



Rapport

présenté et soutenu le : 27 mai 2022 par

Aurélien CORROYER-DULMONT

dans le cadre de la formation Openclassrooms

Ingénieur machine learning

pour le

Projet n°7 « *Développez une preuve de concept* »

**Titre du projet : *Optimisation par Machine Learning des contrôles
qualités patients en radiothérapie au centre de lutte contre le
cancer François Baclesse***

Formation effectuée au sein du :

Centre François Baclesse



I. Thématique et objectif	3
A. Thématique et état de l'art	3
B. Objectifs	4
II. Approche méthodologique	5
A. Jeu de donnée	5
B. Modèle de machine learning - méthode baseline	5
A. Modèle de deep learning - méthode novatrice	6
1. Modèle XBNet	6
2. Modèle LibAUC.....	6
III. Résultats : performances des modèles.....	7
A. Modèle de machine learning - méthode baseline	7
B. Modèle de deep learning - méthode novatrice	8
1. Modèle XBNet	8
2. Modèle LibAUC.....	9
IV. Conclusion	10
V. Référence	10

Figure 1 : Prédiction par application d'un seuil sur une valeur d'indice de complexité (Jazouli et al., 2021).....	3
Figure 2 : Prédiction par machine learning, performance du modèle (Granville et al., 2019) ..	4
Figure 3 : Jeu de donnée permettant la création du modèle de prédiction.....	5
Figure 4 : Jeu de donnée permettant la création du modèle de prédiction.....	5
Figure 5 : Architecture du modèle de deep hybrid learning de la méthode XBNet	6
Figure 6 : Performance du modèle LibAUC sur des datasets médicaux présentant une différence de proportion dans les classes à prédire	7
Figure 5 : Performance de l'approche baseline (modèles de machine learning) : scores d'entrainement et de prédiction.....	8
Figure 6 : Performance de l'approche baseline (RandomForestClassifier) : spécificité et sensibilité des prédictions	8
Figure 7 : Performance de l'approche novatrice n°1 (XBNet) : Spécificité et sensibilité.....	9
Figure 8 : Performance de l'approche novatrice n°1 (LibAUC) : Spécificité et sensibilité	9
Figure 10 : Discussion avec les auteurs de l'article afin d'optimiser les performances de leur modèle sur mes données	10

I. Thématique et objectif

A. Thématique et état de l'art

La radiothérapie (RT) est un traitement local qui consiste à exposer les cellules cancéreuses d'une tumeur à des rayonnements de haute énergie. La problématique est la suivante : La dose planifiée sur le logiciel de préparation des plans de RT (TPS) correspond-elle à celle délivrée au patient sur l'accélérateur ? Pour répondre à cette question, des contrôles qualités (CQ) pré-traitement sont réalisés pour chaque patient. Les CQ sont donc une activité chronophage qui représente un temps d'indisponibilité important pour les machines de traitement.

Certaines équipes de recherche dans des services de physique médicale ont donc tentées de développer des modèles de prédiction des résultats de ces CQ patients en utilisant des métriques caractéristiques des plans de RT, les indices de complexité (Simon, 2021).

La première approche consiste à regarder si une valeur seuil d'un de ces indices de complexité pourrait servir pour la prédiction des résultats de ces CQ. Jazouli et collaborateurs ont ainsi étudié le meilleur seuil du LTMCS pour permettre de n'avoir aucun faux positifs (FP) en minimisant le nombre de faux négatifs (FN) (**Figure 1**). Cette approche est cependant assez simpliste et amène à un nombre de FN assez important sans une grande flexibilité pour des données venant de localisations tumorales différentes (Jazouli, 2021).

