

L'Intelligence Artificielle comme outil de décision pour la Transplantation Hépatique

Mémoire présenté par

Aïcha NUR

Pour l'obtention du Master 1 MIAGE

De l'université

Paris 1 Panthéon - Sorbonne

Année Universitaire : **2021 - 2022**

Date de soutenance: **5 septembre 2022**

Directeur de mémoire : **Irina RYCHKOVA**

Membre du jury: **Nicolas HERBAUT**

Table des matières

Remerciements	3
Résumé	3
Introduction	4
Méthodologie de la recherche	6
Background : La Transplantation Hépatique (TH), une solution médicale qui nécessite un cadrage	7
L'Intelligence Artificielle (IA) : Concepts et quelques définitions	12
Les solutions d'IA proposées dans le cadre de la TH	16
1) Prédire la mortalité en attendant la TH	18
2) Optimisation de l'attribution des organes et de la prédiction de la survie à court terme post-TH (avant réalisation de la TH)	20
a) L'appariement DR et prédiction de survie à court terme avec ANN	20
b) L'appariement DR et prédiction de survie à court terme avec RF	25
c) Comparaison ANN et RF avec des scores de prédiction de survie et LR	26
3) Prédire survie post-TH à long terme et les éventuelles complications	29
a) Prédiction de la survie sur le long terme	29
b) Prédiction des complications post-transplantation	34
i) Prédiction de l'insuffisance rénale aiguë après une TH	34
ii) Prédiction du diabète après une TH	36
iii) Prédiction d'une récurrence chez les patients transplantés	36
Conclusion	38
Glossaire	40
Table des Figures	41
Références	42

Remerciements

Je souhaite remercier tout d'abord madame Irina RYCHKOVA pour son soutien durant cette année et les précieux conseils qu'elle a pu me donner tout au long de la conception de ce mémoire d'état de l'art. Elle a su répondre précisément à mes questionnements et m'a été d'une grande aide du choix de mon sujet à la rédaction de ce mémoire.

Je remercie également ma famille particulièrement ma chère mère NUR Hiba de m'avoir encouragé tout au long de l'année et motivée comme elle l'a toujours fait.

Enfin, je souhaite remercier mon équipe de projet à Ekino, où j'effectue mon alternance en tant que consultante technique, d'avoir toujours été à l'écoute et demander de mes nouvelles concernant l'avancement de mon mémoire même s'il n'avait pas de rapport avec notre domaine d'activité.

Résumé

Depuis 15 ans, on assiste à une stagnation de l'espérance de vie des greffons en particulier sur le long terme [19], en combinant les avantages que pourrait nous apporter l'intelligence artificielle cela permettrait de prolonger davantage la durée de vie des patients mais pas seulement ...

Elle pourrait aider à réduire la pénurie d'organes qui est un problème mondiale en promouvant l'utilisation d'organes de donneurs plus âgés qui n'étaient pas assez utilisés jusqu'à présent. Cela donnera donc accès à la transplantation à plus de patients et l'on pourra sauver encore plus de vies.

L'intelligence artificielle pourrait venir changer les règles établies actuellement par les systèmes d'allocation des organes et améliorer l'espérance de vie grâce à une meilleure estimation de survie des patients post-TH à court (≤ 1 an), moyen (entre 1 et 5 ans) et long terme (> 10 ans), à une optimisation de l'appariement DR et à une prédiction des éventuelles complications de la TH facilitant ainsi les prises en charge des patients.

Introduction

La transplantation hépatique (TH), ou greffe du foie, est une intervention chirurgicale consistant à remplacer un foie malade par un foie sain, prélevé sur un donneur.

Les résultats de la TH se sont nettement améliorés au cours des 3 dernières décennies [19] notamment grâce aux progrès de la technique chirurgicale, aux médicaments immunosuppresseurs et à une meilleure connaissance des mécanismes impliqués dans la préservation de l'organe. Mais derrière ces bonnes nouvelles, se cache une longue liste d'attente de patients toujours plus nombreux face à une pénurie d'organes mondiale et la mortalité des candidats en attente d'un donneur [17].

Il est donc important de trouver des solutions pour gérer cette ressource rare qu'est le greffon en s'assurant d'obtenir les meilleurs résultats post-TH aussi bien sur le court terme que sur le long terme, évitant ainsi des échecs de la TH (risque de perte du greffon voire du patient).

Divers scores ont été développés afin de prédire la mortalité en liste d'attente et la survie post-TH. Pour certains de ces scores, l'objectif principal est de diminuer la mortalité en liste d'attente sans se préoccuper des résultats post-TH. C'est le cas du score Model for End-Stage Liver Disease (MELD) qui est le système de priorisation le plus répandu de nos jours reposant sur le principe du "sickest first" (le plus malade d'abord). Pendant que pour d'autres scores comme le score Survival Outcomes Following Liver Transplantation (SOFT) ou le score Balance of Risk (BAR), le but est d'obtenir le plus grand bénéfice de survie, ce qui revient à accorder la priorité aux patients au meilleur profil clinique sur la liste d'attente appariés aux "meilleurs" greffons délaissant ainsi les plus malades au profit de meilleurs résultats de la TH. Malheureusement, aucun de ces scores n'a été en mesure de combiner l'urgence d'un patient d'être transplanté avec le meilleur bénéfice de survie parmi les candidats de la liste d'attente [4]. En effet, une diminution de la mortalité dans la liste d'attente (constaté depuis l'instauration du score MELD dans les systèmes d'attribution des organes), conduit en contrepartie à de moins bons résultats de survie post-TH; et inversement, l'obtention de meilleurs résultats de survie peut affecter la possibilité d'être transplanté pour les plus malades augmentant ainsi la mortalité en liste d'attente.

Pour pallier ces imperfections ainsi qu'à la pénurie d'organes, l'attribution d'un donneur au candidat le plus approprié sur la liste d'attente communément appelé l'appariement Donneur-Receveur (DR) doit être optimisée. Une meilleure prédiction de la survie en liste d'attente du candidat, de la survie du greffon et/ou du receveur post-TH (à court, moyen et long terme) mais aussi des éventuelles complications qui pourraient survenir après la greffe doit être obtenue afin d'améliorer la prise en charge des patients tout au long de ce processus difficile qu'est la transplantation.

De nos jours, l'Intelligence Artificielle (IA) est utilisée dans de nombreux domaines y compris le domaine médical avec l'utilisation d'algorithmes de Machine Learning (ML) notamment pour l'analyse d'imagerie, le diagnostic de certaines maladies et l'aide à la décision clinique [6]. On peut donc naturellement se demander si l'IA peut intervenir dans les différentes phases de la TH (avant, pendant et après) en tant que support au médecin tout le long de la prise en charge du patient : de l'inscription du patient sur la liste d'attente à la prise en charge post-TH sur le court et long terme.

Ce qui nous amène à la problématique suivante : **L'Intelligence Artificielle peut-elle jouer un rôle d'assistant dans le cadre de la Transplantation Hépatique ?**

Dans la première section, nous allons voir divers scores proposés pour cadrer le système d'attribution des greffons et prédire la survie. Dans la section suivante, nous présenterons brièvement quelques algorithmes ML afin de fournir quelques définitions. Enfin, nous verrons les solutions d'IA proposées dans le cadre de la TH afin d'optimiser l'appariement DR, de mieux prédire la survie avant et après la TH ainsi que les éventuelles complications. Puis nous finirons par une conclusion visant à répondre à notre problématique.

Méthodologie de la recherche

J'ai principalement utilisé le moteur de recherche Google Scholar pour retrouver les articles scientifiques. J'ai aussi utilisé [Miage Scholar](#) en utilisant comme mots clés : "liver transplantation", "donor-recipient pairing", "Artificial intelligence", "machine learning", "long term survival" avec comme limite de publication 2017(inclus), 89 résultats ont été récupérés. Les articles qui concernaient le diagnostic et traitement des maladies du foie en général ont été exclus. Seuls les articles en rapport direct avec la TH ont été retenus en l'occurrence les articles qui s'intéressent aux prédictions de survie, de risque d'effets indésirables et à l'appariement DR par le biais de l'IA dans le cadre de cette transplantation.

Pour mon sujet j'ai eu le plaisir de trouver pas mal d'articles récents, voire même de cette année 2022, on peut donc dire que c'est un sujet en plein cœur de la recherche actuelle.

Au début, j'ai eu des difficultés à trouver les articles qui correspondaient à mon thème car mes mots-clés n'étaient pas assez ciblés. Puis au fur à mesure de mes recherches en précisant par exemple "scores d'attribution des greffons en France", "donor-recipient pairing" ou "liver allocation system" cela m'a permis de trouver des ressources pour chaque section de mon analyse ainsi que pour la rédaction de la partie background.

Concernant la partie background, les articles étaient naturellement moins récents car ils relevaient des solutions mises en place avant l'IA.

Finalement, 21 articles scientifiques ont été sélectionnés pour mon mémoire d'état de l'art provenant des bases comme Elsevier, IEEE, ScienceDirect, SpringerLink, PubMed dont certains ont été récupérés grâce au phénomène de snow-bowling (retrouvés dans les références d'autres articles).

Background : La Transplantation Hépatique (TH), une solution médicale qui nécessite un cadrage

En raison de la pénurie d'organes, un système d'attribution des organes doit être mis en place afin d'assurer la bonne gestion de cette ressource collective rare qu'est le greffon tout en diminuant la mortalité des patients sur la liste d'attente.

En France, le système de répartition des organes repose sur 2 principes :

- le principe d'équité (justice)
- le principe d'utilité (efficience)

Le principe d'utilité consiste à attribuer le greffon au patient qui en a un besoin le plus urgent tout en garantissant un usage optimal du greffon.

On parle du principe "sickest-first" qui consiste à accorder une priorité au patient le plus malade et donc celui qui a le plus haut risque de décès pendant la phase d'attente de la transplantation.

Cependant, le problème est qu'en donnant toujours la priorité aux patients les plus malades cela peut entraver l'accès à des patients dont la maladie n'engage pas le pronostic vital à court terme mais qui ont tout de même besoin d'une transplantation.

L'application exclusive du principe d'utilité n'est donc pas une bonne solution et doit être liée au principe d'équité qui garantit pour chaque candidat une probabilité d'accès au greffon similaire, quelle que soit la nature de sa maladie et sa localisation territoriale.

Pour répondre à la pénurie d'organes, le recours à des organes dits à "critères étendus" (ECD - Expanded Criteria Donors) peut être proposée aux patients pour réduire son temps d'attente sur la liste mais cela relève d'une décision personnelle du patient. Les organes ECD sont des organes provenant de donneurs pour lesquels les critères de sélections ont été élargis. Par exemple il peut s'agir de donneurs âgés de plus de 60 ans ou de donneurs de plus de 50 ans ayant des antécédents d'hypertension artérielle, une créatinine (test sanguin qui montre la fonction rénale) supérieure à la normale, ou encore un décès résultant d'un accident vasculaire cérébral. La décision reviendra au receveur d'accepter ou non cet organe ECD, qui certes peut présenter un taux Donneur Risque Index (DRI - Donor risk index) plus élevé qu'un organe aux critères de sélections standards (SCD - Standard Criteria Donors) mais qui augmente tout de même son espérance de vie comparé aux patients qui restent sur la liste d'attente. Les donneurs SCD sont en général des donneurs âgés de moins de 50 ans ayant souffert d'une mort cérébrale.

Au début de la transplantation, l'allocation du greffon était entièrement entre les mains des équipes médicales. On parle d' "attribution d'équipe" qui peut être décrite comme une approche intuitive et qui repose sur l'expertise et les expériences de l'équipe médicale.

Malheureusement, ce système d'attribution n'était pas le plus efficace d'autant moins que la liste d'attente des patients ayant besoin d'une transplantation ne cessait d'augmenter tandis que la quantité de greffons ne suivait pas la même évolution.

Pour faire face à cette "pénurie d'organes", on est passé d'une "attribution d'équipe" vers une "attribution patient". L' "attribution patient" repose sur des scores d'attribution qui permettent de gérer l'allocation des greffons, plutôt que de se baser seulement sur la décision des équipes médicales dans le cadre de l'attribution d'équipe.

En France, l'attribution des greffons est réglementée par l'Agence de la Biomédecine (ABM). Dans le cadre du foie, l'attribution de ces greffons repose sur un score national d'attribution : le score Foie [22]. Ce score a été mis en place dans le but de réglementer l'allocation des greffons en s'assurant du respect des principes d'équité (durée d'attente de la TH,...) et d'utilité/efficacité (appariement HLA - Human Leukocyte Antigen, différence d'âge entre le donneur et le receveur ...).

Les scores d'attribution sont évolutifs et discutés en permanence pour savoir quels critères prendre en compte dans le calcul du score, si ces derniers jouent un rôle déterministe dans le bon appariement de l'organe entre le donneur et le receveur.

Les greffons sont donc alloués en fonction du score : en cas de dérogation au score, l'équipe de greffe doit justifier son choix.

La TH se compose de 3 phases :

- une phase pré-transplantation (ou préopératoire)
- la transplantation en elle-même (ou peropératoire)
- une phase post-transplantation (ou postopératoire)

Chacune de ces phases comporte une multitude de variables qui interagissent entre elles du côté donneur et receveur. Pendant la phase pré-transplantation, il n'y a que les variables du receveur (âge, sexe, comorbidité, score MELD, taux d'albumine, poids ...). Pendant la phase de la transplantation, il y a les variables du receveur et du donneur (âge, cause du décès, taux de créatinine, stéatose hépatique ...) ainsi que les variables liées à la transplantation en elle-même (temps d'ischémie froide, condition de réalisation de la chirurgie ...). Pendant la phase post-transplantation, les variables du donneur et du receveur convergent ainsi que des variables externes et internes telles que l'immunosuppression, les autres maladies du receveur.

