

الجمهورية الجزائرية الديمقراطية الشعبية

وزارة التعليم العالي والبحث العلمي

UNIVERSITÉ BADJI MOKHTAR - ANNABA
BADJI MOKHTAR – ANNABA UNIVERSITY



جامعة باجي مختار – عنابة

Faculté : Technologie

Département : Electronique

Domaine : Sciences et Technologies

Filière : Télécommunications

Spécialité : Systèmes des télécommunications – Réseaux et télécommunications

Mémoire

Présenté en vue de l'obtention du Diplôme de Master

Thème :

**Elaboration d'un système d'aide au diagnostic du cancer
basé sur l'apprentissage profond**

Présenté par : ABBASSI Imen

SOUIBA AH Hadil

Encadrant : HAFS Toufik

MCA

Université BADJI Mokhtar - ANNABA

Jury de Soutenance :

ZERMI Narima	Grade	Université BADJI Mokhtar - ANNABA	Président
HAFS Toufik	MCA	Université BADJI Mokhtar - ANNABA	Encadrant
NEILI Zakaria	Grade	Université BADJI Mokhtar - ANNABA	Examineur

Année Universitaire : 2022/2023

Remerciements

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui on voudrait témoigner toute notre gratitude.

On voudrait tout d'abord adresser toute notre reconnaissance à notre encadreur, Monsieur Toufik HAFS, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion.

On remercie nos très chères mères, qui nous ont entourées d'amour, d'affection et qui nous ont toujours soutenues et guidées vers le bon chemin, que dieu les protèges.

On remercie nos pères, pour leurs soutiens morales et physiques durant tout notre parcours, que dieu les protèges.

Je tiens à remercier mes grands-parents pour tout l'amour qui m'ont données, et je souhaite que les grands-parents de ma très chère amie et binôme Hadil reposent en paix.

On remercie également toute personne qui était derrière notre réussite et nous soutenait par tous les moyens.

Sommaire

Sommaire	i
Résumé.....	iii
Abstract	iii
ملخص.....	iv
Liste des figures	v
Liste des tableaux	vi
Introduction générale.....	1
Chapitre 01	3
1. Introduction	3
2. Description du Cancer du cerveau	3
3. Deep Learning.....	4
4. Réseau de neurones convolutifs : CNN ou ConvNet	5
5. Les modèles existants du CNN	6
a) Alex Net	6
b) Res Net.....	8
c) GoogLe Net	8
6. Conclusion	10
Chapitre 02	11
1. Introduction	11
2. Modèle proposé.....	11
3. Les couches du modèle	13
a) Couche image input	13
b) Couche de convolution 1	13
c) Couche batch normalisation 1	14
d) Couche correction Relu 1	15
e) Couche max pooling 1.....	15
f) Couche convolution 2	16
g) Couche max pooling 2.....	16
h) Couche convolution 3	16
i) Couche entièrement connectée FC.....	16
j) Couche Softmax	17
k) Couche de sortie : Classification Output	17
4. Les options d'apprentissage Sgdm et Adam	17
a) Option Sgdm: Stochastic Gradient Descent with Momentum	17

b) Option Adam : Adaptive Moment Estimation	18
5. Conclusion	21
Chapitre 03	22
1. Introduction	22
2. Description base de données : IRM – Scanner	22
3. Performance metrics	23
a) Application du modèle pour la base de données	23
4. Résultats obtenus	24
a) Test 01	24
b) Test 02	25
c) Test 03	25
d) Test 04	26
e) Test 05	26
f) Test 06	27
g) Test 07	27
h) Test 08	28
i) Comparaison des résultats entre les deux options d'apprentissage	28
j) Calcul des performances metrics	29
k) Comparaison du modèle proposé avec les modèles prédéfinis	29
5. Conclusion	29
Conclusion générale	31
Référence bibliographique	32
Webographie	32

Résumé

L'utilisation du deep learning s'est répandue dans de nombreux domaines à travers le monde et continue de se développer rapidement, il est de plus en plus utilisé dans le domaine de la santé et de la médecine comme le diagnostic du cancer, offrant de nouvelles opportunités pour améliorer la précision et l'efficacité des méthodes de détection et de classification des tumeurs.

Les algorithmes de deep learning peuvent être formés pour détecter des motifs complexes dans les données médicales, aidant ainsi les médecins à prendre des décisions plus précises et à améliorer les résultats des traitements.

L'objectif de ce projet de fin d'études est d'élaborer un système d'aide au diagnostic du cancer basé sur le deep learning et le réseau de neurones convolutifs CNN à partir des imageries médicales collectées.

Cela constitue un système fiable pour afficher les données sous forme de courbes et une matrice de confusion qui fait le tri des patients malades et normaux.

Mots clés : deep learning, cancer du cerveau, réseau de neurones convolutifs, imageries médicales, matrice de confusion.

Abstract

The use of deep learning has spread in many fields around the world and continues to grow rapidly, it is increasingly used in the field of health and medicine such as cancer diagnosis, providing new opportunities to improve the accuracy and efficiency of tumour detection and classification methods.

Deep learning algorithms can be trained to detect complex patterns in medical data, helping physicians make more accurate decisions and improve treatment outcomes.

The aim of this end-of-study project is to develop a cancer diagnostic support system based on deep learning and the CNN convolutive neural network based on collected medical imagery.

This provides a reliable system for displaying data in the form of curves and a confusion matrix that sorts out sick and normal patients.

Keywords: deep learning, brain cancer, convolutive neural network, medical imaging, confusion matrix.

ينتشر استخدام التعلم العميق في العديد من المجالات حول العالم ويستمر في النمو بسرعة، ويستخدم بشكل متزايد في مجال الصحة والطب مثل تشخيص السرطان، وتوفير فرص جديدة لتحسين دقة وكفاءة طرق الكشف عن الأورام وتصنيفها.

يمكن تدريب خوارزميات التعلم العميق على اكتشاف الأنماط المعقدة في البيانات الطبية، ومساعدة الأطباء على اتخاذ قرارات أكثر دقة وتحسين نتائج العلاج.

الهدف من مشروع نهاية الدراسة هذا هو تطوير نظام دعم تشخيص السرطان بناءً على التعلم العميق والشبكة العصبية التلافيفية لشبكة CNN بناءً على الصور الطبية المجمعة

يوفر هذا نظامًا موثوقًا لعرض البيانات على شكل منحنيات ومصفوفة ارتباط تفرز المرضى والمرضى العاديين.

الكلمات الرئيسية: التعلم العميق، سرطان الدماغ، الشبكة العصبية التلافيفية، التصوير الطبي، مصفوفة الارتباك.

Liste des figures

Figure 1 : Cancer du cerveau [1]	4
Figure 2 : Architecture d'Alex Net.....	7
Figure 3 : Architecture du Resnet50 [2].....	8
Figure 4 : Architecture du Google Net [3].....	10
Figure 5 : Architecture	12
Figure 6 : Allure de la fonction ReLu [4].....	15
Figure 7 : Fonctionnement du mini batch [5]	20
Figure 8 : Architecture du modèle proposé	23
Figure 9 : Courbe d'évolution 1 avec l'option Adam	24
Figure 10 : Matrice de confusion	24
Figure 11 : Courbe d'évolution 2 avec l'option Sgdm.....	25
Figure 12 : Courbe d'évolution 3 avec l'option Adam	25
Figure 13 : Courbe d'évolution 4 avec l'option Sgdm.....	26
Figure 14 : Courbe d'évolution 5 avec l'option Adam	26
Figure 15 : Courbe d'évolution 6 avec l'option Sgdm.....	27
Figure 16 : Courbe d'évolution 7 avec l'option Adam	27
Figure 17 : Courbe d'évolution 8 avec l'option Sgdm.....	28

Liste des tableaux

Tableau 1 : 1ère étape de convolution	14
Tableau 2 : Vecteur de caractéristique après l'application de convolution	14
Tableau 3 : L'application de max pooling	16
Tableau 4 : Comparaison entre les deux options d'apprentissage.....	28
Tableau 5 : Comparaison des performances metrics entre les nombre d'epochs utilisés.....	29
Tableau 6 : Comparaison des modèles prédéfinis.....	29

Introduction générale

Le cancer est un groupe de maladies caractérisées par la croissance incontrôlée et la propagation de cellules anormales. Ces cellules anormales peuvent former des tumeurs et peuvent envahir les tissus voisins ou se propager à d'autres parties du corps par le sang ou le système lymphatique.

Les causes exactes du cancer ne sont pas bien comprises, mais on pense généralement qu'elles résultent d'une combinaison de facteurs génétiques et environnementaux. Voici quelques-unes des causes connues ou présumées du cancer : tabac, alcool, infections, radiations UV, toxines environnementales, mauvaise alimentation et les mutations génétiques héréditaires.

Il est important de noter que de nombreuses personnes qui sont exposées à ces facteurs de risque ne développent pas de cancer, et inversement, certaines personnes qui développent un cancer n'ont aucun facteur de risque connu. Le développement du cancer est un processus complexe et multifactoriel.

Dans le cancer, on différencie deux types de tumeurs : bénigne et maligne. Une tumeur bénigne est une croissance de cellules anormales qui ne se propage pas aux tissus environnants ou d'autres parties du corps.

Une tumeur maligne, également connue sous le nom de tumeur cancéreuse, est une croissance de cellules anormales, elles sont dangereuses et peuvent constituer une menace pour la vie. Elles peuvent former de nouvelles tumeurs (métastases) dans d'autres organes.

De ce fait, le diagnostic du cancer est une procédure complexe qui implique plusieurs étapes typiques telles que la biopsie et les examens d'imagerie.

Et malgré les avancées scientifiques, le cancer reste un défi important pour la santé publique, car de nombreux types de cancer sont toujours difficiles à traiter et peuvent récidiver après le traitement. C'est pourquoi la recherche sur le cancer continue d'être une priorité pour les scientifiques, les cliniciens et les décideurs du monde entier.

C'est dans cette approche que s'inscrit notre travail. L'objectif est de mettre à la disposition des médecins un outil de travail conçu pour comparer par imagerie médicale les tumeurs bénignes et malignes. Nous avons pris l'un des cancers les plus fréquents qui est le cancer du cerveau.

Dans ce mémoire nous présentons trois axes, à travers lesquels nous expliquons comment concevoir un appareil de surveillance :

Le 1^{er} axe présente la problématique générale et les concepts liés au thème.

Le 2^{ème} axe met en avant la méthodologie adoptée dans ce projet.

Le 3^{ème} axe présente le résultat du modèle proposé.

Enfin nous terminons notre mémoire par une conclusion générale.

Chapitre 01

1. Introduction

Dans ce chapitre on va expliquer brièvement les différentes dimensions auxquelles à trait notre travail de projet de fin d'études.

Elles sont scindées en deux axes : l'axe médical qui est la description du type de cancer qu'on veut diagnostiquer, et l'axe technique qui est basé sur le deep learning et le réseau de neurones convolutifs.

2. Description du Cancer du cerveau

Le cancer du cerveau est une maladie dans laquelle des cellules anormales se multiplient de manière incontrôlable dans le cerveau. Il existe différents types de cancers du cerveau, notamment les tumeurs cérébrales primaires qui commencent dans le cerveau lui-même, et les tumeurs cérébrales métastatiques qui se propagent dans le cerveau à partir d'autres parties du corps.

