

Université de Montréal

**Classification de décès neurologique  
par traitement automatique de l'image**

*Par*

Johann Plantin

Faculté de Médecine

Mémoire présenté en vue de l'obtention du grade de Maîtrise

en Sciences Biomédicales

option Médecine Computationnelle

Avril 2023

© Plantin Johann, 2023



Université de Montréal

Faculté de Médecine

---

*Ce mémoire intitulé*

**Classification de décès neurologique  
par traitement automatique de l'image**

*Présenté par*

**Johann Plantin**

*A été évalué par un jury composé des personnes suivantes*

**Dr Michaël Chassé**

Directeur de recherche

**Dr Michaël Sauthier**

Président- rapporteur

**Dr Pascal Laferrière-Langlois**

Membre du jury



## Résumé

Le diagnostic de mort cérébrale est une étape complexe et chronophage lors de l'évaluation des patients en soins intensifs soupçonnés d'être en décès neurologique. Bien que les critères neurologiques cliniques qui déterminent la mort cérébrale soient largement acceptés dans le monde, le diagnostic reste imparfait et l'utilisation de tests auxiliaires tels que la perfusion tomographique cérébrale (CTP) est souvent nécessaire pour le confirmer. L'objectif principal de ce travail était d'explorer la faisabilité de classer la mort cérébrale à partir de scans CTP par le traitement automatique de l'image.

Les scans CTP de l'étude prospective canadienne multicentrique de validation du CTP pour le diagnostic de décès neurologique ont été regroupées à partir de 11 sites participants (INDEX-CTP, ClinicalTrials.gov, NCT03098511). Des caractéristiques spatiales et temporelles ont été extraites en utilisant une combinaison de deux modules de convolution et utilisées pour prédire la mort neurologique. Les performances du modèle ont également été évaluées sur différentes catégories de blessures cérébrales.

Les études de 217 patients ont été utilisées pour entraîner le modèle. Nous rapportons une AUC de 0,79 (IC95 % 0,76-0,82), un score F1 de 0,82 (IC95 % 0,80-0,83), une précision de 0,92 (IC95 % 0,91-0,93), un rappel de 0,76 (CI95 % 0,72-0,79) ainsi qu'une valeur prédictive négative de 0,49 (CI95 % 0,45-0,53). En raison de la petite taille d'échantillon, nous n'avons pas effectué de tests statistiques sur des sous-ensembles de lésions cérébrales, mais avons signalé une valeur prédictive négative du modèle présumé plus élevée sur des blessures cérébrales anoxiques avec 0,82 (CI95 % 0,77-0,87).

Ce modèle montre des preuves préliminaires soutenant la faisabilité de développer un réseau neuronal profond pour classer les patients comateux comme étant neurologiquement décédés ou non. Des recherches supplémentaires sont nécessaires pour valider et améliorer le modèle en utilisant des ensembles de données plus vastes et diversifiés.

**Mots-clés** : mort cérébrale, don d'organes, modèle prédictif, apprentissage automatique, réseaux neuronaux convolutifs, tomographie à perfusion, imagerie cérébrale.

## Abstract

The diagnostic of brain death is a complex and chronophage step when evaluating patients in critical care suspected of being neurologically deceased. Although the clinical neurological criteria that determine brain death are mostly accepted worldwide, the diagnosis remains imperfect and often the use of ancillary tests such as brain computed tomography perfusion (CTP) are required to confirm it. The main objective of this work was to explore the feasibility of classifying brain death from CTP scans using deep learning.

CTP studies from a multicenter prospective diagnostic cohort study with the primary objective of evaluating the diagnostic accuracy of neurological death using CTP were pooled from 11 participating sites (INDEX-CTP, ClinicalTrials.gov, NCT03098511). Spatial and temporal features were extracted using a combination of two convolution modules and used to predict neurological death. The performance of the model was also evaluated on subsets of cerebral injuries.

217 patients' studies were used to train the model. We report an AUC of 0.79 (IC95% 0.76-0.82), a F1 score of 0.82 (IC95% 0.80-0.83), a precision of 0.92 (IC95% 0.91-0.93), a recall of 0.76 (CI95% 0.72-0.79) as well as a negative predictive value of 0.49 (CI95% 0.45-0.53). Due to a lack of sample size, we did not perform statistical tests on subsets of cerebral injury, but report suspected higher model negative predictive value on anoxic cerebral injury with 0.82 (CI95% 0.77-0.87).

This model shows preliminary evidence supporting the feasibility of developing a deep neural network to classify comatose patients as neurologically deceased or not. Additional research is needed to validate and refine the model by employing larger and more diverse datasets.

**Keywords** : brain death, organ donation, predictive model, machine learning, convoluted neural networks, computed tomography perfusion, brain imaging

# Table des matières

## Contents

Résumé .....	5
Abstract .....	6
Table des matières .....	7
Liste des tableaux .....	10
Liste des figures .....	11
Liste des sigles et abréviations .....	12
Liste des annexes .....	13
Remerciements .....	15
Chapitre 1 – Introduction .....	16
1. 1 Détermination du décès neurologique .....	16
1.1.1 Définition du décès neurologique .....	16
1.1.2 Diagnostic clinique de mort cérébrale .....	20
1.1.3 Diagnostic de mort cérébrale: Tests auxiliaires .....	21
1.1.4 Interprétation des test auxiliaires .....	22
1.1.5 Tomodensitométrie à perfusion .....	24
1.1.6 Le CTP comme test auxiliaire .....	24
1.2 L'apprentissage profond .....	25
1.2.1 Introduction à l'intelligence artificielle .....	25
1.2.2 Fondement de l'apprentissage profond .....	27
1.2.3 Limitation des réseaux de neurones artificiels .....	31
1.2.4 Introduction aux réseaux de neurones convolutifs .....	33

1.2.5 Entraînement des réseaux de neurones convolutifs .....	36
1.3 Applications de l'apprentissage profond en médecine .....	40
1.4 Apprentissage profond et imagerie médicale.....	42
1.5 Mise en contexte de la question de recherche.....	43
Chapitre 2 : Matériel et méthodes.....	45
2.1 Hypothèse .....	45
2.2 Objectifs .....	45
2.2.1 Objectif primaire .....	45
2.2.2 Objectif secondaire .....	45
2.3 Source des données .....	46
2.3.1 Introduction à l'étude INDex-CTP .....	46
2.4 Population à l'étude, critères d'inclusion et d'exclusion et sous-groupe d'intérêt .....	48
2.4.1 Analyse de sous-groupes.....	48
2.5 Préparation des données .....	49
2.5.1 Description de la donnée .....	49
2.5.2 Pré-traitement de la donnée.....	50
2.6 Classificateur supervisé: Réseaux de neurones convolutifs.....	52
2.6.1 Introduction à l'architecture .....	52
2.6.2 Entraînement du modèle .....	54
2.6.3 Analyses statistiques .....	55
Chapitre 3 : Résultats sous forme d'article .....	56
3.1 Introduction.....	56
3.2 Classification of neurological death through automatic image processing.....	58
3.2.1 Abstract .....	58



3.2.2 Introduction.....	59
3.2.3 Methods .....	60
3.2.4 Results .....	65
3.2.5 Discussion.....	71
3.2.6 Competing interest.....	73
3.2.7 Data availability .....	73
3.2.8 Bibliography.....	73
3.2.9 Supplementary materials.....	74
Chapitre 4: Discussion .....	78
4.1 Retour sur les objectifs du projet.....	78
4.2 Retour sur les résultats .....	79
4.3 Forces .....	80
4.4 Limitations.....	82
4.4 Implications et directions futures .....	85
4.5 Conclusion .....	88
5: Références bibliographiques.....	89

## Liste des tableaux

## Liste des figures

- Figure 1. - Évolution du nombre de donneurs après déclaration de DN entre 2000 et 2020 à l'échelle mondiale.....
- Figure 2. - Évolution du nombre de donneurs après déclaration de mort circulatoire (DDC) entre 2000 et 2020 à l'échelle mondiale.....
- Figure 3. - Comparatif des différents tests auxiliaires pour la détection du flot, la perfusion et de la fonction au sein du cerveau.....
- Figure 4. - Représentation des relations entre l'intelligence artificielle, l'apprentissage machine et l'apprentissage profond.....
- Figure 5. - Réseau de neurones artificiels performant de simples computations logiques.....
- Figure 6. - Illustration d'un fonctionnement d'une ULS.....
- Figure 7. - Architecture d'un perceptron.....
- Figure 8. - Architecture d'un perceptron multicouche.....
- Figure 9. - Architecture du RNC LeNet-5.....
- Figure 10. - Comparaison entre l'architecture du RNC LeNet-5 (Gauche) et AlexNet (Droite).....
- Figure 11. - Performance des meilleurs modèles en fonction de leur complexité.....
- Figure 12. - Exemple de cartes de perfusions.....
- Figure 13. - Architecture du classificateur supervisé.....
- Figure 14. - Distribution du type de blessures cérébrales.....

## Liste des sigles et abréviations

AM: Apprentissage Machine

AP: Apprentissage Profond

AT : Angiographie par Tomodensitométrie

CTP: Angiographie par Tomodensitométrie avec Perfusion (de l'anglais *Computed Tomography (CT) Perfusion*)

IA: Intelligence Artificielle

DICOM : de l'anglais *Digital Imaging and Communications in Medicine*

DN : Décès Neurologique

DDC: Mort Circulatoire

RNA: Réseaux de Neurones Artificiels

RNC: Réseaux de Neurones Convolutifs

ROC: *Receiver Operating Characteristic Curve*

SE: Compression et excitation (de l'anglais *Squeeze and Excite*)

ULS: Unité Logique de Seuil

## Liste des annexes

*A mon fils, Noah*

## Remerciements

En premier lieu, je tiens à remercier mon superviseur de maîtrise, le Dr Michaël Chassé. Son soutien et sa patience m'ont permis d'avancer avec persévérance face à des problèmes difficiles à surmonter. J'admire énormément sa capacité et son désir de consacrer du temps à ses étudiants malgré toutes ses responsabilités. Le Dr Michaël Chassé est également un excellent mentor qui sait rendre des problématiques compliquées en concepts simples et concis pour que ses étudiants puissent avancer et continuer leur apprentissage. Son soutien a été primordial pour l'accomplissement de ce projet.

Je tiens également à remercier l'équipe de CITADEL qui m'a accueilli avec bonne humeur et professionnalisme. Souvent, je me suis retrouvé dans des difficultés qui semblaient insurmontables, mais il y avait toujours quelqu'un pour m'aider et trouver des solutions créatives.

Je tiens à remercier le Dr Claude Kauffmann qui m'a été d'une aide inestimable tout au long de mon projet. Son expertise, sa patience et son désir d'aider m'ont vraiment aidé à avancer dans les moments difficiles.

Un grand merci va également à mon comité de mentorat qui m'a énormément aidé lors des évaluations de progrès ainsi que lors de nos sessions de discussions. Merci aux Dr Laurent Létourneau-Guillon et Dr Louis Mullie pour leur soutien.

Enfin, je tiens à remercier la personne qui m'a le plus permis d'avancer dans les moments de doute, de fatigue et de découragement. L'amour et le soutien qu'elle m'a apporté pendant ces deux années m'ont donné la force nécessaire pour surmonter les difficultés de la vie et mener à bien ce projet. Merci Julia.

# Chapitre 1 – Introduction

## 1. 1 Détermination du décès neurologique

### 1.1.1 Définition du décès neurologique

Le concept de la mort et de la vie peut parfois sembler abstrait si l'on y accorde réflexion et ouverture d'esprit. Notre compréhension de ce concept oscille entre la culture, la philosophie et la science. La culture peut en effet altérer la définition de la mort, il existe une multitude de cultures et de religions qui ont une vue distincte de ce qu'est la mort et bien souvent, elles distinguent le corps et l'esprit. La mort peut être le passage de notre esprit d'une dimension à une autre, laissant nos composants biologiques et notre matière tangible. La philosophie nous permet de constamment étudier la mort dans un contexte scientifique ou non. La conceptualisation de la personne peut aussi altérer la définition propre de la mort et il existe plusieurs écoles de pensée philosophiques qui différencient la conscience de la personne (1). La définition de la mort a changé à travers le temps. Avant l'introduction de l'assistance respiratoire mécanique, un individu était déclaré mort selon un principe cardio-pulmonaire ce qui laissait moins de place à l'ambiguïté, mais à plus d'erreurs de diagnostic de mort (2). En repoussant les limites concernant le maintien des fonctions et le maintien de la vie d'individus, l'invention de l'assistance respiratoire mécanique en 1952 lors de l'épidémie de poliomyélite (3) a introduit une ambiguïté concernant la définition de la mort et du point de non-retour (2). La nécessité de définir la mort autrement que par approche purement cardio-pulmonaire est devenue nécessaire afin de traiter les patients de façon appropriée. Le concept de décès neurologique (DN) ou de mort par critère neurologique est né en 1959 sous le nom de *coma dépassé* (4) et c'est en 1968 que l'établissement de critères cliniques du coma irréversible sont établis sous le nom de *Brain Death* par le *Harvard Brain Death Criteria* (5). Cet état critique du patient a donné naissance à une myriade de discussions politiques, métaphysiques et scientifiques sur ce qu'est la mort cérébrale, notamment par cette notion de point de non-retour, ou d'irréversibilité (6-10). Malgré ces discussions, désaccords et réflexions, le concept même de DN a de façon générale été accepté



dans le domaine médical (11), malgré certaines dissensions (12). L'importance d'avoir un algorithme défini pour la détermination du décès neurologique était donc fondamentale à sa précision et reproductibilité. C'est en 1995 aux États-Unis qu'un guide conçu spécifiquement à cette fin a été élaboré. Ce guide était constitué de composantes principales au diagnostic de DN, comme l'absence de réflexes du tronc cérébral accompagnée d'un coma ainsi qu'un test d'apnée (13). À la suite du passage du temps, la prise en compte des critiques éthiques, scientifiques, culturelles et religieuses, la définition de décès neurologique a finalement été reconnue globalement non seulement scientifiquement, mais aussi par les grands groupes religieux et les systèmes gouvernementaux (14). Il est intéressant de constater que malgré une reconnaissance du concept général de DN, des variations régionales et internationales persistent en termes de concept algorithmique ainsi qu'en pratique, causant des confusions et difficultés pour conduire des études cliniques randomisées à grande échelle, ou avoir un concept robuste et applicable universellement (2, 15, 16). Ce n'est que récemment qu'une communauté experte concernant plusieurs aspects du concept de DN s'est regroupée dans le but d'établir un consensus international concernant sa définition et son diagnostic afin d'obtenir un consensus ou un guide de recommandations et de suggestions basé sur l'information présentée dans la littérature, ainsi que l'opinion reçue de regroupement d'experts multidisciplinaires regroupant plusieurs concepts importants (2). Le décès neurologique peut prendre plusieurs formes conceptuelles, certaines plus strictes que d'autres, comme la mort cérébrale totale en opposition à la mort du tronc cérébral. La première est une forme plus stricte que la seconde, puisqu'elle a pour prérequis la mort du tronc cérébral et qu'elle inclut la mort biologique et l'absence de fonctions comme l'osmorégulation (2, 17).

Le don d'organes est une raison majeure qui explique le besoin d'une rigueur et précision reproductible lors du diagnostic de décès neurologique. Un individu déclaré mort peut, si compatible, donner un/des organe/s à un autre individu dans le besoin. Plusieurs facteurs expliquent une haute demande d'organes et une faible disponibilité relative. En premier temps, il existe plusieurs barrières naturelles à la transplantation d'organes, comme la compatibilité biologique ou encore la qualité de l'organe en question. Il existe une quantité supérieure d'individus dans le besoin d'organes comparativement aux individus déclarés morts acceptant de

donner leurs organes et aux individus vivants donnant volontairement un organe qui ne causerait pas leur décès (i.e. un rein, un poumon).

Une des conséquences du manque de standardisation du DN est la diminution de la capacité potentielle maximale de don d'organes puisqu'il est plus complexe, chronophage et variable (en termes d'algorithme décisionnel inter/intra national) de diagnostiquer celui-ci comparativement à la mort circulatoire (DDC) (18). En 20 ans (2000-2020), la quantité totale de donneurs diagnostiqués de DN a augmenté considérablement soit d'environ 1.9 fois, et la quantité de donneurs diagnostiqués de DN par million d'habitants a augmenté similairement. En 2020 plus de 40 pays additionnels ont participé au don d'organes comparativement à l'année 2000 (Figure 1). Cette augmentation est en partie due à l'augmentation des incitatifs, le consentement présumé, l'amélioration de la rapidité, qualité et sécurité du transport d'organes (19-21). L'augmentation du don d'organes après diagnostic de DDC a augmenté plus drastiquement que les dons d'organes par DN, mais les chiffres absolus montrent clairement qu'une plus grande proportion d'organes sont donnés sous DN que sous DDC (environ 28 000 et 8000 respectivement en 2020) (Figure 1, Figure 2) (22). Considérant la complexité et la variabilité dans la détermination de DN, il n'est pas étonnant que la standardisation globale de diagnostic de DN occasionne une augmentation du nombre et de la rapidité de disponibilité d'organes. La standardisation ne sera pas absolue, par le simple fait que la déclaration de DN a des composantes intrinsèques politiques, religieuses et culturelles ne permettant pas son application stricte universelle (10, 23, 24).

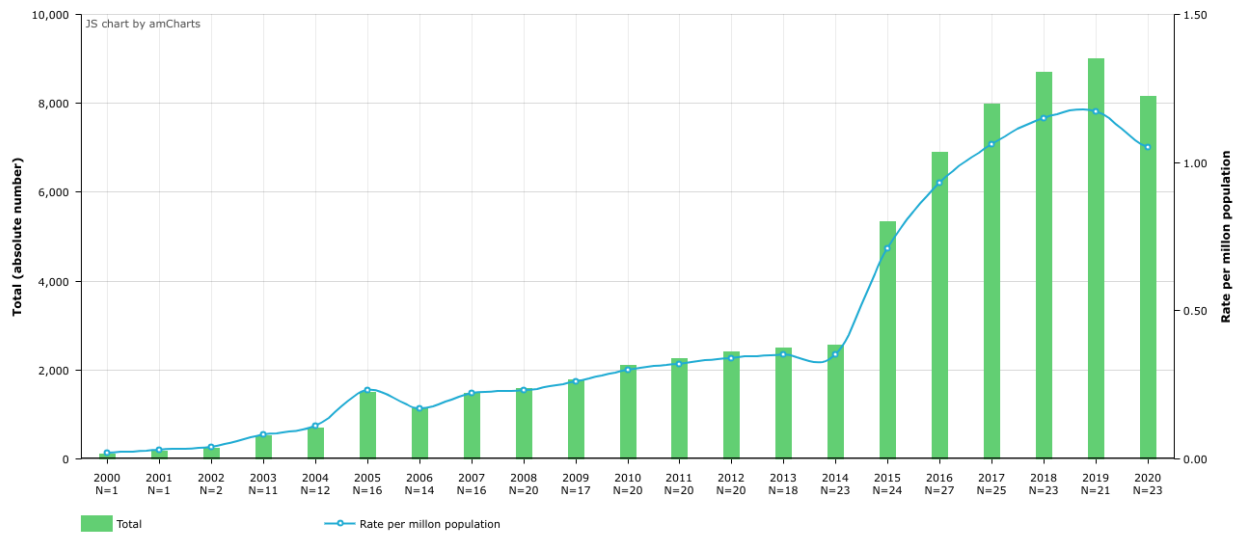
Une raison fondamentale qui nourrit les discussions autour du décès neurologique et du don d'organes provient d'une règle fondamentale, la *Dead Donor Rule* ou règle du donneur mort. Cette règle est absolue et stipule qu'un individu doit être déclaré mort avant le don de ses organes (25) et certaines discussions ont émergé argumentant que la déclaration de DDC brise ce concept fondamental (26). Ces discussions et remises en question soutiennent l'idée que l'établissement de règles/guides afin de diagnostiquer le DN serait de grande utilité afin de non seulement diminuer les barrières au don d'organes, mais aussi d'augmenter le potentiel d'organes disponibles tout en respectant l'éthique scientifique et humaine (2, 18).

**Total Number of Actual donors after brain death (DBD) ,i.e., actual deceased organ donors in whom death has been determined by neurological criteria (Global.2000-2020)**  
 Source: GODT (<http://www.transplant-observatory.org>)



**Figure 1. - Évolution du nombre de donneurs après déclaration de DN entre 2000 et 2020 à l'échelle mondiale (22).**

**Total Number of Actual donors after circulatory death (DCD) ,i.e., actual deceased organ donors in whom death has been determined by circulatory criteria (Global.2000-2020)**  
 Source: GODT (<http://www.transplant-observatory.org>)



**Figure 2. - Évolution du nombre de donneurs après déclaration de mort circulatoire (DDC) entre 2000 et 2020 à l'échelle mondiale (22).**

### 1.1.2 Diagnostic clinique de mort cérébrale

Malgré les débats et discussions concernant certains aspects du diagnostic de DN, les efforts pour standardiser le diagnostic ont permis d'obtenir une algorithmique décisionnelle guidant les professionnels à adopter les nouvelles directives ou modifier leurs critères. Il est accepté que la série de tests cliniques suivants permettent de diagnostiquer le DN :

(1) Il n'y a pas de preuve d'éveil ou de conscience à une stimulation externe maximale, y compris une stimulation visuelle, auditive et tactile nocive; (2) Les pupilles sont fixes, de taille moyenne ou dilatées et non réactives à la lumière; (3) Les réflexes cornéens, oculo-épiphysaires et oculo-vestibulaires sont absents; (4) Il n'y a aucun mouvement facial en réponse à une stimulation nocive; (5) Le réflexe nauséux est absent lors d'une stimulation pharyngée postérieure bilatérale; (6) Le réflexe de toux est absent lors d'une aspiration trachéale profonde; (7) Il n'y a pas de réponse motrice médiée par le cerveau à une stimulation nocive des membres; et (8) Les respirations spontanées ne sont pas observées lorsque les objectifs de l'épreuve d'apnée atteignent un pH <7,30 et une pression partielle de dioxyde de carbone  $\geq 60$  mm Hg. (Traduit de (2)).

Cette série de tests a pour but de démontrer la nature irréversible du coma par la détermination de l'aréflexie du tronc cérébral. En plus de ces tests, plusieurs suggestions et recommandations additionnelles sont ajoutées afin de tenter de standardiser les pratiques internationales lors de l'examen clinique (2). Il existe par exemple différentes pratiques concernant la question de combien d'examens cliniques doivent être effectués afin de diagnostiquer le DN (27). Une standardisation des définitions, approches et techniques afin de diagnostiquer le DN est un atout de haute conséquence dans la diminution d'erreurs de diagnostic de DN, ainsi que l'augmentation de la disponibilité et rapidité de don d'organes.