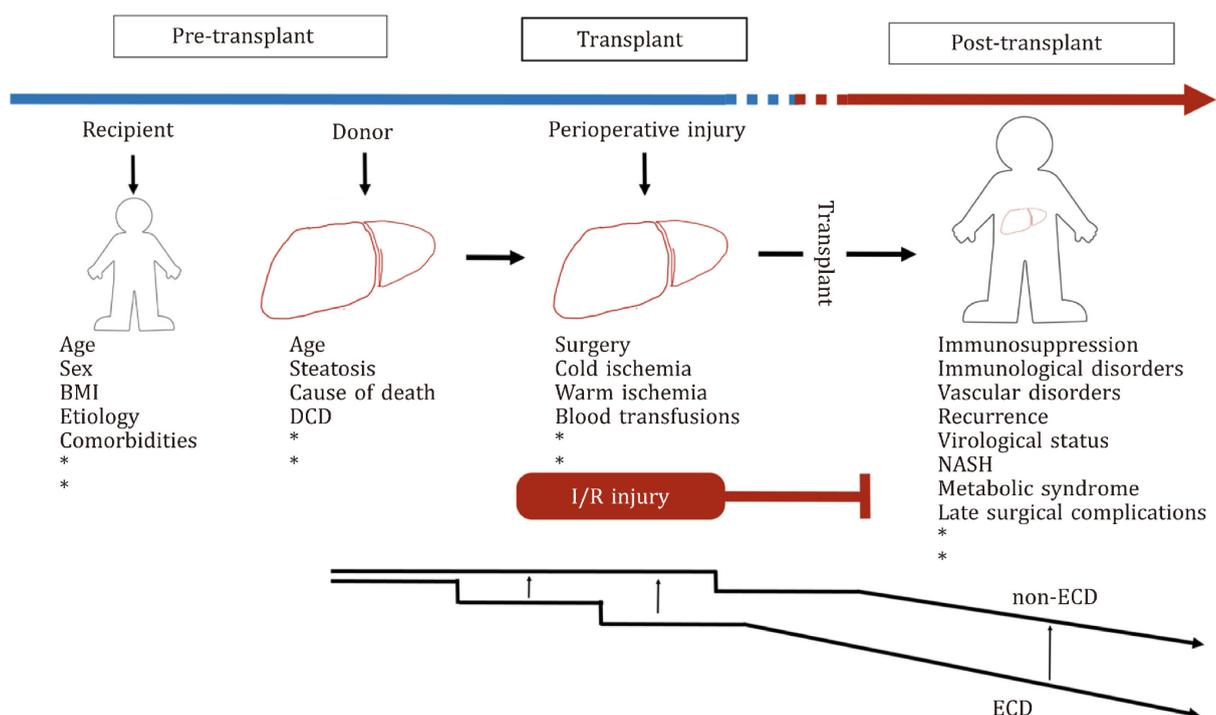


Figure 1 : Les différentes phases et variables de la TH [17]

Des scores ont été mis en place pour prédire la mortalité ou la survie des patients sur différentes périodes et sont actuellement les outils utilisés pour réguler l'attribution des organes en cherchant à augmenter le bénéfice de survie avec la TH.

Le score Foie comme la plupart des scores d'attributions du foie dans le monde utilise le score MELD.

Nous allons voir quelques scores mis en place jusqu'à aujourd'hui, ils peuvent être répartis selon:

- les scores reposant sur des facteurs du receveur
- les scores reposant sur des facteurs du donneur
- les scores reposant sur la combinaison des facteurs du donneur et du receveurs

Les scores reposant sur des facteurs du receveur :

Le score Model for End-Stage Liver Disease (**MELD**) est un score qui permet de prédire la mortalité à 3 mois pour les patients en liste d'attente d'une greffe de foie. [23]

Ce score à été introduit en 2002 aux Etats-Unis par l'Organisation américaine pour le partage des organes, United Network for Organ Sharing (UNOS), afin de cadrer l'attribution des greffons aux patients ayant une cirrhose hépatique (maladie du foie causant une dégénération irréversible des tissus de l'organe). Depuis sa mise en place, la mortalité en liste d'attente a baissé aux Etats-Unis [7] et ce score est aujourd'hui utilisé dans le monde entier.

Il est important de mesurer le score MELD afin de mesurer l'espérance de vie du patient sans la transplantation. Plus le score MELD est élevé, plus le risque de décès en liste d'attente est élevé.

Le calcul du score MELD se base sur 3 composantes biologiques que sont :

- Créatinine : molécule produite par les muscles et éliminée par les reins [24]
- Bilirubine : elle provient de la destruction des vieux globules rouges dans la rate et la moelle osseuse puis est transformée par le foie. Lorsque le foie ne fonctionne plus normalement, cette transformation se fait difficilement et le taux de bilirubine augmente, provoquant une jaunisse. La mesure du taux de bilirubine par une prise de sang permet d'évaluer le fonctionnement du foie [24]
- International Normalized Ratio (INR) : réservé aux patients sous traitements anticoagulants oraux par antivitamines K (AVK)

$$\text{Score MELD} = 10 \times (0,957 \times \text{Log e} (\text{Créatinine en mg/dL}) + 0,378 \times \text{Log e} (\text{Bilirubine en mg/dL}) + 1,120 \times \text{Log e} (\text{INR}) + 0,6431)$$

Le score MELD est fixé à une valeur maximale de 40. Si la valeur de chaque composante est inférieure à 1, la valeur qui sera affectée est 1. Si le patient a été dialysé 2 fois au cours des 7 derniers jours, la valeur de la créatinine doit être fixée à 4,0 mg/dL pour le calcul.

Ce score MELD va donc permettre de donner la priorité à la TH aux patients les plus malades : on parle du principe "sickest first".

Des variantes du score MELD ont ensuite vu le jour comme le score **MELD-Na** qui reprend le score MELD en y ajoutant le taux de sodium.

Les scores reposant sur des facteurs du donneur :

Le score Donor Risk Index (**DRI**) [28] a été développé pour évaluer la qualité du greffon en cherchant à identifier les facteurs prédictifs de l'échec du greffon et ainsi évaluer le risque accru d'utiliser des organes ECD. Il se base sur les caractéristiques du donneur : âge, cause du décès, l'origine ethnique, le type de donneur (don après un décès cardiaque), le type de greffon (partiel/entier), la taille du donneur, la localisation du donneur ainsi que le temps d'ischémie froide.

Un donneur avec un faible indice de risque du donneur (DRI) conviendra à tout receveur, quelle que soit la gravité clinique. Cependant, la combinaison d'un donneur ECD et d'un candidat à haut risque peut donner de mauvais résultats post-TH.

En revanche, le DRI seul n'est pas suffisant pour optimiser un appariement DR car il n'est pas un bon prédicteur de la survie et des éventuels complications post-TH, ne prenant pas en compte les facteurs du receveur.[29]

Dans l'étude [30], le score DRI a été validé et une variante du score DRI est introduite : l' Eurotransplant Donor Risk Index (**ET -DRI**). Il comprend les variables du DRI à l'exception de la taille et de l'origine ethnique du donneur remplacé par 2 autres variables : le taux de gamma glutamyl transférase et l'allocation de sauvetage.

Les scores reposant sur la combinaison des facteurs du donneur et du receveurs :

Des modèles prédictifs ont ensuite été mis en place en prenant en compte non seulement les variables du receveur mais aussi celles du donneur dans l'optique d'optimiser l'appariement DR.

En 2009, le score **D-MELD** a vu le jour, il prend en compte l'âge du donneur en plus du score MELD. Le score D-MELD est le produit simple de l'âge du donneur et le score MELD du receveur au moment de la transplantation. Il a été observé que le score D-MELD montrait une association plus forte avec la probabilité de survie que le score MELD [25]. Plus le score D-MELD est élevé, plus la probabilité de survie est faible.

Le score D-MELD est significatif pour les patients ayant un score MELD ≥ 30 ainsi que les patients recevant un foie d'un donneur âgé de plus de 60 ans. Pour améliorer le bénéfice de survie de ces 2 groupes à risque (probabilité de survie faible dû à un score D-MELD élevé), il faudrait donc associer des donneurs moins âgés aux patients ayant un MELD ≥ 30 . Ainsi, on aura une diminution du score D-MELD et donc une augmentation du bénéfice de survie après transplantation.

Rana et al. [26] ont conçu un score nommé Survival Outcomes Following Liver Transplantation (**SOFT**) qui prédit la survie à 3 mois après la TH, pour compléter la mortalité sur liste d'attente estimée par le score MELD en utilisant à la fois les caractéristiques du donneur et du receveur. SOFT est un score compris entre 0 et 72 points et basé sur 18 facteurs donneurs, receveurs et de transplantation. Parmi les facteurs, on retrouve des facteurs de :

- donneur : âge du donneur, cause du décès, taux de créatinine du donneur
- receveur : âge du receveur, Indice de Masse Corporelle (IMC), nombre de greffes antérieures, chirurgie abdominale, taux d'albumine, dialyse, score MELD etc..

- transplantation : temps d'ischémie froide

Un autre score prenant en compte les facteurs donneurs et receveurs est le score Balance of Risk (**BAR**) [27] qui repose sur moins de facteurs que le SOFT à savoir : âge du donneur, temps d'ischémie froide, score MELD du receveur, réanimation du receveur avant la transplantation, retransplantation et âge du receveur (0–3 points).

En prenant en compte les facteurs donneurs et receveurs, la prédiction de survie post-TH s'est nettement améliorée comparé aux scores ne prenant en considération que les facteurs du receveur ou du donneur.

L'Intelligence Artificielle (IA) : Concepts et quelques définitions

Les algorithmes à base d'IA sont en cours d'évaluation dans la TH, en particulier dans l'attribution d'organes, l'appariement DR, l'analyse de la prédiction de survie avant et après la transplantation.

Dans cette section, nous définirons brièvement des algorithmes d'IA auxquels les études évoquées dans la section suivantes ont eu recours pour accompagner les médecins dans la prise de décision dans le cadre de la TH.

Pour commencer, l'IA peut être définie comme la théorie et le développement de systèmes informatiques qui visent à simuler l'intelligence "humaine". Le Machine Learning (ML) ou apprentissage automatique est une sous-discipline de l'IA qui utilise des techniques statistiques pour donner aux systèmes informatiques la capacité de s'auto-former à partir de données fournies. Les algorithmes adaptent leurs modèles pour améliorer leurs prédictions afin qu'il y ait une amélioration progressive des performances pour une tâche spécifique donnée [12].

Le Deep Learning (DL) ou apprentissage profond est une sous-discipline du ML basée sur des réseaux de neurones artificiels, en anglais Artificial Neural Networks (ANN). Le terme "profond" fait référence au nombre de couches entre l'entrée et la sortie.

L'apprentissage peut-être supervisé, semi-supervisé ou non supervisé. Dans le cadre de l'apprentissage supervisé, les données de formations sont étiquetées (la réponse à la tâche est connue pour ces données). Si les étiquettes sont "discrètes", on parle de classification, si elles sont continues on parle de "régression". Dans notre cas on parlera souvent de classification car il s'agira de choisir entre les classes "survivant" ou "non survivant" pour l'individu ou le greffon.

Dans le cadre de l'apprentissage semi-supervisé, les données de formation ne sont pas toutes étiquetées.

Tandis que pour l'apprentissage non supervisé, aucune donnée d'entrée est étiquetée, l'algorithme doit donc trouver sa propre structure dans les données, généralement une analyse de cluster est utilisée pour découvrir le regroupement des données [12].

Certains modèles ML seront mieux adaptés à la classification des données que d'autres, il est donc nécessaire de choisir le modèle ML approprié pour une question donnée.

Dans notre cadre de la TH, des algorithmes de ML ont été développés et comparés entre eux comme on pourra le voir dans la section "[Les solutions d'IA proposées dans le cadre de la TH](#)".

Parmi ces algorithmes on retrouve :

- Logistic Regression (LR) :

LR est une méthode d'analyse statistique permettant de prédire un résultat binaire, tel que oui ou non, sur la base d'observations antérieures d'un ensemble de données.

Elle prédit une variable dépendante en analysant la relation entre une ou plusieurs variables indépendantes existantes. Dans notre cas, LR peut être utilisée pour prédire si un appariement DR particulier va appartenir ou non à la classe "survivante".

Le modèle LR peut prendre en compte plusieurs critères d'entrée. Dans le cas de l'appariement DR, la fonction logistique pourrait prendre en compte des facteurs donneurs et receveurs parmi ceux évoqués dans la section précédente. En se basant sur des données historiques sur les résultats antérieurs impliquant les mêmes critères d'entrée, il note ensuite les nouveaux cas en fonction de leur probabilité de tomber dans l'une des deux catégories de résultats. [31]

- Decision Tree et Optimal Classification Trees (OCT) :

Les arbres de classification ou arbre de décision sont des structures hiérarchisées de nœuds qui font des prédictions en "divisant les données" séquentiellement en fonction des valeurs de variables indépendantes jusqu'à ce qu'un "nœud feuille" soit atteint [7], on obtient alors une classification comme avoir une insuffisance rénale ou non ou encore une prédiction de la mortalité d'un patient en liste d'attente à partir des observations présent en compte dans l'arbre.

En modélisant les classifications sous forme d'étapes de décision, les arbres de décision peuvent capturer les dépendances non linéaires. Cependant, en cas d'entrée de grande dimension, comme pour la prédiction de survie, les arbres de décision sont extrêmement coûteux en calcul.

L'interprétabilité sous forme de visualisation de la séquence de décision est l'avantage le plus important dans les arbres de décision.

Les arbres de classification optimaux (OCT) offrent ainsi une bonne interprétabilité et une grande précision de prédiction.

- Random Forest (RF) :

RF est une technique d'ensemble capable d'effectuer à la fois des tâches de régression et de classification à l'aide de plusieurs classificateurs d'arbres de décision et d'une technique appelée Bootstrap Aggregation, communément appelée "bagging". Le bagging implique la formation de chaque classificateur d'arbre de décision sur un échantillon de données différent où l'échantillonnage est effectué avec remplacement. L'idée de base derrière cela est de combiner plusieurs classificateurs d'arbre de décision pour déterminer le résultat final plutôt que de s'appuyer sur des classificateurs d'arbre de décision individuels.

RF fonctionne bien avec les données manquantes et a une grande capacité de généralisation. Malheureusement, ils ne peuvent pas gérer une grande quantité de variables, donc une sélection préalable de variables cliniques pertinentes est nécessaire [13].