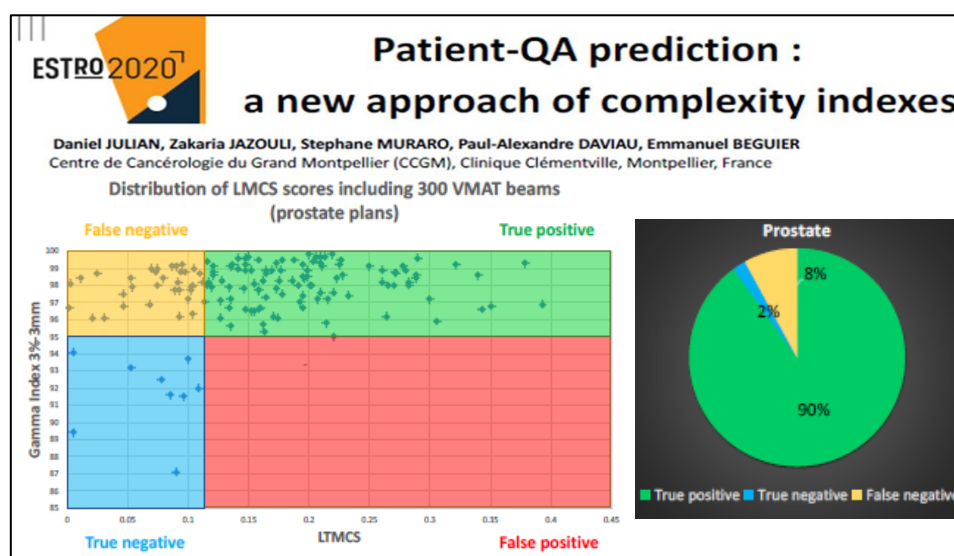


Figure 1 : Prédiction par application d'un seuil sur une valeur d'indice de complexité (Jazouli et al., 2021)

D'autres auteurs ont ainsi utilisés des méthodes de machine learning et de deep learning et ont pu montrer des résultats intéressants (Simon et al., 2021). Cependant, nombre de ces études

n'utilisaient pas des métriques simples à obtenir (Granville, 2019) ou ne s'intéressaient pas au type de traitement de RT VMAT qui est un type de traitement de RT qui nécessite le plus le CQ patient (Interian, 2018; Valdes, 2016). L'étude de Granville et collaborateurs a utilisé des plans de RT VMAT et présentent des résultats avec des taux de FP et FN relativement important comme on peut le voir sur la courbe ROC ci-dessous **Figure 2** :

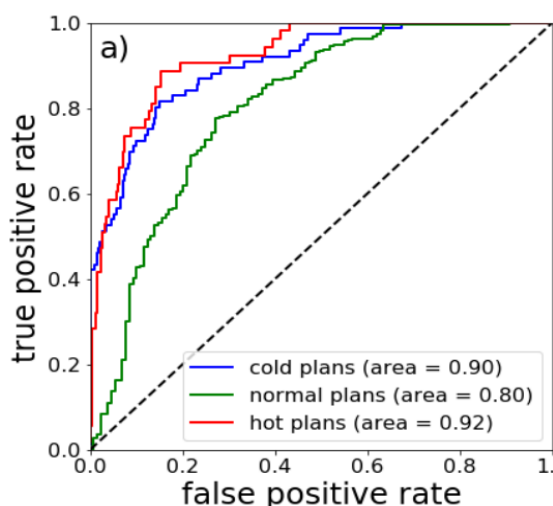


Figure 2 : Prédiction par machine learning, performance du modèle (Granville et al., 2019)

Enfin, ces modèles ne sont malheureusement adaptés qu'aux données du centre et à la machine de traitement d'où sont issues les données d'entraînement. Une application directe de ces modèles de la littérature sans ré-entraînement ou transfert learning est probablement vouée à des performances faibles.

B. Objectifs

L'objectif est de développer des algorithmes de machine (méthode baseline de ce projet) ou deep learning (méthode innovantes/preuve de concept) afin de prédire les résultats du CQ sur les plans de traitement VMAT du centre François Baclesse en utilisant différentes variables quantitatives (indices de complexité) issues des plans de RT antérieurs.

Jeu de données choisi : Pour développer ces modèles, je dispose de 1632 plans de traitement de RT avec à chaque fois six variables issues d'une analyse de la complexité de ce plan de traitement (six premières colonnes du tableau ci-dessous) et le résultat du CQ patient (variable binaire « conforme (1) » ou « non-conforme (0) ») (dernière colonne du tableau ci-dessous)

Figure 3.

