

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

– جامعة أبي بكر بلقايد – تلمسان

Université Aboubakar Belkadi – Tlemcen –

Faculté de TECHNOLOGIE



MEMOIRE

Présenté pour l'obtention du diplôme de MASTER

En : Génie Biomédicale

Spécialité : Imagerie Médicale

Par : ELBACHIR Chaimaa & SITAYEB Norelhouda

Thème

Aide au diagnostic dans l'imagerie cérébrale : Application dans la reconnaissance des AVC.

Soutenu publiquement, Dimanche à 10h le 25 / 06 /2023 , devant le jury composé de :

Mr Behadada Omar	MCB	Université de Tlemcen	Président
Mme Iles Amel	MCA	Université de Tlemcen	Examinatrice
Mme Benchaib Yasmine	MCA	Université de Tlemcen	Encadrante

Remerciement

Pour commencer, nous tenons à adresser nos remerciements à toutes les personnes qui ont contribué au succès de ce mémoire de master, ainsi qu'à toute l'équipe pédagogique de l'université d'Abou Bekr Belkaid de Tlemcen.

Nous souhaitons exprimer nos remerciements à notre encadrante, Mme Benchaib Yasmin. Votre soutien a été inestimable et nous avons particulièrement apprécié votre expertise et votre disponibilité.

Nous remercions également Dr. Mnasri Nabil et Dr. Belnouar Abd elkader, directeur des urgences à l'hôpital Meslem Tayeb, Dr. Ben Ali Ben Naoum, Dr. Marouf Sara, maître de conférences à l'université de Sétif, et doctorante Mokhtari Aicha, Mr. Meddber Maheiddine, manipulateur, et Dr. Dalila Benmahdi, maître de conférences de Sétif.

Nous tenons à remercier tout particulièrement Mr. Behadada Omar, Maître de Conférences à l'université de Tlemcen, d'avoir accepté de présider le jury.

Nous souhaitons adresser nos remerciements à Mme ILES Amel, Maître de Conférences à l'université de Tlemcen, ainsi que pour avoir accepté la lourde tâche d'être examinatrice.

Dédicace

Avant toute chose, je souhaite exprimer ma gratitude envers Dieu pour la réalisation de ce projet.

Je dédie ce travail Avec un profond respect et une immense reconnaissance.

Ma remise de diplôme et ma joie sont dédiées à mon paradis, la source de mon bonheur et de ma joie, ma moitié, ma maman.

Je remercie celui qui m'a fait une femme, ma source de vie, d'amour et d'affection, ma chère mère, ainsi que mon père, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.

Je dédie également ce travail à mon frère Oussama, pour l'amour qu'il me porte.

Ainsi qu'à ma chère grande sœur Assma, qui n'a jamais cessé de me prodiguer ses conseils. Mes chères sœurs Amani, Malek et Nada méritent également ma reconnaissance pour leurs encouragements constants et leur soutien moral.

Je souhaite également exprimer ma gratitude envers ma chère amie Hazar, qui m'a apporté son soutien et son encouragement.

Enfin, je remercie mon binôme Sitayeb Norelhouda d'être toujours présent à mes côtés.

Merci d'être toujours là pour moi

ELBACHIR CHAIMAA

Dédicace

L'histoire arrive à sa conclusion et je lève mon chapeau pour dire adieu aux années passées.

Je rends grâce à Dieu, par la grâce de qui les bonnes actions sont accomplies.

C'est avec une profonde gratitude que je dédie ce travail à une personne très spéciale, celle qui est la prunelle de mes yeux, la source de joie de mon cœur, mon ange de la vie. Ses prières ont été le secret de ma réussite ; ma chère maman.

Je porte son nom avec fierté. À toi, mon cher père, qui a été mon soutien inébranlable et ma source de force. Que Dieu prolonge ta vie et comble tes journées de bonheur.

J'adresse également mes remerciements à mes chers frères, Ben Ali et Ahmed el-Bashir, ainsi qu'à la petite poussin Soumia. Votre présence et votre soutien ont été d'une valeur inestimable dans ma vie.