- Gradient Boosting (GBM) :

GBM est un algorithme ML extrêmement populaire qui a fait ses preuves dans de nombreux domaines. Alors que les forêts aléatoires construisent un ensemble d'arbres indépendants profonds, les GBM construisent un ensemble d'arbres successifs peu profonds et faibles, chaque arbre apprenant et s'améliorant par rapport au précédent [32]. Dans le cadre de la TH, cet algorithme a montré son efficacité dans la prédiction de l'insuffisance rénale post-TH comparé aux autres algorithmes.

- Artificial Neural Networks (ANN) :

Un réseau de neurones artificiels imite l'architecture fonctionnelle des neurones [17] avec trois couches ou plus qui sont interconnectées. La première couche est constituée de neurones d'entrée. Ces neurones envoient des données aux couches plus profondes, qui à leur tour enverront les données de sortie finales à la dernière couche de sortie afin de fournir une classification ou une prédiction.

Toutes les couches internes sont cachées et sont formées par des unités qui modifient de manière adaptative les informations reçues d'une couche à l'autre par une série de transformations [33]. Chaque couche agit à la fois comme une couche d'entrée et de sortie qui permet aux ANN de comprendre des relations plus complexes comme les relations non linéaires très présentes dans le domaine médical.

ANN ont été utilisés dans le cadre de la TH pour prédire les résultats de survie post-TH pour une paire DR en particulier ou encore prédire les éventuelles complications. On aura donc des variables du receveur, donneur et de la transplantation en variables d'entrée qui seront traitées dans les couches internes (cachées) puis la couche de sortie nous fournira les résultats (voir Figure ci-dessous).

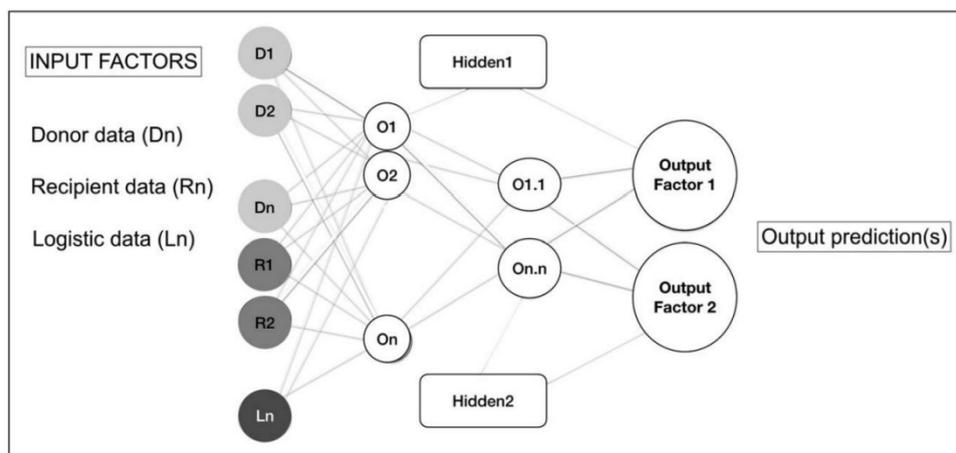


Figure 2 : Structure des ANN service de la TH [13]

- Convolutional Neural Network (CNN) :

Les réseaux de neurones convolutifs capturent les caractéristiques spatiales d'une image [34]. Ils nous aident à identifier un objet avec précision, l'emplacement d'un objet, ainsi que sa relation avec d'autres objets dans une image. CNN sont utilisés pour tout type de problème de prédiction impliquant des données d'image en entrée.

- Recurrent Neural Network (RNN) :

Les réseaux de neurones récurrents ont une connexion récurrente sur les couches cachées. Cette contrainte de bouclage garantit que les informations séquentielles sont capturées dans les données d'entrée permettant ainsi aux RNN d'avoir l'avantage de gérer les données de séquences comparé aux ANN (voir Figure ci-dessous).

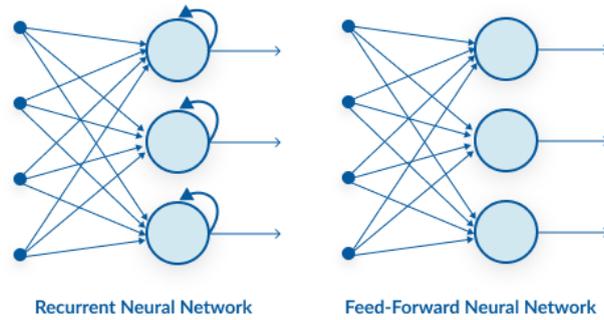


Figure 3 : Différence de structure entre RNN et ANN [34]

- Transformer :

Transformer est un réseau de neurones présenté pour la première fois en 2017 par Google [35]. Il apprend le contexte et donc la signification en suivant les relations dans des données séquentielles comme les mots dans une phrase. Comme les réseaux de neurones récurrents (RNN), Transformer est conçu pour traiter des données d'entrée séquentielles, telles que le langage naturel.

Transformer se base sur un principe appelé attention ou auto-attention. Le mécanisme d'attention fournit un contexte pour toute position dans la séquence d'entrée.

Contrairement aux RNN, le modèle de Transformer traite l'intégralité de l'entrée en une seule fois. Par exemple, si les données d'entrée sont une phrase en langage naturel, le transformateur n'a pas à traiter un mot à la fois comme les RNN. Cela permet une plus grande parallélisation que les RNN et réduit donc les temps de formation.

Les solutions d'IA proposées dans le cadre de la TH

Dans le background de ce mémoire de l'état de l'art, nous avons vu la mise en place de certains scores comme le score MELD afin de prioriser les patients selon la gravité de leur maladie. Mais cela n'est pas suffisant pour optimiser le système d'attribution des organes. En plus de s'assurer que les patients ne meurent pas en liste d'attente, il faut aussi maximiser la réussite de la TH en augmentant le bénéfice de survie si la TH est effectuée. Une TH réussie est non seulement que l'opération soit un succès par le chirurgien mais aussi que le patient ne rejette pas le greffon sur un court (moins d'1 an), moyen (5ans) et long terme (supérieur ou égale à 10 ans) après la TH et qu'il se revoit donc sujet à une retransplantation voir qu'il décède. Des scores ont également été développés comme évoqué dans la partie 2 afin d'estimer la survie du receveur avant la TH comme BAR, SOFT.

Mais le problème aujourd'hui est qu'en se basant sur ces divers scores, nous ne parvenons toujours pas à avoir des estimations précises qui pourraient guider le médecin dans sa prise de décision.

A l'avenir, les algorithmes de ML seraient-ils aptes à jouer un rôle d'assistant dans la prise de décision du médecin ?

Dans les études que nous aborderons, l'IA pourrait potentiellement apporter des réponses dans plusieurs enjeux de la TH notamment :

- l'estimation du bénéfice de survie : en estimant la survie des patients avant [7] et après la TH sur le court [1,8,17] et long terme [2,16]
- l'optimisation du système d'attribution des organes : en essayant de trouver l'appariement DR optimal [1,8,17]
- la prise en charge post-greffe des patients : en prédisant les éventuels complications de la TH [2,9,10,11]

Nous verrons donc à travers les articles sélectionnées pour cet état de l'art si des algorithmes de ML peuvent démontrer une efficacité remarquable dans les différents enjeux présentés au-dessus et ainsi convaincre nos médecins d'y faire appel dans la prise de décision.

Pour mesurer l'efficacité des classificateurs d'IA utilisés, la plupart des études évoquées se basent sur les métriques suivantes :

- AUC : Area Under the Curve équivaut à "Accuracy" ce qui désigne la précision du modèle à donner une bonne classification, autrement dit cela correspond au taux de classification correcte lors de la phase d'entraînement du modèle (nombre de résultats corrects divisé par le nombre total de résultats) [8].
- Sensibility (sensibilité) : La sensibilité d'un test diagnostique est sa capacité à détecter un maximum de malade, dans notre cas elle sera la capacité à détecter un maximum de cas de perte du greffon ou du receveur. Il est utilisé dans nos études car il est plus précis que l'Accuracy dans la prédiction des événements négatifs comme la perte du greffon et/ou du receveur [17].
- AUROC : Area Under the Receiver Operating Characteristic curve (en français l'aire sous la courbe des caractéristiques de fonctionnement du récepteur). Elle est une

mesure de la capacité discriminative du modèle qui est particulièrement adaptée à la classification des classes déséquilibrées. Les valeurs AUROC varient de 0 à 1, et peuvent être interprétées de la sorte :

- supérieur à 0,9 est considéré comme une excellente discrimination
- supérieur à 0,75 est considéré comme une bonne discrimination
- 0,5 équivaut à une estimation aléatoire (comme un “pile ou face”) [3].

1) Prédire la mortalité en attendant la TH

Depuis son instauration en 2002, le score MELD est utilisé dans la plupart des systèmes de hiérarchisation des patients en attente de TH à travers le monde, réduisant ainsi la mortalité en liste d'attente. Cependant, le problème d'accès équitable à la TH n'est pas encore résolu. En effet, encore aujourd'hui, il est difficile de hiérarchiser les patients en attente d'une TH. Si la hiérarchisation des patients sur la liste d'attente d'une TH se fait selon le score MELD qui repose sur le principe du "sickest first", on donnera toujours la priorité aux plus malades. Mais qu'en est-il des patients qui ont une situation clinique moins grave mais qui ont tout de même besoin d'une TH pour se soigner ? Doit-on toujours attendre que leur état se dégrade fortement pour qu'ils se retrouvent prioritaires ?

Sachant que le bénéfice de survie est nettement moindre lorsque les conditions cliniques du receveur sont moins bonnes, on aura ainsi toujours recours à des risques plus élevés de perte du greffon et d'échec de la TH, voire de perte du patient. Un patient qu'on aurait pu sauver si l'on avait réalisé la TH plus tôt lorsque ses conditions cliniques étaient plus favorables.

La hiérarchisation ne se résume donc pas seulement à donner la priorité aux patients qui ont le score MELD le plus élevé au détriment des autres patients mais aussi à s'assurer que les patients en liste d'attente sont aptes à une TH autrement dit que leur maladie n'est pas arrivée à un stade avancé où la TH n'améliore pas leur bénéfice de survie. C'est le cas par exemple de certains patients ayant un stade avancé de la maladie du carcinome hépatocellulaire (CHC).

Une étude propose l'utilisation des arbres de classification optimale (OCT) pour la prédiction de la mortalité à 3 mois en liste d'attente ou du retrait de la liste d'attente d'un candidat, comme solution pour une hiérarchisation plus appropriée des candidats en attente d'une TH [7].

Le problème de prédiction suivant a été posé : Quelle est la probabilité qu'un patient décède ou soit inapte à une TH dans les 3 mois, compte tenu de ses caractéristiques individuelles ?

Le modèle présenté dans cette étude se nomme Optimized Prediction of Mortality (OPOM) : <http://www.opom.online/> [7].

Plus précisément, 2 modèles OPOM ont été utilisés: 1 modèle pour les patients atteints du CHC et 1 modèle pour les patients non atteints du CHC. Ces modèles se sont basés sur des variables dites indépendantes tels que le taux d'albumine, de créatinine, le score MELD (calculé durant les multiples suivis du patient) ou encore son âge. Pour les patients atteints de CHC, le nombre de tumeurs et la somme des tailles des tumeurs s'ajoutent à la liste des variables d'entrée.

L'arbre donnant les meilleurs résultats autrement dit la meilleure prédiction qu'un patient meurt ou devient inapte à une greffe dans les 3 mois (avec une AUC la plus élevée) a fait jusqu'à 10 divisions.

Voici un exemple d'arbre de classification faisant une prédiction à partir de 2 divisions sur la base du score MELD et de la bilirubine :

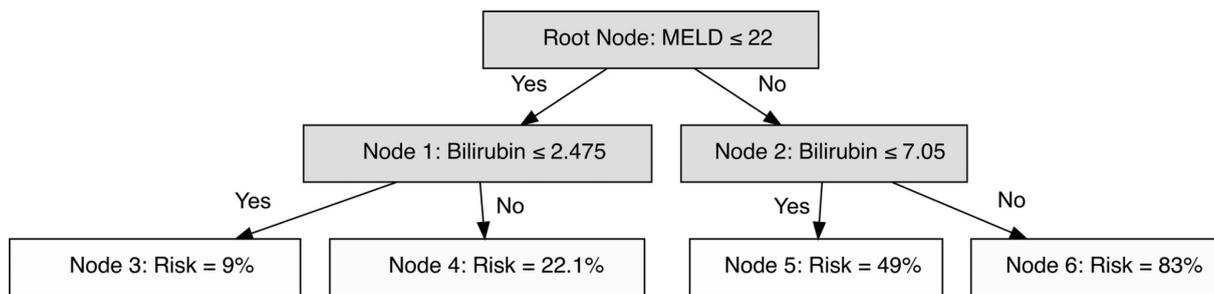


Figure 1 : Arbre de classification sur 2 divisions [7]

Ainsi, un patient ayant un score MELD de 28 et un taux de bilirubine de 6,2 mg/L aura une probabilité égale à 49% de mourir ou d'être inapte à une TH dans les 3 mois d'après cet arbre.

Un simulateur d'attribution d'organes nommé Liver Simulated Allocation Model (LSAM) a été utilisé pour comparer l'attribution d'organes basée sur le score MELD avec celle basée sur l'OPOM. L'utilisation du modèle OPOM a réduit le nombre de décès des candidats en liste d'attente de 23,3% et le nombre de décès post-TH de 1,8% [7].

OPOM a attribué plus de foies aux patients non-CHC et moins aux patients CHC, par rapport au score MELD tout en réduisant le nombre de décès et de retraits sur liste d'attente pour les 2 groupes de patients (non CHC et CHC) [7] ce qui prouve que l'utilisation de l'OPOM permet une attribution équitable du fait de sa bonne prédiction de la mortalité (ou du retrait) sur liste d'attente de la TH.