Les symptômes du cancer du cerveau peuvent varier en fonction de l'emplacement de la tumeur, mais ils peuvent inclure des maux de tête, des troubles de la vision, des convulsions, des nausées et des vomissements, des changements de personnalité et des problèmes de coordination. Le diagnostic du cancer du cerveau implique généralement des examens d'imagerie tels que des IRM ou des scanners, ainsi qu'une biopsie pour déterminer le type de tumeur.

L'évolution du cancer du cerveau dépend du type de tumeur, de sa localisation, de sa taille, de son stade et de nombreux autres facteurs. Les cancers du cerveau peuvent être classés en fonction de leur agressivité et de leur potentiel de croissance, ainsi que de leur réponse aux traitements.

Les tumeurs bénignes du cerveau sont généralement moins agressives et ont tendance à croître plus lentement que les tumeurs malignes. Cependant, même les tumeurs bénignes peuvent être dangereuses si elles se développent dans des zones critiques du cerveau, telles que les régions responsables de la fonction motrice, sensorielle ou cognitive.

Les tumeurs malignes, telles que le glioblastome multiforme, sont considérées comme plus agressives et peuvent se propager rapidement dans d'autres parties du cerveau. Leur croissance rapide peut également rendre le traitement plus difficile et moins efficace.

Les options de traitement peuvent inclure la chirurgie, la radiothérapie et la chimiothérapie. Dans certains cas, une combinaison de ces traitements peut être utilisée. Le pronostic pour le cancer du cerveau dépend également de nombreux facteurs, notamment le stade de la maladie, l'âge et l'état de santé général du patient.

Après avoir expliqué le type de cancer qu'on va diagnostiquer, les parties suivantes vont développer les aspects techniques du travail. ¹

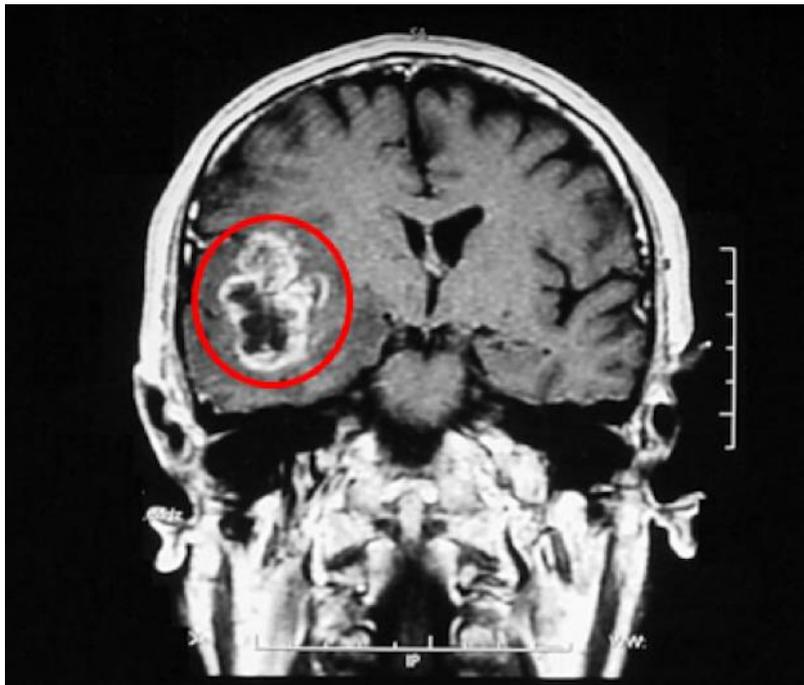


Figure 1 : Cancer du cerveau [1]

3. Deep Learning

L'apprentissage profond est un sous-domaine de l'apprentissage automatique qui utilise des réseaux neuronaux artificiels pour modéliser et résoudre des problèmes complexes. Les réseaux sont composés de nœuds interconnectés, qui sont inspirés par la structure et la fonction des neurones dans le cerveau humain.

¹ Pr SANSON, 2023, Institut du cerveau, 10 mai 2023, <https://institutducerveau-icm.org/fr/tumeurs-cerebrales/>

En apprentissage profond, de multiples couches de nœuds interconnectés sont utilisées pour traiter et analyser de grandes quantités de données. Les couches de nœuds travaillent ensemble pour identifier les modèles et les relations dans les données, ce qui permet au modèle de faire des prédictions ou de prendre des décisions basées sur les données qu'il reçoit.

Il a été utilisé pour résoudre un large éventail de problèmes, y compris la reconnaissance de l'image et de la parole, le traitement du langage naturel et la conduite autonome. Elle s'est avérée particulièrement efficace dans les tâches où les techniques traditionnelles d'apprentissage automatique ont connu des difficultés, comme dans les ensembles de données complexes et de grande dimension.

Les algorithmes d'apprentissage profond nécessitent de grandes quantités de données de formation et de puissance de calcul, mais les avancées en matière de matériel et la disponibilité de grands ensembles de données ont permis de former des modèles d'apprentissage profond qui peuvent surpasser les techniques traditionnelles d'apprentissage automatique. En conséquence, l'apprentissage profond devient un outil de plus en plus important pour résoudre des problèmes complexes dans un large éventail d'industries.²

4. Réseau de neurones convolutifs : CNN ou ConvNet

Un réseau de neurones convolutionnels (ConvNet/CNN) est un type d'architecture de réseau de neurones d'apprentissage profond spécialement conçu pour traiter et analyser les données d'images. ConvNets est inspiré par la structure du cortex visuel dans le cerveau humain, qui est conçu pour identifier les modèles et les caractéristiques dans l'information visuelle.

Dans un CNN, l'image d'entrée est traitée par plusieurs couches de nœuds interconnectés, dont chacun effectue un ensemble d'opérations mathématiques connues sous le nom de convolution. Les couches convolutionnelles extraient de plus en plus de caractéristiques complexes de l'image d'entrée, telles que les bords, les formes et les textures. Les caractéristiques extraites sont ensuite traitées par des couches entièrement connectées, qui

² Deep Learning ou Apprentissage Profond : qu'est-ce que c'est ?, 03 mars 2023, <https://datascientest.com/deep-learning-definition#:~:text=Qu'est%2Dce%20que%20le,des%20r%C3%A9seaux%20de%20neurones%20artificielles.>

combinent les caractéristiques pour faire une prédiction ou une décision finale sur l'image d'entrée.

CNN est particulièrement efficace pour des tâches telles que la classification des images, la détection d'objets et la segmentation, et a été utilisé pour obtenir des performances de pointe sur un large éventail de repères de vision par ordinateur. Ils ont également été appliqués à d'autres domaines, tels que le traitement du langage naturel et la reconnaissance de la parole.

Les systèmes ConvNets exigent beaucoup de calculs, mais les progrès du matériel et la disponibilité de grands ensembles de données ont permis de former des modèles avec des milliards de paramètres. Ces modèles peuvent apprendre à identifier et classer des objets dans des images avec une grande précision.³

5. Les modèles existants du CNN

a) Alex Net

AlexNet est une architecture de réseau neuronal convolutif (CNN) qui a joué un rôle crucial dans l'avancement du domaine de la vision par ordinateur et de l'apprentissage profond. Il a été développé par Alex Krizhevsky, Ilya Sutskever et Geoffrey Hinton en 2012 et a remporté l'ImageNet Large Scale Visual Recognition Challenge (ILSVRC) la même année, surpassant considérablement d'autres méthodes à l'époque.

Voici quelques caractéristiques et contributions clés d'AlexNet :

Architecture : AlexNet se compose de huit couches, dont cinq couches convolutionnelles et trois couches entièrement connectées. Il était plus profond et avait plus de paramètres que les modèles CNN précédents, ce qui lui a permis d'apprendre des fonctionnalités plus complexes.

Activation ReLU : AlexNet a été l'un des premiers modèles à utiliser la fonction d'activation de l'unité linéaire rectifiée (ReLU) au lieu de la fonction sigmoïde traditionnelle. ReLU aide à atténuer le problème de gradient et accélère la formation.

³ Neili Zakaria, 2021, *Aide au Diagnostic et Identification des pathologies pulmonaires par l'analyse et traitement des signaux respiratoires*, Thèse de doctorat, Université BADJI Mokhtar – ANNABA, Algérie.

Chevauchement de la mise en commun : Les couches de mise en commun d'AlexNet ont fait un pas plus important, ce qui a entraîné le chevauchement des régions de mise en commun. Cette approche a permis de réduire les dimensions spatiales tout en conservant plus d'information spatiale. Augmentation des données : pour tenir compte de la taille limitée de l'ensemble de données d'ImageNet, AlexNet a utilisé des techniques d'augmentation des données comme le recadrage aléatoire et le retournement horizontal. Cela a permis d'éviter le débordement et d'améliorer la généralisation.

Utilisation du GPU : AlexNet a été conçu pour tirer parti de la puissance de calcul des unités de traitement graphique (GPU). Il a été l'un des premiers modèles d'apprentissage profond à tirer parti des GPU efficacement, réduisant considérablement le temps de formation.

Impact : Le succès d'AlexNet a démontré la puissance de l'apprentissage profond et des NCN dans les tâches de classification des images. Il a stimulé d'autres avancées dans le domaine et a motivé les chercheurs à explorer des architectures plus profondes et plus complexes.

Depuis le développement d'AlexNet, de nombreuses architectures CNN ultérieures, telles que VGGNet, GoogLeNet et ResNet, ont été développées, s'appuyant sur ses principes et repoussant les limites de la reconnaissance d'image et des tâches de vision par ordinateur.⁴

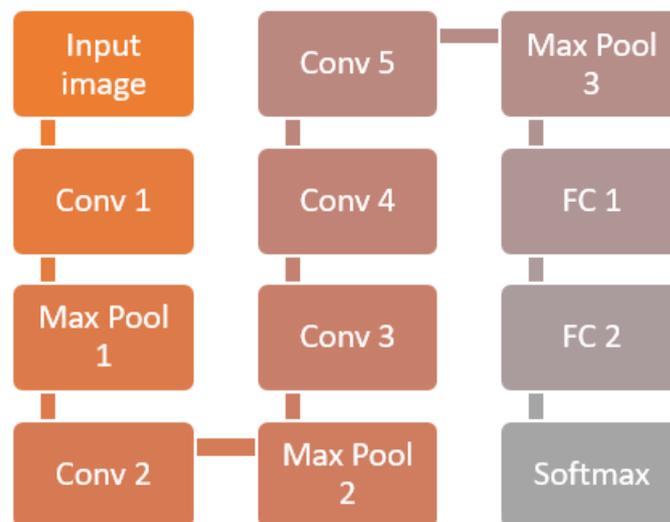


Figure 2 : Architecture d'Alex Net

⁴ Jerry Wei, AlexNet: The Architecture that Challenged CNNs, 03 mars 2023
<https://towardsdatascience.com/alexnet-the-architecture-that-challenged-cnns-e406d5297951>

b) Res Net

ResNet50 est une variante du modèle ResNet. Ce réseau de neurones possède 48 couches de convolution ainsi qu'une couche Max Pool et une couche average Pool. Il compte $3,8 \times 10^9$ opérations en virgule flottante. L'empilage de plus de couches nous donne généralement une meilleure précision. Ce modèle a été pré-entraîné sur plus d'1 million d'images de la base de données Image Net. Il peut classer les images dans 1000 catégories d'objets, comme le clavier, la souris, le crayon et de nombreux animaux : Par conséquent, le réseau a appris des représentations de caractéristiques riches pour un large éventail d'images ce réseau a une taille d'entrée d'image de $224 \times 224 \times 224$. La figure 3 illustre son architecture.