Je tiens à honorer la mémoire de mon grand-père et de ma grand-mère, Ben Abdallah et Hassnia, dont les enseignements et les souvenirs continuent de m'inspirer.

À mes amis, Feryan, Khadija, Bouchra, Ayet, Hajar, Amina et Ikram, je vous suis profondément reconnaissante pour votre amitié sincère et vos encouragements constants.

Un hommage particulier est également rendu à toute la famille Si-Tayeb, qui a toujours été présente pour soutenir mes efforts.

A mon binôme ELBACHIR Chaimaa

Enfin, je souhaite exprimer ma gratitude à tous ceux qui ont contribué de près ou de loin à mon succès. Vos encouragements, vos conseils et votre soutien ont été essentiels tout au long de ce parcours.

Que votre vie soit emplie de bonheur et de succès.

NORELHOUDA SITAYEB

Résumé

Ce projet vise à résoudre la problématique de classification des Accident Vasculaire Cérébral (AVC) en utilisant une approche basée sur la méthode Grey level co-occurrence matrix (GLCM) combinée aux modèles de classification SVM et KNN. L'objectif principal est d'améliorer la détection précoce et la classification précise des AVC à partir d'images médicales.

Dans ce projet, une méthodologie rigoureuse a été suivie. Des images médicales au format DICOM ont été collectées, puis des caractéristiques pertinentes ont été extraites à l'aide de la méthode GLCM, telles que le contraste, la corrélation, l'énergie, l'homogénéité et l'entropie. Ces caractéristiques ont été utilisées comme données d'entrée pour les modèles SVM et KNN.

Les résultats obtenus ont été prometteurs. Le modèle SVM avec le noyau RBF a démontré une sensibilité et une spécificité élevées, avec un taux de classification satisfaisant pour les AVC. Le modèle KNN avec la distance euclidienne et un K égal à 5 a atteint une sensibilité parfaite, bien que sa spécificité soit légèrement inférieure.

Cette étude apporte une contribution significative dans le domaine de la détection précoce et de la classification des AVC. Elle démontre l'efficacité de la méthode GLCM en combinaison avec les modèles SVM et KNN pour obtenir des résultats précis. L'application de cette approche peut avoir un impact considérable sur la prise en charge rapide des patients atteints d'AVC et sur la réduction des complications.

Mots clés : AVC, GLCM, KNN, SVM.

Abstract

This project aims to solve the problem of stroke classification by employing a combined approach of the GLCM method with SVM and KNN classification models. The primary objective is to enhance early detection and accurate classification of strokes based on medical imaging.

The project followed a rigorous methodology that involved collecting DICOM format medical images and extracting relevant features using the GLCM method, including contrast, correlation, energy, homogeneity, and entropy. These extracted features were utilized as input data for the SVM and KNN models.

The obtained results are promising; with the SVM model using the radial basis function (RBF) kernel demonstrating high sensitivity and specificity, resulting in a satisfactory stroke classification rate. The KNN model, employing Euclidean distance and a K value of 5, achieved perfect sensitivity but displayed slightly lower specificity.

Nevertheless, it is crucial to acknowledge certain limitations. Further research utilizing independent datasets is required to validate the findings and assess the generalizability of the approach. Additionally, future enhancements can be explored, such as investigating alternative features or employing feature selection techniques to optimize model performance.

In conclusion, this final year project lays the foundation for advancements in early stroke detection and classification by employing the GLCM method alongside SVM and KNN models. It offers substantial possibilities for enhancing medical care and contributes to the progress of research in this critical healthcare domain.

Keywords: Stroke, GLCM, KNN, SVM.

الملخص

هذا المشروع يهدف إلى حل مشكلة تصنيف السكتات الدماغية باستخدام نهج يعتمد على طريقة GLCM (مصنوفة تكرار مستويات الرمادي) المجتمعة مع نماذج تصنيف SVM (الآلة ذات الدعم النوعي) و KNN (أقرب الجيران K).