OPOM a une prédiction de la mortalité en liste d'attente à 3 mois plus efficace que les 2 variantes de MELD (MELD et MELD-Na) [7] :

- MELD : AUC = 0,823 vs 0,859 avec OPOM
- MELD-Na : AUC = 0,841 vs 0,850 avec OPOM

La plus grande divergence de puissance prédictive entre l'OPOM et MELD a été observée au niveau des patients les plus malades (top 5% et top 20%) comme on peut le voir sur la figure ci-dessous (voir Figure 2)

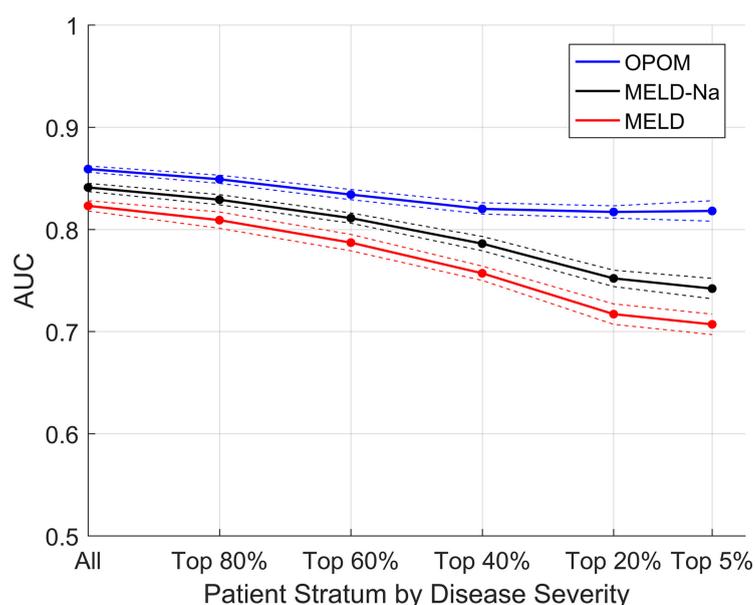


Figure 2 : AUC de prédiction de la mortalité en liste d'attente à 3 mois OPOM vs MELD vs MELD-Na [7]

2) Optimisation de l'attribution des organes et de la prédiction de la survie à court terme post-TH (avant réalisation de la TH)

Dans l'appariement DR beaucoup de variables entrent en jeu, à savoir des variables provenant du donneur, du receveur, ainsi que des variables logistiques et opératoires. Ceci génère une quantité énorme de variables à étudier pour connaître leur influence sur la survie du greffon (plus de 100 variables) [1,17].

Il est essentiel de trouver un appariement DR optimal car il permettrait en conséquence non seulement d'améliorer le bénéfice de survie à court (moins d'un an), moyen (entre 1 et 5 an) et long terme (10 ans et plus) du patient mais aussi de réduire le nombre de retransplantations, ce qui est essentiel surtout avec la pénurie d'organes.

Des articles nous montrent que les algorithmes ML les plus performants dans la recherche de l'appariement DR optimal sont ANN et RF [1,3,8,17].

a) L'appariement DR et prédiction de survie à court terme avec ANN

Avant de former et tester les algorithmes sur des jeux de données, des règles d'exclusion ont été appliquées. Les donneurs vivants ainsi que les receveurs âgés de moins de 18 ans [1,8] ou atteints de la maladie du CHC dans le cas de l'étude appliquée en Angleterre [8] ont été exclus.

De plus, les variables avec un pourcentage de valeurs manquantes (>10%) sont éliminées car elles risquent de réduire la qualité de prédiction du modèle [1,8,17].

Pour s'assurer du bon entraînement des modèles, le processus appelé "ten-fold randomisation" à été appliqué. Ce processus revient à répartir l'ensemble de l'échantillon d'étude aléatoirement en un ensemble de formation (phase d'apprentissage) ou ensemble de validation du modèle (effectué après la phase d'apprentissage) [1,8]. Cette répartition sera répétée 10 fois, à chaque randomisation: 90% des données sont utilisés pour la phase d'apprentissage et les 10% restants pour la phase de test [1,8]. Le but étant d'éviter d'entraîner le modèle toujours sur le même échantillon, de cette manière, le modèle se forme sur un maximum de données différentes.

La probabilité de survie et la probabilité de perte du greffon après la TH est calculée pour chaque paire DR par les algorithmes ANN. Le but étant de maximiser la survie du receveur (et donc du greffon) suite à la TH, l'appariement DR optimal correspondra à la paire DR ayant la plus grande probabilité de survie du greffon et la plus petite probabilité de perte [1,8,17].

Pour cela, 2 algorithmes ont été développés :

- ANN-CCR(CCR = Correct Classification Rate) : cet algorithme a pour objectif de prédire la survie du greffon à 3 mois [8,1] et 12 mois [8] après la TH en essayant de maximiser la probabilité qu'une paire DR appartienne à la classe " survivante".
- ANN-MS(Minimum Sensitivity) : cet algorithme a pour objectif de prédire la perte du greffon à 3 mois [8,17] et 12 mois [8] en essayant de maximiser la probabilité qu'une paire DR appartienne à la classe "non survivante".

Dans le cas de la TH, le taux de succès étant plutôt élevé, la classe " survivante " est majoritaire tandis que la classe " non survivante " est minoritaire [5]. On est donc le cas d'un problème déséquilibré [5,8,15,17] où l'on aura plus de données appartenant à la classe " survivante " que la classe " non survivante " [1], ce qui risque de compromettre la bonne discrimination entre les 2 classes (le modèle pourrait avoir tendance à classer la plupart des paires DR dans la classe " survivante ") [8,15].

Pour contrer cela, l'algorithme ANN-CCR utilise Accuracy car la classe " survivante " est majoritaire; Accuracy (taux de classifications correctes sur le total des classifications) est donc suffisant pour calculer la précision de prédiction de l'algorithme. L'algorithme ANN-MS quant à lui utilise la sensibilité minimum qui correspond au minimum des sensibilités de chaque classe (" survivante " et " non survivante ").

La paire DR qui sera sélectionnée parmi les autres paires sera celle qui aura la plus haute probabilité de survie et la plus basse probabilité de perte.

Mais très vite, des problèmes seront relevés à savoir que :

- L'algorithme sélectionne toujours les paires DR avec la plus grande probabilité de survie et la plus petite probabilité de perte. Cependant entre 98,4% et 98,2% de survie par exemple; il n'y a pas de différence d'un point de vue clinique, les 2 sont de très grande probabilité de succès qui prédisent des résultats prometteurs de la TH. Pourtant, selon le modèle, 98,4% étant supérieur à 98,2%, seule la paire DR ayant 98,4% de probabilité de survie sera retenue au détriment de la paire à 98,2% de survie [1,8,17].
- Par conséquent, les patients ayant un meilleur profil clinique seront toujours privilégiés dans l'appariement DR car la probabilité de succès DR est plus élevée dans leurs cas que dans celui des patients les plus malades. Ces derniers seront malheureusement difficilement appariés par l'algorithme. Le principe du " sickest first " n'est alors plus respecté.[1,8]. Dans cette optique, le principe d'équité est compromis au profit du principe d'utilité.

Pour revenir sur un équilibre entre ces 2 principes d'équité et d'utilité qui permettrait réellement d'optimiser le système d'attribution des organes, un système à été mis en place pour effectuer l'attribution d'un donneur X à un receveur Y. Ce système repose sur la règle suivante: lorsque les probabilités de survie du greffon ne sont pas significativement différentes d'un point de vue clinique, alors l'appariement DR sera attribué en fonction du score MELD afin de prendre en compte la gravité de la maladie des receveurs. Deux probabilités sont significativement différentes si elles ont une différence d'au moins 3% pour l'ANN-CCR et 5% pour l'ANN-MS [1,8]. Ces seuils ont été choisis parmi les écarts-types des probabilités d'appartenance à la classe " survivante " (SD=2,86 % [1], SD = 2,9% [8]) ou " non survivante " (SD=5,56 % [1], SD = 5,5% [8]) du greffon.

Voici un schéma récapitulatif (voir Figure 3) des algorithmes ANN utilisés pour optimiser l'appariement DR dans les études [1,8] :

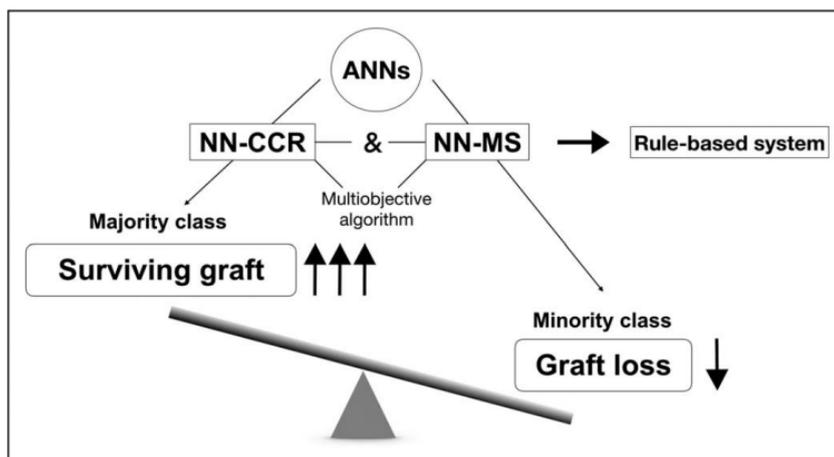


Figure 3 : Schéma de l'appariement DR avec ANN [5]

L'étude 1 de l'article [1] est la première à avoir introduit la combinaison des 2 algorithmes ANN-CCR et ANN-MS comme un modèle capable de prédire la probabilité de survie et de perte du greffon de manière fiable en se basant sur des données provenant de 11 unités de TH en Espagne. L'étude a été effectuée sur un ensemble de 1003 paires de DR des TH réalisées entre janvier 2007 et décembre 2008.

L'étude 2 de l'article [8], quant à elle, se base sur les données de 858 paires de DR (739 paires appartenant à la classe " survivante " et 83 paires à la classe " son survivante ") provenant des TH réalisées de janvier 2002 à décembre 2010 au sein du King's College Hospital (KCH) National Health Service Foundation Trust à Londres, au Royaume-Uni.

Les résultats obtenus pour ces 2 études sont très prometteurs en particulier pour l'étude 2 réalisés en Angleterre [8], voici l'Accuracy de la prédiction de survie et de perte du greffon à 3 mois post-greffe relevé selon les différentes études :

Algorithme	Etude 1 [1]	Etude 2 [8]
ANN-CCR	AUC = 0,8060	AUC = 0,94
ANN-MS	AUC = 0,8215	AUC = 0,94

On remarque que les performances des deux algorithmes ANN dans l'étude 2 basé sur les données de KCH sont plus élevées que dans l'étude 1 basé sur les 11 unités espagnoles. Cela peut s'expliquer par la qualité de la base de données fournie par KCH supérieure à celle utilisée sur l'étude 1 [1,8,17].

En effet, la base de donnée fournie par KCH était plus homogène car elle reposait seulement sur les patients de cet hôpital de Londres (qui proviennent donc de la même région) [8] tandis que l'autre base de données concernant l'étude espagnole est plus variée car elle est multicentrique (étude faite dans plusieurs centres médicaux en Espagne)[1,17]. De plus, la base de données KCH avait un pourcentage de valeurs manquantes inférieur à celui de l'étude multicentrique espagnole. Or, le point faible des modèles ANN est qu'ils sont sensibles aux valeurs manquantes et que cela peut venir altérer l'efficacité de ces derniers [17].

On en déduit par ces résultats que chaque modèle ANN doit être formé dans un but spécifique et dans une population distincte [8,17] :

- but spécifique : en formant des algorithmes ANN dans le but unique de prédire la survie du greffon à 3 mois (ANN-CCR) ou de prédire la perte du greffon à 3 mois (ANN-MS) nous avons pu nous distinguer des autres classificateurs ML comme LR ou bien les scores en particulier pour la prédiction de perte du greffon (voir détails sous partie c).
- population distincte : en formant des algorithmes à partir de données uniformes, il est plus facile pour le modèle d'extraire les fonctionnalités significatives qui lui permettront de faire une bonne classification. On peut donc supposer que les ANN pour l'appariement DR ne peuvent être appliqués que dans des bases de données qui suivent des règles similaires pour la hiérarchisation et l'inclusion des candidats, et qui sont également suffisamment homogènes en ayant les mêmes variables sous les mêmes formats .

A la suite des calculs de probabilités de survie et de perte du greffon à 3 mois pour chaque paire DR, le système développé pour effectuer l'attribution des organes va attribuer l'organe d'un donneur au receveur le plus approprié.

Voici 2 exemples de simulation d'appariement DR réalisé avec ANN entre 5 receveurs avec un score MELD proche et 10 donneurs sélectionnés au hasard. Les receveurs sont regroupés en fonction de leur score MELD car si ce n'était pas le cas, seuls les receveurs ayant les plus petits scores MELD seraient appariés à un donneur au détriment des patients les plus malades. Dans le premier tableau (Figure 4), la simulation est faite avec des receveurs ayant un score MELD compris entre 23 et 27 [8]. Dans le deuxième tableau (Figure 5), la simulation est faite avec des receveurs ayant un score MELD compris entre 34 et 40 [8].