Grâce au cadre présenté par ResNet, il a été possible de former des réseaux neuronaux ultra profonds c'est à dire des réseaux qui peuvent contenir des centaines ou des milliers de couches Pour identifier les caractéristiques de bas, moyen et haut niveau des images tout en obtenant d'excellentes performances. Les ResNets ont été initialement appliqués à la tâche de reconnaissance d'image, mais comme il est mentionné dans le document, le cadre peut également être utilisé pour des tâches de vision non informatique afin d'obtenir une précision bien définie. ResNet a été la chose la plus intéressante qui soit arrivée au monde de la vision par ordinateur et de l'apprentissage profond.

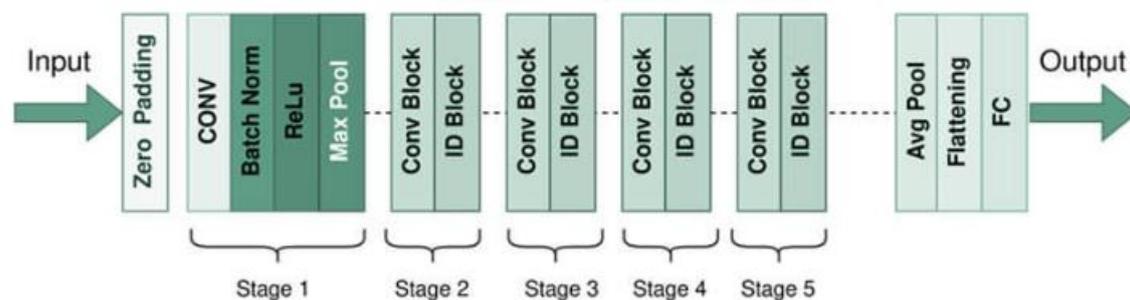


Figure 3 : Architecture du Resnet50 [2]

c) GoogLe Net

GoogLeNet, aussi connu sous le nom d'Inception v1, est une architecture de réseau neuronal convolutif profond introduite par les chercheurs de Google en 2014. Il a été développé pour relever les défis de la formation de réseaux très profonds tout en utilisant efficacement les ressources informatiques.

Voici quelques caractéristiques et contributions clés de GoogLeNet :

Modules de création : GoogLeNet a introduit le concept de "modules de création" dans son architecture. Un module Inception effectue plusieurs convolutions parallèles avec différentes tailles de filtres et concatène les cartes de fonctionnalités résultantes. Cela permet au réseau de capturer des fonctionnalités à plusieurs échelles et résolutions dans la même couche.

Réduction de la dimensionnalité : GoogLeNet utilise des convolutions 1x1 dans les modules Inception pour réduire la dimensionnalité des cartes de fonctionnalités d'entrée avant d'appliquer de plus grandes convolutions. Cela permet de réduire les coûts de calcul tout en maintenant le flux d'information.

Conception de réseau profonde : GoogLeNet est nettement plus profonde que les modèles précédents, malgré sa conception efficace. Il dispose de 22 couches au total, y compris une pile de multiples modules Inception et classificateurs auxiliaires à des couches intermédiaires pour encourager le flux de gradient et lutter contre le débordement.

Mise en commun moyenne mondiale : au lieu d'utiliser des couches entièrement connectées à la fin du réseau, GoogLeNet utilise la mise en commun moyenne mondiale. Cette opération réduit les dimensions spatiales des cartes de caractéristiques à une taille 1x1, suivie d'une couche softmax pour la classification. La mise en commun moyenne mondiale aide à réduire le débordement et le nombre de paramètres.

Efficacité de calcul : GoogLeNet visait à réduire le coût de calcul des réseaux de formation profonde. En utilisant la réduction de dimensionnalité et les convolutions parallèles, il a obtenu un meilleur équilibre entre la précision et l'efficacité de calcul par rapport aux architectures précédentes.

Impact : L'architecture de GoogLeNet a démontré l'efficacité des modules Inception pour capturer des fonctionnalités multi-échelles et inspiré les itérations ultérieures de la famille Inception, telles que Inception v2, v3 et v4. L'architecture GoogLeNet a également inspiré le développement d'autres architectures de réseau qui ont mis l'accent sur l'efficacité, telles que MobileNet et ShuffleNet.

GoogLeNet a apporté des contributions importantes au domaine de la vision par ordinateur, en particulier dans les domaines de la classification des images et de la reconnaissance des objets. Sa conception novatrice et son utilisation efficace des ressources continuent

d'influencer le développement de modèles d'apprentissage profond pour diverses tâches de vision par ordinateur.⁵

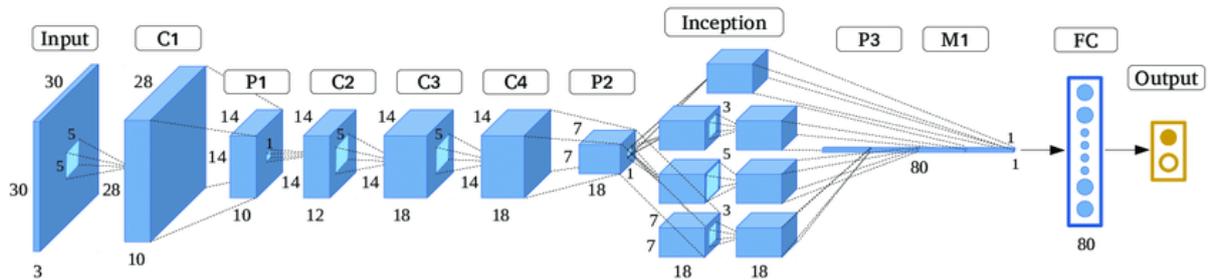


Figure 4 : Architecture du Google Net [3]

6. Conclusion

Après avoir développé brièvement les différents axes auxquelles à trait notre travail de projet de fin d'études, on va présenter le modèle proposé pour la détection du cancer du cerveau et l'expliquer étape par étape.

⁵ Abheer Bhandodker, Inception V1 Architecture Explained, 17 mai 2023
<https://medium.com/@abheerchrome/inception-v1-architecture-explained-454b2eb66baf>

Chapitre 02

1. Introduction

Au cours des dernières décennies, les avancées de la recherche médicale ont conduit à de nouvelles méthodes de diagnostic plus précoces et plus précises, ainsi qu'à des traitements plus ciblés et moins invasifs. Les thérapies ciblées, telles que les thérapies géniques et immunothérapeutiques, ont été développées pour traiter certains types de cancer de manière plus efficace et moins toxique.

En même temps, la prévention et la détection précoce du cancer ont également progressé, avec l'introduction de nouveaux tests de dépistage et la mise en œuvre de programmes de prévention ciblant les facteurs de risque modifiables.

L'objectif principal de ma nouvelle approche est de proposer un nouveau modèle modifié basé sur l'apprentissage profond pour obtenir un meilleur résultat et efficace tout en utilisant des techniques d'amélioration et d'identification du modèle le mieux adapté à ce modèle.

Notre mission est également de mettre à la disposition des médecins un outil de travail conçu pour diagnostiquer rapidement le cancer du cerveau.

2. Modèle proposé

Cette recherche a été faite pour choisir un nouveau modèle appelé : CancNet_AI_SH. Ce modèle sera approprié pour le diagnostic du cancer du cerveau. On procède ainsi dans le sillage des modèles existants de CNN d'où notre modèle sera inspiré donc à partir des modèles prédéfinis c'est à dire prêts à être utilisés, on essaiera de créer un modèle de telle sorte que le nombre de couches soit réduit tout en respectant les grandes couches importantes existantes au point de vue (CNN) qui sont conçues pour le traitement des images et spécifiquement des données de pixels. Le modèle proposé contiendra quinze couches qui vont être décrits aux dépens de leur fonction par la suite. L'un des principaux objectifs de cette recherche est d'obtenir une meilleure précision et des résultats prometteurs, qui peuvent être améliorés et mis en œuvre à l'échelle commerciale.