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

Figure 3 : Jeu de donnée permettant la création du modèle de prédiction

NB : Il est important de noter ici que ces données sont des données de patients du centre, je ne peux donc les fournir en tant que livrables dans ce projet pour des raisons de protection des données patients.

II. Approche méthodologique

A. Jeu de donnée

Un point important dans les données que nous avons (et qui se retrouve très fréquemment dans les données médicales) vient du fait qu'il y a une différence de proportion entre les deux classes à prédire. En effet, comme on peut le voir sur la **Figure 4**, le nombre de CQ non-conforme (« Gamma moyen ne passe pas ») est uniquement de 3.86% ce qui est très faible mais correspond néanmoins à la classe que l'on veut absolument prédire. Il va falloir prendre cette information en compte pour la modélisation.

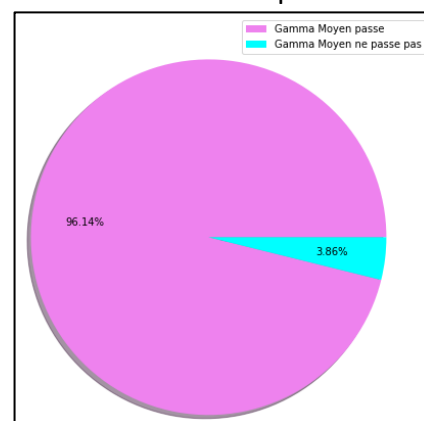


Figure 4 : Proportion des classes du jeu de donnée pour la prédiction

B. Modèle de machine learning - méthode baseline

Dans cette première approche baseline, les modèles suivant ont été testés : LinearDiscriminant, Ridge, KNeighbors, GaussianNB, DecisionTree, SVC, SGD,

LinearRegression et RandomForestClassifier. Une optimisation des hyperparamètres par validation croisée a été effectuée.

A. Modèle de deep learning - méthode novatrice

1. Modèle XBNet

Ce modèle utilise une méthode innovante de deep hybrid learning. Ce modèle est issu de l'article *XBNet: An Extremely Boosted Neural Network* Arxiv. (2021) (Sarkar, 2021). Il consiste à combiner une approche de machine learning (SGD classifieur) avec un réseau de neurone comme on peut le voir dans la **Figure 5** :

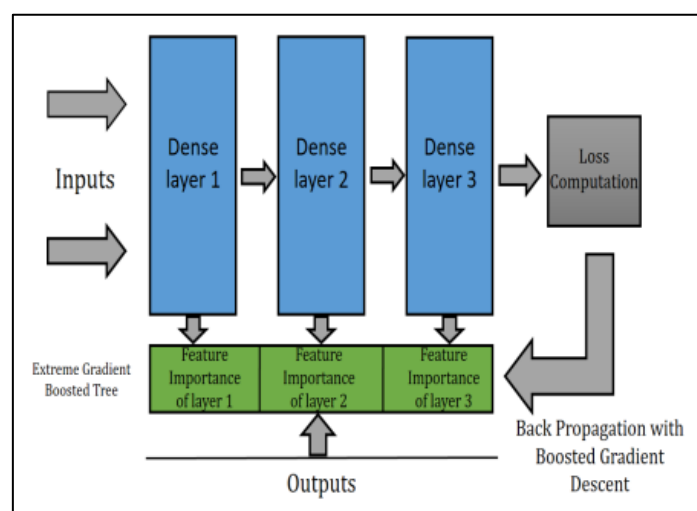


Figure 5 : Architecture du modèle de deep hybrid learning de la méthode XBNet

Tout le code ainsi qu'un tutorial est disponible sur le github de l'auteur, ceci m'ayant permis de tester son approche (<https://github.com/tusharsarkar3/XBNet>).

L'intérêt de cette méthode : cette méthode a montré de grandes performances dans des approches de classifications. Enfin cette méthode utilise du deep hybrid learning ce qui est innovant.