Recipient (MELD)	Donor 1	Donor 2	Donor 3	Donor 4	Donor 5	Donor 6	Donor 7	Donor 8	Donor 9	Donor 10
CCR model: 3-month probability of graft survival										
Recipient 1 (27)	93.19	91.43	93.19	94.26	94.13	93.19	94.26	94.13	94.26	94.27
Recipient 2 (26)	84.37	84.30	84.37	94.25	85.28	84.37	92.95	85.28	92.95	94.26
Recipient 3 (23)	92.95	90.95	92.95	94.26	94.10	92.95	94.26	94.10	94.26	94.27
Recipient 4 (23)	84.35	84.30	84.35	94.24	85.28	84.35	92.95	85.28	92.95	94.26
Recipient 5 (23)	84.18	84.27	84.18	94.09	84.33	84.18	87.05	84.33	87.05	94.13
MS model: 3-month probability of nongraft survival										
Recipient 1 (27)	74.85	76.37	75.60	19.69	25.31	74.16	18.91	76.36	73.31	13.75
Recipient 2 (26)	75.83	76.39	76.11	30.29	41.08	75.58	28.50	76.38	75.24	15.13
Recipient 3 (23)	49.29	75.74	60.16	13.26	13.52	42.41	13.23	75.60	36.03	13.00
Recipient 4 (23)	68.76	76.25	72.39	14.29	15.46	65.63	14.13	76.22	61.67	13.12
Recipient 5 (23)	76.40	76.40	76.40	74.95	75.61	76.40	74.73	76.40	76.38	64.57
Allocation	R3	R1	R3	R3	R3	R3	R3	R1	R3	R1

Figure 4 : Une simulation d'appariement DR avec ANN et des receveurs avec un score MELD moyen [8]

Analysons la première attribution du donneur 1 au receveur 3:

- Le donneur 1 présente la même probabilité de survie du greffon à 3 mois pour le receveur 1 (93,19%) et le receveur 3 (92,95%). En effet, la différence entre les 2 n'est pas significative. Mais, en se reportant sur la probabilité de perte du greffon à 3 mois, le receveur 1 à une probabilité nettement plus élevée que le receveur 3 (74,85% vs 49,29%). Le système attribue donc l'organe du donneur 1 au receveur 3.

Analysons la deuxième attribution du donneur 8 au receveur 1:

- Le donneur 8 présente la même probabilité de survie du greffon à 3 mois pour le receveur 1 (94,13%) et le receveur 3 (94,10%). Elle est plus élevée que pour le reste des receveurs. De même pour la probabilité de perte du greffon à 3 mois, le receveur 1 est à 76,36% et le receveur 3 à 75,60% (le reste des receveurs ont également une probabilité de perte élevée). Le système va donc se reporter au score MELD et attribuer l'organe du donneur 8 au receveur 1 car son score MELD est plus élevé (27 vs 23). Le principe du "sickest first" est respecté.

TABLE 4. A Simulation of D-R Allocation by the ANN Using High-MELD Score Recipients

Recipient (MELD)	Donor 1	Donor 2	Donor 3	Donor 4	Donor 5	Donor 6	Donor 7	Donor 8	Donor 9	Donor 10
CCR model: 3-month probability of graft survival										
Recipient 1 (40)	83.65	84.26	83.64	93.85	84.33	83.64	87.05	84.33	87.05	94.13
Recipient 2 (39)	84.29	84.30	84.29	94.21	85.28	84.29	92.95	85.28	92.95	94.26
Recipient 3 (38)	84.22	84.27	84.22	94.11	84.33	84.22	87.05	84.33	87.05	94.13
Recipient 4 (36)	94.26	94.26	94.26	94.27	94.27	94.26	94.24	94.27	94.25	94.27
Recipient 5 (34)	87.03	85.33	87.03	94.26	92.11	87.03	94.22	92.11	94.22	94.27
MS model: 3-month probability of nongraft survival										
Recipient 1 (40)	76.49	76.40	76.49	75.06	75.57	76.48	74.64	76.40	76.38	64.17
Recipient 2 (39)	75.38	76.38	75.88	23.04	30.81	74.93	21.89	76.37	74.30	14.16
Recipient 3 (38)	76.40	76.40	76.40	75.98	76.17	76.40	75.92	76.40	76.39	72.83
Recipient 4 (36)	73.64	76.35	74.97	16.75	20.06	72.44	16.30	76.33	70.84	13.40
Recipient 5 (34)	76.40	76.40	76.40	75.76	76.06	76.39	75.67	76.40	76.39	71.01
Allocation	R4	R2								

Figure 5 : Une simulation d'appariement DR avec ANN et des receveurs avec un score MELD élevé [8]

Tout d'abord on constate qu'il y a plus de probabilité de perte du greffon élevés chez les receveurs à haut score MELD que chez les autres receveurs avec un score MELD<30. Ceci s'explique par le fait que le risque de perte du greffon est plus élevé lorsque les conditions cliniques du patient sont moins favorables.

Le système attribue l'organe du donneur 1 au receveur 4 car sa probabilité de survie du greffon à 3 mois est plus élevée (94,26% vs [84,22%-87,03%]) et sa probabilité de perte du greffon à 3 mois est plus basse que pour les autres receveurs (94,26% vs [84,22%-87,03%]).

L'organe du donneur 10 est quant à lui attribué au receveur 2 car sa probabilité de survie est élevée tout comme les autres receveurs 1,3,4,5 mais sa probabilité de perte est plus basse pour le receveur 4 (14,16% pour le receveur 2 et 13,40% pour le 4). En revanche, le score MELD est plus élevé pour le receveur 2 (39 vs 36) qui se voit donc attribuer l'organe.

Par ailleurs, une étude basée sur KCH nous propose l'utilisation de données non étiquetées comme solution pour améliorer la classification de la classe minoritaire "non-survivante". Les données non étiquetées seraient les paires DR dites virtuelles qui correspondent en fait aux paires DR qui n'ont pas été effectuées par l'équipe de TH. Ces dernières pourront être utilisées pour représenter la classe "non-survivante" car elles auraient été considérées comme des paires DR conduisant à la perte du greffon par l'expertise de l'équipe TH et des autres scores [15]. En utilisant ces données non étiquetées en plus des données étiquetées, on parlera d'une approche semi-supervisée pour l'appariement DR et la prédiction à court terme.

Avec les ANN, on pourrait donc gérer une masse de données en entrée (toutes les variables donneurs, receveurs, et autour de la transplantations) pour chaque paire DR et faire des prédictions plus juste qu'avec seulement l'utilisation des scores (voir comparaison sous partie c). Tout cela en garantissant que les patients qui nécessitent une greffe de foie de toute urgence sont prioritaires en prenant en compte le score MELD dans le système d'attribution proposé.

De plus, l'attribution des organes se fera de manière objective à l'inverse du médecin qui reste un humain, et qui risque parfois de prendre des décisions subjectives [17] qui mettrait en péril le principe de justice dans l'attribution des organes.

b) L'appariement DR et prédiction de survie à court terme avec RF

L'étude réalisée dans l'article [3] utilise RF dans le but d'optimiser l'appariement DR en améliorant la prédiction de perte du greffon dans les 30 jours.

Cette étude est la première à utiliser RF pour prédire les résultats après une TH.

Le plus avec l'utilisation des RF est la sélection des variables d'entrées (donneur, receveur, et peropératoires) [3,12,17] significatives dans l'appariement DR.

En effet, avec les algorithmes ANN, il est certes possible d'obtenir une très bonne capacité de prédiction mais les ANN ont du mal à sélectionner les variables les plus influentes [13] qui seraient ainsi celles à prendre en compte en premier rang lors de la prise de décision afin d'attribuer tel organe à un receveur en particulier.

RF quant à lui aura une prédiction qui demeure assez bonne mais il fournit plus d'informations sur le poids relatif des variables [13,17]. Cela améliore donc la transparence et l'interprétabilité de l'algorithme comparé au concept de "boîte noire" des ANN [3], ce qui pourrait mettre éventuellement nos médecins en confiance pour l'utilisation de l'IA dans leur prise de décision.

De plus, contrairement aux ANN qui sont sensibles aux données manquantes et qui risquent souvent d'être sujet au surajustement, RF est moins sensible aux données manquantes [3,17].

Ainsi, sur un petit échantillon de 180 TH réalisées entre 2013 et 2015 au centre d'Austin Health à Melbourne en Australie, RF a montré une bonne capacité de prédiction avec une AUROC de 0,787. Une sélection finale des 15 variables (sur les 276 au départ) influençant le résultat de la greffe dans les 30 jours a été effectuée par RF (voir Figure 6).

Parmi ces variables on retrouve surtout des variables du donneur tels que la cause du décès (accident vasculaire cérébral, anoxie, traumatisme, autre), le taux d'albumine sérique, le don après mort cérébrale ou cardiaque, état dans lequel se trouve l'hôpital du donneur, consommation d'alcool (non, quantité inconnue, <1, 2-4, >4 verres par jour), taux d'hémoglobine (Hb), taux de protéines totales, utilisation d'insuline, âge du donneur etc...

TABLE 2.**Overall top 15 predictors with the percentage of missing values in the study data and validation data**

Characteristics	Average rank sum	Missing % in study data	Missing % in validation data
Recipient disease category	1.619	8.89	1.11
Donor serum albumin level	18.836	8.89	36.67
Donor cause of death	20.420	0.56	2.22
Donation after brain or cardiac death	24.931	0	3.33
Donor Hb level	30.375	16.67	45.56
Donor alcohol consumption	30.805	1.67	10
The state in which the donor hospital is located	31.373	0	0
Donor total protein level	32.441	18.89	41.11
Donor insulin usage (dichotomous)	35.011	18.33	72.22
Recipient medical status at activation	36.285	33.89	27.78
Donor pancreas retrieved (dichotomous)	38.166	9.44	0
Donor age	38.412	0	0
Serum herpes simplex antibodies	38.654	12.78	8.89
Donor previous surgery (dichotomous)	41.505	2.78	0
Donor CMV status (dichotomous)	42.083	0	0

Figure 6 : Sélection des 15 principaux facteurs influençant les résultats de la TH dans les 30 jours [3]

En appliquant RF sur seulement ces 15 variables plutôt que les 276 variables initiales, la prédiction est améliorée : AUROC RF = 0,818 (vs AUROC RF = 0,787) [3].

Cependant, n'oublions pas que les classificateurs RF sont peu utiles lors de l'utilisation de bases de données plus volumineuses en raison du grand nombre d'arbres de décision qu'ils peuvent générer [3,17].

c) Comparaison ANN et RF avec des scores de prédiction de survie et LR

Les recherches basés sur ANN et RF ont montré de meilleurs résultats que les scores tels que BAR, SOFT, MELD, DRI [1,3,8,17]. L'étude présenté dans l'article [8] a montré des résultats jusqu'à 10% plus élevés que le score BAR [8]. Ces résultats s'expliquent par le fait que les classificateurs de ML comme les ANN peuvent prendre en compte un grand nombre de variables en entrée qui pourraient potentiellement jouer un rôle important dans l'appariement DR et ainsi donner une meilleure prédiction de la survie ou perte du greffon [1]. Contrairement aux ANN, les scores ne peuvent pas embarquer un grand nombre de variables dans leur calculs. Une sélection limitée de variables au préalable est alors nécessaire, ce qui empêche les scores de donner une prédiction la plus fiable possible.

En effet, la sélection des variables n'est pas sûre car l'appariement DR repose sur des relations complexes. Les ANN pourront prendre en entrée plus de variables de donneurs, de receveurs, de procédure de prélèvement et de transplantation (57 variables ont été prises en compte dans l'étude 1 article[1], 59 variables dans l'étude 2 article [8]) même si l'on ignore le poids exact de chacune de ces variables dans l'appariement DR et la survie du greffon que les scores. Le score MELD ne prend en compte et se limite qu' à 3 variables, le DRI 8 variables et le SOFT 18 variables [21].

De plus, la prédiction de survie avec ces scores basés sur des modèles biostatistiques est moins bonne qu'avec les ANN car les scores reposent sur des principes inadéquats avec les problèmes médicaux [17]. En effet, en plus d'être limités au niveau des variables d'entrée (en ne gardant que celles jugées "significatives" selon chaque score), ces derniers traitent

les variables comme si elles avaient une relation linéaire avec la fonction de survie. Or, en médecine, toutes les variables comptent et ne suivent pas de relation linéaire [1,17].

Les scores ont certes une prédiction adéquate lorsqu'il s'agit de phénomènes fréquents (dans notre cas la survie du greffon à 3 mois : AUC BAR = 0,84 ; AUC SOFT= 0,82 ; AUC MELD= 0,75 ; AUC DRI = 0,65 [8]) mais ils sont incapables de prédire les phénomènes rares [1,17] (dans notre cas la perte du greffon 3 mois : AUC BAR = 0,62 ; AUC SOFT= 0,57 ; AUC MELD = 0,42 ; AUC DRI = 0,49 [1]).

On remarque par la même occasion que dans les deux modèles de prédiction de la survie et de la perte du greffon, les scores qui combinent les variables du donneur et du receveur (scores BAR, SOFT) avaient de meilleurs résultats que les scores prenant en compte seulement les variables du donneur (DRI) ou du receveur (MELD). [1,17]

Tandis qu'avec les ANN, nous avons non seulement de meilleurs résultats pour la prédiction de survie du greffon à 3 mois (ANN-CCR AUC = 0,94 [8]), mais aussi pour la prédiction de perte du greffon à 3 mois grâce à l'algorithme ANN-MS (ANN-MS AUC = 0,94 [8]).

La probabilité de survie et de perte du greffon à 12 mois a également été étudiée dans l'article [8] sur la base de donnée de KCH: même si les résultats sont un peu moins bon comparé à la prédiction de survie à 3 mois, ils restent tout de même pertinents pour une prise en compte dans les décisions cliniques. Comme on peut le voir sur la figure 6 ci-dessus ANN-CCR AUC= 0,78 et ANN-MS AUC = 0,81 [8] ce qui reste supérieur au score BAR placé en 2ème position avec AUC = 0,71.

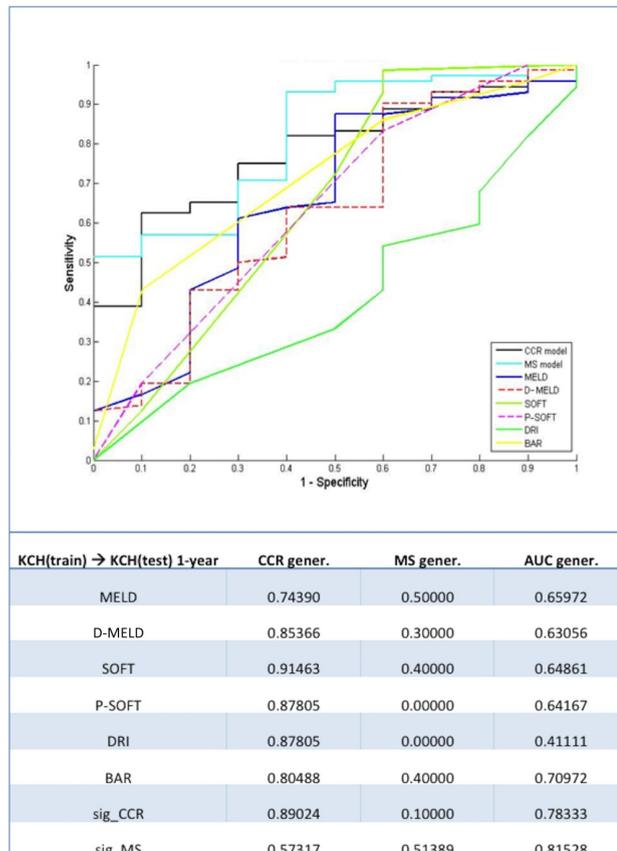


Figure 7 : AUC des algorithmes ANN-CCR et ANN-MS et des scores de prédiction de survie [8]

RF a également de meilleurs résultats dans la prédiction de l'échec du greffon dans les 30 jours comparé aux scores biostatistiques comme on peut le voir sur la figure 8 ci-dessous: AUROC RF = 0,787 vs AUROC DRI = 0,680 et AUROC SOFT = 0,638.

