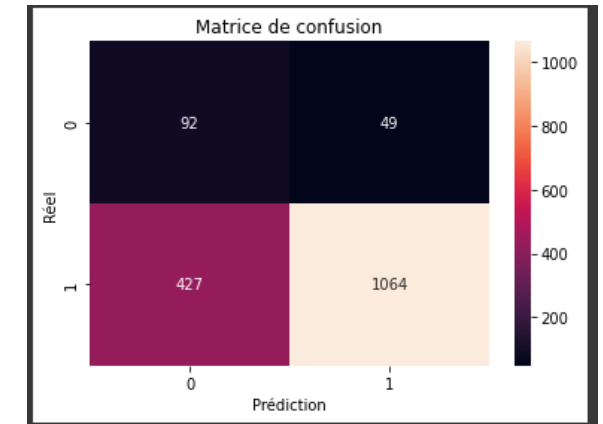
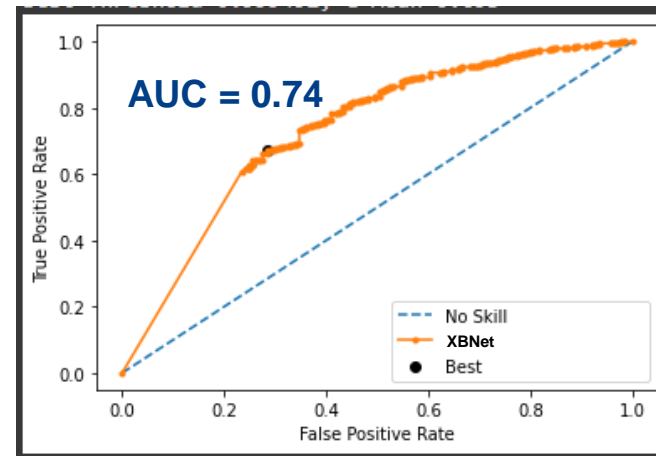
Modèle innovant n°1 - XBNet

• Résultat sur nos données

- https://github.com/AurelienCD/Formation_OCR_Ing_Machine_Learning/blob/main/XBNet.ipynb :



Paramètres optimisés : learning rate, nb de neurone (12 à 2k), Sigmoid vs Softmax