Il existe toutefois des situations de variations dans les approches en fonction de principes religieux, sociétaux ou culturels, ainsi que des limitations matérielles ou en termes de ressources (2). Malheureusement, il n'est pas toujours possible de procéder au standard de diagnostic de DN

uniquement par examen clinique. Il se peut que le patient soit dans une situation où il est par exemple impossible de procéder au test d'apnée, ou bien qu'il/elle soit sous d'autres facteurs de confusion (i.e. déformation faciale, instabilité pulmonaire, médication, etc.(2). Dans une situation comme celle-ci, des tests auxiliaires permettent d'aider le diagnostic de DN. Il est à noter que les tests auxiliaires ne sont pas toujours optionnels, dans certains pays ayant un protocole de diagnostic de DN, le test auxiliaire est une étape obligatoire (15, 28).

### **1.1.3 Diagnostic de mort cérébrale: Tests auxiliaires**

Les tests auxiliaires ont pour but de se rapprocher et refléter le plus possible l'élément suffisant de déclaration du DN, soit l'absence de fonctions du tronc cérébral qui permettent l'état de conscience et certaines fonctions vitales comme la respiration autonome (2). Bien qu'ils ne remplacent pas le standard d'évaluation clinique, ils sont utilisés afin d'aider le diagnostic de DN et ont leurs propres avantages et inconvénients. Il existe trois branches concernant les types de tests auxiliaires: l'évaluation de l'activité, du flot et de la perfusion (29). L'évaluation de l'activité regroupe les tests incluant l'électroencéphalographie, les potentiels évoqués et l'imagerie radionucléide et peut dans certains cas permettre une évaluation fonctionnelle du tronc cérébral, qui est l'élément le plus important pour l'établissement d'un diagnostic juste de DN. Malgré cela, ces tests sont souvent soumis aux mêmes limitations et facteurs de confusion que l'examen clinique et souffrent de variations d'interprétabilité entre les évaluateurs (29) et l'électroencéphalographie n'est plus recommandée comparativement aux autres tests disponibles (2, 30). L'évaluation basée sur le flot cérébral regroupe des techniques d'imageries cérébrales comme l'angiographie par soustraction numérique, le Doppler transcrânien, étude d'angiographie par radionucléides, l'angiographie par tomодensitométrie (AT) et l'angiographie par résonance magnétique. Finalement la branche de tests auxiliaires évaluant la perfusion contient par exemple l'angiographie par tomодensitométrie avec perfusion (CTP) (28) qui est au centre de ce travail. Cette panoplie de tests auxiliaires disponibles mène à un manque de standardisation concernant l'utilisation d'un test auxiliaire universel (31, 32). Inévitablement,

chaque technique d'imagerie a ses avantages et limitations dans son champ d'application. Le développement d'un test auxiliaire moins invasif et de haute précision comme l'AT pourrait ainsi permettre d'augmenter la rapidité de diagnostic de DN sans sacrifier la précision, afin d'accroître la rapidité de disponibilité du don d'organe tout en préservant la règle du donneur décédé (33).

#### **1.1.4 Interprétation des test auxiliaires**

Il est important de définir certains concepts clefs concernant l'interprétation des tests auxiliaires. Le flot, la perfusion et la fonction sont trois concepts fondamentaux qui sont parfois utilisés à tort de façon interchangeable. Le flot représente le mouvement du sang à travers les principaux vaisseaux cérébraux comme l'artère cérébrale antérieure, postérieure et moyenne. La perfusion représente le mouvement du flot vers les vaisseaux sanguins cérébraux de plus petits diamètres, les capillaires. Le flot est un prérequis de la perfusion. En revanche, le flot n'implique pas nécessairement la perfusion (29). Cela est important puisque la perfusion est un prérequis au fonctionnement neurologique du tronc cérébral. Finalement, la perfusion n'implique pas nécessairement la fonction (29). Les examens d'activité comme l'électroencéphalogramme ou les potentiels évoqués permettent une certaine estimation du statut fonctionnel du cerveau et tronc cérébral, mais sont soumises aux mêmes limitations que l'examen clinique. L'angiographie par soustraction numérique et l'AT sont capables de détecter un flot dans les artères principales irrigants le cerveau et le tronc cérébral, en revanche une limitation s'applique dans leur capacité à évaluer la perfusion plus profonde dans ces structures (Figure 3) (29) et plusieurs cas de faux négatifs (pas de détection de flot lorsqu'il y en a) ont été relevés (34). Plus récemment, un test auxiliaire capable d'évaluer qualitativement et quantitativement le flot et de détecter la perfusion au sein de structures profondes comme le cerveau et le tronc cérébral a été investigué: le CTP (35, 36) (Figure 3).

	Flow		Perfusion		Function
	Cerebrum	Brainstem	Cerebrum	Brainstem	
4-vessel angiography	⊕	⊕	⊖	⊖	⊗
Transcranial Doppler	⊕	⊖*	⊗	⊗	⊗
CT-angiography	⊕⊕	⊕	⊖	⊗	⊗
CT-perfusion	(⊕⊕⊕)	(⊕⊕⊕)	(⊕⊕⊕)	(⊕⊕⊕)	⊗
Radionuclide imaging	⊕⊕	⊖	⊖⊖	⊖	⊖
Electroencephalography	⊗	⊗	⊗	⊗	⊖
Evoked Potentials	⊗	⊗	⊗	⊗	⊖
Clinical examination	⊗	⊗	⊗	⊗	⊕⊕⊕

⊗ = not useful

⊖ = may be useful (conditions generally apply)

⊕ = useful

( ) = under study

\* Transcranial Doppler can evaluate flow in the vertebral arteries and the basilar artery (acoustic window permitting).

CT: computed tomography.

**Figure 3. - Comparatif des différents tests auxiliaires pour la détection du flot, la perfusion et de la fonction au sein du cerveau. (29)**

### **1.1.5 Tomodensitométrie à perfusion**

Le CT est une technique de rayons X combinée qui génère de multiples coupes transversales en deux dimensions en parcourant une surface prédéterminée du patient et la combinaison de ces coupes donne un volume (37, 38). Avant l'acquisition d'images du patient, un protocole de scanner est utilisé afin de régler des paramètres importants permettant de contrôler l'épaisseur des coupes, la distance entre les coupes, la résolution de l'image ou encore le nombre de coupes par volume et le nombre de volumes générés. Tous ces paramètres permettent le réglage des images qui vont être générées en fonction du site anatomique d'intérêt ou même de la pathologie d'intérêt (37, 38). La perfusion du CT provient de l'agent de contraste injecté au patient qui absorbe une partie de l'énergie des rayons-X et permet à certaines structures ou fluides d'être mis en avant par rapport à d'autres. L'avantage de la perfusion est non seulement la mise en avant de régions ou composants, mais aussi le déplacement de l'agent à travers l'espace d'intérêt (38). Les images générées sont en échelle de gris et sont standardisées par l'échelle Hounsfield. Cette échelle est définie comme l'atténuation de la valeur du rayon-X dans un voxel moins l'atténuation de l'eau (qui est la référence 0) multiplié par 1000 (39). L'utilité majeure du CTP est la génération de cartes de perfusions, qui sont des cartes de couleurs indicatrices de plusieurs paramètres comme le volume sanguin cérébral, le débit sanguin cérébral, le temps de transit moyen ainsi que le temps jusqu'au pic (40). Le CTP est souvent utilisé pour l'évaluation d'accidents vasculaires cérébraux aigus (40-44).

### **1.1.6 Le CTP comme test auxiliaire**

Quelques études pilotes suggèrent que le CTP a une sensibilité similaire à l'AT ou un score AUC (aire sous la courbe ROC) supérieur à des fins de diagnostic de DN comparativement à l'AT (45, 46). Plus récemment le CTP est reconnu comme un atout additionnel lorsqu'utilisé conjointement avec l'AT, augmentant potentiellement la sensibilité de cette dernière (36). Malgré un manque



de validation à grande échelle de la précision du CTP à prédire le DN, il a un potentiel d'accessibilité, de simplicité en termes d'exécution et de qualité d'interprétation de flot et perfusion. L'étude prospective canadienne multicentrique INDex-CTP (*CT-Perfusion for Neurological Diagnostic Evaluation*) d'une cohorte de 333 patients de validation du CTP pour le diagnostic de DN a été mise sur pied (ClinicalTrials.gov, NCT03098511) afin de combler le manque de validation du CTP à plus grande échelle et potentiellement l'établir comme test auxiliaire de référence. L'objectif primaire de cette étude était de comparer la précision du CTP à celle de l'examen clinique complet pour la détermination du DN. L'étude INDex-CTP a aussi plusieurs objectifs secondaires comme: a) l'évaluation du risque de la procédure (CTP) chez les patients suspectés d'être décédés neurologiquement; b) l'évaluation de la fiabilité d'interprétation du scan entre les évaluateurs (radiologistes); c) l'évaluation de la précision de diagnostic de l'AT comparativement à l'examen clinique et au CTP;

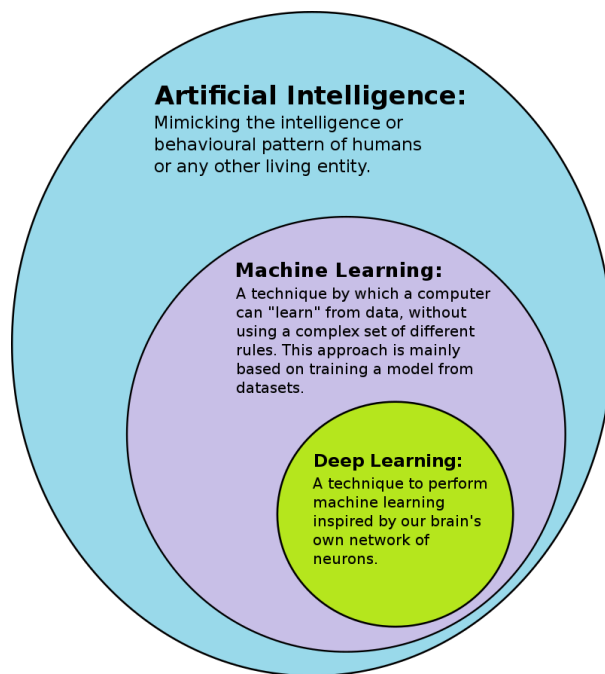
Une description plus approfondie de la structure d'INDex-CTP sera présentée dans la section matériel et méthode.

## **1.2 L'apprentissage profond**

### **1.2.1 Introduction à l'intelligence artificielle**

L'apprentissage profond (AP) prend racine de l'apprentissage machine (AM) qui prend à son tour racine de l'intelligence artificielle (IA) (Figure 4) (47). Beaucoup considèrent Alan Turing comme le père de l'IA (48), mais c'est en 1955 que le *Dartmouth Research Project* donne naissance à la première définition de l'IA : "Faire en sorte qu'une machine se comporte d'une manière qui serait qualifiée d'intelligente si un humain se comportait ainsi" (Traduction libre de (49)). Plusieurs définitions de l'IA sont présentes en fonction du domaine d'application et il est parfois difficile d'arriver à une définition standardisée, même au sein d'un même domaine d'application (50). Il

est potentiellement plus utile de visualiser l'IA comme l'un des quatre systèmes artificiels d'intelligence suivant: [a) Système pensant comme un humain; b) Système agissant comme un humain; c) Système pensant rationnellement; d) Système agissant rationnellement] (Traduit de (51)). Le manque de standardisation de la définition de l'IA a même induit un effort de AI Watch pour regrouper les différentes définitions interdisciplinaires entre 1955 et 2019 et tenter d'établir une définition opérationnelle de l'IA (52). Pour ce travail, nous avons gardé la définition de l'IA suivante: "capacité d'un système à interpréter correctement les données externes, à apprendre de ces données et à utiliser ces apprentissages pour atteindre des objectifs et des tâches spécifiques grâce à une adaptation flexible" (53).



**Figure 4. - Représentation des relations entre l'intelligence artificielle, l'apprentissage machine et l'apprentissage profond. (47)**

Tout comme l'IA, l'AM souffre également d'une grande variabilité dans la nature de sa définition. En général, la discipline de l'AM se situe dans un axe entre les statistiques, l'informatique ainsi que d'autres domaines potentiels en fonction de son application. L'AM est utilisé dans la vie quotidienne et a une pléthore de champs d'applications comme le traitement d'images, la reconnaissance vocale, la robotique, la modélisation prédictive ou encore le traitement des langues naturelles. Le domaine d'entraînement de l'AM le plus utilisé est l'apprentissage

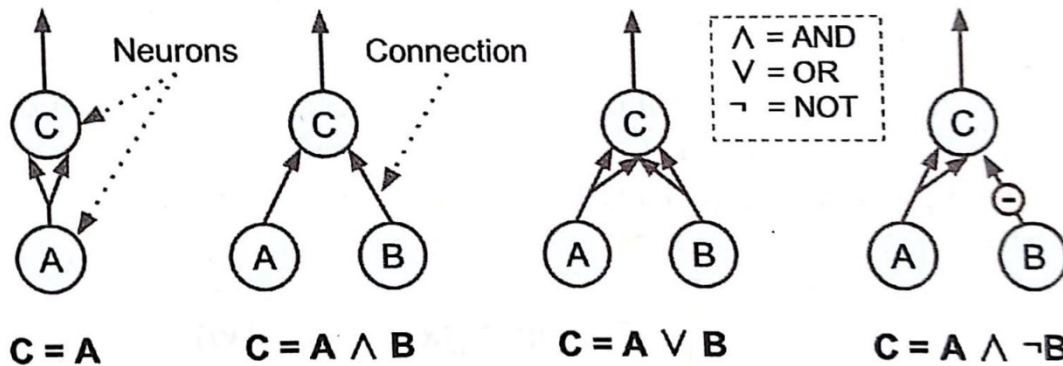
supervisé (54), où le programme reçoit en entrée un vecteur de variables ( $x_i$ ) afin de prédire une sortie ( $y_{pred,i}$ ) et va minimiser la différence entre la valeur prédite ( $y_{pred,i}$ ) et la vraie valeur ( $y_i$ ) associée au vecteur  $x$ . De façon générale, un problème d'apprentissage tourne autour du principe d'une tâche évaluée par une métrique de façon itérative sur chaque entraînement (54). Bien qu'il existe un vaste champ d'application de l'AM, ce mémoire présent se focalise sur l'imagerie médicale cérébrale. Le champ d'application de l'AM de ce travail converge donc vers le traitement d'images.

Le traitement d'images est aussi un domaine qui donne naissance à plusieurs sous-domaines comme la classification d'images, la détection d'objets, la segmentation, le suivi d'un objet dans l'espace et bien d'autres. Bien qu'il soit possible d'accomplir certaines tâches de traitement d'images sans nécessairement utiliser l'AP (55), il est généralement accepté que l'AP présente un avantage considérable sur l'AM dans ce domaine. Une des raisons expliquant cet avantage réside dans le fait que l'AM est limitée dans l'interprétation de données dans leur forme brute et qu'un lourd travail d'extraction de variables (transformation de pixels en un vecteur de variable) est nécessaire afin d'accomplir une tâche de classification ou reconnaissance de motifs (56). La différence majeure entre l'AM et l'AP réside donc dans le fait que l'AP est capable d'extraire des variables hautement complexes sans nécessiter l'expertise humaine, ainsi que d'établir des relations entre elles. Cet avantage permet donc à l'AP de trouver des fonctions complexes afin de prédire/classer une variable dépendante à partir de variables indépendantes.

### **1.2.2 Fondement de l'apprentissage profond**

Tout comme beaucoup d'inventions scientifiques, l'apprentissage profond prend naissance grâce au biomimétisme sous forme de réseau de neurones artificiels (RNA). C'est en 1943 que ce concept apparaît avec Warren S.McCulloch et Walter Pitts qui modélisent une version computationnelle simplifiée du fonctionnement biologique de neurones dans l'accomplissement de computations complexes en utilisant une logique propositionnelle (57) (figure 5) (58).

Similairement aux neurones biologiques, les réseaux de neurones artificiels fonctionnent grâce à un système de communication entre plusieurs couches de neurones.



**Figure 5. - Réseau de neurones artificiels performant de simples computations logiques**  
Adaptée de (58).

En 1957, le Perceptron, une des architectures les plus simples mais fondatrices a été inventé par Frank Rosenblatt (58). Cette nouvelle architecture composée d'unités (neurones) nommées *Threshold Logic Unit* ou l'Unité logique de seuil (ULS) fonctionne sur un même principe que le concept de McCulloch et Pitts, à la différence que les entrées et sorties du réseau ne sont plus des valeurs binaires, mais des nombres. Les inputs ( $\mathbf{x}$ ) sont maintenant associés à des poids ( $\mathbf{w}$ ) qui vont par la suite être sommés ( $z = \mathbf{x}^T \mathbf{w}$ ) pour donner une première valeur numérique à l'ULS.  $z$  va ensuite être passé à travers une fonction de seuil (*Step*) qui va permettre de donner la sortie suivante:  $h_{\mathbf{w}}(x) = \text{Step}(z)$  (Figure 6). La fonction principale utilisée pour le Perceptron est la *Heaviside Step Function* (58) qui stipule que si la somme pondérée entrée ( $z$ ) est inférieure à 0, alors la sortie du perceptron sera 0, sinon elle sera 1. Cette architecture fait penser à un modèle de classification binaire comme la régression logistique, ce qui est tout à fait applicable lorsqu'en présence d'une seule ULS, ou plusieurs (figure 7). Lorsque plusieurs ULS sont mis en commun, on parle de *Fully Connected Layer* ou couche dense.

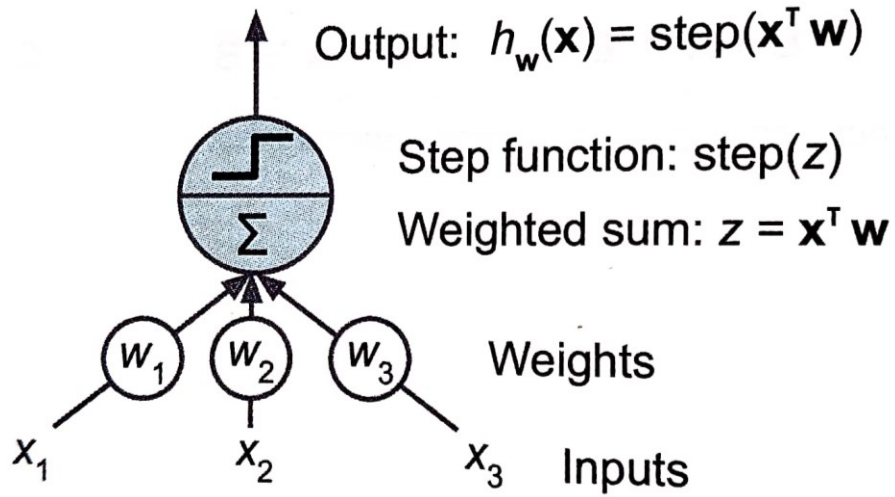


Figure 6. - Illustration d'un fonctionnement d'une unité logique de seuil Adaptée de (58).

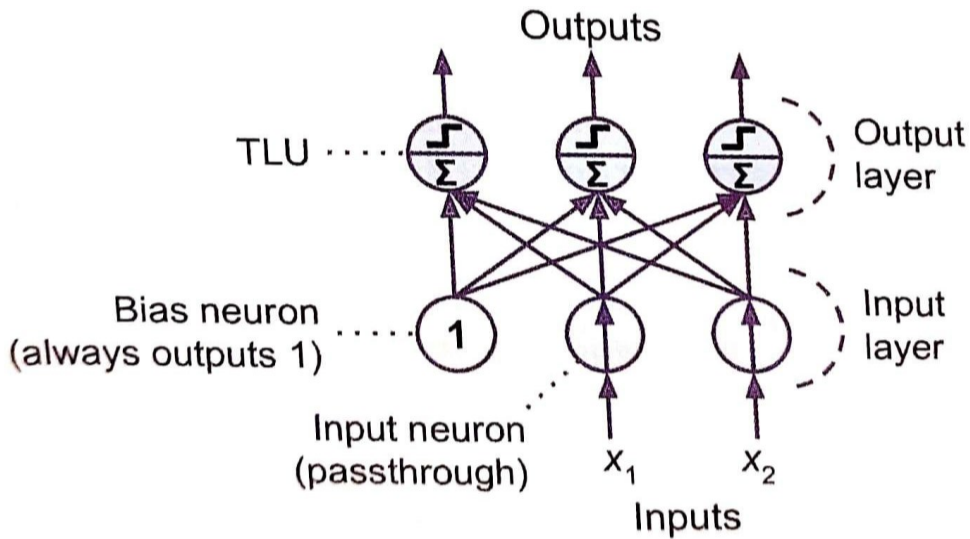
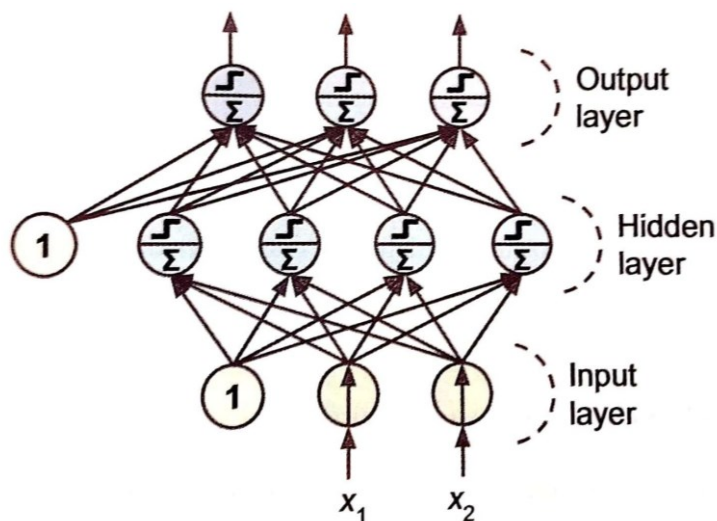


Figure 7. - Architecture d'un perceptron. Adaptée de (58).

Il est par la suite possible d'effectuer des tâches dites « d'entraînement ». Cela est majoritairement basé sur la loi de Hebb, qui fonctionne sous un principe de renforcement lorsque deux neurones dans deux couches différentes sont activés, leur connexion devrait être renforcée et vice-versa (59) (58). Lorsque le Perceptron donne une prédiction en sortie, l'algorithme permet de renforcer les connexions à travers les poids ( $w$ ) qui diminuent l'erreur de prédiction. A chaque itération d'entraînement, la prédiction est comparée à la vraie valeur et les poids sont ajustés afin de renforcer les connexions qui auraient donné une bonne prédiction. Certaines limitations du perceptron à une couche, comme le fait qu'il est fonctionnel uniquement sur des données séparables linéairement, ont donné naissance au fameux problème XOR, l'opérateur booléen qui vérifie deux conditions mutuellement exclusives (0,1) ou (1,0) (60). Ce problème a donné naissance au *Multilayer Perceptron* ou Perceptron Multicouche, qui ajoute une ou plusieurs couches de neurones artificiels, permettant de sortir de la limitation de données séparables linéairement et donne la capacité au Perceptron Multicouche de résoudre des problèmes non linéaires. La figure 8 illustre l'architecture d'un perceptron multicouche, contenant tout comme le perceptron à une couche, les neurones d'entrées et de sorties, mais aussi une couche cachée.