Lorsque les scores étaient associés à RF, les performances étaient meilleures par rapport aux performances des scores sans association à RF : AUROC DRI + RF = 0,697, AUROC DRI + MELD + RF = 0,764 vs AUROC DRI = 0,680 [3].

Plus intéressant encore était la capacité de RF à sélectionner les 15 variables les plus influentes améliorant ainsi la prédiction pour RF et ANN : AUROC RF = 0,818 (vs 0,787 avec les 276 variables), AUROC ANN = 0,835 (vs 0,734 avec les 276 variables).

TABLE 3.
Comparison of AUC-ROC values of different models created during the study

Characteristics used	AUC-ROC (95% CI)
DRI	0.680 (0.669-0.690)
SOFT score	0.638 (0.632-0.645)
Neural network with all the factors	0.734 (0.729-0.739)
Random forest with all the factors	0.787 (0.782-0.793)
DRI characteristics in random forest	0.697 (0.688-0.705)
DRI characteristics and MELD score in random forest	0.764 (0.756-0.771)
Random forest with characteristic selection (Top 15)	0.818 (0.812-0.824)
Neural network with random forest characteristic selection (Top 15)	0.835 (0.831-0.840)

DRI factors: age, cause of death, race, partial/split, height, regionality, cold ischemia time; MELD factors: recipient creatinine, bilirubin and INR; SOFT score factors-age, BMI, number of previous transplants, previous abdominal surgery, albumin, dialysis before transplantation, intensive care unit pretransplant, admitted to hospital pretransplant, MELD, life support pretransplant, encephalopathy, portal vein thrombosis, ascites pretransplant, portal bleed 48 h pretransplant, donor age, donor cause of death from cerebral vascular accident, donor creatinine, national allocation, cold ischemia time.

Figure 8 : AUROC des différents algorithmes ANN , RF et des scores de prédiction de survie [3]

Une étude a comparé plusieurs méthodes d'apprentissage dont LR, Support Vector Machine (SVM), Multilayer perceptron(MLP) et RF [4] concernant la survie du greffon à 3 mois, 1 an, 2 ans et 5 ans. LR s'est démarqué par rapport aux autres méthodes y compris MLP et RF en ayant la performance la plus élevée pour la prédiction à 5 ans (AUC LR = 0,654 vs AUC MLP = 0.599 et AUC RF = 0.644). Cela ne remet pas en cause les autres méthodes d'apprentissage mais en appliquant l'IA sur des bases de données aussi volumineuses que celles de l'UNOS dans cette étude, cela montre à quel point la qualité de la base de donnée est importante pour avoir des bonnes performances comme on a pu voir avec les ANN dans les études [8] et [1].

3) Prédire survie post-TH à long terme et les éventuelles complications

Jusqu'à présent, les recherches abordées dans la partie 1) et partie 2) de l'analyse, nous ont montré que les algorithmes de ML peuvent jouer un grand rôle dans:

- la prédiction de la mortalité en attendant la TH
- la prédiction de la survie sur 3 mois et 12 mois post-TH (avant réalisation de la TH)
- la recherche de la paire DR optimal pour de meilleurs résultats de la TH et de meilleures conditions de vie des receveurs.

Mais la projection ne s'est pas faite sur le long terme, ce qui reste un enjeu important. En effet, qu'en est-il de la survie du malade après réalisation de la TH ? Comment l'IA peut accompagner le médecin dans la prise en charge des patients après la TH ?

Il faut savoir que la prédiction de la survie sur le long terme et les éventuelles complications après la TH sont plus difficiles à prévoir car d'autres facteurs non liés à la greffe peuvent survenir au cours de la vie du patient.

Les études que nous aborderons dans cette partie 3), se focalise sur :

- la prédiction de la survie après la TH sur le long terme [2,16]
- prédiction des éventuelles complications après la TH comme:
 - l'insuffisance rénale aiguë [10,11]
 - le diabète [18]
 - une récurrence du carcinome hépatocellulaire (CHC) [9,20]

a) Prédiction de la survie sur le long terme

Pour développer des algorithmes de ML qui prédisent la survie du patient sur le long terme, il a fallu des données de patients qui ont été suivi sur le long terme de manière continu afin de récolter un maximum de données pertinentes et nécessaires pour la formation des modèles abordés dans l'article [2] et [16].

Les 2 études présentent des algorithmes de DL pour la prédiction de la survie post-TH sur le long terme. Dans l'article [16], un modèle MLP est développé pour prédire la survie des patients de 6 mois à 13 ans après la TH. Dans l'article [2], 3 modèles de réseaux de neurones en plus de MLP ont été développés pour la prédiction de la survie du patient après chaque suivi clinique dans un délai de 1 an (court terme) ou de 5 ans (moyen terme). Les modèles développés sont Recurrent Neural Network (RNN), Temporal Convolutional Network (TCN) et le modèle Transformer qui sortira vainqueur parmi les 3.

Comme pour la plupart des études abordées dans les sections précédentes, les patients de moins de 18 ans ont été exclus des jeux de données des études [2,16] ainsi que les patients ayant reçu plusieurs greffes d'organes (en revanche, l'étude de l'article [16] inclut les retransplantations).

Les patients qui ont survécu mais qui ont eu moins de 5 ans des données de suivi au moment de la collecte des données ont été exclus de l'étude [2].

L'étude menée dans l'article [2] a deux particularités :

- c'est la 1ère étude à faire des prédictions de résultats TH multiclassées. En effet, elle ne se contente pas de prédire la survie ou non du patient. Il y a 5 sorties possibles: d'une part la survie du patient, d'autre part la non survie due à une mort causé par une ou plusieurs des 4 catégories de causes de décès suivantes: causes cardiovasculaires, échec de la greffe, infection et cancer.
- elle se base sur des données longitudinales (qui varient avec le temps vue qu'elles sont mise à jour lors de chaque suivi) [2]

La base de données Scientific Registry of Transplant Recipients (SRTR) provenant des États-Unis a été utilisée pour la phase de formation et de validation en interne. Une validation externe a également été effectuée sur la base de données University Health Network (UHN) de l'Ontario au Canada. [2]

Les suivis de la base SRTR sont enregistrés à 6 mois après la TH, puis annuellement. Pour UHN, des résultats de laboratoire sanguin à partir du jour de la TH, avec une surveillance plus étroite au début, puis des données de suivi tous les 1 à 3 mois en fonction de l'âge et de la fonction du greffon sont à disposition.[2]

Le modèle Transformer, qui s'est révélé être le plus performant lors de phase de validation interne avec SRTR, a été entraîné en plus sur UHN (Transfer Learning) avant la validation externe afin de tester la généralisabilité des fonctionnalités apprises pas le modèle.[2]

Pour SRTR et UHN, les 4 modèles de DL (Transformer, RNN, TCN et MLP) ont montré une efficacité plus grande de la prédiction de la survie ou de la cause du décès du patient à 1 an et 5 ans lors de chaque visite comparé à la régression logistique (LR). En comparant avec les 4 modèles DL, Transformer est le plus performant avec une AUROC=0,804 à 1 an et 0,733 à 5 ans avec SRTR, une AUROC=0,803 à 1 an et 0,713 à 5 ans avec UHN. Les résultats avec UHN sont d'autant améliorés lorsqu'il y a un transfert d'apprentissage (Transfer Learning) : AUROC=0,807 à 1an et 0,722 à 5 ans.

Ces résultats s'expliquent par la complexité du modèle Transformer qui semble bien gérer les données de séquences, encore mieux que le TCN ou RNN (SRTR AUROC à 1 an = 0,733 et AUROC à 5 ans = 0,670; UHN AUROC à 1 an = 0,772 et AUROC à 5 ans = 0,665). Le MLP est quant à lui moins bon car il lui manque l'intégration des données de séquences (voir Figure 9).

	SRTR (internal validation test set)		UHN (cross-validation)	
	1-year outlook	5-year outlook	1-year outlook	5-year outlook
Transformer	0.804 (0.795-0.854)*	0.733 (0.729-0.769)*	0.803 (0.787-0.836)	0.713 (0.692-0.751)
With transfer learning	0.807 (0.795-0.842)*	0.722 (0.705-0.764)*
Temporal convolutional network	0.768 (0.755-0.806)	0.702 (0.687-0.734)	0.779 (0.750-0.798)	0.672 (0.647-0.693)
Recurrent neural network	0.733 (0.723-0.760)	0.670 (0.658-0.695)	0.772 (0.758-0.797)	0.665 (0.640-0.678)
Multilayer perceptron	0.727 (0.714-0.751)	0.660 (0.651-0.684)	0.764 (0.747-0.781)	0.657 (0.640-0.669)
Logistic regression	0.714 (0.689-0.745)	0.648 (0.632-0.667)	0.754 (0.744-0.780)	0.650 (0.608-0.664)

Data are AUROC (99% CI) for model performance across all outcomes. Transformer with transfer learning indicates training on SRTR followed by fine-tuning on UHN data. AUROC=area under the receiver operating characteristic curve. SRTR=Scientific Registry of Transplant Recipients. UHN=University Health Network. *Top score in each category.

Figure 9 : AUROC de la prédiction de survie à 1 an et 5 ans selon les différents modèles [2]

La prédiction se fait sur 1 an et 5 ans à chaque visite de suivi dans le but d'accompagner le médecin dans la prise en charge du patient après la TH. Ainsi, le médecin pourra préparer

chaque visite en étant guidé par ces algorithmes et prendre les mesures nécessaires comme une prescription de dépistage de cancer si cette complication venait à être prédite par les algorithmes de DL. A l'avenir, on pourra même adapter la posologie des médicaments immunosuppresseurs afin de trouver le bon équilibre entre le maintien de la santé du greffon et d'éviter les complications métaboliques liées au surdosage de ces médicaments [2].

Le modèle Transformer a pu prédire un risque élevé d'événements cardiovasculaires pour un patient 18 après la TH. Ce qui s'est avéré malheureusement vrai, le patient est décédé 5 ans plus tard d'une insuffisance cardiaque. Ce patient avait été transplanté pour une cholangite biliaire primitive sans CHC et n'aurait pas été considéré comme à risque plus élevé de cancer ou complications cardiovasculaires sur la seule base du jugement clinique. Mais si l'on met en place le modèle Transformer comme assistant dans la prise en charge des suivis post-TH; on voit bien que le médecin aurait pu remarqué lors de la visite au cours de la 18ème année après la greffe que le patient passait de 17% de risque de d'événements cardiovasculaires dans un délai d'1 an à 28% dans un délai de 5 ans (voir Figure 10). Le médecin aurait alors pu accroître sa surveillance et adapter la posologie des médicaments en fonction de cette observation afin de réduire au maximum ce risque. D'autant plus, que ce risque ne cessait d'augmenter au fur à mesure des suivis:

- sur les prédictions à 1 an on est passé de 17% à la 18ème année post-TH à 32% à la 22ème année post-TH)
- sur les prédictions à 5 ans on est passé de 28% à la 18ème année post-TH à 35% à la 22ème année post-TH)

La faible différence entre la prédiction à 1 an et à 5 ans d'événements cardiovasculaires au cours de la 22ème année de TH, en plus de la valeur prédominante par rapport aux autres risques, étaient des indicateurs à prendre en compte de toute urgence. Au cours de cette 22ème année, le patient avait développé une fibrillation auriculaire et est décédé d'une insuffisance cardiaque, 23 ans après la transplantation [2].

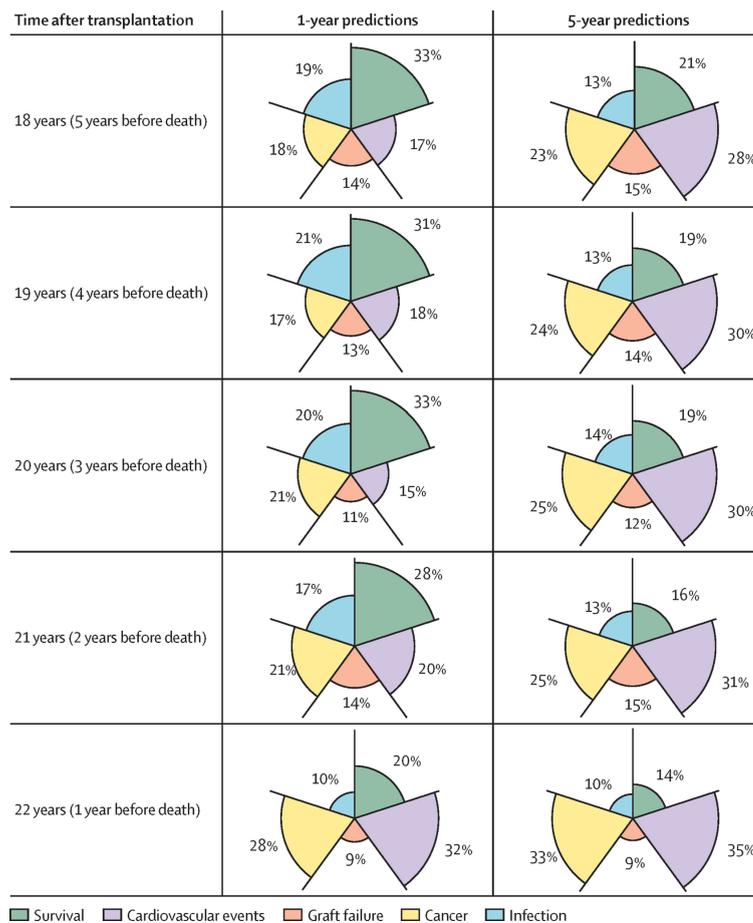


Figure 10 : Exemple de résultats de prédiction fait par le modèle Transformer pour un patient 5 ans avant son décès [2]

Cet exemple d'application nous montre d'ores et déjà la grande aide que pourrait nous apporter l'IA dans la prise en charge des patients post-TH sur le long terme en maximisant la qualité et la durée de vie des receveurs.