2. Modèle LibAUC

Cette méthode focalise l'optimisation d'un modèle de deep learning par la métrique AUC (area under the curve, relatif à la spécificité et à la sensibilité). Ce modèle est issu de l'article : *Large-scale Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification* Arxiv. (2022) (Yuan, 2022).

L'ensemble du code de cette méthode est disponible sur le github de l'auteur : <https://github.com/Optimization-AI/LibAUC>.

L'intérêt de cette méthode : ce modèle cherche à optimiser au maximum l'AUC en minimisant le nombre de faux positifs (FP) et négatifs (FN) ce qui est exactement mon principal objectif, un modèle "bon" en termes de score global de prédiction mais présentant des FP et FN n'est pas applicable dans ce projet et ne sera pas pertinent à utiliser en routine clinique. De plus, ce modèle a été développé sur des données médicales qui présentent la plupart du temps une disproportion de répartition des données dans les différentes classes que l'on veut prédire, comme on peut le voir sur la **Figure 8** :

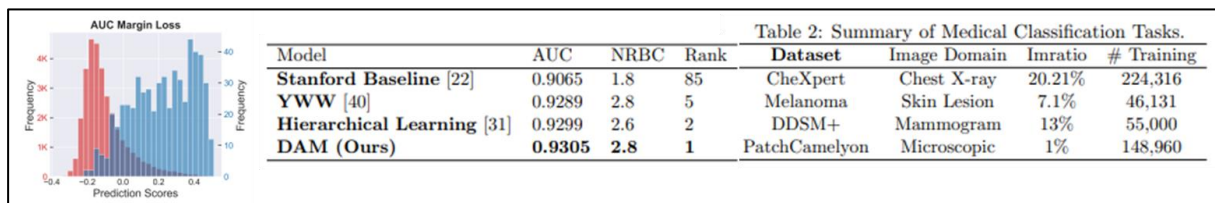


Figure 6 : Performance du modèle LibAUC sur des datasets médicaux présentant une différence de proportion dans les classes à prédire

Pour l'optimisation, ce modèle ne prend pas en compte l'AUC square loss (**Équation 1**) mais l'AUC margin loss. En effet les auteurs considèrent que cette première fonction n'est pas optimale sur des données facilement distinguables et sur des données bruitées. L'AUC margin loss va introduire une marge (notée m dans l'**Équation 1** ci-dessous) et résoudre un problème d'optimisation du minimal et du maximal plus robuste face aux cas cités précédemment.

Équation 1 : Calcul de l'AUC square loss

$$\begin{aligned}
 A_S(\mathbf{w}) &= \mathbb{E}[(1 - h_{\mathbf{w}}(\mathbf{x}) + h_{\mathbf{w}}(\mathbf{x}'))^2 | y = 1, y' = -1] \\
 &= \underbrace{\mathbb{E}[(h_{\mathbf{w}}(\mathbf{x}) - a(\mathbf{w}))^2 | y = 1]}_{A_1(\mathbf{w})} \\
 &\quad + \underbrace{\mathbb{E}[(h_{\mathbf{w}}(\mathbf{x}') - b(\mathbf{w}))^2 | y' = 1]}_{A_2(\mathbf{w})} + \underbrace{(1 - a(\mathbf{w}) + b(\mathbf{w}))^2}_{A_3(\mathbf{w})} \\
 &= A_1(\mathbf{w}) + A_2(\mathbf{w}) + \max_{\alpha} \{2\alpha(1 - a(\mathbf{w}) + b(\mathbf{w})) - \alpha^2\},
 \end{aligned}$$

Équation 2 : Calcul de l'AUC margin loss

$$\begin{aligned}
 A_M(\mathbf{w}) &= A_1(\mathbf{w}) + A_2(\mathbf{w}) \\
 &\quad + \max_{\alpha \geq 0} 2\alpha(m - a(\mathbf{w}) + b(\mathbf{w})) - \alpha^2.
 \end{aligned}$$

III. Résultats : performances des modèles

A. Modèle de machine learning - méthode baseline

Comme on peut le voir dans la **Figure 7** ci-dessous, dans mon approche baseline, l'algorithme RandomForestClassifier donne les meilleurs scores de prédiction et d'entraînement.