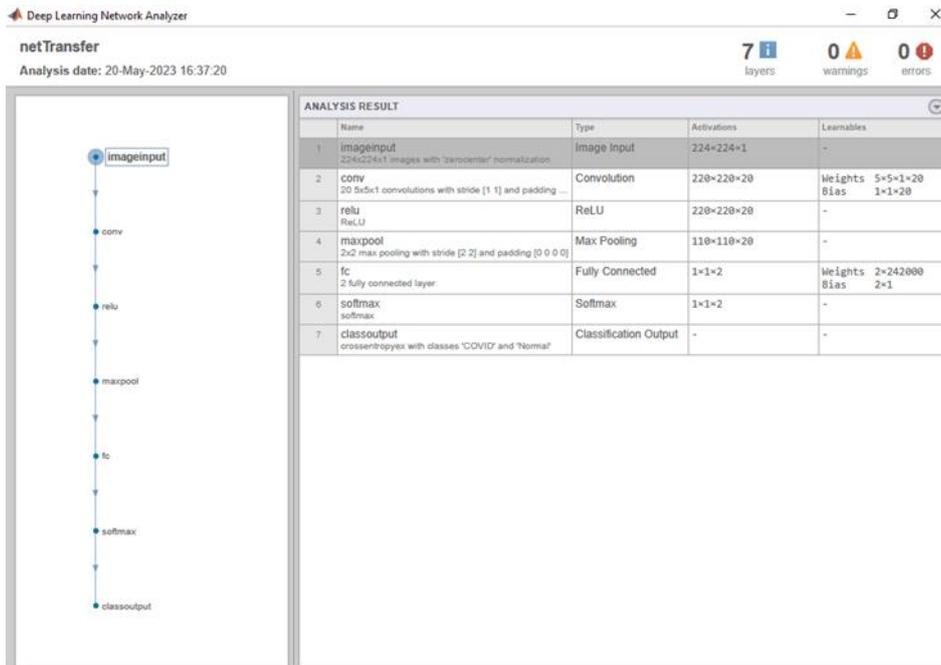


Figure 1 : Les couches du modèle proposé

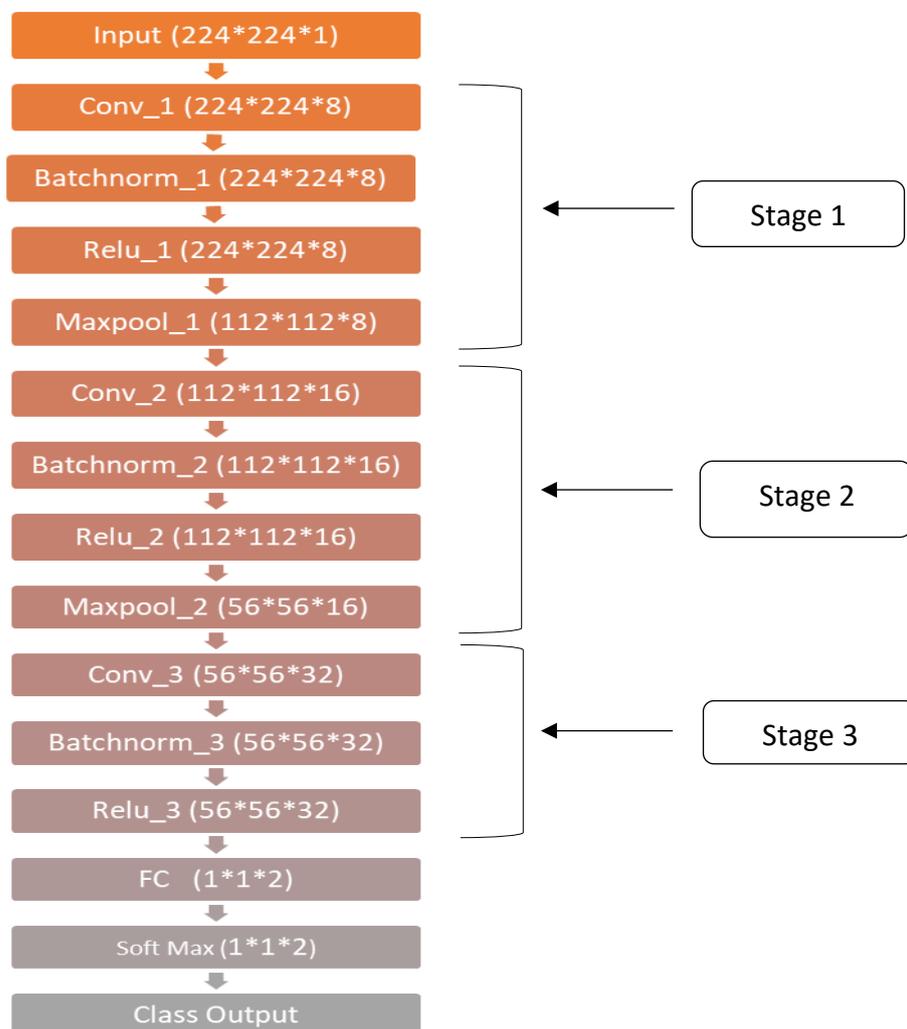


Figure 2 : Architecture du modèle

3. Les couches du modèle

a) Couche image input

Cette couche définit la taille de l'image qui, est dans notre cas 224 224 1, qui est la hauteur, la largeur et la taille du canal. Les données numériques sont faites d'images en niveaux de gris, pour que la taille du canal de la couleur soit 1.

Dans une image couleur, la taille du canal est 3, ce qui correspond aux valeurs RGB, qui va ensuite être convertie en une image NG de taille 224 224 1.

Le Train Network peut mélanger automatiquement les données au début de chaque époque lors de l'apprentissage.

b) Couche de convolution 1

Le filtre size dans la couche convolutive correspond à la hauteur et la largeur des filtres utilisés dans la fonction d'apprentissage lors du balayage des images. Dans notre cas, le nombre 3 signifie que la taille du filtre est 224 par 224 avec 8 profondeurs.

Le nombre de neurones qui se connectent à une même région d'entrée est identifié par numFiltres, il détermine le nombre des cartes d'entités.

Pour ajouter un remplissage à la carte des entités en entrée, on utilise la paire nom-valeur 'Remplissage'. Le taux d'apprentissage et la foulée peuvent être définis à l'aide de convolution2dLayer.

La 1^{ère} figure indique l'entrée de la couche de convolution, la 2^{ème} est le masque de convolution, c'est une convolution 3x3 en raison de la forme du masque.

66	65	63	62	60	55
66	65	63	61	60	72
66	64	62	60	59	107
68	73	55	59	97	191
72	64	49	79	137	231
74	54	56	122	195	255

INPUT

1	0	1
0	1	0
1	0	1

Masque

L'opération de convolution est effectuée plusieurs fois en déplaçant ce masque sur toute la matrice ou l'image d'entrée, on multiplie la matrice avec le masque et on additionne le résultat. Ce résultat va à vecteur des caractéristiques. Nous faisons ce processus pour couvrir toutes les parties de l'image.

66*1	65*0	63*1	62	60	55
66*0	65*1	63*0	61	60	72
66*1	64*0	62*1	60	59	107
68	73	55	59	97	191
72	64	49	79	137	231
74	54	56	122	195	255

INPUT*Masque

36		

Feature Map

Tableau 1 : 1ère étape de convolution

On répète le processus jusqu'à ce qu'on couvre toute l'image.

66*1	65*0	63*1	62	60	55
66*0	65*1	63*0	61	60	72
66*1	64*0	62*1	60	59	107
68	73	55	59	97	191
72	64	49	79	137	231
74	54	56	122	195	255

36	35	34	38
35	36	37	49
36	36	41	64
35	40	54	85

Tableau 2 : Vecteur de caractéristique après l'application de convolution

c) Couche batch normalisation 1

Est une couche de normalisation par lots, les gradients et l'activation propagées dans le réseau seront normalisées par ces couches, ce qui donne une optimisation plus simple à l'entraînement du réseau.

On utilise des couches de normalisation entre les couches convolutives et les couches correction ReLu pour accélérer l'apprentissage du réseau et minimiser la sensibilité à l'initialisation du réseau.

Les images doivent être normalisées de taille 224 224 1.⁶

d) Couche correction Relu 1

ReLU (Rectified Linear Units) désigne la fonction réelle non-linéaire définie par

$$ReLU(x) = \max(0, x) \quad (1)$$

La couche de correction ReLU remplace donc toutes les valeurs négatives reçues en entrées par des zéros. Elle joue le rôle de fonction d'activation.

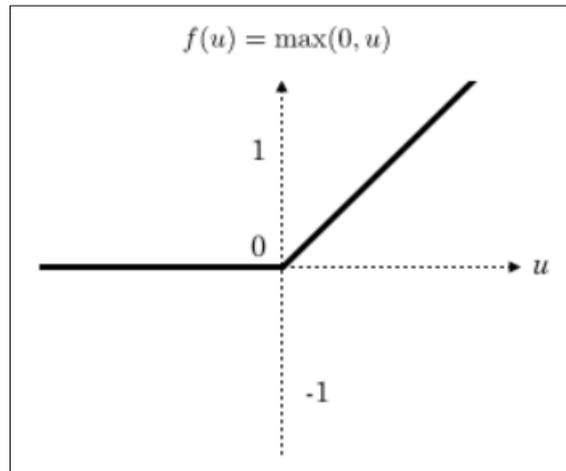


Figure 6 : Allure de la fonction ReLu [4]

e) Couche max pooling 1

La couche de max pooling est une des quatre couches dans un réseau de neurones convolutif. Cette couche permet de sous-échantillonner une image lors de sa classification par un réseau de neurones. Le but de cette couche est de réduire la taille des images sans modifier les features importantes de cette image.

La couche Max-pool va prendre en entrée des listes de caractéristiques et en fonction de la taille que l'on souhaite va retourner une matrice de taille choisie. Pour ce faire, l'image va être découpée.

Max-Pool 2x2 va rendre une matrice de taille 2x2, Max-Pool 3x3 va rendre une matrice de taille 3x3. L'image va être séparée en zone égale, Max-Pool va alors garder seulement la valeur

⁶ Fonctionnement CNN, 18 février 2023, <https://fr.mathworks.com/discovery/convolutional-neural-network-matlab.html>

maximale. La matrice de sortie possède le même nombre de caractéristiques de celle en entrée mais le max-pool n'a conservé que les valeurs maximales.⁷

Cette couche permet donc de réduire le nombre de paramètres et donc de calcul dans le réseau de neurones. Il y a donc une augmentation de l'efficacité du réseau et nous évitons ainsi le sur-apprentissage par le réseau de neurones.

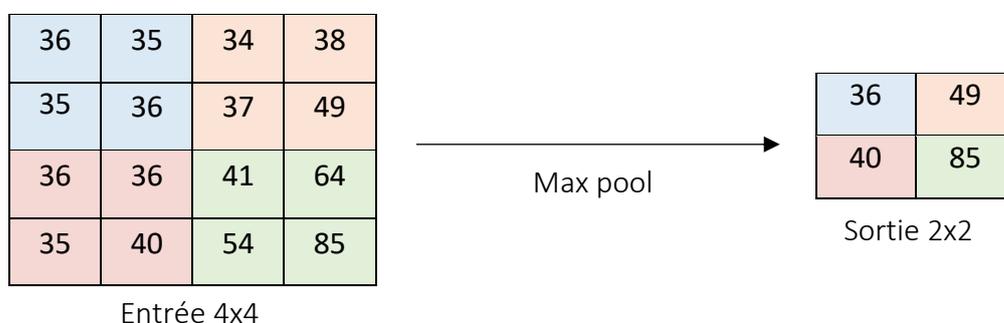


Tableau 3 : L'application de max pooling

f) Couche convolution 2

On garde le même principe de la convolution 1 mais on augmente le nombre des masques (à la place 8 on met 16 masques).

g) Couche max pooling 2

Nous avons réduit la taille des images en 2, la taille était 112, 112 et est devenu 56, 56 avec le même masque de 16.

h) Couche convolution 3

On garde le même principe de la convolution 2 mais on augmente le nombre de masques (à la place 16 on met 32 masques).

i) Couche entièrement connectée FC

Ces couches sont placées en fin d'architecture de CNN et sont entièrement connectées à tous les neurones de sorties (d'où le terme fully-connected). Après avoir reçu un vecteur en entrée, la couche FC applique successivement une combinaison linéaire puis une fonction d'activation dans le but final de classifier l'input image. Elle renvoie enfin en sortie un vecteur de taille d

⁷ Classez et segmentez des données visuelles, 08 mai 2023, <https://openclassrooms.com/fr/courses/4470531-classez-et-segmentez-des-donnees-visuelles>

correspondant au nombre de classes dans lequel chaque composante représente la probabilité pour l'input image d'appartenir à une classe.

j) Couche Softmax

La fonction softmax, ou fonction exponentielle normalisée, est une généralisation de la fonction logistique qui prend en entrée un vecteur $z = (z_1, \dots, z_k)$ de K nombres réels et qui en sort un vecteur $\sigma(z)$ de K nombres réels strictement positifs et de somme 1.⁸

k) Couche de sortie : Classification Output

La dernière couche est la couche de classification. Cette couche utilise les probabilités renvoyées par la fonction d'activation softmax pour chaque entrée pour affecter l'entrée à l'une des classes mutuellement exclusives et calculer la perte. Pour créer une couche de classification, on utilise classification layer.