**Figure 8. - Architecture d'un perceptron multicouche.** Adaptée de (58).

Un perceptron multicouche peut avoir plusieurs couches cachées, et la définition de “profond” dans l’AP reste vague. En 1990, un réseau possédant deux couches cachées était considéré comme profond, mais de nos jours la majorité des réseaux dépasse la douzaine de couches, voire des centaines (58). Des difficultés dans l'entraînement de ces modèles ont donné naissance en 1985 à un algorithme d’apprentissage qui a révolutionné leurs performances, la *Backpropagation* ou propagation inverse (61). Intuitivement, l'algorithme de propagation inverse fonctionne de manière similaire à la descente de gradient, cette dernière permettant de diminuer l’erreur de prédiction en modifiant les paramètres du modèles incrémentalement. Une fois que la prédiction est faite, l’erreur est calculée et la proportion de l’erreur pour chaque connexion est mesurée en remontant dans les connexions antérieures du réseau. Une fois que les proportions de l'erreur de chaque connexion sont mesurées, l'algorithme procède à une descente de gradient afin d’optimiser les paramètres en question. L’algorithme de propagation inverse a introduit une série de problèmes d'optimisation comme le *Vanishing/Exploding gradient* qui se traduit par une disparition (rapprochement de 0) ou explosion (rapprochement de l’infini) du poids d’une certaine ligne de connexion, causant la difficulté de l’algorithme à converger vers la solution optimale. Ceci a causé le désintérêt de la communauté envers l’AP jusqu’en 2010 ou une solution élégante a été proposée en remplaçant l’initialisation aléatoires des poids par une initialisation normalisée, ce qui permettait une convergence du modèle plus rapide (62). Suite à cette solution, une pléthore d'améliorations concernant l’initialisation des paramètres, la normalisation ainsi que du type de fonction d’activation utilisée ont été proposées (63, 64).

### **1.2.3 Limitation des réseaux de neurones artificiels**

Maintenant que les concepts principaux de l’AP ont été introduits (i.e. Capacité d’analyser des données hautement complexes, architecture fondamentale, apprentissage par propagation antisens), il est important d’établir certaines limitations qui ont inspiré l’invention de l’architecture au centre de ce travail, les réseaux de neurones convolutifs (RNC). Concernant la reconnaissance d’image, la complexité paramétrique d’un réseau de neurones artificiel peut vite

devenir difficile à gérer. Généralement, une image est relativement grande en termes de nombre de pixels totaux (variables à l'entrée), composée souvent de centaines voire centaines de milliers (i.e. 512x512 pour un CT scan). Une architecture contenant par exemple une centaine de neurones dans la première couche cachée et recevant une coupe d'un CT scan aurait déjà plus de 26 millions de paramètres (512x512x100).

Un risque non négligeable lors d'un apprentissage est le surapprentissage. Ce concept est lorsqu'un modèle a une quantité excessive de paramètres par rapport au problème en question ou lorsqu'il n'y a pas assez de données. Le modèle va apprendre en ajustant ses paramètres pour accommoder les données d'entraînement, mais ne sera pas capable de généraliser ses performances de prédiction sur d'autres jeux de données. Une solution à ce problème est donc d'augmenter considérablement la taille du jeu de données (tout en faisant attention à la qualité des données pour éviter des biais) ou de choisir un modèle plus simple et ainsi diminuer la complexité paramétrique. Un autre problème est le fait que plus un modèle est complexe, plus l'espace mémoire et le temps d'apprentissage sont élevés, ce qui peut être un facteur limitant et consommer beaucoup de ressources (65). Une autre limitation importante à considérer est le fait que les réseaux de neurones artificiels ont une capacité très limitée en termes de variance translationnelle et distorsion locale de variables d'entrée (65) (i.e. Le chiffre 1 écrit tout à gauche d'une matrice 28x28 devait toujours être reconnu comme un 1 si écrit tout à droite de cette même matrice). Finalement, une image est composée de pixels qui ne sont pas indépendants (i.e. le pixel délimitant la rayure noire de la rayure blanche d'un zèbre contient une information spatiale intrinsèque, ces deux pixels noir et blanc voisins font partie d'une information commune et sont donc dépendants). C'est aussi une des difficultés des réseaux de neurones artificiels qui n'ont pas spécifiquement la capacité de conserver l'information topologique des variables d'entrées (65).



## 1.2.4 Introduction aux réseaux de neurones convolutifs

Ces limitations ont donc mené à l'introduction au concept de réseaux de neurones convolutifs en 1998 avec la première architecture, LeNet-5 (Figure 9) (65). Un RNC est similaire à un RNA profond classique conceptuellement, soit l'extraction de variables et la formulation d'une fonction explicative de haute complexité afin de faire une prédiction. Les RNC proposent certaines améliorations par rapport au RNA notamment certains concepts fondamentaux comme les *local Receptive Fields* ou champs récepteurs locaux, la convolution, le partage de poids ainsi que le *Subsampling* ou sous-échantillonnage. Ces concepts permettent d'adresser en partie les limitations soulevées antérieurement (56, 58, 65).

Premièrement, il est important de discuter du principe fondamental des RNC, la convolution. La convolution est une opération mathématique qui fait glisser une fonction sur une image afin d'extraire des variables (58). Au sein des RNC, on parle souvent de couches de convolutions. Ce sont donc des fonctions "filtres" qui passent au travers de l'entrée (i.e. image initiale) et qui produisent des *Features Maps* ou cartes de variables en sortie. Ces cartes de variables permettent l'extraction de variables de plus en plus complexes, grâce à l'addition de couches de convolutions additionnelles.

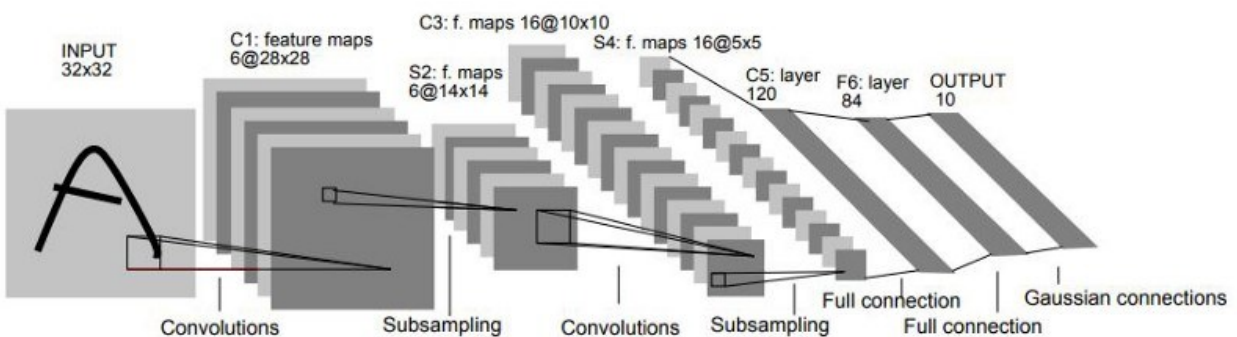


Figure 9. - Architecture du réseau de neurone convolutif LeNet-5. Provenant de (65).

Habituellement les couches cachées superficielles (i.e. A gauche, plus proches de l'entrée du réseau) vont permettre l'extraction de variables de bas niveau comme les lignes horizontales et verticales, puis plus les couches cachées gagnent en profondeur (i.e. A droite, plus éloignée de l'entrée du réseau), plus le type de variables extraites va gagner en complexité (i.e. yeux, visages, arbres etc...). Chaque couche de convolution contient habituellement plusieurs filtres de convolutions, donnant naissance à plusieurs cartes de variables par couche de convolution. Cette structure hiérarchique de couches de convolutions capables d'extraire des variables de plus en plus complexes dans leurs champs récepteurs locaux respectif est l'attribut principal des RNC. Un autre atout important des RNC est le partage de poids. Lorsqu'une carte de variable est créée à partir d'un filtre de convolution, chaque pixel (neurone) dans cette carte possède le même poids. Grâce à ce partage de poids entre tous les neurones d'une même carte de variable, cela réduit drastiquement la complexité paramétrique du modèle et donne aussi la capacité au modèle de reconnaître une variable/motif n'importe où dans l'entrée (58), d'où la haute capacité des RNC en termes d'invariance translationnelle.

Comme discuté précédemment, les RNA souffrent d'une haute complexité paramétrique, qui peut occasionner des demandes en termes de ressource mémoire considérable. Une des solutions à ce problème introduit par les RNC est la diminution drastique de la complexité paramétrique du modèle par le sous-échantillonnage. Conceptuellement, le sous-échantillonnage fonctionne similairement à la convolution. Outre la diminution de la complexité et l'utilisation en mémoire, le sous-échantillonnage permet aussi un certain degré d'invariance translationnelle (58).

Une fois que plusieurs couches de convolution et de sous-échantillonnage se terminent, la structure d'un RNC est de nouveau comparable au RNA. La dernière couche du RNC, qui contient les variables de haut niveau extraites et condensées vont être aplaties en un vecteur 1D afin de créer les couches denses discutées précédemment. La dernière couche contenant le vecteur en question représente donc les variables de haut niveau extraites grâce aux étapes précédemment décrites et permet d'accomplir la tâche finale de prédiction. Plusieurs architectures célèbres ont suivi LeNet-5 en améliorant significativement les performances ou innovant certains concepts. En

2012, AlexNet a gagné la compétition *ImageNet ILSVRC*. Doté d'une architecture similaire à LeNet-5, AlexNet se différencie par son degré de profondeur supérieur, diverses techniques de régularisation ainsi qu'une fonction d'activation neuronale utilisant la *Rectified Linear Unit* ou Unité Linéaire Rectifiée (Figure 10).

Layer	Type	Maps	Size	Kernel size	Stride	Activation
Out	Fully connected	–	10	–	–	RBF
F6	Fully connected	–	84	–	–	tanh
C5	Convolution	120	1 × 1	5 × 5	1	tanh
S4	Avg pooling	16	5 × 5	2 × 2	2	tanh
C3	Convolution	16	10 × 10	5 × 5	1	tanh
S2	Avg pooling	6	14 × 14	2 × 2	2	tanh
C1	Convolution	6	28 × 28	5 × 5	1	tanh
In	Input	1	32 × 32	–	–	–

Layer	Type	Maps	Size	Kernel size	Stride	Padding	Activation
Out	Fully connected	–	1,000	–	–	–	Softmax
F10	Fully connected	–	4,096	–	–	–	ReLU
F9	Fully connected	–	4,096	–	–	–	ReLU
S8	Max pooling	256	6 × 6	3 × 3	2	valid	–
C7	Convolution	256	13 × 13	3 × 3	1	same	ReLU
C6	Convolution	384	13 × 13	3 × 3	1	same	ReLU
C5	Convolution	384	13 × 13	3 × 3	1	same	ReLU
S4	Max pooling	256	13 × 13	3 × 3	2	valid	–
C3	Convolution	256	27 × 27	5 × 5	1	same	ReLU
S2	Max pooling	96	27 × 27	3 × 3	2	valid	–
C1	Convolution	96	55 × 55	11 × 11	4	valid	ReLU
In	Input	3 (RGB)	227 × 227	–	–	–	–

**Figure 10. - Comparaison entre l'architecture du RNC LeNet-5 (Gauche) et AlexNet (Droite)**

Provenant de (58).

Les techniques de régularisation et hyperparamètres seront discutées plus en détail dans la prochaine section mais pour résumer, AlexNet utilise une technique d'abandon aléatoire (*dropout*) où 50% des neurones dans les couches F9 et F10 (Figure 10 à droite) sont inactifs durant l'entraînement. Afin d'augmenter la quantité de données en entrée et potentiellement réduire le surapprentissage, une technique d'augmentation des données a été appliquée avant d'entraîner AlexNet, générant des variantes des images sources à l'aide de rotations, inversions et changements de contraste (66).

En 2014, Zisserman et Simonyan de l'université d'Oxford souhaitent étudier l'impact de la profondeur d'un RNC traditionnel (similaire à LeNet-5) sur sa performance. Ce modèle, *VGGNet*, n'impliquait aucun module d'inception ou de nouveauté conceptuelle et utilisait une série de 2 à 3 convolutions suivi d'un sous-échantillonnage, allant jusqu'à 19 couches de convolution (pour VGG19), soit 144 millions de paramètres et 512 cartes de variables. L'équipe a démontré l'importance de la profondeur d'une RNC et la capacité de ces réseaux "simples mais profonds" à

obtenir des performances supérieures à *GoogLeNet* (67). L'architecture VGG19 est au centre de ce travail et plus de détails seront fournis sur la nature du modèle ainsi que son utilisation dans ce projet à la section 2.7.

À la suite de l'implémentation de VGG, beaucoup de travaux ont fait surface en introduisant de nouvelles architectures et concepts (58, 68-72). La figure 11 récapitule la performance de plusieurs modèles incluant VGG16 et VGG19 en fonction d'une métrique incluant le nombre de paramètres, d'opérations et la taille (73).

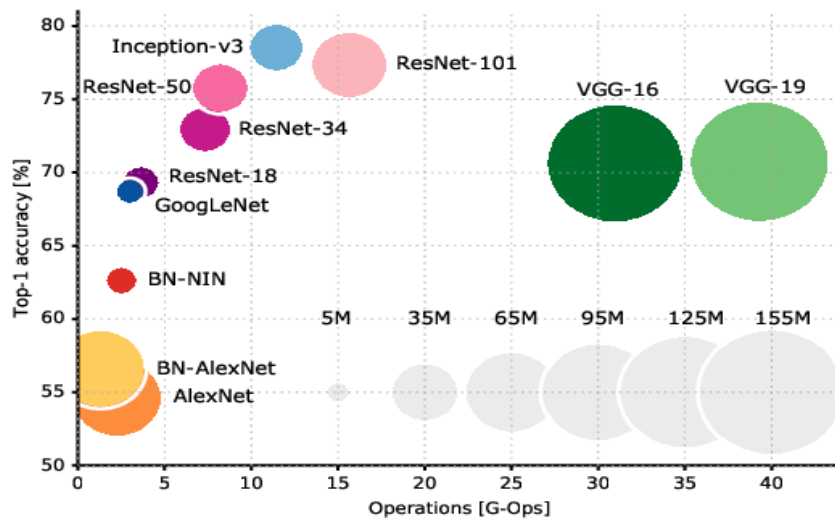


Figure 11. - Performance des meilleurs modèles en fonction de leur complexité.

Adapté de (73).

### 1.2.5 Entraînement des réseaux de neurones convolutifs

Les données de ce travail proviennent d'images dans un contexte clinique de soins critiques, et cela introduit une variabilité dans la qualité et la reproductibilité des images. Il est important de garder cela en considération lors de l'évaluation des résultats du modèle. Bien que l'architecture soit le pilier fondamental d'un RNC, la manière d'apprendre et sa source de données sont toutes

aussi importantes. L'élément crucial à tout modèle d'apprentissage est la qualité des données qui lui sont données en entrée. La qualité des données peut être évaluée de différentes manières en posant les bonnes questions, comme la liste non exhaustive et non hiérarchique suivante:

a) D'où proviennent les données ? b) Existe-t-il un biais important dans les données ? c) Les différentes sources dont proviennent les données sont-elles comparables entre elles ? d) Ont-elles été prélevées de façon rigoureuse ? e) Les données doivent-elles être normalisées ou standardisées ? Y a-t-il des valeurs manquantes, aberrantes? f) Sont-elles obsolètes ou toujours d'actualité? (74)

Dépendamment du domaine d'application, d'autres questions peuvent être posées afin de déterminer la nature et la qualité des données. Le modèle va apprendre à partir des informations fournies, il est donc crucial que celles-ci lui permettent de généraliser ses prédictions sur la population cible à partir de l'échantillon.

Outre la qualité des données, la manière dont un modèle apprend est aussi extrêmement importante. On parle d'hyperparamètres lorsque l'on décrit un élément ou configuration externe au modèle qui n'est pas estimé à partir des données; leur but est d'aider le modèle à estimer les paramètres de la fonction (75). Concernant les RNC, une panoplie d'hyperparamètres sont d'importance, certains plus que d'autres dépendamment du contexte, mais il est toujours important de considérer une approche concernant leur réglage car ils permettent bien souvent d'adresser certains problèmes ou limitations lors de l'entraînement.

Généralement en apprentissage supervisé, le but d'un modèle est de trouver les paramètres qui diminuent au maximum sa fonction de perte. Une manière d'atteindre cet objectif est la descente de gradient. Cette technique est capable de trouver une solution quasi-optimale en modifiant les paramètres itérativement et analysant l'impact de cette modification sur la variation de la fonction de perte. La descente de gradient introduit aussi quelques hyperparamètres comme le choix du type de descente de gradient ou encore le *Learning Rate* ou taux d'apprentissage. Les techniques principales de descente de gradient sont le *Batch Gradient Descent* ou descente de gradient complète, le *Stochastic Gradient Descent* ou descente de gradient stochastique et le *Mini-batch Gradient Descent* ou descente de gradient par mini-lots (58). Dans le contexte de RNC,

la descente de gradient souffre d'une lenteur considérable et cette limitation a donné naissance à de multiples optimisations de la descente de gradient comme l'optimisation d'inertie (76), le gradient d'accélération (77), l'algorithme d'AdaGrad (78) ainsi que l'optimisation d'Adam (79). Conceptuellement, cette optimisation de la descente de gradient fonctionne en prenant en compte la direction et l'inertie de gradients précédents, permettant une convergence plus rapide vers l'optimum minimal de la fonction de perte.

Le taux d'apprentissage est un hyperparamètre additionnel crucial lors de l'entraînement d'un modèle, et plus particulièrement un modèle de haute complexité comme un RNC. Il représente l'amplitude d'ajustement des paramètres du modèle à chaque itération d'entraînement permettant d'ajuster les paramètres de façon contrôlée et il existe une multitude d'options concernant la stratégie d'apprentissage (80). Le taux d'apprentissage est un hyperparamètre intégré dans les techniques de descentes de gradient décrites plus haut.

Le *Batch size* ou taille de lot représente la taille du lot de données fournies au modèle à chaque itération d'entraînement. C'est non seulement un hyperparamètre important pour l'optimisation du temps d'entraînement et l'impact sur les ressources computationnelles comme la mémoire vidéo vive (VRAM) mais aussi pour la performance du modèle. Il existe différentes écoles de pensée concernant la meilleure approche, certains décrivant un maximum de 32 instances par lot (81), d'autres décrivant des tailles de lots vastement supérieures allant jusqu'à 8192 instances par lot (82, 83).

Le nombre d'itérations (ou *Epoch*) est un hyperparamètre à considérer et représente le nombre de fois où le modèle va passer à travers le jeu d'entraînement. Afin d'éviter le surapprentissage, un arrêt prématuré peut-être utilisé lorsque le taux d'erreur sur le jeu de validation commence à augmenter ou le modèle peut aussi être configuré pour être sauvegardé uniquement lorsque l'erreur sur le jeu de validation diminue (58).

Les hyperparamètres de régularisation comme le *dropout* permettent de diminuer le risque de surapprentissage du modèle en assignant une probabilité  $p$  pour chaque neurone d'être neutralisé à chaque itération d'entraînement, forçant le modèle à gagner en flexibilité (84).

Deux hyperparamètres plus reliés à l'architecture du modèle sont le choix du nombre de couches ainsi que le nombre de neurones par couche du réseau. Il n'existe pas de règles absolues, mais en général, un modèle profond (plus de couches cachées) est préféré à un modèle large (plus de neurones par couche) (58). Concernant plus spécifiquement les RNC, ils possèdent les hyperparamètres des RNA profonds, mais en possèdent quelques-uns qui augmentent leur degré d'optimisation, par exemple la dimension et nature du filtre de convolution ou encore le *pas*, qui représente le nombre de colonnes sautées sur la matrice à chaque itération du filtre de convolution.

Le choix du nombre de filtres appliqués à chaque couche de convolution fait aussi partie des éléments modifiables. Il n'existe pas de règles en or, mais il est généralement accepté que le nombre de cartes de variables augmentent avec la profondeur du réseau. Le rationnel derrière cela est que de plus en plus de variables de haut niveau (i.e. Visages) sont extraites à partir de variables de bas niveau (i.e. Cercles, lignes, angles) (58).

Le type de modèle ou d'algorithme choisi ainsi que l'optimisation des hyperparamètres sont des aspects cruciaux à prendre en compte afin de préserver un équilibre entre la variance et le biais. Le choix d'un modèle complexe (qui contient une haute dimension paramétrique) et une optimisation trop élevée des hyperparamètres peut aisément mener à un problème de surapprentissage qui va limiter la capacité de généralisation du modèle (modèle à variance élevée). En opposition au problème de surapprentissage, le choix d'un modèle trop simple pour un problème donné peut causer un problème de biais, qui représente la limitation ou l'incapacité d'un modèle à inférer avec précision compte tenu de la faible complexité paramétrique du modèle. Récemment, des modèles sur-paramétrés surpassent les limitations du surapprentissage et permettent d'obtenir des performances élevées, stimulant ainsi les discussions sur l'équilibre entre le biais et la variance (85, 86).

Lors de l'entraînement d'un modèle, la validation croisée (*cross validation*) est souvent sollicitée. Cette technique est basée sur la division d'un jeu de données en deux parties, un jeu de données d'entraînement et un jeu de données de validation. C'est une technique à un seul paramètre :  $k$ , qui représente le nombre de fois ou le jeu de données va être divisé. A titre

d'exemple, un choix de  $k=5$  va diviser un jeu de données de 1000 entrées 5 fois en 800 entraînements et 200 validations (5 jeux au total). Chaque jeu est ensuite entraîné et validé indépendamment et permet d'obtenir une meilleure évaluation de la capacité de généralisation du modèle. Cela est dû au fait que chacun des 5 jeux de 200 validations contient des entrées indépendantes des autres jeux et permet de valider le modèle sur de nouvelles données (87).

Les éléments listés ne sont pas une liste exhaustive de tous les hyperparamètres applicables à l'AP ou les RNA et RNC, mais sont en général les premiers éléments qui passent sous la loupe lors de l'optimisation d'un modèle. Il est important de garder en tête qu'un modèle n'a pas forcément besoin d'être optimisé de façon exhaustive, et parfois une approche simple et réfléchie en termes d'optimisation peut faire gagner en temps sans sacrifier précision. Finalement, le plus important est de relativiser le modèle et son optimisation en fonction du problème en question, ainsi qu'à la qualité et la nature des données disponibles à l'entraînement du modèle. Les hyperparamètres utilisés pour ce travail sont listés en 2.6.2.

### **1.3 Applications de l'apprentissage profond en médecine**

Le monde de la médecine est constamment en évolution et l'innovation technologique est un vecteur de grande importance concernant sa trajectoire. La priorité absolue reste le bien-être et le soin des patients, et il est surtout question d'identifier quels outils technologiques peuvent être utilisés à ces deux fins fondamentales.(88) Dans les années 90, le terme *Big Data* fait surface afin de désigner des données excessivement larges et complexes, difficilement stockables ou traitables avec les systèmes en place (89, 90). Toutes ces données à disposition promettent de répondre à énormément de questions et problèmes pour l'humanité, mais il est nécessaire en premier temps de stocker, trier, nettoyer, mettre à jour et analyser ces données gigantesques. Ce défi a donc propulsé les recherches informatiques de l'avant afin de combler ces besoins et le domaine de la santé a subi une révolution en termes d'accessibilité, stockage et traitement des données. L'AM au sein du système de santé explose en termes d'accessibilité et de performance



permettant dépendamment de son utilisation de réduire les coûts du système, augmenter la qualité du rapport médecin-patient et créer une médecine personnalisée pour combler les besoins de chaque individu.