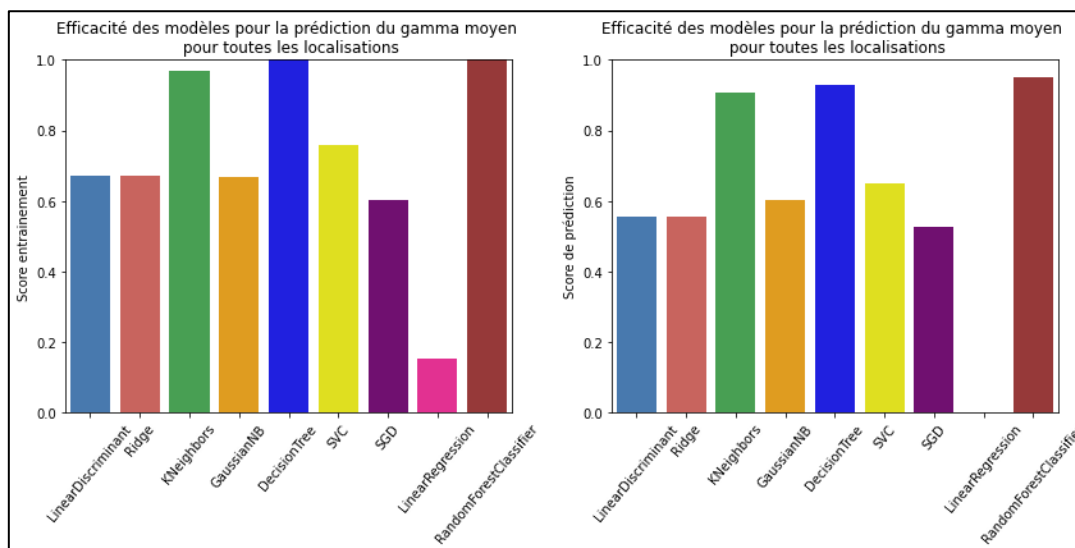


Figure 7 : Performance de l'approche baseline (modèles de machine learning) : scores d'entraînement et de prédiction

Ces scores conduisent à un AUC de 0.95 et un nombre de FP et FN très intéressant **Figure 8** :

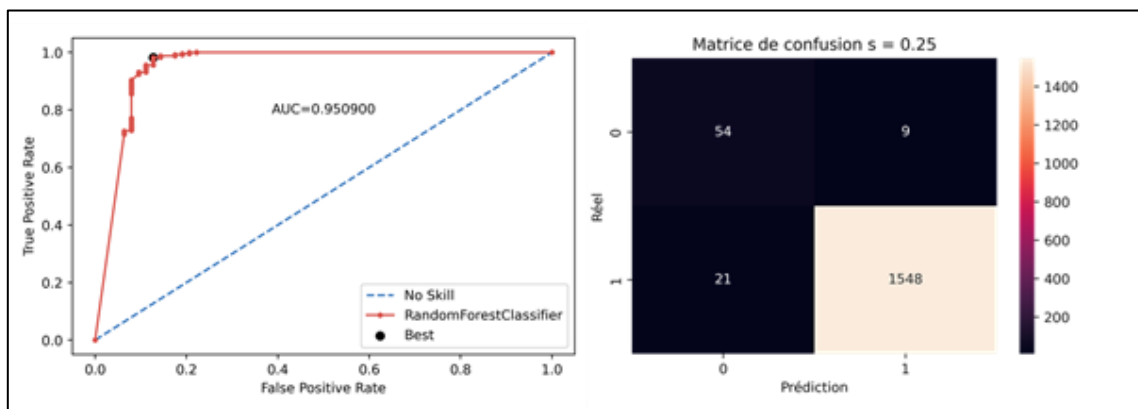


Figure 8 : Performance de l'approche baseline (RandomForestClassifier) : spécificité et sensibilité des prédictions

B. Modèle de deep learning - méthode novatrice

1. Modèle XBNNet

Les paramètres suivant ont été optimisés pour de meilleurs résultats avec mes données : learning rate, nombre de neurones par couche (12 à 2000) et les fonctions de sortie Sigmoid vs Softmax. Les autres paramètres ne sont pas modifiables. Après optimisation des hyperparamètres, le F1-score et l'accuracy les plus élevés ont été de 0.77 et 0.65 respectivement. Comme on peut le voir sur la **Figure 9**, cela se traduit par un score d'AUC et un nombre de FN et FP moyen.