4. Les options d'apprentissage Sgdm et Adam

Dans cette partie, nous allons présenter les deux méthodes d'optimisation dans l'apprentissage profond. Les optimiseurs sont des algorithmes ou des méthodes modernes utilisées pour modifier les attributs du réseau de neurones tels que les points et le taux d'apprentissage afin de minimiser les pertes.

a) Option Sgdm: Stochastic Gradient Descent with Momentum

L'idée de base derrière l'approximation stochastique remonte à l'algorithme de Robbins - Monro des années 1950, la descente de gradient stochastique est devenue une méthode d'optimisation importante dans l'apprentissage automatique. Cet optimiseur met à jour les paramètres du réseau (poids et biais) pour minimiser la fonction de perte. Le facteur d'impulsion supplémentaire permet de réduire l'oscillation qui peut apparaître le long de la trajectoire de descente la plus abrupte vers l'optimum.⁹

L'ajout d'un terme de Momentum à la mise à jour des paramètres est un moyen de réduire cette oscillation. Cette mise à jour est la suivante :

⁸ Fonction Softmax, 08 mai 2023, https://fr.wikipedia.org/wiki/Fonction_softmax

⁹ Vitaly Bushaev, 2017, Stochastic Gradient Descent with momentum, 08 mai 2023, <https://towardsdatascience.com/stochastic-gradient-descent-with-momentum-a84097641a5d>

$$\theta_{l+1} = \theta_l - \alpha \nabla E + \gamma(\theta_l - \theta_{l-1}) \quad (2)$$

Où γ détermine la contribution de l'étape de gradient précédente à l'itération actuelle. Vous pouvez spécifier cette valeur à l'aide de l'option de formation Momentum. Pour former un réseau neuronal à l'aide de l'algorithme de descente de gradient stochastique avec Momentum, spécifiez « Sgdm » comme premier argument d'entrée de training options et pour spécifier la valeur initiale du taux d'apprentissage α , utilisez l'option de formation Initial Learn-Rate.

Cet algorithme utilise un seul taux d'apprentissage pour tous les paramètres. Il peut également obtenir de nouveaux minima.

b) Option Adam : Adaptive Moment Estimation

C'est un algorithme d'optimisation qui peut être utilisé à la place de procédure classique de descente de gradient stochastique pour mettre à jour les poids de réseau itératifs en fonction des données d'apprentissage.

Adam décrit la combinaison des avantages des deux extensions de la descente de gradient stochastique spécifiquement (Ada Grad) et (RMS prop).

Adam utilise également la moyenne des seconds moments des gradients (la variance non centrée). C'est un algorithme populaire dans le domaine de l'apprentissage profond car il obtient rapidement de bons résultats. Adam est relativement facile à configurer la ou les paramètres de configuration par défaut fonctionnent bien sûr sur la plupart des problèmes. Lors de l'introduction de l'algorithme, certains avantages intéressants de l'utilisation d'Adam sur des problèmes d'optimisation sont indiqués comme suit :

- Simple à mettre en œuvre.
- Efficace en termes de calcul.
- Peu de besoins en mémoire.
- Bien adapté aux problèmes importants en termes de données et / ou de paramètres.
- Convient aux objectifs non stationnaires.
- Convient aux problèmes avec des gradients très bruyants / ou clairsemés.
- Invariant au redimensionnement diagonal des dégradés.

- Les hyper-paramètres ont une interprétation intuitive et nécessitent généralement peu de réglages.

Concernant cette option d'apprentissage, cet optimiseur conserve une moyenne mobile par élément des gradients de paramètres et de leur valeur au carré.

$$Ml = \beta_1 ml - 1 + (1 - \beta_1) \nabla E(\theta l) \quad (3)$$

$$Vl = \beta_2 vl - 1 + (1 - \beta_2) [\nabla E(\theta l)]^2 \quad (4)$$

On peut spécifier les taux de décroissance β_1 et β_2 à l'aide des options d'entraînement Gradient Decay factor respectivement. Adam utilise les moyennes mobiles pour mettre à jour les paramètres du réseau comme :

$$\theta l + 1 = \theta l - \frac{\alpha m_l}{\sqrt{v_l + \epsilon}} \quad (5)$$

C'est une méthode qui calcule les taux d'apprentissage adaptatifs pour chaque paramètre et stocke une moyenne à décroissance exponentielle des gradients carrés. Cet Algorithme conserve également une moyenne des gradients passés qui décroît de façon exponentielle (mt) semblable au Momentum (Mt) et (Vt). Si les gradients sur de nombreuses itérations sont similaires, l'utilisation d'une moyenne mobile du gradient permet aux mises à jour des paramètres de prendre de l'élan dans une certaine direction. Si les gradients contiennent principalement du bruit, alors la moyenne mobile du gradient devient plus petite, et donc les mises à jour des paramètres deviennent également plus petites. La mise à jour complète d'Adam comprend également un mécanisme permettant de corriger un biais qui apparaît au début de la formation.

Cet optimiseur utilise également le concept de moment adaptatif en ajoutant des fractions de gradients précédents au gradient actuel.¹⁰

Enfin, Adam est devenu assez répandu, il est pratiquement accepté pour être utilisé dans les réseaux neuronaux d'entraînement et spécifié comme premier argument d'entrée.

¹⁰ Jeremy Zhang, 2020, Optimisation Algorithm — Adaptive Moment Estimation(Adam), 13 mai 2023, [https://towardsdatascience.com/optimisation-algorithm-adaptive-moment-estimation-adam-92144d75e232#:~:text=Adaptive%20Moment%20Estimation%20\(Adam\)%20is,gradients%20v%20%2C%20similar%20to%20momentum.](https://towardsdatascience.com/optimisation-algorithm-adaptive-moment-estimation-adam-92144d75e232#:~:text=Adaptive%20Moment%20Estimation%20(Adam)%20is,gradients%20v%20%2C%20similar%20to%20momentum.)

Suivant les types d'options, ces deux algorithmes utilisent les mêmes paramètres qui sont représentés comme suit :

a. Nombre d'épochs

Désigne un passage sur l'ensemble d'entraînement complet, une epoch correspond à un apprentissage sur toutes les données, plus ce nombre est grand plus on obtient une bonne précision.

b. Learning rate LR

Le taux d'apprentissage, indiqué par le symbole α , est un hyper-paramètre utilisé pour régir le rythme auquel un algorithme actualise ou apprend les valeurs d'une estimation de paramètre. En d'autres termes, le taux d'apprentissage régule les poids de notre réseau neuronal concernant le gradient de perte.

c. Mini batch size

Le mini-lot est un nombre fixe d'exemples de formation qui est inférieur à l'ensemble de données réel. Ainsi, à chaque itération, nous formons le réseau sur un groupe différent d'échantillons jusqu'à ce que tous les échantillons de l'ensemble de données soient utilisés.¹¹

Dans le diagramme ci-dessous, nous pouvons voir comment la descente de gradient mini-batch fonctionne quand la taille de mini-batch est égale à deux :

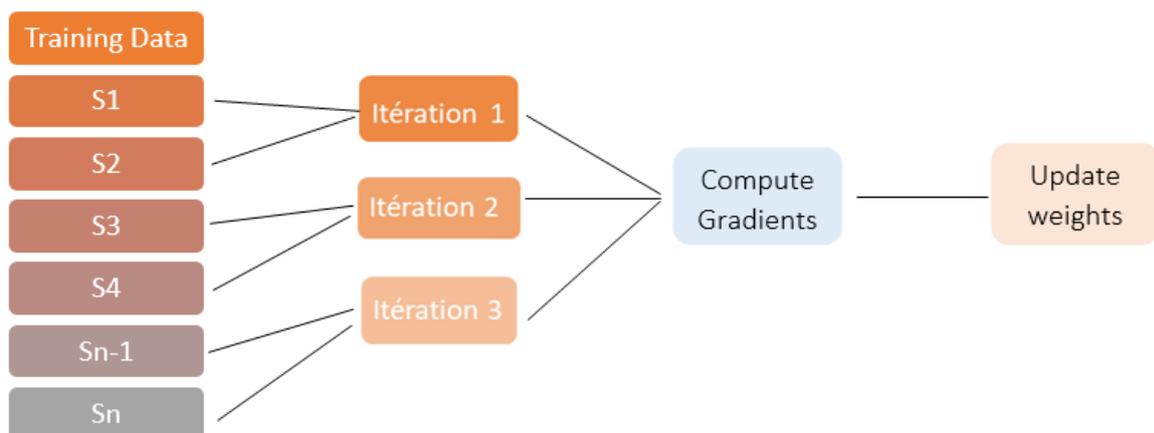


Figure 7 : Fonctionnement du mini batch [5]

¹¹ Panagiotis Antoniadis, Differences between Epoch, Batch, and Mini-batch, 16 mars 2023, <https://www.baeldung.com/cs/epoch-vs-batch-vs-mini-batch#:~:text=The%20mini%2Dbatch%20is%20a,of%20the%20dataset%20are%20used.>

5. Conclusion

Après avoir expliqué le fonctionnement des couches du réseau et les différentes options d'apprentissage appliquées, on va présenter les résultats du modèle proposé et faire une petite interprétation.

Chapitre 03

1. Introduction

Après avoir décrit le modèle qu'on a proposé dans le chapitre précédent, ce dernier axe sera dédié à la mise en œuvre de son application pour le diagnostic du cancer.

Pour cela on va utiliser une catégorie d'images sélectionnées à partir de d'une base de données qui est l'IRM ou Scanner.

Ensuite, on va présenter les résultats obtenus par l'application du modèle sur la base de données.

A la fin, on va faire une comparaison sur les résultats obtenus avec plusieurs paramètres.

2. Description base de données : IRM – Scanner

Cet ensemble de données est constitué d'images scannées du cerveau de patients chez qui un diagnostic de tumeur cérébrale a été établi ou pas. L'ensemble de données proposé est composé de 4506 images qui se répartissent entre 2426 pour les patients infectés par une tumeur cérébrale, et 2080 images pour les patients non infectés par une tumeur cérébrale. Ces données ont été collectées de plusieurs sources [6]. Les images de cette base de données sont au format JPG et de résolution variable. La figure suivante démontre des exemples de notre base de données.