Une limitation indéniable pour l'instant est la forme et qualité des données, qui peuvent varier fortement dans le milieu de système de santé (91). Il n'est pas surprenant que l'AP soit un outil très attrayant dans le domaine médical, plus particulièrement en imagerie médicale (i.e. Traitement d'image, programme d'aide au diagnostic, interprétation d'image, fusion d'image, le recalage d'image, la segmentation d'image, etc...) (92). L'AP donne naissance à des outils qui permettent aux professionnels de la santé concernés de diminuer leur temps de diagnostic et d'augmenter la précision de leur diagnostic, ou encore permettent d'améliorer des stratégies prophylactiques. Malgré l'existence de plusieurs techniques d'AP afin de répondre aux besoins de la *computer vision* ou vision artificielle, les réseaux de neurones convolutifs reviennent souvent au centre de l'attention (92).

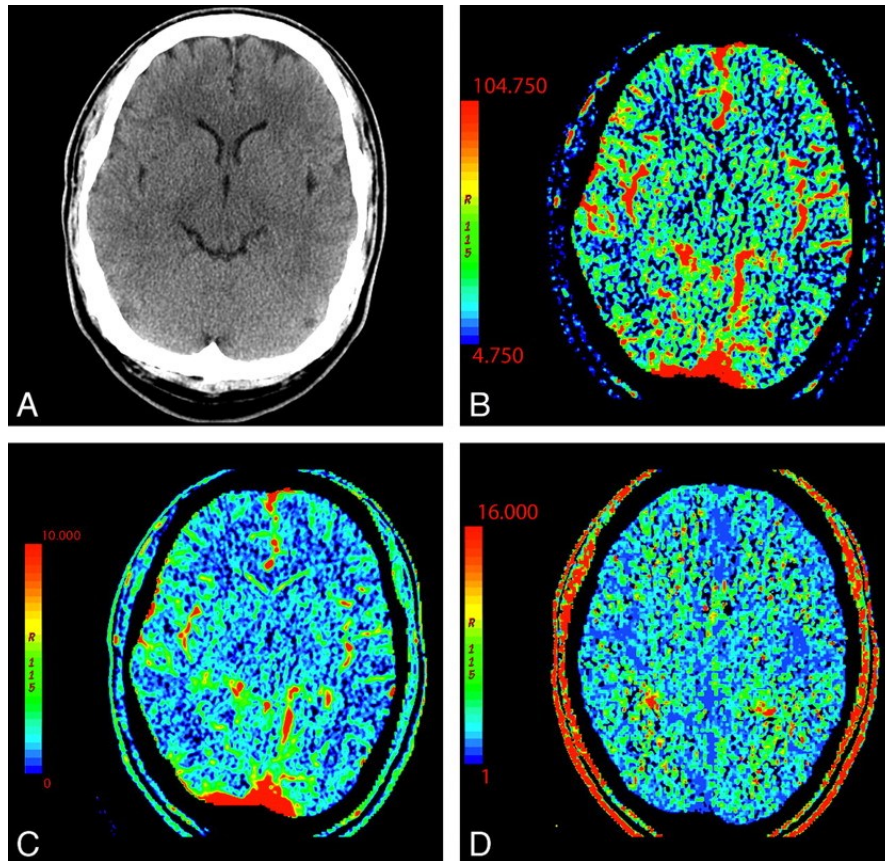
Bien qu'il existe toujours un débat entre l'approche radiomique et l'approche par AP, la différence majeure étant la manière dont les variables sont extraites (48). Les réseaux de neurones convolutifs restent sur le podium en termes de leur utilisation afin de répondre à des classifications d'images médicales, par exemple la majorité des méthodes de détection de COVID-19 à partir de scanner de tomodensitométrie du thorax sont de RNC (93). Kundu R *et al.* 2022 utilisent deux stratégies d'apprentissages non discuté précédemment dans ce travail. La première est en fait la concaténation de 3 RNC au sein de la même architecture (Inception-v3, ResNet35 et DenseNet201) qui reçoivent chacun l'entrée (les CT-scans) et produisent chacun une prédiction qui sera ensuite pondérée. Cette technique est connue sous le nom d'*Ensemble Learning* ou apprentissage par ensemble (93). La seconde stratégie en question est nommée *Transfert Learning* ou apprentissage par transfert, qui est une stratégie d'initialisation des paramètres des couches supérieures du réseau, accélérant la vitesse d'apprentissage du modèle (94).

## 1.4 Apprentissage profond et imagerie médicale

Concernant l'imagerie cérébrale, plusieurs publications intéressantes ont démontré la capacité des RNC et des réseaux de neurones récurrents (une variante des RNC qui inclut une composante temporelle) à effectuer un grand nombre de tâches de haute complexité, comme la segmentation de métastases cérébrales (95), l'identification de pathologies cérébrales comme la maladie d'Alzheimer (96) ainsi que l'identification d'hémorragies intracrâniennes (97-99).

Kumaravel et al. (2021) ont démontré qu'en combinant l'utilisation d'une architecture de RNC relativement simple pré-entraîné pour l'extraction de variables (AlexNet) avec un classificateur simple (*Support Vector Machine* ou machine de support vectoriel) sur des scans CT sans contraste, l'équipe était capable d'obtenir des résultats exemplaires pour de détecter la présence et la localisation du sang dans l'espace intracrânien (99).

Le CTP n'est pas une image en 2 dimensions, mais de multiples volumes à travers le temps (3 dimensions spatiales et 1 dimension temporelle) et ces volumes sont habituellement utilisés afin de générer des cartes de perfusion qui condensent l'information en plusieurs images 2D représentant la localisation et le mouvement de l'agent de contraste à travers le temps (100) (Figure 12 du mémoire et Annexe SIV de l'article). Certains travaux utilisent cette troisième dimension spatiale afin de détecter des dysfonctionnements au sein de capillaires cérébraux en utilisant un RNC 3D (101) ou encore la classification des différents types d'hémorragies cérébrales (102). Plusieurs approches différentes existent afin de créer un modèle d'AP capable d'interpréter une image. La nature de la question de recherche peut pousser vers des modèles plus ou moins complexes, mais il est crucial de garder en tête que la tâche la plus importante avant de se lancer dans l'apprentissage d'un modèle est la qualité des données et l'étape de prétraitement des données (103). Une limitation considérable dans le domaine est la qualité des étiquettes, plus particulièrement lorsque les données proviennent de multiples sources et qu'il n'est pas toujours simple de vérifier la qualité et la méthodologie utilisée lors de l'identification des étiquettes (104).



**Figure 12. – Exemples de cartes de perfusions.** Scan CT sans contraste (a) et cartes de perfusion générée par CTP, Flux Sanguin Cérébral (B), Volume Sanguin Cérébral (C), Temps de Transit Moyen (D). Adaptée de (100)

## 1.5 Mise en contexte de la question de recherche

Le diagnostic de décès neurologique est une étape cruciale et complexe lors de l'évaluation d'un patient suspecté d'être décédé neurologiquement. En somme, l'examen clinique a pour but d'évaluer certaines fonctions critiques du patient comme les fonctions respiratoires ou les réflexes visuels.

Malgré une standardisation du diagnostic clinique de mort cérébrale de plus en plus acceptée, il demeure chronophage et complexe à établir. Il est dans le bénéfice du patient que ce diagnostic soit établi avec la plus haute certitude possible et c'est pour cette raison qu'un minimum de deux cliniciens doivent être en accord sur le statut du patient. Plusieurs tests auxiliaires peuvent être utilisés en clinique comme outils de soutien décisionnel au diagnostic de décès neurologique, comme le scanner de tomodensitométrie à perfusion, qui permet de visualiser le mouvement du flot sanguin et de la perfusion au sein de structures clef du cerveau, comme le tronc cérébral. Une des limitations de l'utilisation du CTP comme test auxiliaire est qu'il requiert l'expertise spécifique d'un radiologiste formé pour ce test pouvant augmenter le temps nécessaire au diagnostic de décès neurologique (45). La raison pour laquelle le temps est un facteur crucial dans le diagnostic du décès neurologique est le don d'organes. La règle d'or du don d'organes, la *Dead Donor Rule*, stipule qu'un individu doit être déclaré mort avant le don. Le don ne peut donc en aucun cas causer le décès du patient (10, 25, 45). Cette règle primordiale implique qu'une rigueur extrême doit être portée au diagnostic de DN et cela peut donc engendrer un temps considérable avant de diagnostiquer un individu comme mort neurologiquement. Une problématique découlant de cela est le fait que certains patients ont besoin d'une transplantation d'urgence, et quelques jours ou même heures peuvent faire toute la différence entre la vie et la mort.

L'utilisation du traitement automatique de l'image est de plus en plus utilisée en médecine et en radiologie. Dans ce travail, nous explorons la capacité d'un modèle d'apprentissage profond à classifier des images de CTP de patients aux soins intensifs atteints de blessure cérébrale sévère en coma profond. Un avantage à long terme de l'automatisation du traitement de scan CTP dans ce contexte serait de pouvoir offrir un outil de dépistage afin de pouvoir identifier des patients à risque de DN, ou encore d'offrir un soutien décisionnel aux radiologistes lors de l'interprétation des images.

## Chapitre 2 : Matériel et méthodes

### 2.1 Hypothèse

Le traitement automatique de scans CTP par le traitement automatique de l'image permet la classification du décès neurologique chez des patients aux soins intensifs atteints de blessure cérébrale sévère en coma profond.

### 2.2 Objectifs

#### 2.2.1 Objectif primaire

L'objectif primaire de ce travail est d'entraîner un modèle d'apprentissage profond afin de classer des patients à risque de décès neurologique en tant que *Vivant* ou *Décès neurologique*.

#### 2.2.2 Objectif secondaire

Évaluer et comparer la capacité prédictive du modèle en fonction du type de blessure cérébrale.

## 2.3 Source des données

### 2.3.1 Introduction à l'étude INdEx-CTP

INdEx-CTP est une étude de cohorte diagnostic prospective multicentrique ayant pour objectif principal de déterminer la précision diagnostic de détermination de décès neurologique du CTP comparativement au diagnostic clinique complet. Ultimement, le but de cette étude est donc de déterminer si le CTP peut être utilisé comme test auxiliaire de référence lorsque l'évaluation clinique complète ne peut être accomplie. L'étude comprend 333 patients  $\geq$  18 ans souffrant de blessure cérébrale et admis aux soins intensifs. Les patients doivent présenter un score de 3 sur l'échelle de coma de Glasgow malgré une? sédation doit être interrompue depuis au moins 6 heures.

#### 2.3.1.1 Critères d'exclusions pour l'étude principale

Les critères d'exclusion de l'étude principale sont les suivants :

Les patients présentant les contre-indications à CT-perfusion qui suivent seront exclus de l'étude : a) Patientes enceintes, b) Allergie à l'agent de contraste, c) Le clinicien refuse l'inclusion à cause d'une blessure rénale

Les patients ou l'évaluation neurologique clinique complète ne peut être faite dû à l'un des facteurs de confusion suivants seront exclus de l'étude : a) Fracture cervicale au-dessus de la C6, b) Blessure faciale significative limitant l'examen du nerf crânien, c) Hypothermie inférieure à 34 degrés Celsius, d) Utilisation de barbituriques intraveineux depuis l'admission, e) Choc non ressuscitant, f) Dysfonctionnement des nerfs périphériques, des muscles ou blocage neuromusculaire pouvant expliquer l'absence de réponse, g) Lésion cérébrale anoxique inférieure à 24h (ou 72h si présence d'hypothermie thérapeutique), h) L'examen du test d'apnée est contre-indiqué par le clinicien responsable, i) Tout autre anomalie identifiée comme facteur de confusion potentielle par le clinicien responsable, j) Absence totale de réflexes du tronc cérébral, k) Test d'apnée positif. (Traduit de l'étude INdEx-CTP, ClinicalTrials.gov, NCT03098511).

### 2.3.1.2 Angiographie par Tomodensitométrie avec Perfusion et interprétation

Une fois inclus dans l'étude, les patients ont été soumis à un protocole de CTP afin de capturer la totalité de l'espace cérébral. Un volume de 40 ml de contraste a été injecté au patient et l'image a été capturée selon le protocole de l'étude (ClinicalTrials.gov, NCT03098511). Les images ont été de-identifiées, envoyées par les centres correspondant à un laboratoire de post-traitement des images utilisant le programme (Olea Sphere 3.0, Olea Medical, Cambridge, USA) afin de standardiser les techniques de traitement pour toutes les images, indépendamment des sites. Les scans de CTP ont été analysés et interprétés par deux radiologistes à l'aveugle entre eux et à l'aveugle par rapport aux informations cliniques/démographiques des patients. Lors de désaccord par rapport à l'interprétation du scan, un consensus a dû être atteint pour l'interprétation finale. Afin de définir le DN par CTP, une diminution conséquente qualitative du débit sanguin cérébral ainsi que du volume sanguin cérébral dans une section transversale entière du tronc cérébral sur deux coupes axiales consécutives de 5 mm devait être mesurée.

### 2.3.1.3 Détermination du décès neurologique clinique

Dans les deux heures suivant l'acquisition de l'angiographie par tomodensitométrie avec perfusion (CTP), deux spécialistes, tels que des neurologues, des neurochirurgiens ou des intensivistes, ont réalisé un examen clinique complet pour diagnostiquer le décès neurologique (DN). Les cliniciens ont mené cet examen en aveugle, sans connaissance préalable des résultats du CTP ou des évaluations cliniques des autres praticiens.

L'examen clinique est basé selon un standard canadien (105). Un patient était considéré comme décédé neurologiquement si tous les éléments suivants étaient présents:

- 1) Patient souffre d'une étiologie reconnue comme causant le décès neurologique.
- 2) Absence de facteurs de confusion pouvant simuler le décès neurologique.
- 3) Absence totale de réflexes du tronc cérébral.

#### 4) Test d'apnée positif.

Les deux cliniciens devaient ensuite déclarer les patients comme décédés neurologiquement ou non. Dans l'éventualité où les cliniciens étaient en désaccord concernant leur diagnostic, un consensus devait être trouvé. Si le consensus n'était pas obtenu entre les deux cliniciens, un troisième clinicien était sollicité afin de procéder à un examen clinique additionnel afin d'aider l'obtention d'un consensus. Dans le scénario rare où aucun consensus n'était obtenu après cela, un test auxiliaire additionnel autre que le CTP était utilisé afin d'obtenir l'accord des trois cliniciens.

## **2.4 Population à l'étude, critères d'inclusion et d'exclusion et sous-groupe d'intérêt**

Nous avons effectué une étude transversale nichée dans l'étude de cohorte INDex-CTP décrite en 2.3.1 La population cible était donc les patients  $\geq 18$  ans souffrant d'une blessure cérébrale et admis aux soins intensifs, ayant un score de 3 sur l'échelle de coma de Glasgow et ayant cessé la sédation depuis au moins 6 heures. Tous les patients de l'étude principale étaient éligibles à l'exception des cas suivants : déviation du protocole rendant le patient non éligible dans l'analyse principale de l'étude INDex-CTP; absence d'images ou de perfusion; retrait du consentement ou absence/anonymisation de champs DICOM (Digital Imaging and Communications in Médecine) nécessaires à l'entraînement des modèles.

### **2.4.1 Analyse de sous-groupes**

Nous avons analysé les patients en fonction du type de blessure cérébrale (i.e. blessure ischémique, hémorragique, anoxique, traumatique ou autre) puisqu'il était envisageable que le



modèle obtienne de meilleures performances lors de l'analyse d'images provenant de différentes blessures cérébrales.

## 2.5 Préparation des données

### 2.5.1 Description de la donnée

Chaque étude (série d'images pour un patient lors d'une séance) était associée à une étiquette binaire (0 pour DN- et 1 pour DN+). Cette étiquette de haute qualité était basée sur le standard de référence de diagnostic de DN, soit le diagnostic clinique de mort du tronc cérébral (détails en 2.3.1.3). Nous avons travaillé avec des images médicales de haute résolution, en 4 dimensions et qui incorporent une composante temporelle. Un scan de CTP permet de suivre un agent de contraste dans l'espace (cerveau) et dans le temps. La nature même de l'image (ou plutôt, volume) comprenait donc une dimension temporelle. Une séance de scan de CTP ne générerait pas un seul volume, mais une quantité prédéterminée de volumes, et chaque volume était à un temps  $t + t_{vol}$  du précédent, ou  $t_{vol}$  était environ le temps nécessaire à la génération d'un volume. Un volume était généré en concaténant  $z$  tranches axiales (2-Dimensions) d'images cérébrales l'une après l'autre. Une fois le volume terminé, cette étape est répétée  $n$  fois, générant ainsi  $n$  volumes de  $z$  tranches (3-Dimensions) acquis à travers le temps (4-Dimensions). Comme décrit dans la section 2.3.1.2, ces volumes ont été envoyés à un laboratoire de post-traitement afin de générer des cartes de perfusions comme la carte de *flot sanguin cérébral*, *volume sanguin cérébral*, *temps de transit moyen*, *temps de seuil maximal* ou encore *premier moment de transit* qui incorporent directement la dimension temporelle. Pour ce travail, les cartes de perfusions n'ont pas été utilisées et seuls les volumes bruts ont été requis.

Les volumes avaient des dimensions variables entre les patients. A titre d'exemple, un patient pouvait avoir les dimensions suivantes: 512 x 512 x 22 pour chacun de ses 30 volumes, comparativement à un patient Y qui avait 480x480x20 pour chacun de ses 25 volumes. Cela

impliquait donc la nécessité de transformer les données afin de fournir une entrée constante à un modèle et synchronisé de façon spatio-temporelle.

## 2.5.2 Pré-traitement de la donnée

### 2.5.2.1 Calibration des dimensions

Les dimensions spatiales et temporelles pouvaient varier entre les patients à l'étude. Ceci était expliqué par des différences intrinsèques aux patients (i.e. Dimensions crânio-cérébrales) ou encore par les protocoles de scan qui pouvaient varier sensiblement (i.e. nombres de tranches, profondeur du cerveau sur l'axe, temps d'acquisition). Afin de calibrer les dimensions à travers tous les patients de l'échantillon, il fallait en premier temps s'assurer que les dimensions spatiales et temporelles étaient synchronisées entre tous les scans. Il est important de préciser que toute étape d'ajustement de dimension causait inévitablement une distorsion de l'image/volume source, ce qui ne signifiait pas nécessairement que l'information importante était perdue, mais c'était un aspect à prendre en compte car un excès de distorsion pouvait en effet causer une perte d'information. Afin de synchroniser les dimensions spatio-temporelles des images, plusieurs paramètres étaient à prendre en compte.

1) *l'Espace\_Interpixel*. Ce paramètre représentait la distance entre le centre de deux pixels sur une plane 2D. 2) *l'Epaisseur\_de\_Coupe*, qui représentait la profondeur de chaque coupe (en mm) lors du balayage. 3) Le *Temps\_Acquisition* permettait de calculer le temps nécessaire à la génération d'un volume. 4,5) Les dimensions *X* et *Y* représentaient la résolution de l'image. 6) La dimension *Z* représentait la profondeur du volume.

Afin de réduire un maximum la distorsion appliquée aux images, les valeurs les plus fréquentes qui étaient retrouvées dans les attributs DICOMs étaient utilisées comme référence pour certains paramètres. À titre d'exemple, concernant *l'Espace\_Interpixel* pour l'aspect spatial, si les valeurs les plus fréquentes pour ce paramètre étaient 1 mm x 1mm et une étude avait des valeurs de 0.5 mm x 0.5 mm, en appliquant un *zoom out* de 50% aux *l'Espace\_Interpixel* de chaque coupe (par technique d'interpolation), nous avons transformé les valeurs de ces paramètres à 1 mm x 1 mm.

Pour l'*Epaisseur\_de\_Coupe*, la valeur la plus haute a été conservée (et non la valeur la plus fréquente). Dans un cas où la valeur la plus élevée d'*Epaisseur\_de\_Coupe* était de 5 mm et où une étude avait 2.5 mm, il fallait éliminer 1 coupe sur 2 afin de retomber sur 5 mm d'épaisseur puisque cette étude générerait 2 coupes de 2.5 mm pour balayer le même volume que les paramètres de références (5 mm). Dans cet exemple, après traitement, les volumes auraient donc des voxels de 1 mm x 1mm x 5mm.

Concernant le paramètre *Temps\_Acquisition*, il était possible de déduire le temps nécessaire à la génération de chaque volume en soustrayant le champ DICOM *Temps\_Acquisition* de la première coupe à la dernière coupe du même volume. Ceci était important puisque certains protocoles de CTP pouvaient générer plus de volumes que d'autres pour un même temps. Ceci pouvait être un problème puisque le modèle avait besoin de dimensions identiques. Si une étude produisait 30% de volumes en plus, il y aurait donc 30% d'informations en plus à traiter, et à des temps différents. Afin de remédier à ce problème, nous avons conservé le *Temps\_Acquisition* le plus bas et synchronisé les autres études par rapport à cette référence en procédant à une technique de *sous-échantillonnage*. Par exemple, si la valeur de référence pour *Temps\_Acquisition* était de 0.5 sec par volume et l'étude 2 avait un *Temps\_Acquisition* de 0.25 sec par volume et produisait donc 2 fois plus de volumes que la valeur de référence. Ainsi pour cette étude, 50% des volumes ont été éliminés. Une première problématique était qu'une proportion relativement élevée d'études n'avait pas de *Temps\_Acquisition* par coupe, mais par volume. Il n'était donc pas possible de calculer le temps exact nécessaire à la génération d'un volume sans inclure le temps de battement entre les deux volumes. Dans ces cas ci, nous avons imputé le *Temps\_Acquisition* par la moyenne échantillonnage. Une deuxième problématique est que le temps nécessaire à la génération d'un volume n'est pas identique à travers une même étude. Nous avons donc pris le temps moyen de génération d'un volume par étude.

Après cette étape de synchronisation, les dimensions ( $x, y, z$ ) des volumes étaient différentes. Concernant les dimensions  $x, y$  une technique d'interpolation a été utilisée afin d'ajuster pour la taille nécessaire d'entrée au modèle, soit (128,128). Concernant la dimension  $z$ , une technique de *sous-échantillonnage* a été utilisée afin d'accommoder la dimension  $z$  la plus petite.

Chaque image a été normalisée à l'aide de la technique de normalisation de Hounsfield et un seuil sur les unités de Hounsfield (H.U) a été appliqué. Chaque pixel en dessous d'une valeur minimale de H.U de 40 a été défini à 40 et chaque pixel au-dessus de la valeur de 600 a été défini à 600. Cela a été considéré comme augmentant la représentation du contraste dans le cerveau (106). Afin d'initialiser l'entrée du modèle, nous avons projeté l'image 2D sur un axe RVB (Rouge, Vert, Bleu) (3 dimensions) pour imiter la structure d'une image RVB. L'image grise a été projetée 3 fois sur la dimension RVB. Pour chaque tranche de chaque volume, nous les empilons les unes sur les autres pour générer une série temporelle 2D contenant des informations sur le mouvement de l'agent de contraste à travers une seule tranche, ou un même espace. Une fois cela fait pour chaque tranche, nous avons obtenu une entrée de forme (20,128,128,3,16) pour chaque patient, où 20 représente le nombre de volumes, (128,128) la résolution de chaque image 2D, 3 l'image en niveaux de gris projetée sur l'axe RVB, et 16 est le nombre de coupes. Des détails supplémentaires sont fournis en 2.6.1 concernant l'architecture et la dimension des données d'entrée au modèle.