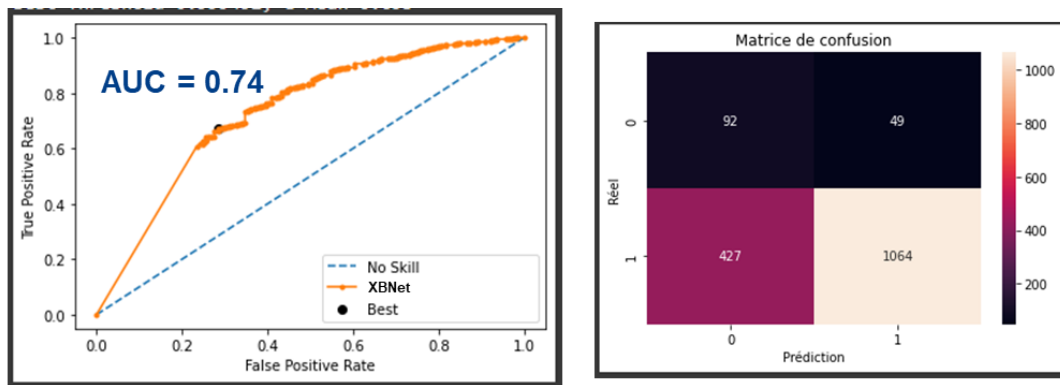


Figure 9 : Performance de l'approche novatrice n°1 (XBNet) : Spécificité et sensibilité

Une potentielle raison de la performance moyenne de ce modèle pourtant prometteur pourrait être la non prise en charge des données disproportionnées pourtant capitale dans notre cas comme vu précédemment.

2. Modèle LibAUC

Les paramètres suivant ont été optimisés pour de meilleurs résultats avec mes données : batch size, gamma, weight decay, margin, learning rate, nombre de couches et nombre de neurones par couche (6 à 800). Comme le montre la **Figure 10**, les résultats sont supérieurs à ceux du modèle de deep hybrid learning (0.83 vs 0.74). Cependant un nombre important de FP et FN est retrouvé.

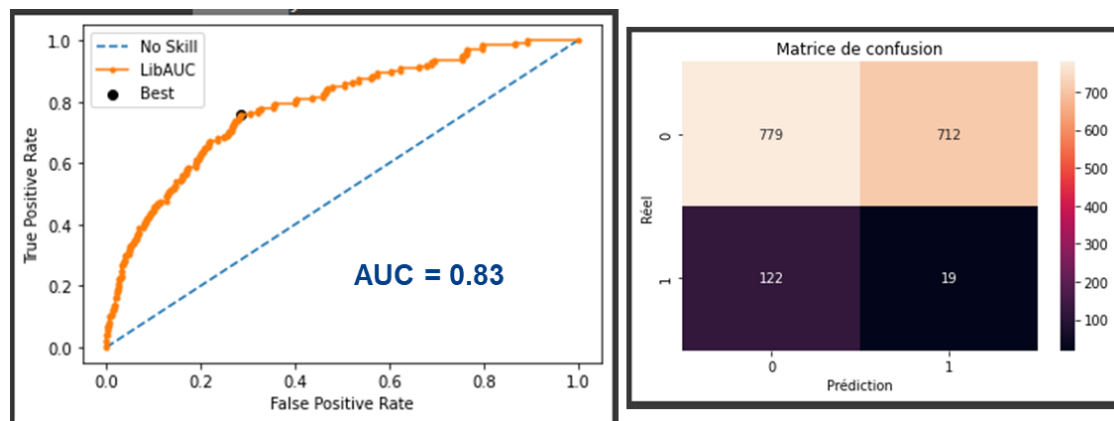


Figure 10 : Performance de l'approche novatrice n°2 (LibAUC) : Spécificité et sensibilité

Afin d'essayer d'améliorer les performances du modèle j'ai contacté les auteurs de l'article qui m'ont rapidement répondu : **Figure 11**. Nous sommes en train de voir si nous pouvons améliorer leur modèle pour être plus adapté à notre problématique.