Dans l'article [16], un modèle MLP a été développé pour prédire la survie des patients de 6 mois à 13 ans après la transplantation. Le modèle MLP s'est formé à partir de 27 variables d'entrée sur le donneur, le receveur ainsi que la transplantation. MLP donne une variable de sortie binaire appelée GRF_STAT. Cette variable vaut :

- 0 en cas de prédiction de survie du receveur / du greffon
- 1 en cas de prédiction de non survie du receveur / du greffon

Pendant la phase de formation, la fonction d'activation qui fut utilisée dans les couches cachées était la fonction sigmoïde.[16]

Le nombre de couche cachées (a) nécessaire à été calculé à partir de la formule suivante:

$$a = (I + O) / 2$$

où I désigne le nombre de variables d'entrée (ici 27) et O le nombre de classes (ici 2). On obtient donc 14 couches cachées. Il faut savoir que plus le nombre de couches cachées est grand, plus la phase d'entraînement est longue et plus le résultat est précis, les marges d'erreur étant réduites.

Contrairement à l'étude menée dans l'article [2], la probabilité de survie était globale à tous les receveurs. En effet, chaque année après TH, la probabilité de survie était obtenue grâce à la formule suivante :

$$\text{pb de survie} = \frac{(\text{nombre de patients vivants au début}) - (\text{nombre de patients décédés à la fin})}{(\text{nombre de patients vivants au début})}$$

Un classement des variables les plus influentes dans la survie des patients post-TH est représenté dans la figure 11 grâce au modèle MLP.

Rank	Score	Attribute
1	89.9	Age_Don
2	88.6	BMI_TCR
3	87.1	Gender_Don
4	85.3	Final_Albumin
5	83.5	Creat_Don
6	81	SGOT_Don
7	78.2	SGPT_Don
8	76.9	Clin_Infect_Don
9	75.3	Final_Bilirubin
10	72.1	Final_Ascites
11	70	Final_INR
12	69.4	Gender
13	67.8	Diabetes_Don
14	67.5	Don_TY
15	66.6	Init_Age
16	64.2	TBili_Don
17	63.8	EXC_HCC
18	61.9	Final_MELD_PELD_LAB_Score
19	60	Final_MELD_or_PELD
20	58.4	Final_Serum_Creat
21	58.1	Final_Serum_Sodium
22	57.2	Non_Hrt_Don
23	55.9	Enceph_TCR
24	54.7	Malig_TRR
25	53.5	Malig_TCR
26	51.6	Num_Prev_Tx
27	50.5	TX_Liv

Figure 11 : Classement des variables d'entrée selon leur poids lors de la phase d'entraînement [16]

On peut voir que si l'on se concentre sur les 5 premières variables :

- du côté du donneur : son âge, son genre, son taux de créatinine jouent un rôle important dans la prédiction de la survie post-TH. Plus l'âge du donneur est grand, plus la probabilité de survie du greffon est faible, des médecins ont également montré que lorsque le donneur était une femme et le receveur, un homme les résultats de la TH étaient moins bon [16].
- du côté du receveur : son poids, son dernier taux d'albumine sont déterminants. Les chances de survie des patients obèses sont inférieures à celles des patients non obèses. [16]

Le modèle a prouvé une meilleure performance que le score MELD (voir Figure 12). On peut expliquer cela par le fait que le score MELD repose que sur des variables pré-transplantation tandis que le modèle MLP repose sur des variables pré et post-transplantation. Comme on peut le voir sur le graphique, le modèle MLP s'aligne quasiment sur la probabilité de survie réelle (obtenue à partir des enregistrements de la

base de données UNOS) de 6 mois post-TH jusqu'au 13 ans tandis que le score MELD se trouve loin derrière.

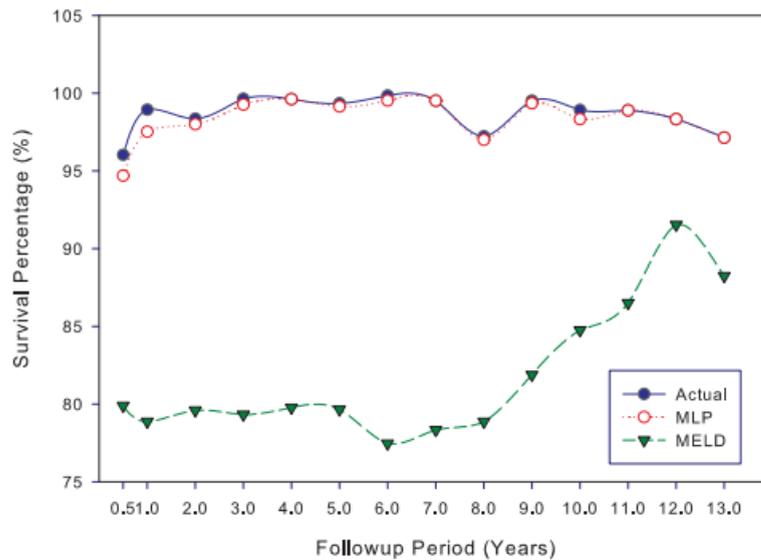


Figure 12 : Comparaison de la probabilité de survie de 6 mois à 13 ans post-TH avec MLP et le score MELD au la probabilité de survie actuel [16]

Comme tout modèle de DL, les limites de ces études [2,16] demeurent les données manquantes qui peuvent altérer la précision des résultats. Ce qu'on peut aussi appeler des données censurées est un handicap évident pouvant causer des résultats médiocres avec des AUC proches de 0,5 comme on peut le voir dans l'étude [14] appliqué sur la base de donnée UNOS.

b) Prédiction des complications post-transplantation

i) Prédiction de l'insuffisance rénale aiguë après une TH

L'insuffisance rénale aiguë (IRA) fait partie des complications les plus fréquentes post-TH. C'est une diminution rapide de la fonction rénale se manifestant souvent par une diminution de la production des urines ainsi que par une élévation des taux sanguins de créatinine et d'urée.

Cette complication post-TH est associée à une faible survie du greffon et une mortalité plus élevée du patient [11].

Dans les études [10] et [11], les chercheurs ont montré l'efficacité des algorithmes ML dans la prédiction de l'IRA post-TH. Les 2 études ont exclu les cas pédiatriques de leur échantillon de données. L'étude [11] a gardé les TH avec donneur vivant et décédé à l'inverse de l'étude [10] où seules les TH avec donneur décédé ont été conservées.

Tous 2 mettent en évidence GBM comme le modèle le plus performant (AUROC = 0,90 [11], AUC = 0,76 [10]) suite à une comparaison avec plusieurs modèles comme : SVM, RF, LR [10,11] et MLP [11].

Les facteurs prédominants dans la survenue de l'IRA post-TH sont surtout des variables peropératoires. GBM ainsi que RF ont mis en évidence le temps d'ischémie froide et la saturation veineuse mixte en oxygène (SvO2) moyenne [10,11] comme les variables les plus importantes dans la prédiction de l'IRA post-TH [11]. Le temps d'ischémie froide est relevé dans l'étude [11] car il y a une nette différence entre celle d'un greffon provenant d'un donneur vivant ou décédé. En effet, le temps d'ischémie froide ne serait peut-être pas vu comme critère discriminant si l'on ne regroupait pas les donneurs vivants et décédés ensemble. En revanche, un faible taux de SvO2 signifie une mauvaise alimentation en oxygène des principaux organes dont le rein ce qui peut favoriser une IRA. Il s'avère que lors de la TH, le clampage de la veine cave inférieure et les saignements durant l'opération entraînent un apport insuffisant d'oxygène.

Par conséquent, SvO2 peut être un facteur important dont le bon ajustement durant l'opération pourrait réduire l'incidence de l'IRA post-TH. D'autres essais prospectifs pourraient évaluer l'effet de l'optimisation de la SvO2 peropératoire pour réduire le risque d'IRA post-TH [11] et venir confirmer ou infirmer cette hypothèse.

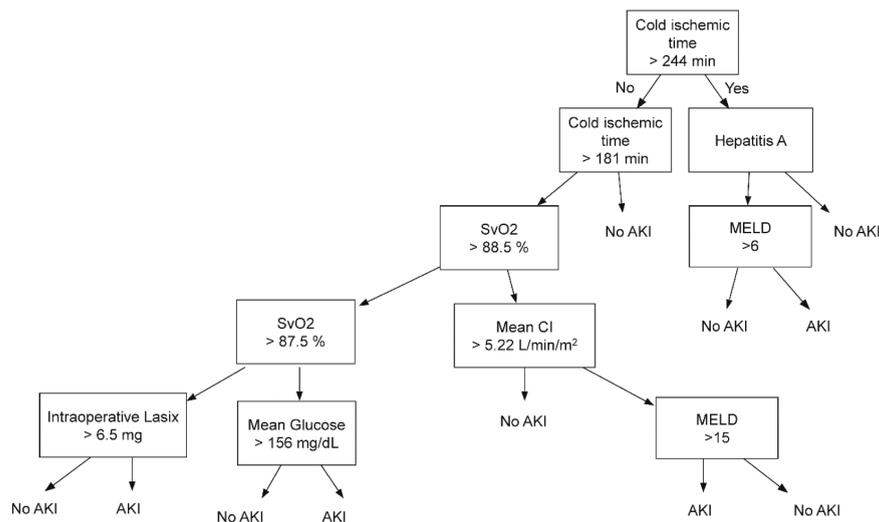


Figure 13 : Arbre de décision des patients IRA et non IRA sur la base du temps d'ischémie, SvO2, score MELD et autres variables [11]

L'étude [10], qui ne prend pas en compte les donneurs vivants, relève 5 principales variables par la méthode de SHAP comme contribuant au diagnostic de l'IRA post TH faite par le modèle GBM à savoir une bilirubine indirecte préopératoire élevée, un faible débit urinaire peropératoire, une longue durée d'anesthésie, un faible nombre de plaquettes préopératoires et une stéatose du greffon de grade NASH CRN 1 et supérieur [10].

Les échantillons étant de petites tailles (moins de 1000), ils seraient intéressants d'appliquer à l'avenir ces études à des échantillons de plus grandes tailles. De plus, si ces études étaient multicentriques, nous pourrions découvrir d'autres facteurs prédisant l'IRA post-TH et généraliser les observations relevées afin de mettre en place un système de prédiction universelle qui guidera le médecin lors de la prise en charge du patient.

En attendant, chacun de ces articles [10] et [11], évoquent la mise en place de leur propre système de prédiction basé sur GBM afin de prédire l'IRA à la fin de la chirurgie dont voici celui mis en place par les chercheurs de l'article [10] : <http://wb.aidcloud.cn/zssy/aki.html> .

ii) Prédiction du diabète après une TH

Le diabète survient chez certains patients au courant de la 1ère année post-TH. Le diabète post-TH tout comme le diabète préexistant avant la greffe n'ont pas d'effet sur la survie à court terme mais réduisent la survie sur le long terme. En effet, les patients atteints du diabète post-TH avaient une survie significativement réduite avec un risque de décès 55% plus élevé à 10 ans que les greffés n'ayant pas développé le diabète [18]. Il convient donc de savoir quels sont les facteurs qui favorisent le développement du diabète post-TH ? Ceci afin d'apporter des réponses efficaces qui limiteraient le développement de cette maladie après la greffe et ainsi améliorer la survie sur le long terme des patients post-TH.

Dans l'article [18], des algorithmes ML ont été mis en place afin de répondre à cette question et RF s'est avéré le plus performant avec un taux d'erreur plus petit dans la prédiction du risque de diabète 1 an après la TH que les algorithmes tels SVM, GBM et ANN.

Avec le modèle RF, 88% des données SRTR récoltées ont été classées correctement [18].

A l'inverse de l'IRA qui était favorisé par des facteurs peropératoires, RF a permis de relever des facteurs postopératoires comme prédictifs de l'apparition du diabète post-TH. L'étude [18] nous montre que l'utilisation du sirolimus comparé au tacrolimus et autres immunosuppresseurs augmente fortement le risque de développer un diabète post-TH (+33% avec sirolimus vs tacrolimus). L'immunosuppression est donc un facteur clé pour le diabète post-TH.

Parmi les facteurs prédictifs, on retrouve également des facteurs propres au receveurs: le sexe masculin, âge (plus le receveurs est âgé, plus il y a de chance qu'il développe un diabète post-TH) ainsi que l'IMC (le risque de diabète post-TH est augmenté de 13% avec l'obésité du receveur). En revanche, les facteurs propres aux donneurs n'ont pas montré de différences significatives [18].

Par conséquent, une solution pour réduire l'incidence du diabète post-TH serait d'éviter dans la mesure du possible la prescription du sirolimus sur le long terme car son utilisation chronique conduit à une résistance à l'insuline.

iii) Prédiction d'une récurrence chez les patients transplantés

Bien que la TH soit le meilleur traitement pour les patients atteints d'un CHC localisé, une récurrence survient chez 6 à 18% des patients [20] compromettant ainsi leur durée et conditions de survie post-TH.

Des études ont tenté de stratifier le risque de récurrence chez les patients transplantés en raison de CHC [9, 20] en développant des modèles tels que le modèle i-RAPIT (Risk Assessment of Post Liver Transplantation) [9] (AUC = 0,87). Ceci afin d'accompagner le

médecin dans la prise en charge des patients post-TH et accroître leur surveillance auprès des patients à haut risque de récurrence après la transplantation notamment en leur prescrivant plus d'imagerie voir en leur fournissant une chimiothérapie pour prévenir le cancer avant que des signes de récurrence ne soient constatés [9].