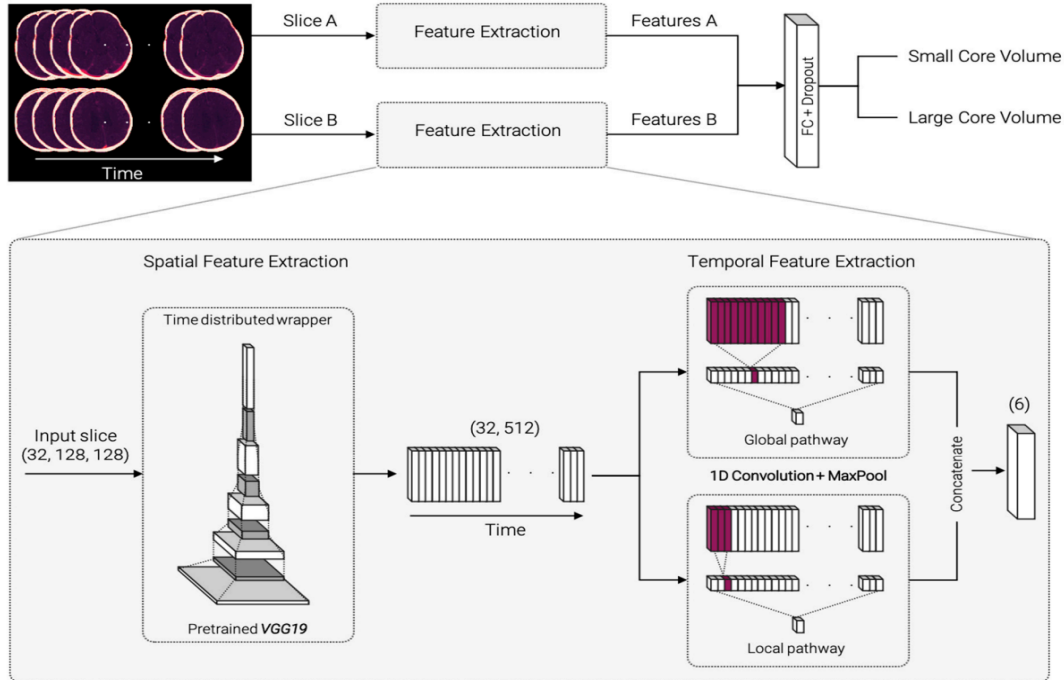
## **2.6 Classificateur supervisé: Réseaux de neurones convolutifs**

### **2.6.1 Introduction à l'architecture**

Cette section a pour but d'introduire conceptuellement l'architecture du modèle. L'entraînement du modèle avec les hyperparamètres choisis est décrit en 2.6.2.

La difficulté majeure de ce travail était de travailler avec des images en 4 dimensions (3 dimensions spatiales et 1 dimension temporelle). Nous avons décidé d'utiliser une approche permettant l'interaction avec les dimensions spatiales et temporelles afin de tenter d'extraire un maximum de variables pertinentes des images. Bien qu'il existe plusieurs approches pour les tâches de classifications d'images (107, 108) une grande partie de la communauté utilisent des RNC, donnant donc accès à une pléthore d'architectures (109).

Pour l'architecture du réseau, nous nous sommes inspirés du modèle décrit à la figure 13 (110). L'algorithme du modèle fonctionne comme suit:



**Figure 13: Architecture du classificateur supervisé.** Provenant de (110)

Pour chaque patient et pour chaque niveau de coupe, les coupes correspondantes des volumes subséquents du même espace cérébral ont été concaténées pour créer une matrice 3D de coupes d'un même espace à des temps différents (*Slice A* et *B* sur la figure 13). Cela a donc créé autant de *coupes* qu'il n'y avait de coupes après l'étape de calibration (2.5.2.1) et chaque *coupe* était de profondeur égale au nombre de volumes présents après l'étape de calibration. Chaque matrice était ensuite passée à travers un bloc d'extraction de variables spatiales (grâce à un RNC pré-entraîné, soit VGG-19) suivis par deux blocs de convolutions 1D en parallèle afin d'extraire les variables temporelles.

Cette architecture nous a permis de traiter les images brutes (c'est-à-dire sans l'utilisation de cartes de perfusion) ainsi que de capturer l'information spatiale ainsi qu'une partie de l'information temporelle. Le bloc d'extraction de variables par le RNC VGG-19 a permis pour chaque coupe dans *coupe A* d'extraire l'information spatiale sous forme de cartes de variables

grâce aux étapes de convolutions et de concaténer ces mêmes cartes. Cette étape a été faite pour chaque *coupe* et a créé une large matrice de cartes de variables organisées à travers le temps pour chaque *coupe* (i.e. *Mat-A*, *Mat-B* etc...). Pour chaque *Mat*, le bloc d'extraction de variables temporelles a procédé à deux blocs de convolutions 1D en parallèle. Le premier bloc utilisait 3 filtres de convolution 1D de *Kernel\_size* = 11, ce qui permettait la capture de variables temporelles diffusées dans un temps plus long puisque le kernel traversait plus de cartes de variables spatio-temporelles, donc plus de temps. Inversement, le deuxième bloc utilisait 3 filtres de convolution 1D de *Kernel\_size* = 3, et permettait la capture de variables temporelles à court terme. Après l'extraction des variables spatio-temporelles pour chaque matrice, les variables ont été concaténées et passées à un réseau de neurones artificiels standard et a permis une classification binaire par *Softmax*, qui consiste en la conversion d'un vecteur numérique en un vecteur a probabilité.

## 2.6.2 Entraînement du modèle

Cette section a pour but d'introduire brièvement les méthodes de validations et de développement qui ont permis d'arriver au modèle final. Une description détaillée est fournie dans la section présentant l'article au chapitre 3.

Afin d'éviter le risque de surapprentissage, nous avons choisi d'utiliser des hyperparamètres prédéfinis sur la base de suggestions dans la littérature. En plus de cela, notre modèle prenait plus de 20 heures à entraîner et utilisait entièrement la mémoire de notre GPU de 16GB. Un entraînement sur une dizaines de combinaisons d'hyperparamètres aurait pris un temps et des ressources considérables.

Le modèle a été entraîné par validation croisée ( $k=5$ ) stratifiée. Cela a permis d'obtenir des métriques d'évaluations sur 5 nouveaux jeux de données (*VAL*) puisque qu'à chaque itération, les nouveaux composants (*ENTRAINEMENT*) du modèle ont été testés à 5 reprises sur les données de *VAL* correspondante. La validation croisée est souvent utilisée afin d'évaluer la capacité de

généralisation du modèle sur de nouvelles données avant de le déployer sur les données *TEST* (87). Toutes les couches du modèle transféré n'étaient pas entraînaibles. Le modèle a été entraîné pendant 300 époques en utilisant une bibliothèque de points de contrôle (Keras.checkpoint), et ne sauvegardait que les modèles basés sur la perte minimale sur le jeu de validation. Le modèle a été entraîné en utilisant une taille de lot de 32 [78], un taux d'apprentissage de 0,001 (111) en utilisant l'optimiseur Adam [76, 103], un taux d'abandon aléatoire de 0,4 [104] et la fonction de perte a été mesurée en utilisant l'entropie croisée binaire. Le rationnel derrière cette décision était que l'entropie croisée (ou plus particulièrement l'entropie croisée binaire) est une métrique hautement utilisée lors de l'optimisation d'algorithme à classification binaire (112).

### **2.6.3 Analyses statistiques**

L'objectif primaire a été évalué par analyse AUROC, scores F1, la précision, le rappel et la valeur prédictive négative ainsi que par une matrice de confusion. Les intervalles de confiances de ces scores ont été obtenus par rééchantillonnage avec remplacement sur les 5 valeurs obtenues lors de la validation croisée.

Concernant l'objectifs secondaire, nous avons procédé à une analyse de performance du modèle au sein des niveaux du type de blessure cérébrale. Les mesures d'évaluation ont été calculées par rééchantillonnage avec remplacement sur les 5 valeurs obtenues lors de la validation croisée.

## **Chapitre 3 : Résultats sous forme d'article**

### **3.1 Introduction**

L'article suivant est en cours de révision par les auteurs et les co-auteurs en vue de sa soumission à la revue Plos Digital Health. Johann Plantin est l'auteur principal ayant entièrement conçu le code (à l'exception de l'architecture du modèle qui est inspiré d'un manuscrit) et réalisé la préparation des données ainsi que les analyses statistiques. Une partie significative du formatage et du nettoyage des données a été faite par Claude Kauffmann, qui a aussi grandement aidé à la conceptualisation et à l'approche de pré-traitement des données. Dr. Michaël Chassé est le dernier auteur et a contribué à la mise en place du projet ainsi qu'à sa supervision. Dr. Louis-Antoine Mullie et Dr. Laurent Létourneau-Guillon, en tant que parrain, ont fourni des suggestions en fonction de leur expertise afin d'améliorer le travail.





## **3.2 Classification of neurological death through automatic image processing**

Authors: Johann Plantin, MSc; Claude Kauffmann PhD; Louis-Antoine Mullie MD MSc; Laurent Létourneau-Guillon MD; Jai Jai Shiva Shankar, MD ; Michaël Chassé, MD PhD

### **3.2.1 Abstract**

The diagnostic of brain death is a complex step when evaluating patients in critical care suspected of being neurologically deceased. Although the clinical neurological criteria that determine brain death are mostly accepted worldwide, the diagnosis remains imperfect and often the use of ancillary tests such as brain computed tomography are required to confirm it. The main objective of this work was to explore the feasibility of classifying brain death from CTP scans using deep learning.

CTP studies from INdEx-CTP (ClinicalTrials.gov, NCT03098511) were pooled from 11 participating sites. Spatial and temporal features were extracted using a combination of two convolution modules and used to predict neurological death. The performance of the model was also evaluated on subsets of cerebral injuries.

217 patients' studies were used to train the model. We report an AUC of 0.79 (IC95% 0.76-0.82), a F1 score of 0.82 (IC95% 0.80-0.83), a precision of 0.92 (IC95% 0.91-0.93), a recall of 0.76 (CI95% 0.72-0.79) as well as a negative predictive value of 0.49 (CI95% 0.45-0.53). Due to a lack of sample size, we did not perform statistical tests on subsets of cerebral injury but report a suspected higher model negative predictive value on anoxic cerebral injury with 0.82 (CI95% 0.77-0.87).

This model shows preliminary evidence supporting the feasibility of developing a deep neural network to classify comatose patients as neurologically deceased or not. Additional research is needed to validate and refine the model by employing larger and more diverse datasets.

### 3.2.2 Introduction

An accurate determination of death is crucial to protect patients and respect the fundamental rule of organ donation, the dead donor rule (26). Although the ground truth of brain death can still be up to debate depending on legislation, cultures or countries (10, 14, 15), it is only recently that an expert community on several aspects of the concept of brain death has come together to establish an international consensus in its definition for a diagnosis (2). Although the determination of brain death has seen some significant improvement in its standardization across countries and cultures and organ donation can be seen rising slightly in Canada (113) organ donation remains at a bottleneck and there remains a gap between the supply and the demand (114, 115). Recent advances in the field of artificial intelligence and availability of big data have enabled great advances in the field of radiology and medicine (116, 117). It is generally admitted that the golden standard for diagnosis of brain death is the clinical examination (2, 29). Unfortunately, it is not always possible to perform the standard diagnostic of brain death through this way. In these situations, ancillary tests can help with the diagnosis. Computed Tomography Perfusion (CTP) is one such test that can measure blood flow in specific areas of the brain to generate perfusion maps and make a diagnosis (36, 45, 46).

Convolutional neural networks (CNNs) are among the most important model architecture for the classification of medical images. For example, the majority of COVID-19 detection methods from chest computed tomography scans are based on CNNs (93) and CNNs are also able to perform highly complex tasks, such as the segmentation of brain metastases (95), the identification of brain pathologies such as Alzheimer's (96) or the identification of intracranial hemorrhage (97-99).

The application of deep learning is widespread and increasingly useful as technology improves and models are used in different medical settings. In critical care, the diagnosis of brain death could benefit from the application of computer vision. One potential could be the early identification of potentially brain death patients to clinical professionals, who could then assess the patient in priority. This could potentially lead to an increase in the amount of accurately diagnosed patients and therefore increase the potential number of organs to be donated.

Additionally, the use of deep learning models in computer vision on CTP scans could also be used as a supporting tool for radiologist during their interpretation, making their task easier and faster without compromising precision. Deep learning has not been applied to the classification of brain death in the critical care setting and its exploration is a first step in assessing its usefulness and potential.

The primary objective of this research was to create a classification model that can predict brain death using CTPs from patients in the intensive care unit who are experiencing severe brain injuries and are in a profound coma state.

The secondary aim of this study was to examine prediction errors according to brain injury types, with the goal of identifying whether the model's predictive capability varies based on the varying types of brain damage.

### **3.2.3 Methods**

#### 3.2.3.1 Population and outcomes

We conducted a transversal study nested in the Canadian clinical trial INDeX-CTP (ClinicalTrials.gov, NCT03098511). INDeX-CTP was a multicenter prospective diagnostic cohort study with the primary objective of evaluating the diagnostic accuracy of neurological death using CTP compared to the standard (complete clinical diagnosis). The goal of this study was to determine if CTP can be used as an auxiliary reference test when complete clinical evaluation cannot be performed.

Within two hours following the acquisition of CTP, two specialists such as neurologists, neurosurgeons, or intensivists conducted a comprehensive clinical examination to diagnose neurological death (ND). The clinicians conducted this examination blindly, without prior knowledge of the CTP results.

The two clinicians then had to declare the patients as neurologically dead or not. If the clinicians disagreed regarding their diagnosis, a consensus had to be reached. If consensus was not reached between the two clinicians, a third clinician was consulted to conduct an additional

clinical examination to assist in obtaining consensus. In the rare scenario where no consensus was obtained after this, an additional auxiliary test other than CTP was used to obtain agreement from the three clinicians.

All patients included in the study suffered devastating brain injury and were suspected of being neurologically deceased. Patients included in the study were  $\geq 18$  years old and admitted to the intensive care unit with either: traumatic brain injury, anoxic brain injury, ischemic or hemorrhagic stroke or other causes of brain injury that lead to a Glasgow Coma Scale of 3. In addition, patients had to have ceased sedation for at least 6 hours to be included in the study.

All patients from INDex-CTP were eligible except for the following cases: deviation from the protocol that made the patient ineligible for the main analysis of the INDex-CTP study; contraindications to the CT-perfusion; confounding factors that had the potential to hinder full clinical examination; absence of images or perfusion; withdrawal of consent or absence/anonymization of DICOM (Digital Imaging and Communications in Medicine) fields necessary for the training of the models; patients with corrupted images.

The outcome for this study was the brain death status prediction. These predictions were then compared to the gold standard of brain death diagnosis, the clinical examination, which served as the ground truth.

#### 3.2.3.2 Data and preprocessing

The data were images derived from brain CTPs from 11 individual sites with usable data stored in DICOM (Digital Imaging and Communications in Medicine) format. As these images originated from different sites with variations in acquisition protocols, preprocessing was performed to improve spatial and temporal synchronization. These differences could arise from intrinsic factors such as craniocerebral dimensions of the patients or variations in the scan protocols such as the number of slices, depth of brain coverage, acquisition time, acquisition resolution or image resolution. The parameters needed to be considered to synchronize the spatiotemporal dimensions of the images were the interpixel spacing, slice thickness, acquisition time, the X and Y dimensions of the 2D image (representing image resolution), and the Z

dimension representing the volume's depth. A summary of the DICOM fields stratification is reported in Supplementary SI.

To minimize the distortion applied to the images, we conducted spatial preprocessing using the most frequent values found in the DICOM attributes. were used as a reference. For instance, if the most common values for interpixel spacing (x,y) and slice thickness(z) were 0.4mm(x) x 0.4mm(y) and 5mm(z) respectively, techniques such as interpolation and elimination of slices were used to ensure that all volumes had the same dimensions of 0.4mm x 0.4mm x 5mm. Each slice would be interpolated based on these most frequent values, [0.4, 0.4]. Regarding the slice thickness, the largest slice thickness value would be used as reference. For instance, if the largest slice thickness value were 5mm and a study had slice thickness values of 2.5, then one slice out of two would be dropped. This standardization of dimensions helped make the data comparable across all patients and minimize the potential impact of the variations in the image dimensions on the study outcome.

Regarding temporal synchronization, we used the DICOM field AcquisitionTime, which is the time the acquisition of data that resulted in this image started. Using this field, it was possible to deduce the time required to generate each volume by subtracting the AcquisitionTime field of the first slice of a volume to the last slice of the same volume. As some imaging study protocols could generate more volumes than others for the same time frame, we subsampled volumes based on the slowest value of acquisition speed. For instance, if the slowest acquisition speed was 0.5 s/volume and another study had 0.25s/volume, then one volume out of two would be dropped for this study. In some cases, the AcquisitionTime DICOM field was not available for each slice, but each volume. This meant that it was not possible to have the true value of the speed of acquisition since there is some time in between the acquisition of each volume. In these cases, the acquisition speed was inferred from the mean acquisition speed of the sample.

After these steps, the resolution of the images was varying due to the interpolation, the total number of slices and volumes were also varying because of the different acquisition protocols. To remedy this, each 2D image (slice) was resized to the input required by the pre-trained model (128x128). Regarding the varying number of slices, the slices were subsampled

based on the smallest remaining value (i.e., If the minimum number of slices in the sample was 16, and a study has 22, then 8 will be dropped from this study). The same rational goes for volume numbers.

At this step, each image now had the same resolution, number of slices and number of volumes, and the spatiotemporal synchronization was improved. Each image was then normalized using a Hounsfield Unit normalization technique and a threshold on Hounsfield Units was applied. When using thresholding, each pixel under a minimum H.U value of 40 was set to 40 and each pixel above the value of 600 was set to 600. This was thought to increase the representation of the contrast within the brain (106). To create the input for our model, we had to cast the 2D image on an RGB (Red, Green, Blue) axis to mimic the structure of an RGB image which led to the greyscale image being projected 3 times on the RGB dimension. For each slice of each volume, we stack them on top of each other to generate a 2D timeseries contained the information regarding the movement of the contrast agent across a single slice. Once this was done for each slice, we ended up with an input of shape (20, 128, 128, 3, 16) for each patient, where 20 represents the number of volumes, 128,128 the resolution of each 2D image, 3 the greyscale image cast on the RBG axis, and 16 is the number of slices.

### 3.2.3.3 Model architecture and training

The model architecture, based on a recently published study and summarized in Supplementary SII, consisted in two CNNs in succession that enabled the extraction of spatial and temporal features (110). The model took as input  $n$  2D timeseries of shape (20, 128, 128, 3), where  $n$  was the number of slices present in each input (16 as described in 3.2.3.2). Each 2D timeseries was passed through the first CNN, the spatial feature extractor VGG19, and created a set of feature maps that were then passed through the second CNN. This CNN was essentially a temporal feature extractor as it performed 1D convolutions on the spatial feature maps extracted by VGG19. This new set of spatiotemporal features were then concatenated with every other set of features generated for each  $n$  slices for that patient. The vector was then passed to a single fully connected layer containing 20 neurons and then produced a binary output using a sigmoid function.

Regarding model training, we used 5-fold cross validation stratified on label and cerebral injury type (Supplementary SIII). The model was trained for 300 epochs using a checkpoint library from keras (version 2.2.4), only saving models based on minimum validation loss. The model was trained using a batch size of 32 (81), a learning rate of 0.001 using the Adam optimizer (79, 111), a dropout rate of 0.4 (84) and the loss function was measured using binary cross entropy.

All the code for pre-processing, training as well as analysis was written in python (v 3.7.4) and libraries such as Keras (version 2.2.4) and TensorFlow (version 1.14.0) were used for deep learning. Jupyter lab was used as our web-based interactive development environment (version 3.3.2)

#### 3.2.3.4 Statistical analysis

For our primary objective, we reported the area under the ROC curve (AUC score), F1 score, precision, recall as well as negative predictive value (NPV). Confidence intervals (95% CI) for these metrics were generated by bootstrapping 1000 times with replacement on the 5 scores obtained after training on the 5 folds. The threshold used to transform the output of the sigmoid activation was optimized on the F1 score. The rationale behind this was that both the precision and recall were important, and considering we were in the case of an unbalanced dataset, we thought it relevant to optimize on this metric.

Regarding the secondary objective, the very limited amount of data after stratification based on cerebral injury made it challenging to have a robust representation of predictive capabilities based on these subsets of injuries. We also reported F1 score, precision, recall and NPV, with the same approach for confidence intervals. For both objectives, we also report the ROC curves, confusion matrices as well as the training and validation loss plots during training of each fold.



### 3.2.4 Results

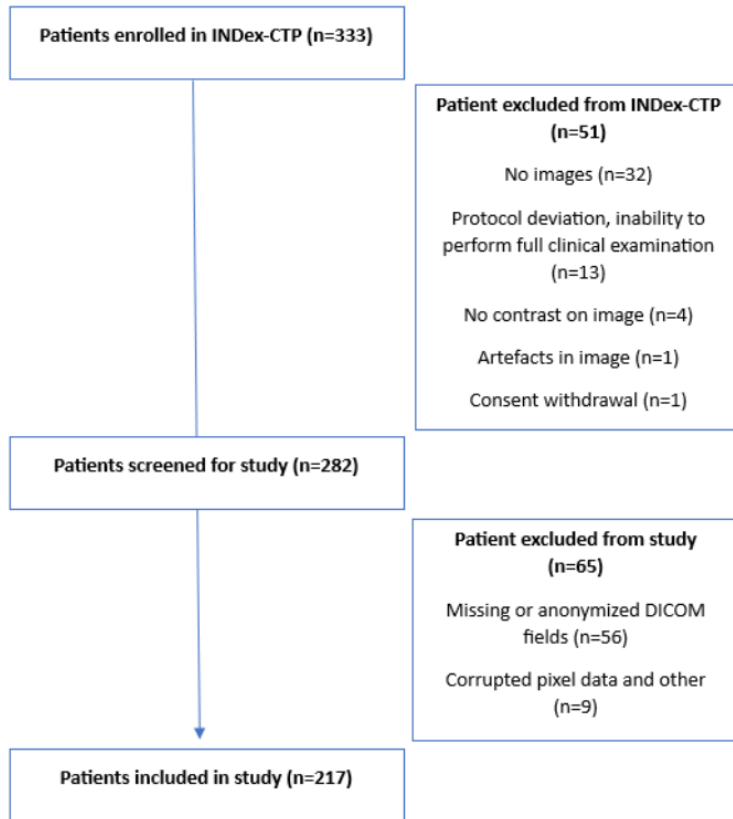
The demographic information as well as the stratification of patients in function of site and cerebral injury type are reported in table 1.