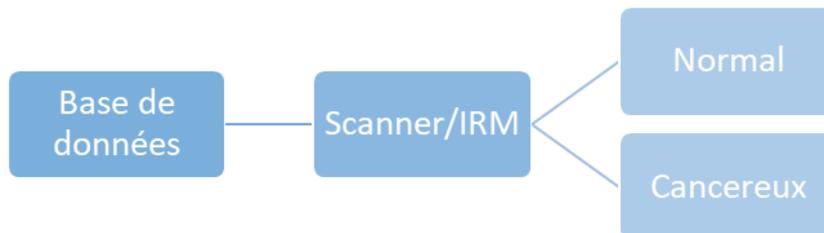


Figure 1 : Diagramme de la base de données

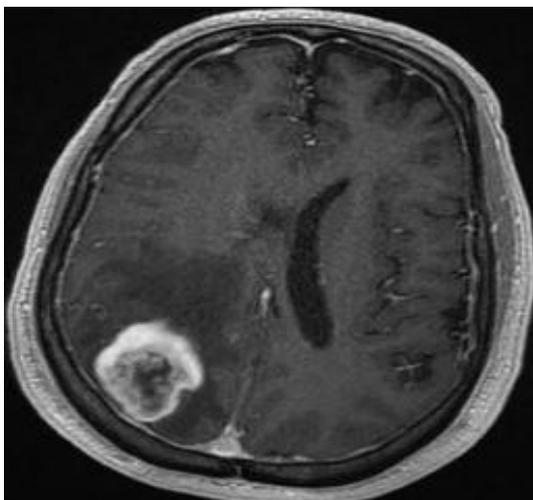


Figure 2 : Image patient malade
La base de données

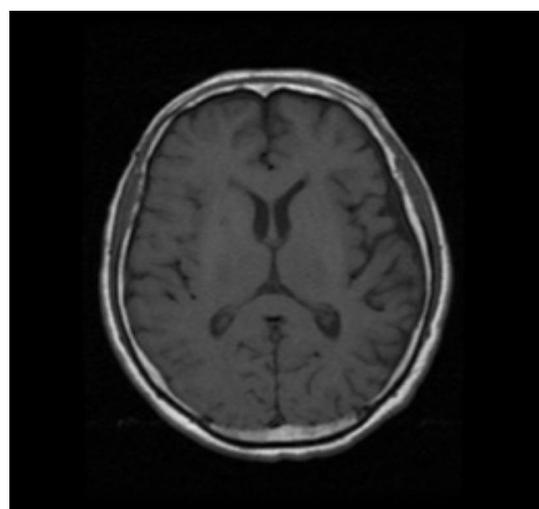


Figure 3 : Image patient normal
La base de données

3. Performance metrics

L'analyse des performances du modèle proposé est évaluée à partir de la matrice de confusion en fonction de l'exactitude, de la précision, Recall et du score F1. Cette performance est déterminée en utilisant les termes vrais positifs (TP), faux positif (FP), vrai négatif (TN) et faux négatifs (FN). Le taux de détection correct des photographies atteintes parmi toutes les images est appelé Recall. Le score F1 et la mesure combinée de la précision et du rappel qui montre combien de fois la valeur prédite est exacte. Ce score est également connu comme la moyenne harmonique de p et r en mathématiques. La précision est une mesure de l'efficacité du fonctionnement d'un modèle ou d'un système : (c'est à dire le nombre de fois ou le modèle a correctement prédit le résultat réel et doit être calculé).¹² Les formules mathématiques pour déterminer la précision sont exprimées dans les équations ci-dessous :

- Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$ (6)

- Précision = $\frac{TP}{TP+FP}$ (7)

- Sensitivity/recall = $\frac{TP}{TP+FN}$ (8)

- Specificity = $\frac{TN}{TN+FP}$ (9)

- F1-score = $\frac{2 \times \text{précision} \times \text{recall}}{\text{précision} + \text{recall}}$ (10)

a) Application du modèle pour la base de données

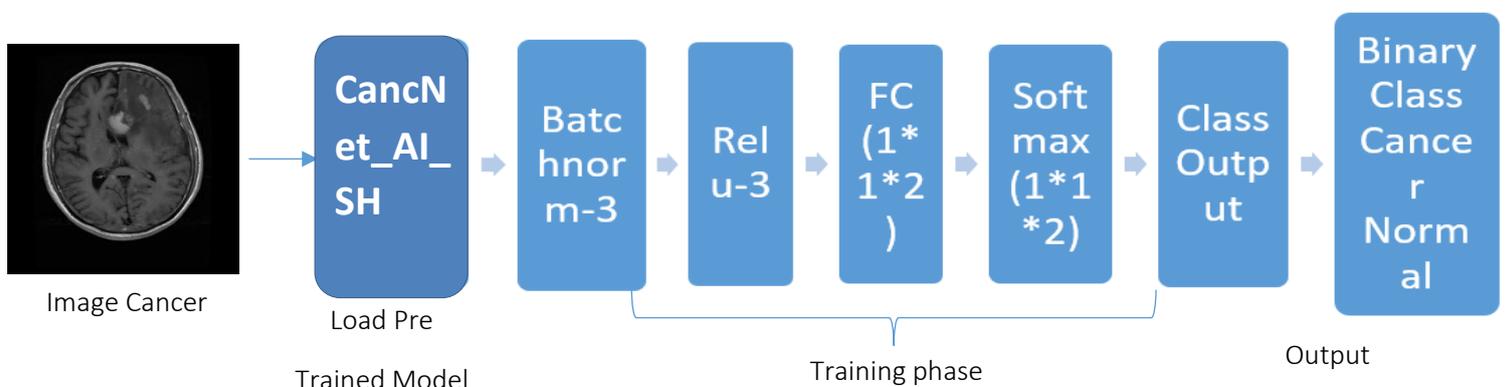


Figure 8 : Architecture du modèle proposé

¹² Fezari.A, Boudrar.N, 2022, *Détection et analyse du covid_19 dans les images médicales à l'aide de techniques d'apprentissage profond*, Mémoire de fin d'études, Université BADJI Mokhtar – ANNABA, Algérie.

4. Résultats obtenus

Dans ces expériences, nous avons utilisé notre ensemble de données de scanner du cancer du cerveau préparé pour examiner l'efficacité de notre modèle proposé. Les performances de classification de ce modèle pour le problème de classification binaire (Normal /Covid) sur les images radiographiques sont fournies dans le figure 12. Les figures montrent les courbes d'évolutions sur plusieurs époques et sur les deux options 'Adam' et 'Sgdm'. Nous présentons également une comparaison des performances de notre modèle pour les deux options dans la figure 11 et une comparaison de notre modèle avec les modèles prédéfinis dans la figure 13.

a) Test 01

Suivant les paramètres : NE = 10, LR = 0.0002, TP = 0.9 et l'option Adam

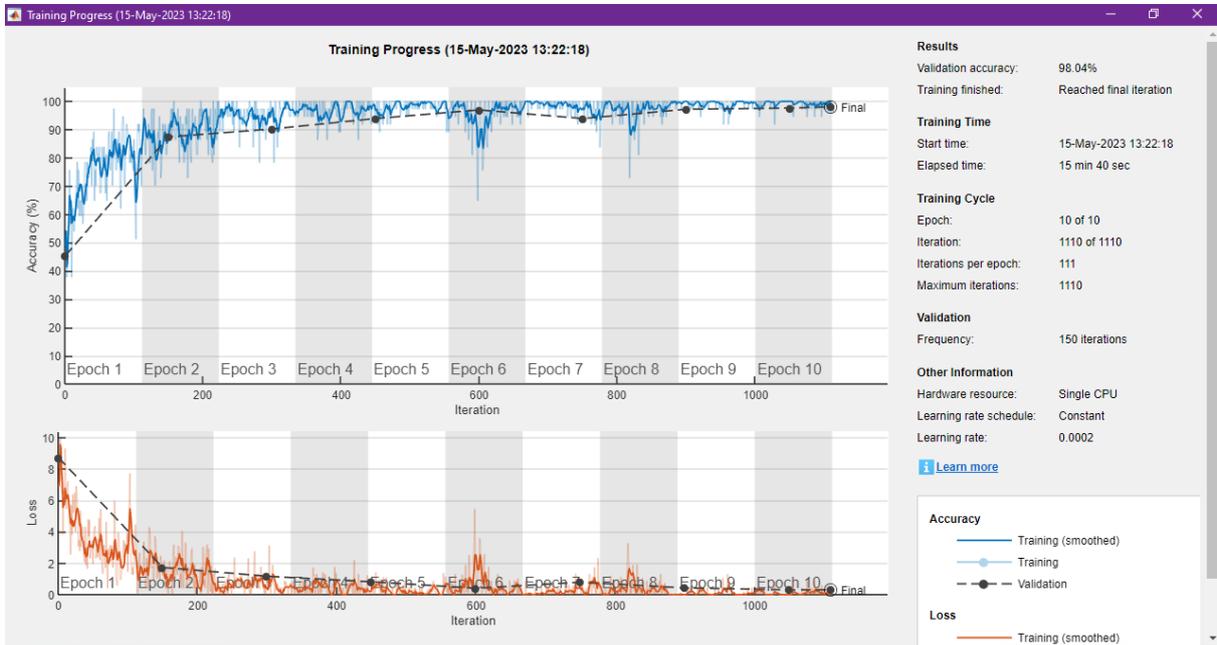


Figure 9 : Courbe d'évolution 1 avec l'option Adam,

Sa matrice de confusion est la suivante :