L'immunosuppression à base de sirolimus a été associée à une amélioration de la survie à 5 ans chez les patients qui étaient atteints du CHC [20] ce qui vient relever un défi difficile car nous avons vu dans l'étude [18] que le sirolimus à long terme augmentait le risque de développer un diabète post-TH.

Le modèle i-RAPIT utilise les réseaux de neurones CapsNet, un type de neurones artificiels basé sur le CNN. Il explore plusieurs types de données pour identifier les marqueurs cliniques qui prédisent la récurrence après une greffe de foie pour le CHC [9].

Le modèle pourra par la même occasion modifier le système d'allocation pour les patients atteints de CHC avec une taille tumorale plus élevée (> 5 cm voir >2 cm en France selon le score Foie mis en place par l'ABM). Actuellement, ces derniers ne sont pas éligibles à la TH de par leur tumeur qui est trop grande et risquant de donner des résultats post-TH médiocres. Les algorithmes du DL pourraient donc venir changer le sort de ses patients [9].

Conclusion

De nombreux domaines ont fait appel à l'IA y compris le domaine médical avec l'utilisation d'algorithmes de ML notamment pour l'analyse d'imagerie, le diagnostic de certaines maladies et l'aide à la décision clinique.

Tout au long de ce mémoire, nous nous sommes intéressés au potentiel de l'IA dans le cadre de la transplantation d'organes plus particulièrement de la TH. Différentes applications de ML ont été explorées répondant chacune à des enjeux importants de la TH :

- Avant la TH :
 - la prédiction de la mortalité en attendant la TH
 - la prédiction de la survie sur 3 mois et 12 mois post-TH
 - l'optimisation de l'appariement DR pour de meilleurs résultats de la TH et de meilleures conditions de vie des receveurs
- Après la TH :
 - la prédiction de la survie après la TH sur le long terme
 - la prédiction des éventuelles complications après la TH comme l'insuffisance rénale aiguë, le diabète, une récurrence du CHC

Les résultats obtenus avec l'IA ont dépassé les scores actuels, on pourrait donc se dire qu'il est évident que les équipes médicales doivent faire appel à l'IA dans leur prise de décision tout le long de la prise en charge du patient : de l'inscription du patient sur la liste d'attente à la prise en charge post-TH sur le court et long terme.

Cependant la plupart des modèles proposés présentent des limites :

- Les données manquantes peuvent altérer la précision des résultats en particulier pour les modèles de DP comme ANN qui sont sensibles à la qualité des données.
- Les résultats sont propres aux bases de données à partir desquelles ils ont été créés et formés, et sont encore difficilement généralisables aux données des autres centres de santé. Les modèles ne sont pas universels.
- Malgré les capacités avancées des techniques d'IA à adopter une pensée "humaine", le jugement clinique d'une équipe de transplantation expérimentée peut ne pas être reproductible.

Des travaux futurs sont nécessaires pour améliorer la qualité des données disponibles, en particulier sur les grandes bases de données comme UNOS où la quantité d'information est énorme [4]. Les modèles d'IA vus jusqu'à présent nécessitent une validation dans des bases de données autres que celles où elles ont été formés afin de les affiner et s'assurer que ces modèles sont efficaces. Une formation des modèles ML avec une accumulation de données internationales serait nécessaire pour former un modèle capable d'évaluer avec précision les patients au cas par cas [12].

La survie des transplantés pourrait un jour être améliorée par l'intégration de l'IA dans la pratique clinique, mais une évaluation plus approfondie grâce à la collaboration de médecins en transplantation et d'experts en IA est d'abord nécessaire.

Dans un avenir proche, les modèles basés sur l'IA pourraient ainsi être utilisés à l'échelle mondiale comme un outil de décision pour les médecins.

En attendant, une application récente a été développée nommée E-Stella [36] qui permet de classer et mesurer avec précision le taux de stéatose d'un foie, en vue d'une transplantation, grâce à l'IA. Elle est en cours de validation clinique par les équipes médicales.

Certains problèmes indépendamment de la TH sont à prendre en compte notamment la confiance des équipes médicales qui n'est pas encore gagnée. Jusqu'ici, les décisions cliniques s'appuient entièrement sur l'expertise des médecins, on acceptera alors avec réserve les recommandations de ces nouvelles technologies qui sont censées jouer un rôle d'assistant et non faire le travail des médecins au risque de les déresponsabiliser.

On peut donc cloturer ce mémoire en se demandant à quel point peut-on faire confiance à l'IA dans les décisions de transplantations ?

Glossaire

TH : Transplantation Hépatique
MELD : Model for End-Stage Liver Disease
SOFT : Survival Outcomes Following Liver Transplantation
BAR : Balance of Risk
DR : Donneur-Receveur
IA : Intelligence Artificielle
ML : Machine Learning
DL : Deep Learning
ECD : Expanded Criteria Donors
DRI : Donor risk index
SCD : Standard Criteria Donors
HLA : Human Leukocyte Antigen
ET -DRI : Eurotransplant Donor Risk Index
IMC : Indice de Masse Corporelle
ANN : Artificial Neural Networks
LR : Logistic Regression
OCT : Optimal Classification Trees
RF : Random Forest
GBM : Gradient Boosting
CNN : Convolutional Neural Network
TCN : Temporal Convolutional Network
RNN : Recurrent Neural Network
AUC : Area Under the Curve
AUROC : Area Under the Receiver Operating Characteristic curve
CHC : Carcinome Hépatocellulaire
OPOM : Optimized Prediction of Mortality
LSAM : Liver Simulated Allocation Model
CCR : Correct Classification Rate
MS : Minimum Sensitivity
SD : Standard Deviation
KCH : King's College Hospital
MLP : Multilayer perceptron
SVM : Support Vector Machine
ABM : Agence de Biomédecine
IRA : Insuffisance Rénale Aiguë
UNOS : United Network for Organ Sharing
UHN : University Health Network
SRTR : Scientific Registry of Transplant Recipients

Table des Figures

- Partie Background et IA définition :

Figure 1 : Les différentes phases et variables de la TH [17] -----	8
Figure 2 : Structure des ANN service de la TH [13] -----	14
Figure 3 : Différence de structure entre RNN et ANN [34] -----	15

- Partie Analyse :

Figure 1 : Arbre de classification sur 2 divisions [7] -----	19
Figure 2 : AUC de prédiction de la mortalité en liste d'attente à 3 mois OPOM vs MELD vs MELD-Na [7] -----	19
Figure 3 : Schéma de l'appariement DR avec ANN [5] -----	22
Figure 4 : Une simulation d'appariement DR avec ANN et des receveurs avec un score MELD moyen [8] -----	23
Figure 5 : Une simulation d'appariement DR avec ANN et des receveurs avec un score MELD élevé [8] -----	24
Figure 6 : Sélection des 15 principaux facteurs influençant les résultats de la TH dans les 30 jours [3] -----	26
Figure 7 : AUC des algorithmes ANN-CCR et ANN-MS et des scores de prédiction de survie [8] -----	27
Figure 8 : AUROC des différents algorithmes ANN , RF et des scores de prédiction de survie [3] -----	28
Figure 9 : AUROC de la prédiction de survie à 1 an et 5 ans selon les différents modèles [2] -----	30
Figure 10 : Exemple de résultats de prédiction fait par le modèle Transformer pour un patient 5 ans avant son décès [2] -----	32
Figure 11 : Classement des variables d'entrée selon leur poids lors de la phase d'entraînement [16] -----	33
Figure 12 : Comparaison de la probabilité de survie de 6 mois à 13 ans post-TH avec MLP et le score MELD au la probabilité de survie actuel [16] -----	34
Figure 13 : Arbre de décision des patients IRA et non IRA sur la base du temps d'ischémie, SvO2, score MELD et autres variables [11] -----	35

Références

- [1] - Briceño, Javier and al. "Use of artificial intelligence as an innovative donor-recipient matching model for liver transplantation: results from a multicenter Spanish study." *Journal of hepatology* 61.5 (2014): 1020-1028.
- [2] - Nitski, Osvald and al. "Long-term mortality risk stratification of liver transplant recipients: real-time application of deep learning algorithms on longitudinal data." *The Lancet Digital Health* 3.5 (2021): e295-e305.
- [3] - Lau, Lawrence and al. "Machine-learning algorithms predict graft failure after liver transplantation." *Transplantation* 101.4 (2017): e125.
- [4] - Guijo-Rubio, David and al. "Statistical methods versus machine learning techniques for donor-recipient matching in liver transplantation." *Plos one* 16.5 (2021): e0252068.
- [5] - Briceño, Javier, María Dolores Ayllón, and Rubén Ciria. "Machine-learning algorithms for predicting results in liver transplantation: the problem of donor-recipient matching." *Current Opinion in Organ Transplantation* 25.4 (2020): 406-411.
- [6] - Veerankutty, Fadl H. and al. "Artificial Intelligence in hepatology, liver surgery and transplantation: Emerging applications and frontiers of research." *World Journal of Hepatology* 13.12 (2021): 1977.
- [7] - Bertsimas, Dimitris and al. "Development and validation of an optimized prediction of mortality for candidates awaiting liver transplantation." *American Journal of Transplantation* 19.4 (2019): 1109-1118.
- [8] - Ayllón, María Dolores, and al. "Validation of artificial neural networks as a methodology for donor-recipient matching for liver transplantation." *Liver Transplantation* 24.2 (2018): 192-203.
- [9] - He, Tiancheng and al. "An imageomics and multi-network based deep learning model for risk assessment of liver transplantation for hepatocellular cancer." *Computerized Medical Imaging and Graphics* 89 (2021): 101894.
- [10] - Zhang, Yihan and al. "An explainable supervised machine learning predictor of acute kidney injury after adult deceased donor liver transplantation." *Journal of translational medicine* 19.1 (2021): 1-15.
- [11] - Lee, Hyung-Chul and al. "Prediction of acute kidney injury after liver transplantation: machine learning approaches vs. logistic regression model." *Journal of clinical medicine* 7.11 (2018): 428.
- [12] - Khorsandi, Shirin E. and al. "Artificial Intelligence in Liver Transplantation." *Transplantation Proceedings*. Elsevier Vol. 53. No. 10 (2021)

- [13] - Briceño, Javier. "Artificial intelligence and organ transplantation: challenges and expectations." *Current Opinion in Organ Transplantation* 25.4 (2020): 393-398.
- [14] - Guijo-Rubio, David and al. "Modelling survival by machine learning methods in liver transplantation: application to the UNOS dataset." *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, Cham, (2019)
- [15] - Pérez-Ortiz, M., and al. "Synthetic semi-supervised learning in imbalanced domains: Constructing a model for donor-recipient matching in liver transplantation." *Knowledge-Based Systems* 123 (2017): 75-87.
- [16] - Raji, C. G., and SS Vinod Chandra. "Long-term forecasting the survival in liver transplantation using multilayer perceptron networks." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47.8 (2017): 2318-2329.
- [17] - Briceño, Javier, Rafael Calleja, and César Hervás. "Artificial intelligence and liver transplantation: Looking for the best donor-recipient pairing." *Hepatobiliary & Pancreatic Diseases International* (2022).
- [18] - Bhat, Venkat and al. "New-onset diabetes and preexisting diabetes are associated with comparable reduction in long-term survival after liver transplant: a machine learning approach." *Mayo Clinic Proceedings*. Elsevier Vol. 93. No. 12 (2018)
- [19] - Durand, François. "How to improve long-term outcome after liver transplantation?." *Liver International* 38 (2018): 134-138.
- [20] - Berenguer, Marina and al. "Posttransplant management of recipients undergoing liver transplantation for hepatocellular carcinoma. Working Group Report from the ILTS Transplant Oncology Consensus Conference." *Transplantation* 104.6 (2020): 1143-1149.
- [21] - Schlegel, A., and al. "Risk assessment in high-and low-MELD liver transplantation." *American Journal of Transplantation* 17.4 (2017): 1050-1063.
- [22] - Guide du score Foie, Agence de la BioMedecine (2019)
https://www.agence-biomedecine.fr/IMG/pdf/guide_score_foie_v3.pdf
- [23] - Wiesner, Russell and al. "Model for end-stage liver disease (MELD) and allocation of donor livers." *Gastroenterology* 124.1 (2003): 91-96.
- [24] - Institut national du cancer <https://www.e-cancer.fr/>
- [25] - Halldorson, J. B., and al. "D-MELD, a simple predictor of post liver transplant mortality for optimization of donor/recipient matching." *American Journal of Transplantation* 9.2 (2009): 318-326.
- [26] - Rana A, Hardy MA, Halazun KJ, and al. "Survival outcomes following liver transplantation (SOFT) score: a novel method to predict patient survival following liver transplantation." *American Journal of Transplantation* 8.12 (2008): 2537-2546.

[27] - Dutkowski, Philipp and al. "Are there better guidelines for allocation in liver transplantation?: A novel score targeting justice and utility in the model for end-stage liver disease era." *Annals of surgery* 254.5 (2011): 745-754.

[28] - Feng, Sandy and al. "Characteristics associated with liver graft failure: the concept of a donor risk index." *American journal of transplantation* 6.4 (2006): 783-790.

[29] - Rosenberger, Laura H. and al. "Donor risk index predicts graft failure reliably but not post-transplant infections." *Surgical infections* 15.2 (2014): 94-98.

[30] - Braat, Andries E. and al. "The Eurotransplant donor risk index in liver transplantation: ET-DRI." *American Journal of Transplantation* 12.10 (2012): 2789-2796.

[31] - LR : <https://www.techtarget.com/searchbusinessanalytics/definition/logistic-regression>

[32] - GBM : http://uc-r.github.io/gbm_regression

[33] - ANN : <https://www.techopedia.com/definition/5967/artificial-neural-network-ann>

[34] <https://www.analyticsvidhya.com/blog/2020/02/cnn-vs-rnn-vs-mlp-analyzing-3-types-of-neural-networks-in-deep-learning/>

[35] - Vaswani, Ashish and al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

[36] - E-Stella : <https://stellasurgical.com/e-stella/>