**Table 1. Patient demographics**

	<b>Sample (n=217)</b>
<b>Median Age (std)</b>	58 (15.1)
<b>Sex (F) n(%)</b>	101 (46)
<b>Neurological Status (Alive) n(%)</b>	54 (25)
<b>Race n(%)</b>	
Caucasian	163 (75)
Asian	23 (11)
Unknown	18 (8)
First Nations	9 (4)
Black or African American	2 (1)
Other	2 (1)
<b>Cerebral injury n(%)</b>	
Hemorrhagic stroke	113 (52)
Anoxic	45 (21)
Traumatic brain injury	35 (16)
Other	14 (6)
Ischemic stroke	10 (5)
<b>Scan site n(%)</b>	
01	82 (38)
14	39 (18)
07	35 (16)
05	16 (7)
17	10 (5)
02	9 (4)
11	9 (4)
09	6 (3)
04	5 (2)
12	4 (2)
10	1 < (1)
18	1 < (1)

The study includes 333 patients aged 18 years or older with brain injuries who were admitted to the intensive care unit. Patients were required to have a Glasgow Coma Scale score of 3 and at least 6 hours without sedation. Out of the 333 patients enrolled in INdEx-CTP, 51 were excluded from the main study for several reasons (protocol deviation, insufficient images or perfusion, consent withdrawal). From the remaining 282 patients, 65 were excluded from this work due to reasons such as anonymization or missing crucial DICOM fields like AcquisitionTime (Figure 1).

Figure 1. Study flowchart



The distribution of cerebral injury types was as follows: Hemorrhagic stroke (H.Stroke, 52%), Anoxic (21%), Traumatic brain injury (TBI, 16%), Other (6%), and Ischemic Stroke (I.Stroke, 5%). Sample variables such as label, site, and cerebral injury type stratification across training folds are reported in Supplementary SIII. Each fold was trained on 173 patients and validated on 44 patients. The training and validation loss plots for each fold are reported in SIV.

For each fold, the AUROC curve is depicted in Figure 2. The overall AUC score obtained with bootstrapping was 0.79 [0.76, 0.82]. An F1 score of 0.82 [0.80, 0.83], precision of 0.92 [0.91, 0.93], recall of 0.76 [0.72, 0.79], and NPV values of 0.49 [0.45, 0.53] were reported (Table 2).

The threshold optimization for these scores was based on the F1 score, and varying thresholds for each of the 5 training folds are reported (0.68, 0.84, 0.63, 0.71, 0.87). The confusion matrix across the overall training folds revealed most accurate predictions when inferring Brain Death and very low instances of misclassified Brain Deaths (Figure 3). The confusion matrix across the overall training fold for stratified cerebral injuries also indicated a majority of correct predictions when inferring Brain Death and very few erroneously labeled Brain Deaths (Figure 4). In terms of performance metrics for different types of cerebral injury, we report the highest F1 scores and precision metrics for H.Stroke (0.82 and 0.96, respectively) and Anoxic (0.85 and 0.89, respectively) levels, as shown in Table 3.

The F1 scores obtained during each training fold for each cerebral injury type demonstrated high variation between folds, particularly for TBI (ranging from 0.44 to 0.88), I.Stroke (ranging from NA to 1), and Other (ranging from 0.66 to 1) levels.

Fig 2. Roc curves for each of the 5 training folds

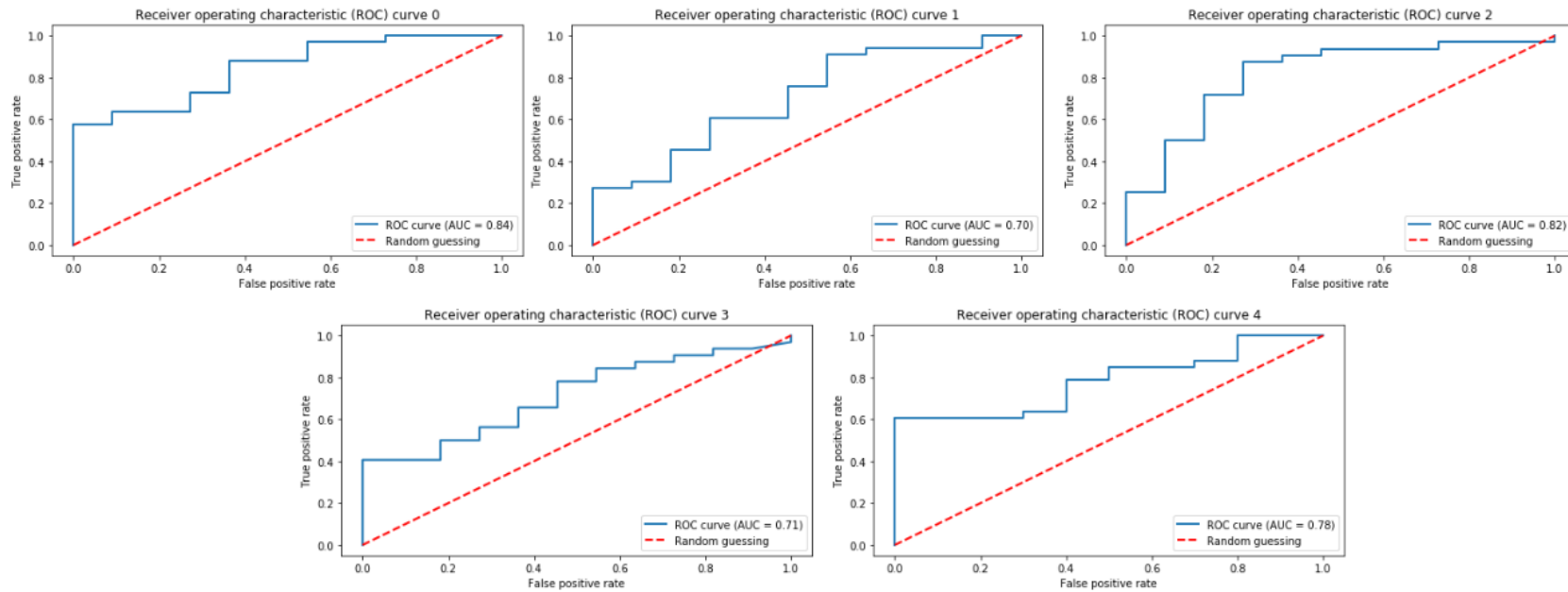


Table 2. Model performance across training folds

	<b>AUC</b>	<b>F1</b>	<b>Precision</b>	<b>Recall</b>	<b>npv</b>
<b>Scores</b>	0.79 [0.76, 0.82]	0.82 [0.80, 0.83]	0.92 [0.91, 0.93]	0.76 [0.72, 0.79]	0.49 [0.45, 0.53]

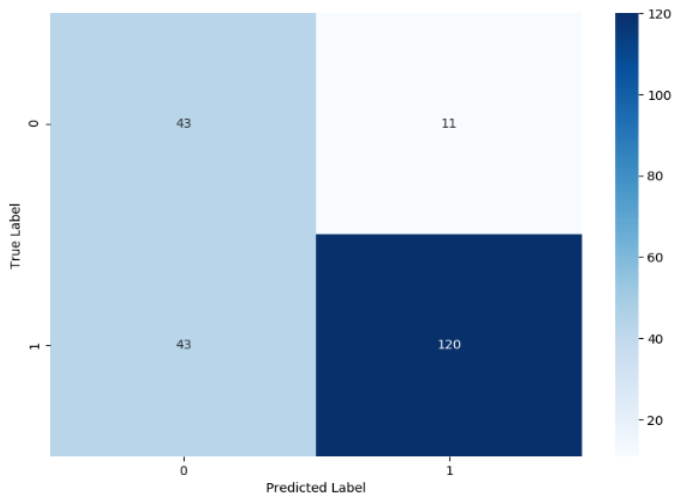
Scores were obtained by bootstrapping with replacement over the validation scores in the 5 training folds, with a 95% confidence interval.

Table 3. Model performance across injury type

	<b>Cerebral injury type</b>	<b>F1</b>	<b>Precision</b>	<b>Recall</b>	<b>npv</b>
<b>Scores</b>	H.Stroke	0.82 [0.80, 0.84]	0.96 [0.94, 0.97]	0.75 [0.73, 0.78]	0.41 [0.37, 0.45]
	Anoxic	0.85 [0.82, 0.87]	0.89 [0.85, 0.94]	0.81 [0.77, 0.85]	0.82 [0.77, 0.87]
	TBI	0.61 [0.55, 0.68]	0.76 [0.69, 0.83]	0.58 [0.51, 0.64]	0.49 [0.44, 0.54]
	Other	0.74 [0.70, 0.78]*	0.87 [0.78, 0.96]*	0.79 [0.74, 0.83]*	0.63 [0.54, 0.73]*
	i.Stroke	1.0*	1.0*	0.78 [0.64, 0.91]*	-

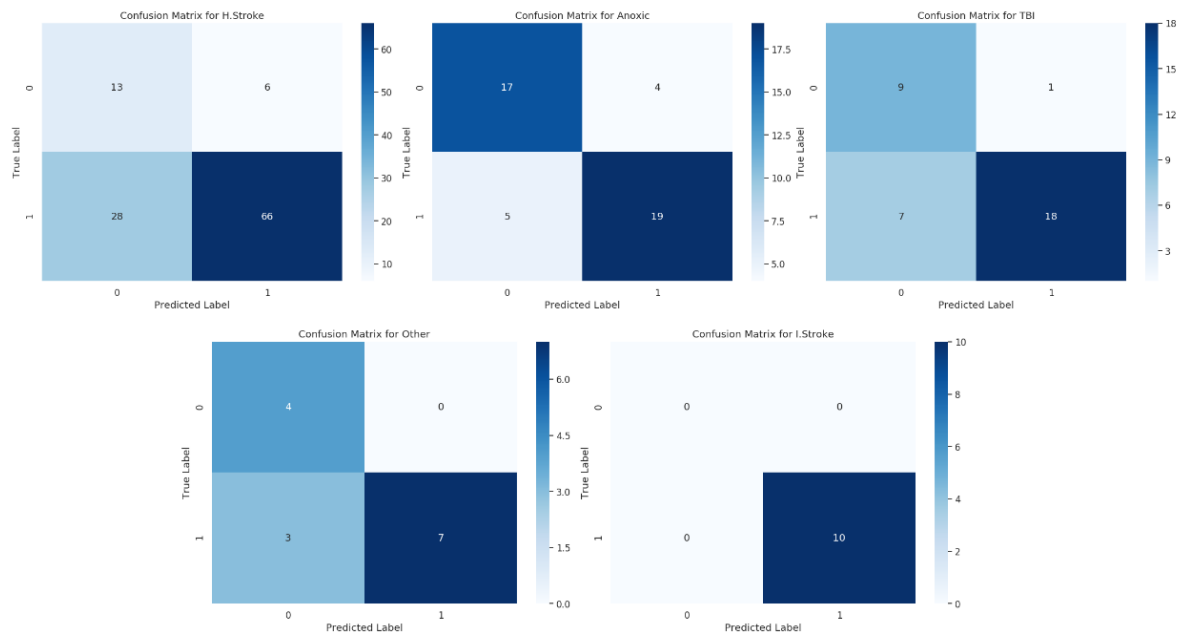
Scores were obtained by bootstrapping with replacement over the validation scores in the 5 training folds, with a 95% confidence interval. Values with an asterisk indicate either: 1) Less than 5 values were available for bootstrapping (i.e. 2 out of 5 folds returned npv values of NaN); 2) Only 1 value was available, therefore not displaying any confidence intervals.

Fig 3. Confusion matrix



Confusion matrix across all training folds. 1 indicates Brain death and 0 Alive.

Fig 4. Confusion matrix in function of cerebral injury



Confusion matrix across all training folds for different cerebral injuries. 1 indicates Brain Death and 0 Alive.

### 3.2.5 Discussion

In this work, we use an open-source architecture using two blocks of convolution to extract spatial and temporal features (110) of CTP scans from individuals suffering several cerebral injuries to classify their status of neurological death.

We found that the model achieved high precision and F1 scores when all cerebral injury types were combined (0.92 and 0.82, respectively). This high precision translates to a high positive predictive value, where the model will accurately classify patients as deceased 92% of the time. Conversely, we report relatively lower negative predictive values for all cerebral injury types combined (0.49), suggesting some calibration limitations. This indicates that the model may incorrectly classify a patient as alive. It is important to note that overall, we report a recall of 0.76 and a specificity of 0.80. In the parent INDeX-CTP study, we observed a recall of 90% and a specificity of 75% among human radiology readers (118). Empirically, the performance of our model appears to be comparable to that of humans using similar images and underscores the current practice of not relying solely on ancillary testing for brain death diagnosis, but also incorporating clinical evaluations. Although we chose to optimise the threshold on the F1 score, it would be worthwhile to test optimizing on different metrics such as recall. The decision of threshold optimization would largely depend on the application of the model and the question at hand. Optimizing on recall would force the model to minimise the number of false negatives, therefore potentially increasing the number *Brain Dead* classifications.

In terms of cerebral injury type analysis, the limited sample size after stratification hinders our ability to draw definitive conclusions between the levels. However, the performance of the model for H.Stroke, TBI, and Anoxic levels can be cautiously interpreted. Notably, the model appears to perform considerably better in the Anoxic level, with the negative predictive value being higher than all other levels (0.82).

One major benefit of our approach was that the raw CTP images were used directly without going through perfusion mapping. Although some degree of pre-processing was required, very little human intervention was required after some initial file organization (the program pre-processed the images based on an algorithm). This could indicate that in future works, an

improved model could be used in clinical care directly after the acquisition of the CTP. This could help clinicians and radiologist establish the diagnosis of neurological death.

In a paradoxical manner, the principal strength of the method also presents its most significant weakness. While the technique allows for the acquisition of scans from multiple sites without necessitating the costly and labor-intensive generation of perfusion maps, it introduces certain limitations. Notably, despite a well-defined INDEX-CTP acquisition protocol, there exist variations in image generation across studies. For example, inconsistencies arise from differing slice spacing and unavailable parameters, leading to variable acquisition resolutions. Furthermore, the scan initiation location and volume coverage vary across sites and patients, which can impact the model's ability to extract spatial features and determine explicative factors for brain death. Consequently, the model's learning process may either be enhanced through generalization across diverse acquisition protocols or hindered by the necessity to learn non-explicative variables.

Another limitation arises from the unavailability of true acquisition times for all studies, resulting in approximately 44% of acquisition times being inferred from the sample mean (Supplementary I). Additionally, both the acquisition time and time between acquisitions lack consistency within and between studies. This variability may introduce temporal distortions across studies, despite efforts to synchronize temporally during pre-processing. Since VGG 19 was a deep model, it had an extremely large number of parameters per patient, which diluted the explicative features within this large parameter space and induced overfitting after 100 or 200 epochs of training depending on training fold (Supplementary IV). This problem was addressed by saving models on minimal validation loss but did not address the issue that the model could be learning from non-explicative features. The use of Grad-CAM algorithm could be beneficial to open this black box and unfold what features the model has learning on (119, 120), although there are no validated techniques that enable explicability analyses on 4D models. Some newer models are able to achieve top performances using a smaller parameter space and higher parameter efficiency such as EfficientNetV2 (70), as such it could be relevant to try replacing VGG19 with a lighter model. Some works also use 3D CNNs to detect stalled brain capillaries by



converting the 3D volumes to mp4 videos (101), it would therefore be interesting to explore a more complex model that would fully incorporate the 3D and temporal components of a CTP.

Another prominent limitation of this study is the sample size. During each training fold, the model was trained on merely 173 images and validated on 44. When combined with the issue of class imbalance (despite stratified training based on label and cerebral injury), the interpretation of these results warrants caution. Furthermore, the absence of a test set complicates the evaluation of the model's generalization potential. Consequently, it is recommended that more data be collected and tested with this model prior to drawing conclusions about its true generalization performance. In conclusion, this study offers preliminary evidence supporting the feasibility of developing a deep neural network to classify comatose patients as neurologically deceased or not. Our results indicate that the proposed model can differentiate between these two groups with promising accuracy, potentially contributing to a more precise and objective determination of neurological death in clinical settings. However, we acknowledge the limitations of the current study, including the relatively small sample size, class imbalance, and the absence of a test set, which may result in overfitting. Further research is necessary to validate and refine the deep neural network by employing larger and more diverse datasets to improve its generalizability across various patient populations and healthcare settings. Despite these limitations, our work represents a novel approach to processing clinical CTP images of patients suspected of neurological death determination, allowing for classification based on brainstem criteria.

### **3.2.6 Competing interest**

The author of this manuscript has no conflicts of interest to declare.

### **3.2.7 Data availability**

Access to the data is restricted to the research team by the ethical board according to provincial laws.

### **3.2.8 Bibliography**

The bibliography is synchronized with the memoire, please refer to that section.

### 3.2.9 Supplementary materials

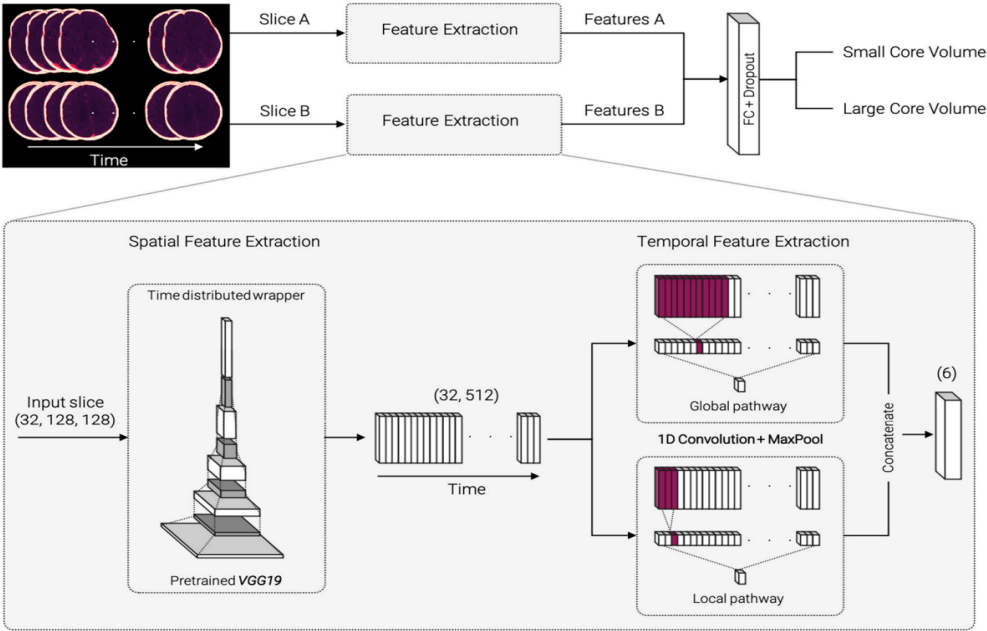
#### Supplementary SI: Distribution of DICOM parameters.

	Sample (n=217)
<b>PixelSpacing n(%)</b>	
[0.449219, 0.449219]*	42 (19)
[0.390625, 0.390625]	39 (18)
[0.430000, 0.430000]	21 (10)
[0.46875, 0.46875]	21 (10)
[0.488281, 0.488281]	10 (5)
[0.466796875, 0.466796875]	8 (4)
[0.470703125, 0.470703125]	7 (3)
[0.47265625, 0.47265625]	5 (2)
[0.39453125, 0.39453125]	4 (2)
[0.41796875, 0.41796875]	3 (1)
[0.4296875, 0.4296875]	3 (1)
[0.43359375, 0.43359375]	3 (1)
Other values < 1%	54 (24)
<b>SliceThickness (mm) n(%)</b>	
5*	180 (82)
1.5	37 (18)
<b>Volume per second availability (Yes) n(%)</b>	
	121 (56)
<b>Acquisition Time (Volume/sec) n(%)</b>	
0.773406	95 (44)
0.78125	68 (32)
0.719424*	14 (6)
0.75188	13 (6)
0.862069	7 (3)
0.746269	6 (3)
0.724638	4 (2)
0.806452	3 (1)
0.854701	2 (1)
0.757576	2 (1)
0.793651	1 (<1)
0.787402	1 (<1)
0.740741	1 (<1)

\* The asterisk represents values that were kept as reference for data preprocessing in section

3.2.3.2.