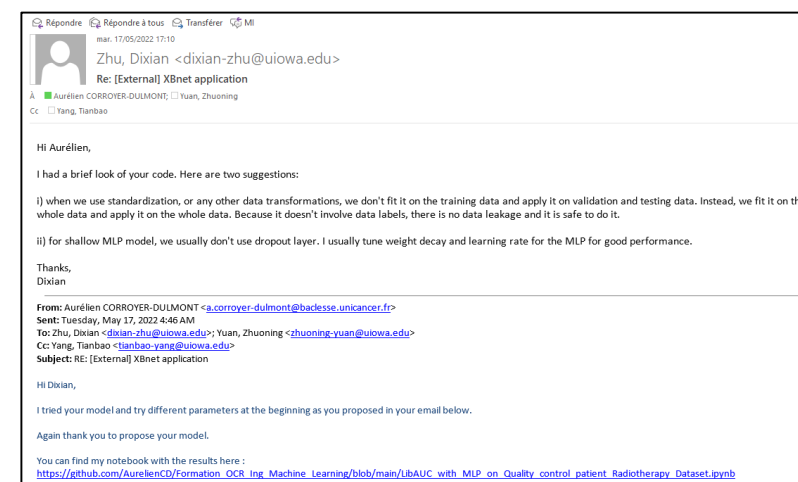
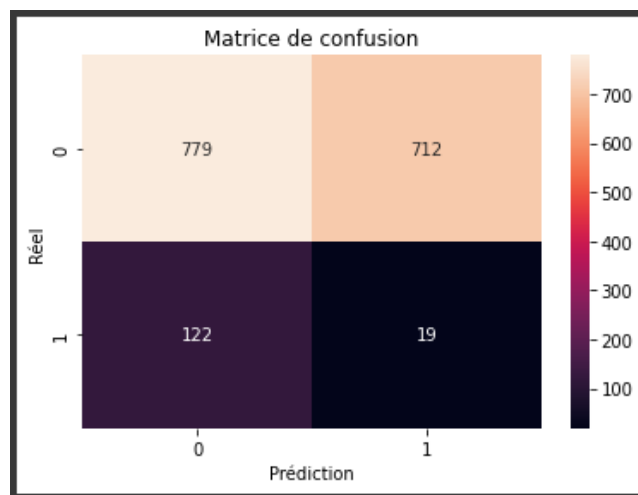
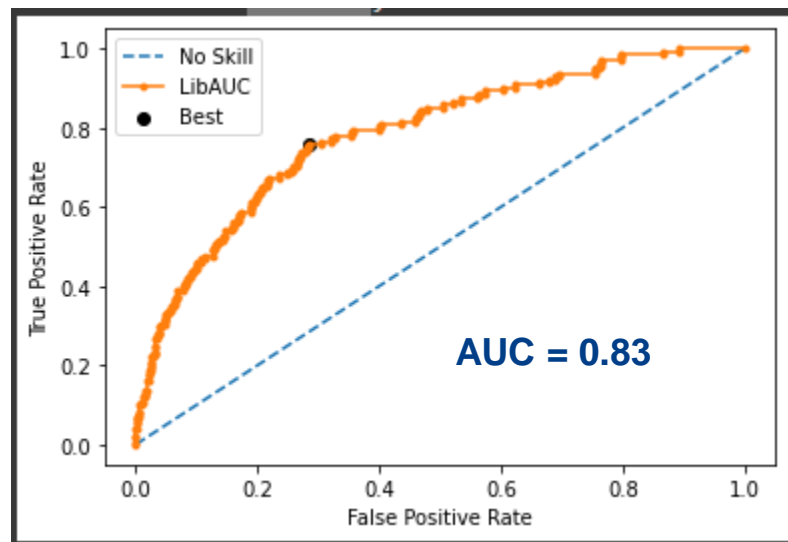
Modèle innovant n°2 - LibAUC

• Résultat sur nos données

- https://github.com/AurelienCD/Formation_OCR_Ing_Machine_Learning/blob/main/LibAUC.ipynb :

```
epoch: 94, train_loss: 0.188928, train_auc:0.766076, val_auc:0.832542, test_auc:0.732489, lr:0.003000
epoch: 95, train_loss: 0.086640, train_auc:0.785897, val_auc:0.832542, test_auc:0.732489, lr:0.003000
epoch: 96, train_loss: -0.381157, train_auc:0.780588, val_auc:0.832384, test_auc:0.732489, lr:0.003000
epoch: 97, train_loss: 0.117723, train_auc:0.783294, val_auc:0.832384, test_auc:0.732340, lr:0.003000
epoch: 98, train_loss: 0.176427, train_auc:0.768713, val_auc:0.832384, test_auc:0.732340, lr:0.003000
epoch: 99, train_loss: 0.093677, train_auc:0.754099, val_auc:0.832225, test_auc:0.732340, lr:0.003000
```

Paramètres optimisés : learning rate, gamma, weight decay, margin, nb de neurone (6 à 800)



CONCLUSION

- **Rappel de la problématique :**
 - Développer une solution baseline pour prédire des résultats de CQ en RT
 - Faire une recherche bibliographique pour trouver de nouvelles approches adaptées à la problématique
 - Comparer les performances de ces approches innovantes à la solution en baseline
- **Résultats :**
 - Le modèle de ML avec RFC() propose de bons résultats (AUC 0,95)
 - Deux modèles innovants de deep et deep hybrid learning ont été testés (XBNet et LibAUC)
 - Ces modèles n'offrent pas de meilleurs résultats que la solution développée précédemment (AUC = 0,74 et 0,83 respectivement)



Projet 7 - Développez une preuve de concept

Aurélien Corroyer-Dulmont, PhD
Ingénieur imagerie médicale