Figure 10 : Matrice de confusion



b) Test 02

Suivant les paramètres : NE = 10, LR = 0.0002, TP = 0.9 et l'option Sgdm

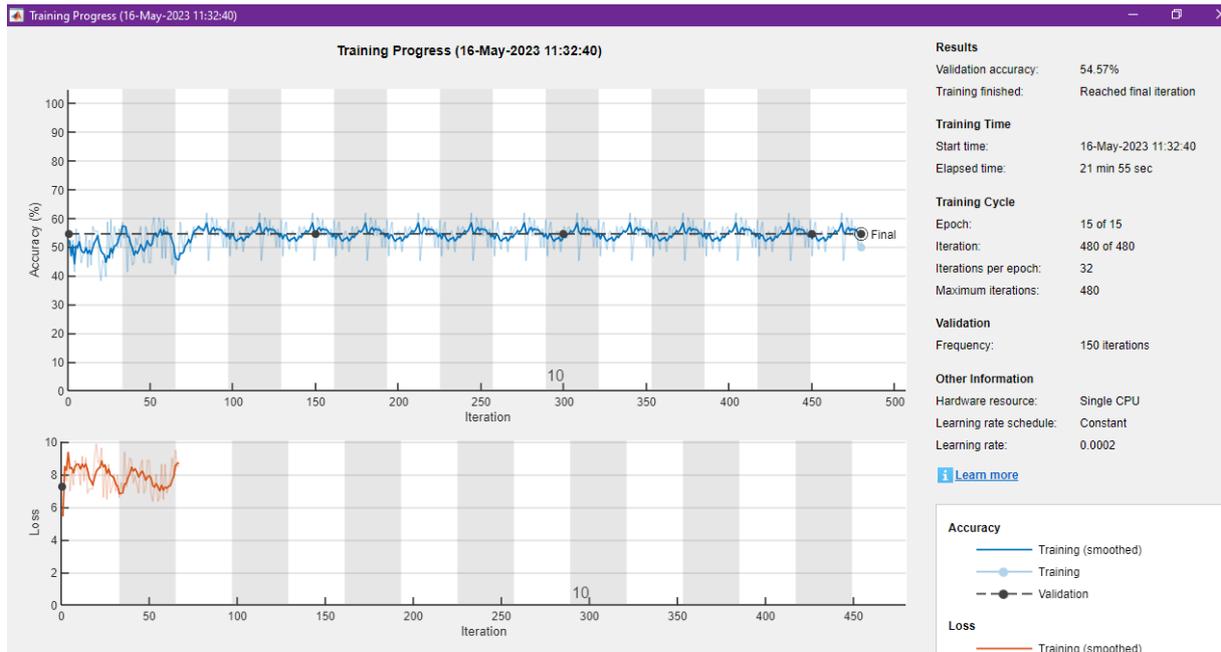


Figure 11 : Courbe d'évolution 2 avec l'option Sgdm

c) Test 03

Suivant les paramètres : NE = 15, LR = 0.0002, TP = 0.9 et l'option Adam

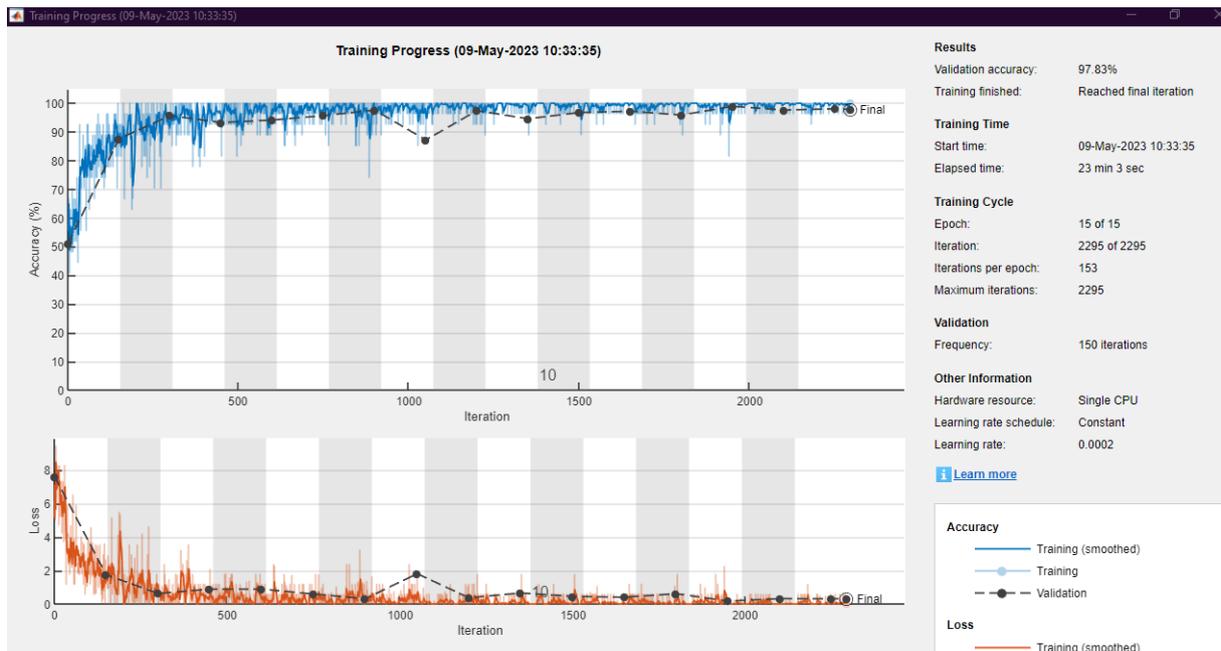


Figure 12 : Courbe d'évolution 3 avec l'option Adam

d) Test 04

Suivant les paramètres : NE = 15, LR = 0.0002, TP = 0.9 et l'option Sgdm

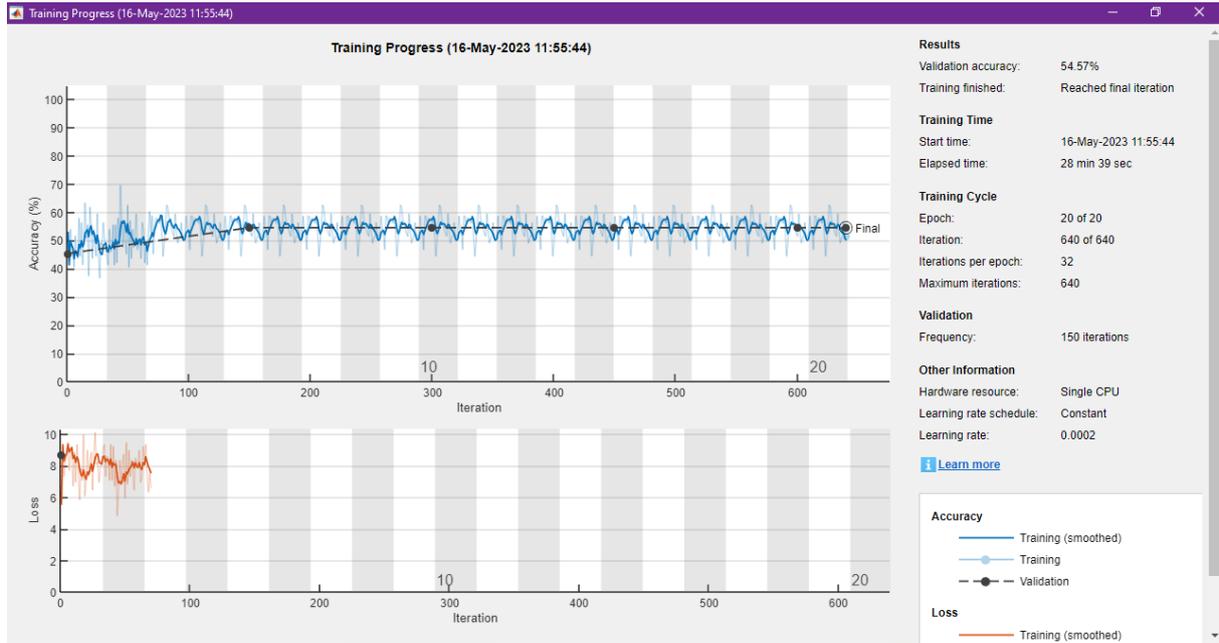


Figure 13 : Courbe d'évolution 4 avec l'option Sgdm

e) Test 05

Suivant les paramètres : NE = 20, LR = 0.0002, TP = 0.9 et l'option Adam

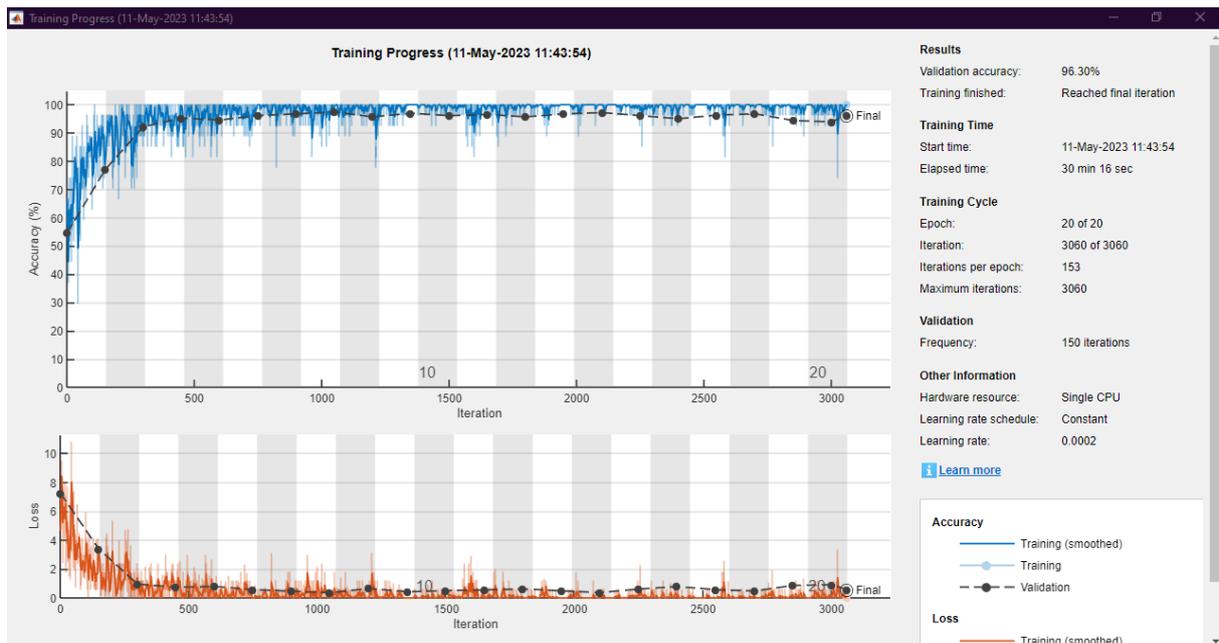


Figure 14 : Courbe d'évolution 5 avec l'option Adam

f) Test 06

Suivant les paramètres : NE = 20, LR = 0.0002, TP = 0.9 et l'option Sgdm

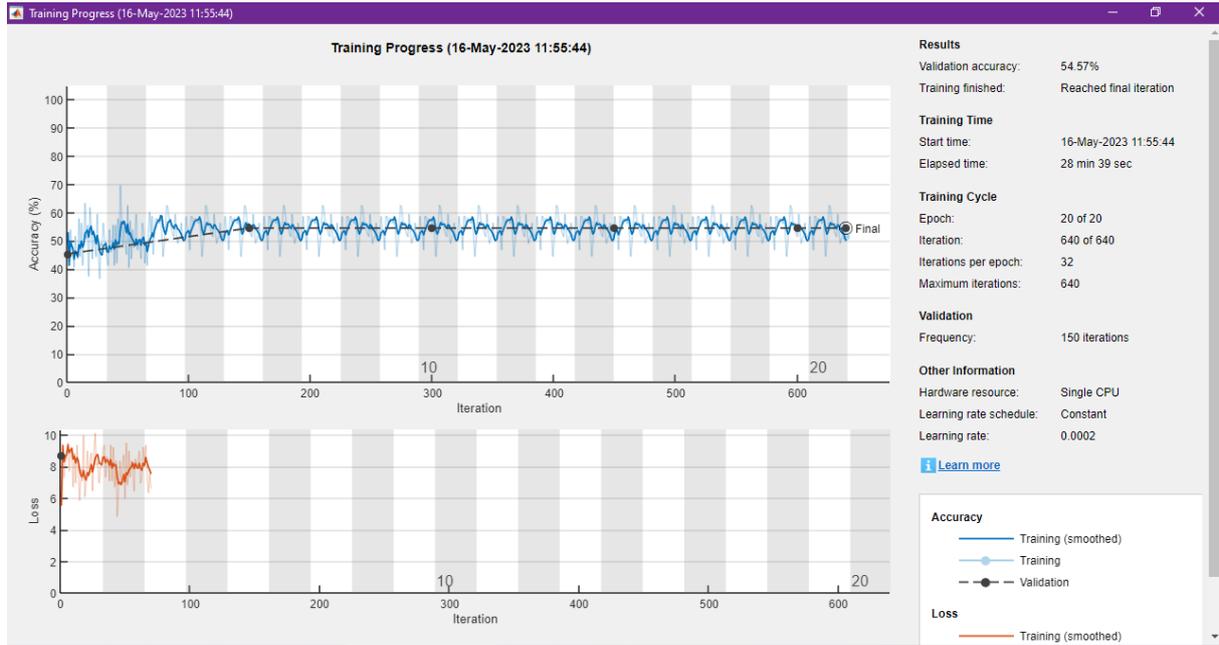


Figure 15 : Courbe d'évolution 6 avec l'option Sgdm

g) Test 07

Suivant les paramètres : NE = 35, LR = 0.0002, TP = 0.9 et l'option Adam

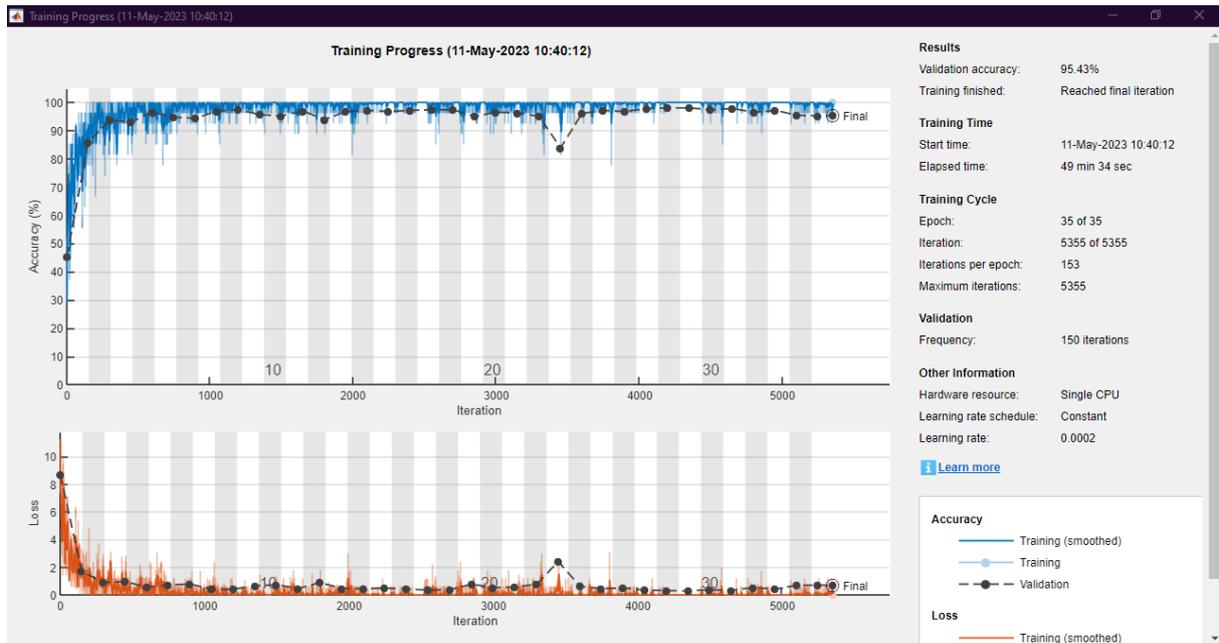


Figure 16 : Courbe d'évolution 7 avec l'option Adam

h) Test 08

Suivant les paramètres : NE = 35, LR = 0.0002, TP = 0.9 et l'option Sgdm

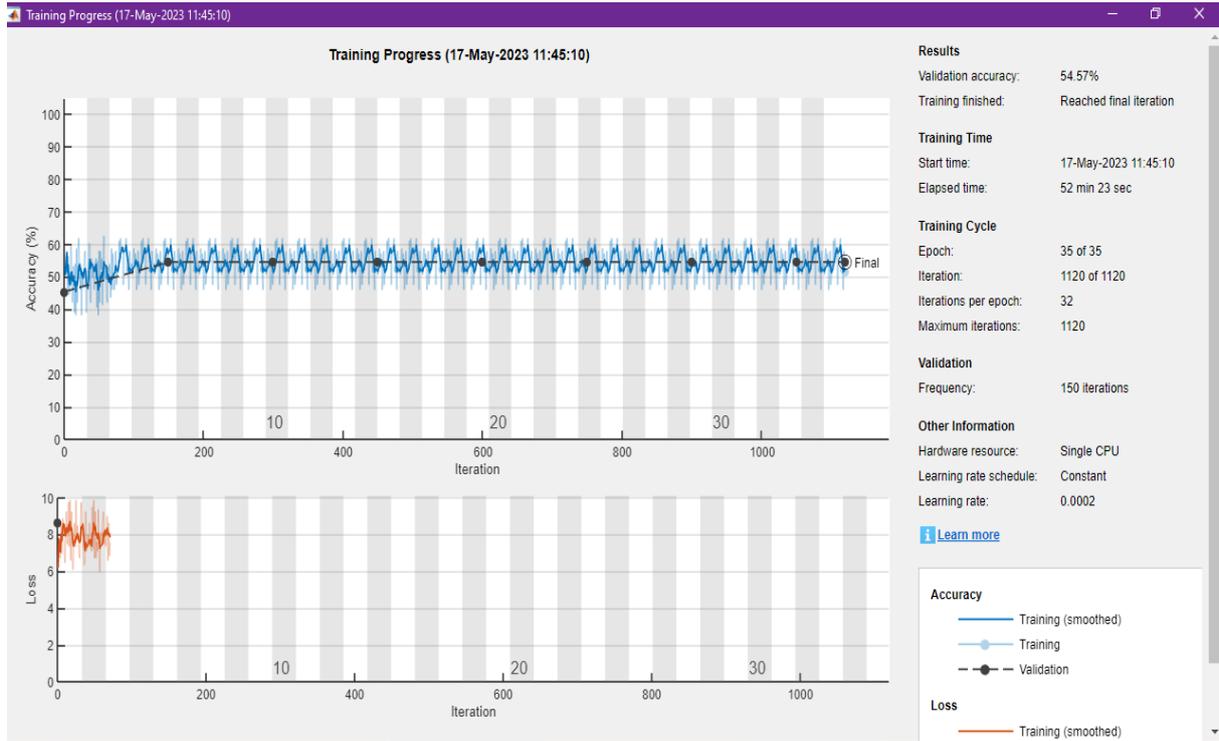


Figure 17 : Courbe d'évolution 8 avec l'option Sgdm

i) Comparaison des résultats entre les deux options d'apprentissage

M-Epochs	ADAM	SGDM
10	98.48%	54.57%
15	97.83%	54.57%
20	96.30%	54.57%
35	95.43%	54.57%

Tableau 4 : Comparaison entre les deux options d'apprentissage

j) Calcul des performances metrics

M-Epochs	10	15	20	35
Accuracy	98.40%	97.8%	96.3%	78.47%
Précision	98.35%	96.12%	93.94%	72.60%
Sensitivity/recall	98.74%	100%	99.63%	97.24%
Specificity	98.02%	95.16%	92.3%	55.94%
F1-Score	98.52%	98.02%	96.7%	83.13%

Tableau 5 : Comparaison des performances metrics entre les nombre d'epochs utilisés

k) Comparaison du modèle proposé avec les modèles prédéfinis

Modèle	Accuracy
GoogLe Net	93.78%
Alex Net	92.37%
Resnet50	94.25%
CancNet_BN_FA	98.40%

Tableau 6 : Comparaison des modèles prédéfinis

5. Conclusion

Dans ce chapitre, on a exécuté le modèle qu'on a proposé sur plusieurs nombre d'epochs et en utilisant les deux options d'apprentissage 'Adam' et 'Sgdm', en de ça on a fait une comparaison entre les modèles prédéfinis tels que GoogLeNet, Alex Net et Resnet50 avec le modèle CancNet_AI_SH. Le résultat :

- Le test a donné un bon résultat en utilisant l'option Adam et sur un nombre d'epochs égal à 10, il ne prend pas beaucoup de temps pour faire l'apprentissage et nous donne à la fin un meilleur diagnostic.
- Selon le 1er tableau, on peut dire que l'option Adam donne de meilleurs résultats que l'option Sgdm.
- Le modèle CancNet_AI_SH est plus fiable que les modèles prédéfinis dans ce diagnostic.

- Les performances metrics de 10 sont supérieurs que des autres nombres d'epochs, c'est-à-dire que dans ce modèle plus le nombre d'epochs est petit plus on obtient de meilleurs résultats.

Conclusion générale

Ce mémoire avait pour ambition de créer un modèle en apprentissage profond pour faire un diagnostic du cancer du cerveau par imagerie médicale.