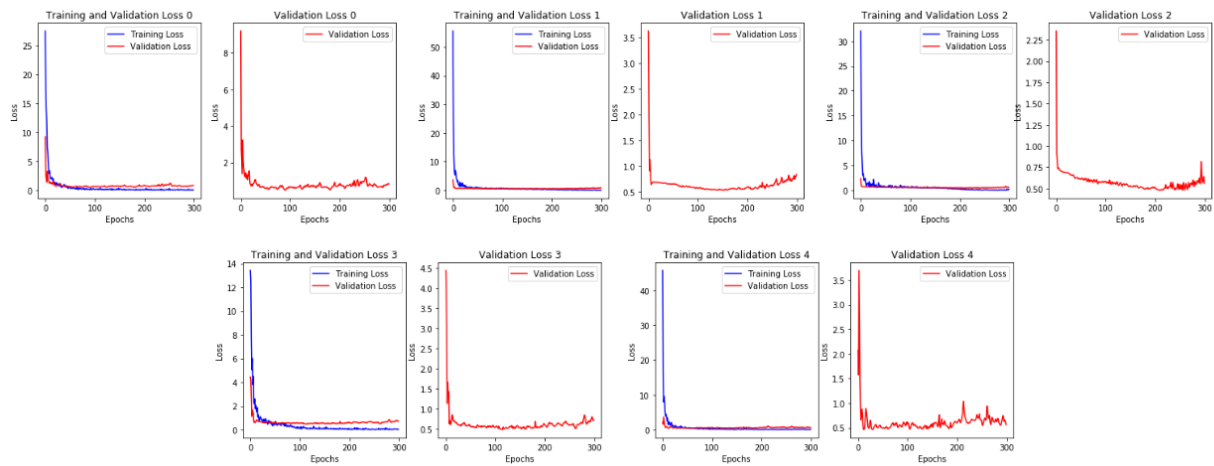
**Supplementary SII: Architecture of the model (106).** This figure represents closely the model used in this study. The exception is the output which is Brain Death vs No Brain Death as opposed to the output of this figure. In addition, the input slice is (20, 128, 128) as opposed to (32, 128,128). The architecture was taken from the open source: <https://github.com/AndreasMittermeier/stroke-perfusion-CNN>



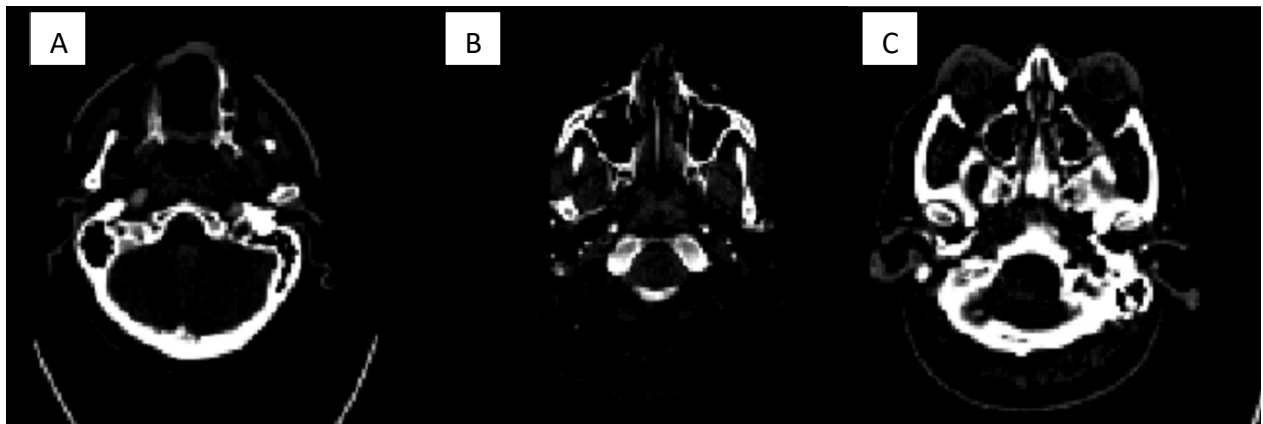
**Supplementary SIII: Stratification of variables of interest across the 5 Train/Val folds.**

	Train 1	Val 1	Train 2	Val 2	Train 3	Val 3	Train 4	Val 4	Train 5	Val 5
<b>Neurological death status</b>	43	11	43	11	43	11	43	11	44	10
<b>Alive n(%)</b>	(25)	(25)	(25)	(25)	(25)	(25)	(25)	(25)	(25)	(23)
<b>Cerebral injury subtypes</b>										
<b>n(%)</b>										
Hemorrhagic stroke	89 (51)	24 (54)	91 (53)	22 (50)	91 (53)	22 (50)	92 (53)	21 (49)	89 (51)	24 (54)
Anoxic	36 (21)	9 (20)	33 (19)	12 (27)	36 (21)	9 (21)	37 (21)	8 (19)	38 (22)	7 (16)
Traumatic brain injury (TBI)	31 (18)	4 (9)	30 (17)	5 (11)	29 (17)	6 (14)	23 (13)	12 (28)	27 (16)	8 (19)
Other	10 (6)	4 (9)	10 (6)	5 (11)	12 (7)	4 (9)	14 (8)	0 (0)	11 (6)	3 (6)
Ischemic stroke	7 (4)	3 (6)	9 (5)	0 (0)	6 (3)	2 (5)	8 (5)	2 (4)	9 (5)	1 (2)
<b>Sites n(%)</b>										
01	65 (38)	17 (39)	66 (38)	16 (36)	63 (36)	19 (44)	69 (40)	13 (30)	65 (38)	17 (39)
14	31 (18)	8 (18)	30 (17)	9 (20)	32 (18)	7 (16)	31 (18)	8 (19)	32 (18)	8 (18)
07	27 (16)	8 (18)	30 (17)	5 (11)	30 (17)	5 (12)	27 (16)	8 (19)	26 (16)	8 (18)
05	14 (8)	2 (5)	13 (8)	3 (7)	12 (7)	4 (9)	12 (7)	4 (10)	13 (8)	2 (5)
17	9 (5)	1 (2)	8 (5)	2 (5)	9 (6)	4 (9)	8 (5)	2 (5)	9 (5)	1 (2)
11	8 (4)	1 (2)	6 (3)	4 (9)	7 (4)	2 (5)	8 (5)	1 (2)	8 (4)	1 (2)
02	6 (3)	3 (7)	5 (3)	3 (7)	6 (3)	0 (0)	7 (4)	2 (5)	8 (4)	3 (7)
09	5 (3)	1 (2)	5 (3)	1 (2)	5 (3)	1 (2)	4 (2)	2 (5)	5 (3)	1 (2)
04	4 (2)	1 (2)	4 (2)	1 (2)	5 (3)	0 (0)	4 (2)	2 (5)	4 (2)	1 (2)
12	2 (1)	2 (5)	4 (2)	0 (0)	4 (2)	0 (0)	4 (2)	0 (0)	2 (1)	2 (5)
10	1 (1)	0 (0)	1 (1)	0 (0)	0 (0)	1 (2)	1 (1)	0 (0)	1 (1)	0 (0)
18	1 (1)	0 (0)	1 (1)	0 (0)	1 (1)	0 (0)	0 (0)	1 (2)	1 (1)	0 (0)

### Supplementary SIV: Train and Validation loss for each fold



### Supplementary SV: First slice of three randomly selected patients from different sites



## Chapitre 4: Discussion

### 4.1 Retour sur les objectifs du projet

L'objectif principal de ce projet était d'évaluer la performance de classification de mort neurologique par traitement automatique d'images de scans CTP. Les 217 patients inclus dans ce travail souffraient de blessures cérébrales sévères et ont tous été évalués par le standard de détermination de décès neurologique, soit l'examen clinique. Il s'agissait donc d'une situation d'apprentissage supervisé avec une étiquette de haute qualité, mais avec peu de données disponibles pour entraîner ce type de modèle. Le modèle utilisé afin d'accomplir cette tâche était deux blocs de réseaux de neurones convolutifs séquentiels. Le premier de ces blocs était à des fins d'extraction de variables spatiales tandis que le deuxième était à des fins d'extraction de variables temporelles.

Puisque notre échantillon contenait plusieurs types de blessures cérébrales (Figure 14), notre objectif secondaire était d'évaluer les différences de performance du modèle en fonction des niveaux de cette variable.

L'application de l'apprentissage profond en radiologie, notamment pour le diagnostic de mort cérébrale dans le contexte des soins critiques, présente un potentiel considérable. La vision par ordinateur pourrait faciliter l'identification des patients en état de mort cérébrale lors du dépistage, augmentant ainsi le nombre d'organes disponibles pour les transplantations. De plus, les modèles d'apprentissage profond pourraient servir d'outil d'assistance aux radiologistes, améliorant leur efficacité sans sacrifier la précision. Cette étude représente une première étape pour évaluer l'utilité et le potentiel de cette approche.

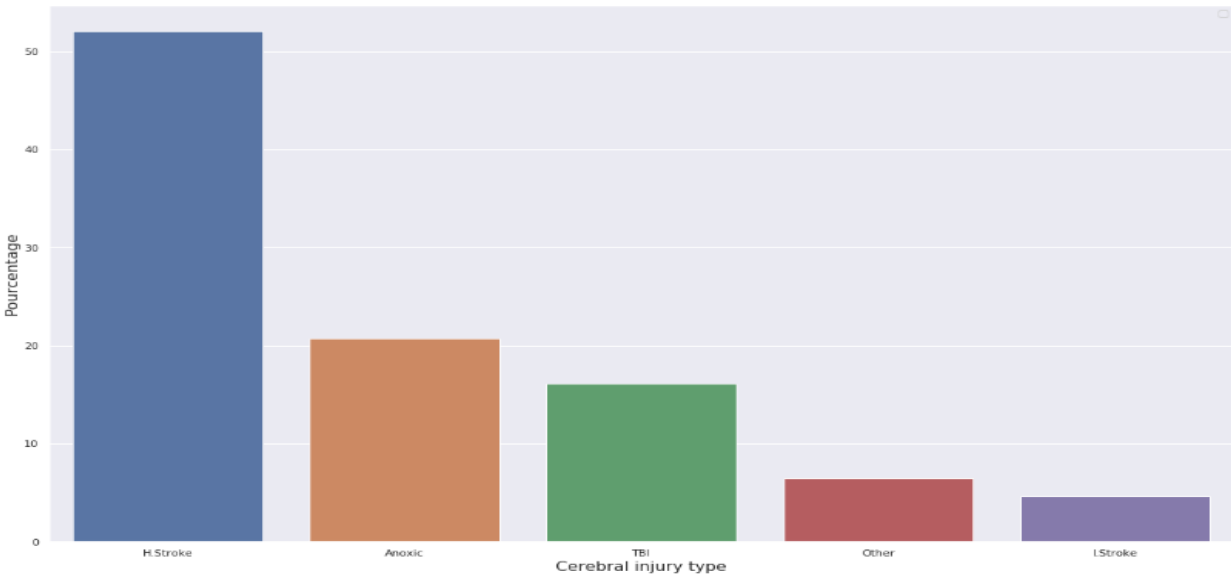
## 4.2 Retour sur les résultats

Les résultats de ce projet montrent une performance intéressante. Nous avons observé une valeur prédictive positive de 0.92 au détriment du rappel qui se situe à 0.76. Cela se traduit par une classification de DN appropriée dans 92% des cas mais que 24% des patients ayant la condition d'intérêt (décès neurologique) seront incorrectement classifiés comme non décédés. Nous observons aussi une valeur prédictive négative de 0.49, indiquant que le modèle réduisait de moitié la probabilité de déclarer un patient comme décédé par rapport au risque prévalent de base. Cela suggère une certaine difficulté à extraire les informations de l'image associées au statut *VIVANT* versus *décédé* dans notre cohorte.

Nous constatons aussi que le seuil d'optimisation utilisé entre les 5 différents jeux d'entraînement/validation ne sont pas similaires (0.68, 0.84, 0.63, 0.71, 0.87). Ces différences suggèrent qu'à chaque entraînement, le modèle fait varier son seuil fortement afin de maximiser le score F1. Si les données étaient homogènes à travers les 5 jeux d'entraînement, une valeur relativement proche du seuil aurait été observée puisque le modèle aurait probablement des poids similaires. Cela suggère une variabilité dans les données et au sein du jeux d'entraînement, augmentant les risques de problèmes de généralisation et supportant la limitation en lien avec la taille d'échantillon.

En ce qui concerne l'évaluation des différents types de lésions cérébrales, les résultats obtenus à partir de notre échantillon stratifié sont limités en termes de généralisation dû à la taille d'échantillon réduite et à leur distribution (Article Annexe SIII et Figure 14). Toutefois, les résultats suggèrent que la performance du modèle au sein du niveau anoxique présente une meilleure performance globale, surtout quand comparée à la valeur prédictive négative du niveau H.stroke (0.82 vs 0.41 respectivement).

Figure 14 : Distribution du type d'injures cérébrales



H.stroke: Accident vasculaire cérébral hémorragique; Anoxic : Anoxique, TBI : Traumatisme crânien, Other : Autre, I.Stroke : Accident vasculaire cérébral ischémique

### 4.3 Forces

Une des forces de cette étude provient de l'utilisation directe des scans CTP multicentriques de l'étude principale (INDex-CTP) malgré des variations dans les protocoles d'acquisition. Très souvent en clinique, les données ne sont pas générées à des fins d'apprentissage profond, et le fait de pouvoir inclure des scans utilisant des protocoles d'acquisition différents (comme l'espace entre les coupes, le type d'agent de contraste utilise, la localisation de la première coupe et le volume total balayé) permet d'élargir le spectre d'utilisation d'images. Cette force est aussi une limitation primordiale du travail, qui sera discuté en 4.4.

Une force additionnelle de ce travail est l'utilisation des images brutes de CTP sans passer par la génération de cartes de perfusion. Habituellement, les scans CTP permettent la construction de cartes de perfusions reflétant plusieurs indicateurs comme le volume sanguin cérébral, le débit sanguin cérébral, le temps de transit moyen ainsi que le temps jusqu'au pic (40). Ces cartes permettent l'interprétation de la perfusion cérébrale du patient par le radiologiste. Dans un



contexte hospitalier, il pourrait être avantageux de pouvoir utiliser directement les images d'un patient arrivant aux soins intensifs pour lequel nous avons besoin d'établir un premier dépistage de risque de décès neurologique.

Une autre force de ce projet est l'automatisation de la synchronisation spatiale et temporelle des images. Après l'étape chronophage de réorganisation des fichiers DICOM, la vérification de la structure, de l'ordre et de la qualité des images, l'étape de synchronisation (décrite en 2.6.2.1) est automatisée et permet la comparaison des images entre elles. Il est aussi possible que cela puisse favoriser la généralisation du modèle sur plusieurs sites avec des protocoles d'acquisition pouvant varier sensiblement. Ce point est un gain de temps considérable, mais est aussi fortement limité par l'imperfection intrinsèque de la méthode de synchronisation, décrite en 4.4.

Une autre force provient de la rareté des images ainsi que de la haute qualité méthodologique de l'étude INdEx-CTP. Le fait d'avoir autant d'images de patients en DN confirmées par le standard de diagnostic de DN est un atout primordial, spécifiquement lors d'un apprentissage supervisé comme celui-ci. La qualité des données et de l'étiquette font partie des facteurs les plus importants afin de donner l'opportunité au modèle d'apprendre optimalement. Malgré quelques différences non négligeables concernant les protocoles d'acquisition inter site et intra site, il existe néanmoins une directive spécifique sur plusieurs composants du protocole et l'étiquette de DN est haute qualité.

## 4.4 Limitations

Plusieurs limitations de ce travail sont à noter.

Étant donné que les scans CTP de l'étude principale (INDex-CTP) n'étaient pas effectués à des fins d'apprentissage profond, les différences qui existaient entre les protocoles d'acquisition ont complexifié la tâche en termes d'extraction de variables explicatives. Cette limitation est habituellement surmontée par la génération de cartes de perfusion qui permettent de prendre en compte ces variations de façon optimale. Il est possible qu'un modèle fonctionne mieux sur des cartes de perfusions puisqu'elles résument le mouvement de l'agent de contraste à travers le temps, et les images sont traitées de façon plus précise qu'avec l'étape de prétraitement utilisée dans ce projet.

Un exemple pour illustrer cela est le fait que nous n'avons pas corrigé pour les variations entre l'espace qui s'éparait chaque coupe. Nous avons dû travailler de cette façon puisque cette information était manquante pour une proportion de patients (ces champs ayant été anonymisés)

La figure de l'article en annexe SV de l'article montre aussi une limitation potentielle. En effet, nous remarquons que les divers scans ne débutent pas au même endroit au niveau du crâne. Puisque le scan ne débute pas aux mêmes endroits, il s'agit d'un défi supplémentaire lié à la synchronisation spatiale entre les patients.

Une limitation additionnelle au sein de ce travail vient du fait que la référence pour le diagnostic de DN dans l'étude INDex-CTP était la mort du tronc. Il est possible qu'il y ait eu quelques erreurs de classification faites par les cliniciens. Étant donné que les étiquettes de notre projet utilisent le diagnostic de DN d'INDex-CTP, ces erreurs potentielles ont peut-être réduit les capacités d'apprentissage de notre modèle.

L'optimisation de l'apprentissage sur le score F1 pourrait ne pas être la meilleure approche. Il serait intéressant d'essayer d'optimiser sur différentes métriques comme la sensibilité. Cette décision serait majoritairement dépendante du contexte d'application du modèle. Par exemple, l'optimisation sur la sensibilité forcerait le modèle à minimiser ses fausses prédictions négatives

(dire qu'un patient est vivant alors qu'il est décédé neurologiquement), ce qui permettrait d'augmenter la quantité de patients identifiés comme décédé neurologiquement, au risque d'augmenter le nombre de mauvaises prédictions positives (dire qu'un patient est mort alors qu'il est vivant).

La contrainte majeure de cette étude réside dans la limitation des données disponibles. Avec un échantillon de 217 patients et une procédure d'apprentissage utilisant 5 validations croisées, répartissant les données en 173 examens d'entraînement et 44 examens de validation, nous sommes confrontés à une situation où non seulement l'échantillon est de petite taille et déséquilibré en termes de répartition des étiquettes, mais également insuffisant pour conserver un jeu de données TEST. Cette faible taille d'échantillon entrave non seulement l'évaluation finale du modèle sur un jeu de données TEST, mais restreint également l'analyse statistique. Afin d'évaluer l'impact de cette limitation potentielle, il serait intéressant d'analyser la performance du modèle en utilisant une cohorte limitée artificiellement. Un graphique montrant la performance du modèle en fonction du pourcentage d'échantillon utilisé pourrait fournir une information intéressante afin d'explorer cette limitation.

Cette problématique est courante dans le domaine biomédical et souligne la nécessité de développer des méthodes d'apprentissage adaptées aux contextes pauvres en données, telles que le *Few-shot Learning*, ou des approches de transfert d'apprentissage. Néanmoins, il est encourageant de constater que le présent travail a atteint un certain niveau d'apprentissage malgré ces contraintes. Les recherches futures devraient s'orienter vers l'exploration de méthodes d'apprentissage adaptées aux situations de faible disponibilité de données, comme cela peut être le cas pour les maladies rares. Ainsi, pour développer un modèle d'apprentissage performant, il sera crucial d'exploiter de manière optimale les données limitées disponibles. Plusieurs indices montraient une instabilité dans la performance du modèle à travers les 5 étapes d'entraînement, comme la haute variation de la valeur de seuil optimisant le score F1, ou encore la volatilité de la courbe de fonction de perte.

Il est aussi important de mentionner que le modèle n'a pas été optimisé sur plusieurs hyperparamètres afin de diminuer le risque de surapprentissage. Une autre raison à cela est que

l'entraînement du modèle prenait en moyenne 20 heures sur une NVIDIA V100 16GB et utilisait la quasi-totalité de la mémoire GPU. Bien qu'une optimisation sur différents hyperparamètres puisse améliorer la performance du modèle, il est probable que la limitation majeure soit causée par la quantité restreinte de données ainsi que par les variations discutées plus tôt. Il est difficile d'évaluer le vrai potentiel du modèle en termes de performances généralisables.

Le choix d'utiliser VGG-19 comme architecture d'extraction de variables spatiales était basé sur un modèle libre accès qui utilisait déjà celui-ci (110). Initialement, nous avons considéré explorer les différences de performances entre deux modèles, soit VGG-19 et EfficientNetV2-L. Ce dernier est un modèle plus récent que VGG-19 et utilise des techniques non utilisées par VGG-19, une complexité paramétrique inférieure, et une meilleure performance sur les compétitions de classifications d'images (110). Pour des raisons de faisabilité, de nécessité de limiter les comparaisons multiples compte tenu de notre taille d'échantillon et de la grande complexité de l'étape de pré-traitement, nous avons opté pour un modèle mieux publié (VGG19). C'est dans une optique similaire que nous avons décidé de ne pas faire d'optimisation sur les hyperparamètres et d'utiliser un modèle déjà supporté dans la littérature (110). Un transfert d'apprentissage sur un modèle entraîné uniquement sur des images CTP aurait par ailleurs possiblement été un atout, mais ce type de modèle n'est à ce jour pas disponible.

Finalement, une limitation supplémentaire de ce projet est le manque d'analyse d'explicabilité sur les variables apprises par le modèle qui lui permet de faire ses prédictions. La visualisation des variables apprises par les réseaux de neurones est courante grâce aux cartes de saillances pour des images 2D (119, 120), voir 3D, mais reste une approche non validée dans le contexte de réseaux de neurones en 4 dimensions. Cela est assez limitant en termes de généralisation et d'interprétabilité des résultats puisque nous sommes dans l'incapacité de conclure si le modèle utilise des variables réellement explicatives, comme le signal de perfusion. Sans cette information, il n'est pas possible de garantir que le modèle a appris des variables spécifiques au DN.

## 4.4 Implications et directions futures

Plusieurs étapes peuvent être envisagées afin d'améliorer ce travail ou les approches futures concernant cette question de recherche.

Une grande partie du travail a été consacrée à la réorganisation de la structure des répertoires contenant les fichiers DICOM car la structure de stockage des images variaient d'un site à un autre, complexifiant donc l'automatisation. Une standardisation de la méthodologie de stockage des images pourrait grandement aider cette étape.

Bien que le protocole d'INDEX-CTP indique une section minimale à scanner (8 cm d'une région du cerveau), les scans ne couvraient pas forcément les mêmes régions à travers les sites et les patients et ne commençaient pas à la même région anatomique. Il serait potentiellement avantageux de sélectionner un point de repère dans les images afin de les recadrer manuellement, mais cela implique une étape de pré-traitement manuelle chronophage et nécessitant une certaine expertise en radiologie. Additionnellement, une uniformisation de la région cérébrale où le scan commence pourrait augmenter la comparabilité des images.

Le RNC responsable de l'extraction de variables spatiales dans notre modèle est VGG19, qui est un réseau très profond et qui est constitué de millions de paramètres. Il serait intéressant de tester la capacité d'autres RNC à extraire des variables spatiales, comme EfficientNetV2 qui a nettement moins de paramètres que VGG19 mais atteint des performances remarquables lors de compétitions de classification d'images (70). Un modèle avec moins de paramètres pourrait potentiellement atteindre de meilleures performances, surtout dû au fait que la taille de notre échantillon était relativement basse.

Afin d'améliorer la capacité de généralisation du modèle, il serait bénéfique de travailler avec plus de données. Bien que la nature et la qualité des données provenant de la cohorte d'INDEX-CTP soient rares, il n'est pas impossible que d'autres études similaires fassent surface à l'avenir et permettent d'augmenter la quantité de données disponibles. Autrement, l'utilisation de

techniques comme l'augmentation artificielle de données pourrait permettre de combler cette limitation (110, 121).

Il serait aussi intéressant d'intégrer d'autres variables au classificateur, comme des variables d'examens cliniques, ou encore des résultats de laboratoires. Récemment, un travail a démontré la faisabilité d'utiliser ces variables dans un modèle d'AP afin de détecter des donneurs d'organes potentiels (122). La combinaison entre ce modèle avec celui de notre travail pourrait potentiellement améliorer les capacités prédictives de DN. Ce concept s'appelle l'apprentissage par ensemble (*Ensemble Learning*) et permet de combiner les prédictions de plusieurs modèles afin d'arriver à une prédiction plus robuste (93, 123).

Dans des travaux futurs, il serait intéressant de comparer la performance du modèle sur ces images brutes à un autre modèle entraîné sur des cartes de perfusions. Il est possible qu'un modèle fonctionne mieux sur des cartes de perfusions puisqu'elles résument le mouvement de l'agent de contraste à travers le temps en plusieurs indicateurs, et les images sont calibrées de façon plus précise qu'avec l'étape de prétraitement utilisée dans ce projet.

L'erreur humaine lors du diagnostic de DN ou lors de l'interprétation du scan par le radiologiste pourrait être une source intéressante d'amélioration. Il serait par exemple pertinent d'explorer les prédictions du modèle lorsqu'il y a une situation de désaccord en les cliniciens entre eux, ou entre les cliniciens et l'interprétation du scan par le radiologiste. Afin d'augmenter l'utilité du modèle dans ces situations de désaccord et augmenter son interprétabilité, il serait avantageux d'utiliser des cartes de saillance. Ces cartes permettent de visualiser les régions d'une image qui ont été sollicités par le modèle afin d'établir une prédiction (119, 120). En ayant accès aux variables apprises par le modèle, on pourrait même concevoir qu'un modèle puisse faire découvrir au radiologiste des régions d'intérêts au DN. En faisant cela, un médecin ou radiologiste pourrait éventuellement analyser les sections d'importances retenue par le modèle afin d'informer son diagnostic.