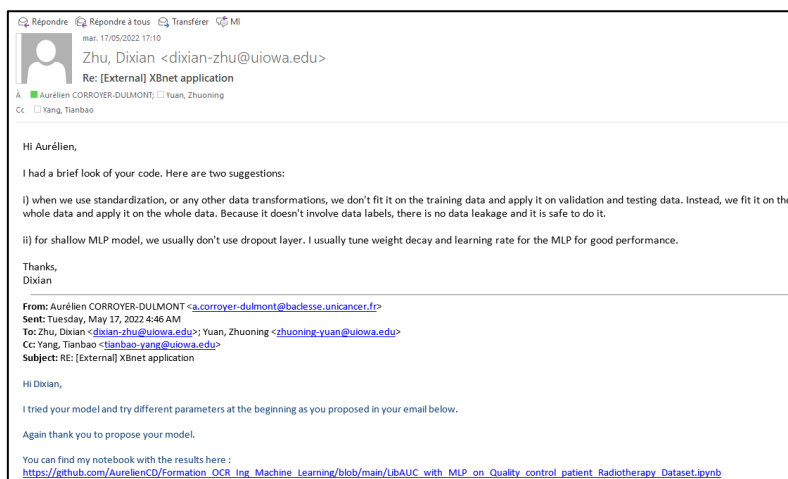


Figure 11 : Discussion avec les auteurs de l'article afin d'optimiser les performances de leur modèle sur mes données

IV. Conclusion

L'objectif de ce projet était de développer une preuve de concept sur l'utilisation d'algorithmes innovants sur une problématique réelle de mon entreprise et de comparer ces nouvelles approches à des baselines déjà obtenues. Malgré des algorithmes prometteurs au regard de ma problématique et des spécificités qu'ils proposaient, ces nouvelles méthodes ne donnent pas de résultats supérieurs à ceux obtenus avec ma méthode précédente.

Je vais être à l'affut des futures publications sur le sujet afin de voir si je peux encore améliorer mon modèle de prédiction mais en attendant j'ai déployé une API sur le cloud utilisant mon approche de RandomForestClassifier pour une utilisation en routine clinique. [Lien de l'API](#)

V. Référence

- Granville, D. A., Sutherland, J. G., Belec, J. G., & La Russa, D. J. (2019). Predicting VMAT patient-specific QA results using a support vector classifier trained on treatment plan characteristics and linac QC metrics. *Physics in Medicine and Biology*, 64(9).
- Interian, Y., Rideout, V., Kearney, V. P., Gennatas, E., Morin, O., Cheung, J., ... Valdes, G. (2018). Deep nets vs expert designed features in medical physics: An IMRT QA case study. *Medical Physics*, 45(6), 2672–2680.
- Jazouli, Z. L., Muraro, S., & Julian, D. (2021). Patient-QA prediction: a new approach of complexity indexes. *Physica Medica*, 92(218), S87.
- Sarkar, T. (2021). XBNet : An Extremely Boosted Neural Network, 1–20. Retrieved from <http://arxiv.org/abs/2106.05239>
- Simon, L., Robert, C., & Meyer, P. (2021). Artificial intelligence for quality assurance in radiotherapy. *Cancer/Radiotherapie*, 25(6–7), 623–626.
- Valdes, G., Scheuermann, R., Hung, C. Y., Olszanski, A., Bellerive, M., & Solberg, T. D. (2016). A mathematical framework for virtual IMRT QA using machine learning. *Medical Physics*, 43(7), 4323–4334.
- Yuan, Z., Yan, Y., Sonka, M., & Yang, T. (2022). Large-scale Robust Deep AUC Maximization: A New Surrogate Loss and Empirical Studies on Medical Image Classification, 3020–3029.