Il a fallu dans un premier temps présenter les deux grands axes de notre travail qui sont l'axe théorique c'est-à-dire donner une idée sur le cancer qu'on a voulu étudier ainsi que l'axe technique qui est l'apprentissage profond et le réseau de neurones convolutifs CNN, on a défini aussi les couches du réseau convolutif CNN que notre modèle utilise et les examiner une par une.

Après avoir créé le modèle, on a lancé les tests sur deux options d'apprentissage qui sont 'Adam' et 'Sgdm' et sur plusieurs nombres d'epochs. En parallèle, on a effectué une comparaison entre les modèles prédéfinis et le modèle proposé.

A partir des résultats obtenus, on a constaté que CancNet_AI_SH donne de meilleurs résultats en utilisant l'option 'Adam' et sur un petit nombre d'epochs qui est égal à 10.

Ce résultat obtenu est intéressant car il permet d'effectuer le diagnostic en un temps plus court et arriver ainsi à des résultats de meilleure qualité.

Cette réalisation nous a permis d'élargir notre connaissance dans ce domaine d'intelligence artificielle et l'intérêt porté par les utilisateurs.

Référence bibliographique

- [1] <https://www.imaios.com/fr/e-anatomy/structures-anatomiques/artere-ophtalmique-116801436>
- [2] <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758>
- [3] https://www.researchgate.net/figure/GoogleNet-like-architecture_fig7_320723863
- [4] https://www.researchgate.net/figure/Allure-de-la-fonction-ReLU_fig21_344266838
- [5] <https://www.baeldung.com/cs/epoch-vs-batch-vs-minibatch#:~:text=The%20mini%2Dbatch%20is%20a,of%20the%20dataset%20are%20used.>
- [6] <https://www.kaggle.com/datasets/preetviradiya/brian-tumor-dataset?resource=download> . Dernier accès le 20/05/2023

Webographie

- Pr SANSON, 2023, Institut du cerveau, 10 mai 2023, <https://institutducerveau-icm.org/fr/tumeurs-cerebrales/>
- Deep Learning ou Apprentissage Profond : qu'est-ce que c'est ?, 03 mars 2023, <https://datascientest.com/deep-learning-definition#:~:text=Qu'est%2Dce%20que%20le,des%20r%C3%A9seaux%20de%20neurones%20artificielles>
- Abheer Bhandodker, Inception V1 Architecture Explained, 17 mai 2023 <https://medium.com/@abheerchrome/inception-v1-architecture-explained-454b2eb66baf>
- Fonctionnement CNN, 18 février 2023, <https://fr.mathworks.com/discovery/convolutional-neural-network-matlab.html>

- Classez et segmentez des données visuelles, 08 mai 2023, <https://openclassrooms.com/fr/courses/4470531-classez-et-segmentez-des-donnees-visuelles>
- Fonction Softmax, 08 mai 2023, https://fr.wikipedia.org/wiki/Fonction_softmax
- Vitaly Bushaev, 2017, Stochastic Gradient Descent with momentum, 08 mai 2023, <https://towardsdatascience.com/stochastic-gradient-descent-with-momentum-a84097641a5d>
- Jeremy Zhang, 2020, Optimisation Algorithm — Adaptive Moment Estimation(Adam), 13 mai 2023, [https://towardsdatascience.com/optimisation-algorithm-adaptive-moment-estimation-adam-92144d75e232#:~:text=Adaptive%20Moment%20Estimation%20\(Adam\)%20is,gradie%20nts%20v%20%2C%20similar%20to%20momentum.](https://towardsdatascience.com/optimisation-algorithm-adaptive-moment-estimation-adam-92144d75e232#:~:text=Adaptive%20Moment%20Estimation%20(Adam)%20is,gradie%20nts%20v%20%2C%20similar%20to%20momentum.)