Dans le contexte du don d'organes, le temps ainsi que le rapport offre/demande sont deux composantes précieuses qui bien souvent dictent si un patient va pouvoir être sauvé (18, 19). Le traitement automatique de l'image est de plus en plus utilisé dans le contexte médical et en

radiologie (55, 92, 107), il était donc intéressant d'explorer si cet outil pouvait être utile à des fins de classification de DN. Étant donné la complexité et le temps nécessaire au diagnostic de DN, il est envisageable qu'un processus de dépistage automatisé puisse accélérer la détection de donneur potentiel dans un contexte de soin intensif. Un tel outil aurait aussi le potentiel d'assister les radiologistes et cliniciens lors d'un manque de consensus sur l'état neurologique d'un patient. Il est aussi envisageable qu'il puisse servir de dépistage dans des hôpitaux avec moins d'expertise ou de ressources.

## 4.5 Conclusion

Le diagnostic de décès neurologique est une étape complexe cruciale pour la protection du patient mais aussi pour le don d'organes qui a souvent pour prérequis la déclaration de DN du patient. L'étape de diagnostic de DN est méthodique et rigoureuse, mais n'est malheureusement pas parfaite.

Dans ce mémoire, nous avons examiné la capacité d'un modèle de traitement d'images automatique à classifier des images de tomodensitométrie par perfusion provenant de patients souffrant de lésions cérébrales sévères, en distinguant les cas de survie et de décès neurologiques. Malgré les contraintes mentionnées dans ce travail, notre modèle a démontré des performances se rapprochant de celles d'un expert humain dans le domaine médical.

Dans cette perspective, il convient de souligner l'importance des travaux futurs visant à étudier les méthodes les plus appropriées pour gérer les données médicales rares, déséquilibrées et issues de la pratique clinique quotidienne. De plus, il serait pertinent d'explorer davantage les techniques d'augmentation de données, d'apprentissage par transfert et de *Few-shot Learning* afin d'améliorer la précision et la robustesse de notre modèle face à des contextes de données limitées.

En outre, il serait intéressant d'évaluer la capacité de notre modèle à généraliser ses performances à d'autres types de lésions cérébrales ou d'affections neurologiques. Cela permettrait de valider son utilité et son applicabilité dans un contexte clinique plus large, contribuant ainsi à améliorer la prise en charge des patients et à orienter de manière plus précise les décisions thérapeutiques.

Enfin, une collaboration étroite avec les professionnels de la santé et une intégration réussie des modèles d'apprentissage automatique dans les systèmes de santé existants sont cruciales pour garantir l'adoption et l'impact positif de ces technologies sur les soins aux patients.



## 5: Références bibliographiques

1. McMahan J. The metaphysics of brain death. *Bioethics*. 1995;9(2):91-126.
2. Greer DM, Shemie SD, Lewis A, Torrance S, Varelas P, Goldenberg FD, et al. Determination of Brain Death/Death by Neurologic Criteria: The World Brain Death Project. *JAMA*. 2020;324(11):1078-97.
3. Wunsch H. The outbreak that invented intensive care. *Nature*. 2020.
4. Mollaret P, Goulon M. [The depassed coma (preliminary memoir)]. *Rev Neurol (Paris)*. 1959;101:3-15.
5. A definition of irreversible coma. Report of the Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death. *JAMA*. 1968;205(6):337-40.
6. (1981) PsCftSoEPiMaBaBR. Defining death: a report on the medical, legal and ethical issues in the determination of death.: United States; 1981.
7. Cole DJ. The reversibility of death. *J Med Ethics*. 1992;18(1):26-30; discussion 1-3.
8. Lamb D. Reversibility and death: a reply to David J. Cole. *J Med Ethics*. 1992;18(1):31-3.
9. Lizza JP. Defining death for persons and human organisms. *Theor Med Bioeth*. 1999;20(5):439-53.
10. den Hartogh G. When are you dead enough to be a donor? Can any feasible protocol for the determination of death on circulatory criteria respect the dead donor rule? *Theor Med Bioeth*. 2019;40(4):299-319.
11. Pallis C. ABC of brain stem death. Reappraising death. *Br Med J (Clin Res Ed)*. 1982;285(6352):1409-12.
12. Byrne PA, O'Reilly S, Quay PM. Brain death--an opposing viewpoint. *Jama*. 1979;242(18):1985-90.
13. Practice parameters for determining brain death in adults (summary statement). The Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*. 1995;45(5):1012-4.
14. Lewis A, Bakkar A, Kreiger-Benson E, Kumpfbeck A, Liebman J, Shemie SD, et al. Determination of death by neurologic criteria around the world. *Neurology*. 2020;95(3):e299-e309.
15. Wahlster S, Wijdicks EF, Patel PV, Greer DM, Hemphill JC, 3rd, Carone M, et al. Brain death declaration: Practices and perceptions worldwide. *Neurology*. 2015;84(18):1870-9.
16. Greer DM, Wang HH, Robinson JD, Varelas PN, Henderson GV, Wijdicks EF. Variability of Brain Death Policies in the United States. *JAMA Neurol*. 2016;73(2):213-8.
17. Munakomi S, Al Khalili Y. Brainstem Death. *StatPearls*. Treasure Island (FL)2022.
18. Da Silva IR, Frontera JA. Worldwide barriers to organ donation. *JAMA Neurol*. 2015;72(1):112-8.
19. Abouna GM. Organ shortage crisis: problems and possible solutions. *Transplant Proc*. 2008;40(1):34-8.
20. Zuniga-Fajuri A. Increasing organ donation by presumed consent and allocation priority: Chile. *Bull World Health Organ*. 2015;93(3):199-202.

21. Mantecchini L, Paganelli F, Morabito V, Ricci A, Peritore D, Trapani S, et al. Transportation of Organs by Air: Safety, Quality, and Sustainability Criteria. *Transplant Proc.* 2016;48(2):304-8.
22. Transplantation GOoDa. Those 2000-2020 data are based on the Global Observatory on Donation and Transplantation (GODT) data, produced by the WHO-ONT collaboration: World Health Organization; [Available from: <http://www.transplant-observatory.org/summary/#>.
23. Taylor RM. Reexamining the definition and criteria of death. *Semin Neurol.* 1997;17(3):265-70.
24. Neiders I, Dranseika V. Minds, brains, and hearts: an empirical study on pluralism concerning death determination. *Monash Bioeth Rev.* 2020;38(1):35-48.
25. Truog RD, Miller FG. The dead donor rule and organ transplantation. *N Engl J Med.* 2008;359(7):674-5.
26. Dalle Ave AL, Sulmasy DP, Bernat JL. The ethical obligation of the dead donor rule. *Med Health Care Philos.* 2020;23(1):43-50.
27. Wijdicks EF. Brain death worldwide: accepted fact but no global consensus in diagnostic criteria. *Neurology.* 2002;58(1):20-5.
28. Kramer AH. Ancillary testing in brain death. *Semin Neurol.* 2015;35(2):125-38.
29. Plourde G, Briard JN, Shemie SD, Shankar JJS, Chasse M. Flow is not perfusion, and perfusion is not function: ancillary testing for the diagnosis of brain death. *Can J Anaesth.* 2021;68(7):953-61.
30. Heran MK, Heran NS, Shemie SD. A review of ancillary tests in evaluating brain death. *Can J Neurol Sci.* 2008;35(4):409-19.
31. Shemie SD, Hornby L, Baker A, Teitelbaum J, Torrance S, Young K, et al. International guideline development for the determination of death. *Intensive Care Med.* 2014;40(6):788-97.
32. Lewis A, Liebman J, Kreiger-Benson E, Kumpfbeck A, Bakkar A, Shemie SD, et al. Ancillary Testing for Determination of Death by Neurologic Criteria Around the World. *Neurocrit Care.* 2021;34(2):473-84.
33. Zampakis P, Panagiotopoulos V, Kalogeropoulou C, Karachaliou M, Aretha D, Sioulas N, et al. Computed tomography angiography scoring systems and the role of skull defects in the confirmation of brain death. *Sci Rep.* 2021;11(1):15081.
34. Combes JC, Chomel A, Ricolfi F, d'Athis P, Freysz M. Reliability of computed tomographic angiography in the diagnosis of brain death. *Transplant Proc.* 2007;39(1):16-20.
35. Shankar JJ, Banfield JC. Comments on Shemie et al.: international guideline development for the determination of death. *Intensive Care Med.* 2015;41(3):571.
36. Sawicki M, Solek-Pastuszka J, Chamier-Cieminska K, Walecka A, Walecki J, Bohatyrewicz R. Computed Tomography Perfusion is a Useful Adjunct to Computed Tomography Angiography in the Diagnosis of Brain Death. *Clin Neuroradiol.* 2019;29(1):101-8.
37. Patel PR DJO. CT Scan. Treasure Island (FL): StatPearls; 2023 [updated 2023 Jan 2. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK567796/>.
38. Bibb R, Eggbeer D, Paterson A. 2 - Medical imaging. In: Bibb R, Eggbeer D, Paterson A, editors. *Medical Modelling (Second Edition)*. Oxford: Woodhead Publishing; 2015. p. 7-34.
39. Kamalian S, Lev MH, Gupta R. Chapter 1 - Computed tomography imaging and angiography – principles. In: Masdeu JC, González RG, editors. *Handbook of Clinical Neurology*. 135: Elsevier; 2016. p. 3-20.

40. Christensen S, Lansberg MG. CT perfusion in acute stroke: Practical guidance for implementation in clinical practice. *J Cereb Blood Flow Metab.* 2019;39(9):1664-8.
41. Sabarudin A, Subramaniam C, Sun Z. Cerebral CT angiography and CT perfusion in acute stroke detection: a systematic review of diagnostic value. *Quant Imaging Med Surg.* 2014;4(4):282-90.
42. Konstas AA, Wintermark M, Lev MH. CT perfusion imaging in acute stroke. *Neuroimaging Clin N Am.* 2011;21(2):215-38, ix.
43. Lin L, Bivard A, Parsons MW. Perfusion patterns of ischemic stroke on computed tomography perfusion. *J Stroke.* 2013;15(3):164-73.
44. Kauw F, Heit JJ, Martin BW, van Ommen F, Kappelle LJ, Velthuis BK, et al. Computed Tomography Perfusion Data for Acute Ischemic Stroke Evaluation Using Rapid Software: Pitfalls of Automated Postprocessing. *J Comput Assist Tomogr.* 2020;44(1):75-7.
45. Shankar JJ, Vandorpe R. CT perfusion for confirmation of brain death. *AJNR Am J Neuroradiol.* 2013;34(6):1175-9.
46. Shankar JJS, Stewart-Perrin B, Quraishi AU, Bata I, Vandorpe R. Computed Tomography Perfusion Aids in the Prognostication of Comatose Postcardiac Arrest Patients. *Am J Cardiol.* 2018;121(7):874-8.
47. Wikipedia. Deep Learning [Available from: [https://en.wikipedia.org/wiki/Deep\\_learning](https://en.wikipedia.org/wiki/Deep_learning).
48. Wagner MW, Namdar K, Biswas A, Monah S, Khalvati F, Ertl-Wagner BB. Radiomics, machine learning, and artificial intelligence-what the neuroradiologist needs to know. *Neuroradiology.* 2021;63(12):1957-67.
49. McCarthy J, Minsky ML, Rochester N, Shannon CE. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine.* 2006;27(4):12.
50. Wang P. On Defining Artificial Intelligence. *Journal of Artificial General Intelligence.* 2019;10:1-37.
51. Kok J, Boers E, Kosters WA, Putten Pvd, Poel M. Artificial intelligence: definition, trends, techniques, and cases. *Knowledge for sustainable development: an insight into the Encyclopedia of life support systems.* 12002. p. 1095-107.
52. Samoili S, Lopez Cobo M, Gomez Gutierrez E, De Prato G, Martinez-Plumed F, Delipetrev B. AI WATCH. Defining Artificial Intelligence. 2020.
53. Kaplan A, Haenlein M. Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons.* 2019;62(1):15-25.
54. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science.* 2015;349(6245):255-60.
55. Rajendran P, Madheswaran M. Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm 2010 January 01, 2010:[arXiv:1001.3503 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2010arXiv1001.3503R>.
56. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436-44.
57. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. 1943. *Bull Math Biol.* 1990;52(1-2):99-115; discussion 73-97.
58. Geron A. *Hands-on machine learning with Scikit-Learn, Keras & TensorFlow.* 2nd ed. 1005 gravenstein highway north sebastopol, United States of America: O'Reilly Media, Inc.; 2019.

59. Wang M, Yang X, Qian Y, Lei Y, Cai J, Huan Z, et al. Adaptive Neural Network Structure Optimization Algorithm Based on Dynamic Nodes. *Curr Issues Mol Biol.* 2022;44(2):817-32.
60. Mishra A, Cha J, Kim S. Single Neuron for Solving XOR like Nonlinear Problems. *Comput Intell Neurosci.* 2022;2022:9097868.
61. David ER, James LM. Learning Internal Representations by Error Propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations: MIT Press;* 1987. p. 318-62.
62. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Yee Whye T, Mike T, editors. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; Proceedings of Machine Learning Research: PMLR;* 2010. p. 249--56.
63. He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification2015 February 01, 2015:[arXiv:1502.01852 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2015arXiv150201852H>.
64. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift2015 February 01, 2015:[arXiv:1502.03167 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2015arXiv150203167I>.
65. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE.* 1998;86(11):2278-324.
66. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Association for Computing Machinery.* 2017;60(6):84-90.
67. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition2014 September 01, 2014:[arXiv:1409.556 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2014arXiv1409.1556S>.
68. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition2015 December 01, 2015:[arXiv:1512.03385 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2015arXiv151203385H>.
69. Boesch G. Deep Residual Networks (ResNet, ResNet50) – 2022 Guide: Viso Suite; 2022 [Available from: <https://viso.ai/deep-learning/resnet-residual-neural-network/>].
70. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks2019 May 01, 2019:[arXiv:1905.11946 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2019arXiv190511946T>.
71. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning2016 February 01, 2016:[arXiv:1602.07261 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2016arXiv160207261S>.
72. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks2017 September 01, 2017:[arXiv:1709.01507 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2017arXiv170901507H>.
73. Canziani A, Paszke A, Culurciello E. An Analysis of Deep Neural Network Models for Practical Applications2016 May 01, 2016:[arXiv:1605.07678 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2016arXiv160507678C>.

74. Gudivada V, Apon A, Ding J. Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *International Journal on Advances in Software*. 2017;10:1-20.
75. Brownlee J. What is the Difference Between a Parameter and a Hyperparameter? : *Machine Learning Process*; 2017 [updated 17 June 2019. Available from: <https://machinelearningmastery.com/difference-between-a-parameter-and-a-hyperparameter/>.
76. Polyak BT. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*. 1964;4(5):1-17.
77. Nesterov Y, editor A method for unconstrained convex minimization problem with the rate of convergence  $o(1/k^2)$ 1983.
78. Duchi J, Hazam E, Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *The Journal of Machine Learning Research*. 2011;12:2121-59.
79. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization2014 December 01, 2014:[arXiv:1412.6980 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6980K>.
80. Senior A, Heigold G, Ranzato M, Yang K, editors. An empirical study of learning rates in deep neural networks for speech recognition. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing; 2013 26-31 May 2013.
81. Masters D, Luschi C. Revisiting Small Batch Training for Deep Neural Networks2018 April 01, 2018:[arXiv:1804.07612 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2018arXiv180407612M>.
82. Hoffer E, Hubara I, Soudry D. Train longer, generalize better: closing the generalization gap in large batch training of neural networks2017 May 01, 2017:[arXiv:1705.08741 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2017arXiv170508741H>.
83. Goyal P, Dollár P, Girshick R, Noordhuis P, Wesolowski L, Kyrola A, et al. Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour2017 June 01, 2017:[arXiv:1706.02677 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2017arXiv170602677G>.
84. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*. 2014;15:1929-58.
85. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proc Natl Acad Sci U S A*. 2019;116(32):15849-54.
86. Rocks JW, Mehta P. Memorizing without overfitting: Bias, variance, and interpolation in overparameterized models. *Phys Rev Res*. 2022;4(1).
87. Brownlee J. A Gentle Introduction to k-fold Cross-Validation: *Machine Learning Process*; 2018 [updated August 3 2020. Available from: <https://machinelearningmastery.com/k-fold-cross-validation/>.
88. George EI, Brand TC, LaPorta A, Marescaux J, Satava RM. Origins of Robotic Surgery: From Skepticism to Standard of Care. *JSLs*. 2018;22(4).
89. Bhardwaj R, Nambiar AR, Dutta D, editors. A Study of Machine Learning in Healthcare. 2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); 2017 4-8 July 2017.

90. Mallappallil M, Sabu J, Gruessner A, Salifu M. A review of big data and medical research. *SAGE Open Med.* 2020;8:2050312120934839.
91. Cai L, Zhu Y. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal* 2015;14:2.
92. Razzak MI, Naz S, Zaib A. Deep Learning for Medical Image Processing: Overview, Challenges and Future 2017 April 01, 2017:[arXiv:1704.06825 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2017arXiv170406825R>.
93. Kundu R, Singh PK, Ferrara M, Ahmadian A, Sarkar R. ET-NET: an ensemble of transfer learning models for prediction of COVID-19 infection through chest CT-scan images. *Multimed Tools Appl.* 2022;81(1):31-50.
94. Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. *BMC Med Imaging.* 2022;22(1):69.
95. Charron O, Lallement A, Jarnet D, Noblet V, Clavier JB, Meyer P. Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. *Comput Biol Med.* 2018;95:43-54.
96. Lin K, Jie B, Dong P, Ding X, Bian W, Liu M. Convolutional Recurrent Neural Network for Dynamic Functional MRI Analysis and Brain Disease Identification. *Front Neurosci.* 2022;16:933660.
97. Kuo W, Hne C, Mukherjee P, Malik J, Yuh EL. Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. *Proc Natl Acad Sci U S A.* 2019;116(45):22737-45.
98. Lee JY, Kim JS, Kim TY, Kim YS. Detection and classification of intracranial haemorrhage on CT images using a novel deep-learning algorithm. *Scientific Reports.* 2020;10(1):20546.
99. Kumaravel P, Mohan S, Arivudaiyanambi J, Shajil N, Venkatakrishnan HN. A Simplified Framework for the Detection of Intracranial Hemorrhage in CT Brain Images Using Deep Learning. *Curr Med Imaging.* 2021;17(10):1226-36.
100. Lui YW, Tang ER, Allmendinger AM, Spektor V. Evaluation of CT Perfusion in the Setting of Cerebral Ischemia: Patterns and Pitfalls. *American Journal of Neuroradiology.* 2010;31(9):1552.
101. Solovyev R, Kalinin AA, Gabruseva T. 3D convolutional neural networks for stalled brain capillary detection. *Comput Biol Med.* 2022;141:105089.
102. Ker J, Singh SP, Bai Y, Rao J, Lim T, Wang L. Image Thresholding Improves 3-Dimensional Convolutional Neural Network Diagnosis of Different Acute Brain Hemorrhages on Computed Tomography Scans. *Sensors (Basel).* 2019;19(9).
103. Cortes C, Jackel L, Chiang W-p. Limits on Learning Machine Accuracy Imposed by Data Quality. 2000.
104. Sanchez-Pinto LN, Luo Y, Churpek MM. Big Data and Data Science in Critical Care. *Chest.* 2018;154(5):1239-48.
105. Shemie SD, Doig C, Dickens B, Byrne P, Wheelock B, Rocker G, et al. Severe brain injury to neurological determination of death: Canadian forum recommendations. *CMAJ.* 2006;174(6):S1-13.

106. Lev MH, Gonzalez RG. 17 - CT Angiography and CT Perfusion Imaging. In: Toga AW, Mazziotta JC, editors. Brain Mapping: The Methods (Second Edition). San Diego: Academic Press; 2002. p. 427-84.
107. He K, Gan C, Li Z, Rekik I, Yin Z, Ji W, et al. Transformers in Medical Image Analysis: A Review 2022 February 01, 2022:[arXiv:2202.12165 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2022arXiv220212165H>.
108. Jang J, Hwang D, editors. M3T: three-dimensional Medical image classifier using Multi-plane and Multi-slice Transformer. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 18-24 June 2022.
109. Khan A, Sohail A, Zahoor A, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. Artificial Intelligence Review. 2020;53(8):5455-516.
110. Mittermeier A, Reidler P, Fabritius MP, Schachtner B, Wesp P, Ertl-Wagner B, et al. End-to-End Deep Learning Approach for Perfusion Data: A Proof-of-Concept Study to Classify Core Volume in Stroke CT. Diagnostics (Basel). 2022;12(5).
111. Stewart M. Simple Guide to Hyperparameter Tuning in Neural Networks Towards Data Science 2019 [Available from: <https://towardsdatascience.com/simple-guide-to-hyperparameter-tuning-in-neural-networks-3fe03dad8594>].
112. Ramos D, Franco-Pedroso J, Lozano-Diez A, Gonzalez-Rodriguez J. Deconstructing Cross-Entropy for Probabilistic Binary Classifiers. Entropy (Basel). 2018;20(3).
113. Information ClfH. Annual Statistics on Organ
- Replacement in Canada: Dialysis, Transplantation and Donation, 2010 to 2019.: Ottawa, ON: CIHI; 2020.; 2020 [Available from: [https://secure.cihi.ca/free\\_products/corr-dialysis-transplantation-donation-2010-2019-snapshot-en.pdf](https://secure.cihi.ca/free_products/corr-dialysis-transplantation-donation-2010-2019-snapshot-en.pdf)].
114. Lewis A, Koukoura A, Tsianos GI, Gargavanis AA, Nielsen AA, Vassiliadis E. Organ donation in the US and Europe: The supply vs demand imbalance. Transplant Rev (Orlando). 2021;35(2):100585.
115. Bastani B. The present and future of transplant organ shortage: some potential remedies. J Nephrol. 2020;33(2):277-88.
116. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. Jama. 2018;319(13):1317-8.
117. Hinton G. Deep Learning-A Technology With the Potential to Transform Health Care. Jama. 2018;320(11):1101-2.
118. Chassé M, Neves Briard J, Yu M, L PC, S WE, D'Aragon F, et al. Clinical evaluation and ancillary testing for the diagnosis of death by neurologic criteria: a cross-sectional survey of Canadian intensivists. Can J Anaesth. 2022;69(3):353-63.
119. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization 2016 October 01, 2016:[arXiv:1610.02391 p.]. Available from: <https://ui.adsabs.harvard.edu/abs/2016arXiv161002391S>.
120. Zhang Y, Hong D, McClement D, Oladosu O, Pridham G, Slaney G. Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. J Neurosci Methods. 2021;353:109098.

121. Chlap P, Min H, Vandenberg N, Dowling J, Holloway L, Haworth A. A review of medical image data augmentation techniques for deep learning applications. *J Med Imaging Radiat Oncol.* 2021;65(5):545-63.
122. Sauthier N, Bouchakri R, Carrier FM, Sauthier M, Mullie L-A, Cardinal H, et al. Automated screening of potential organ donors using a temporal machine learning model. *Scientific Reports.* 2023;13(1):8459.
123. Nguyen DK, Lan CH, Chan CL. Deep Ensemble Learning Approaches in Healthcare to Enhance the Prediction and Diagnosing Performance: The Workflows, Deployments, and Surveys on the Statistical, Image-Based, and Sequential Datasets. *Int J Environ Res Public Health.* 2021;18(20).