الهدف الرئيسي هو تحسين التشخيص المبكر والتصنيف الدقيق للسكتة الدماغية من خلال الصور الطبية في هذا المشروع، تم اتباع منهجية صارمة. تم جمع صور طبية بتنسيق DICOM ، ثم تم استخراج سمات ذات صلة باستخدام طريقة GLCM ، مثل التباين والترابط والطاقة والتجانس والانتروبيا. تم استخدام هذه السمات كبيانات مدخلة لنماذج SVM و KNN.

تم الحصول على نتائج مشجعة. أظهرت نموذج SVM بنواة RBF حساسية وتحديد عاليين، مع معدل تصنيف مرض السكتة الدماغية مرضاً مقبولاً. تم تحقيق حساسية مثالية بواسطة نموذج KNN باستخدام المسافة الإقليدية وقيمة K تساوي 5، على الرغم من أن تحديد هذا النموذج أقل بقليل..

ومع ذلك، من المهم أن نأخذ في الاعتبار بعض القيود. يتطلب التحقق من النتائج وتقييم تعميم النهج دراسات إضافية على مجموعات بيانات مستقلة. بالإضافة إلى ذلك، يمكن مراعاة تحسينات مستقبلية، مثل استكشاف سمات أخرى أو استخدام تقنيات اختيار السمات لتحسين أداء النماذج.

في الختام، يفتح هذا المشروع طريقاً للتقدم في التشخيص المبكر وتصنيف السكتة الدماغية من خلال استخدام طريقة GLCM بالاشتراك مع نماذج SVM و KNN. إنه يقدم إمكانيات هامة لتحسين الرعاية الصحية ويساهم في تقدم البحث في هذا المجال الحيوي للصحة.

الكلمات المفتاحية: السكتة الدماغية، GLCM، SVM، KNN .

Liste des acronymes

AVC	Accident Vasculaire Cérébral
AVCi	Accident Vasculaire Cérébral Ischémique
AVCh	Accident Vasculaire Cérébral Hémorragique
DSC	Débit Sanguin Cérébral
ACI	Artère Carotide Interne
AV	Artère Vertébrale
ATP	Adénosine Triphosphate
ACoP	Artère Communicante Postérieure
ACP	Artère Cérébrale Postérieure
ACM	Artère Cérébrale Moyenne
HTA	Hypertension Artérielle
RR	Risque Relatif
SMR	Standardized Mortality Ratio
ARM	Angiographie Par Résonance Magnétique
HSA	Hémorragie Sous-Arachnoïdienne
AIT	Accident Ischémique Transitoire
RTPA	Recombinant Tissue Plasminogen Activator
IRM	Imagerie par Résonance Magnétique ou Images de Résonance Magnétique
T1	Séquence T1
T2	Séquence T2
T2-FLAIR	Séquence T2 avec annulation du signal des liquides
GLCM	Grey level co-occurrence matrix
LCR	Liquide Céphalo-Rachidien
ASPECTS	Alberta CT Stroke Program Early
ROI	Regions D'Intérêts
LNP	Local Neighbourhood Pattern
LBP	Local Binary Patterns
LTP	Local Ternary Patterns
LWP	Local Wavelet Patterns
LDEP	Local Diagonal Extrema Patterns

PACS	Partial Anterior Circulation Syndrome
LACS	Lacunar Syndrome
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
EM	Expectation-Maximization
FODPSO	Fractional-Order Darwinian Particle Swarm Optimization
RF	Random Forest
FCM	Fuzzy C-Mean
LPE	Ligne de Partage des Eaux
ADC	Apparent Diffusion Coefficient
RBF	Fonction De Base Radiale

Table des figures

Figure 1.1: Schéma d'un cerveau humain dans sa boîte crânienne.	4
Figure 1.2: Communication nerveuse entre les organes.	5
Figure 1.3: Représentation en coupe horizontale du cerveau présentant les diverses structures anatomiques.	6
Figure 1.4: Irrigation sanguine du cerveau.	6
Figure 1.5: Accident vasculaire cérébrale ischémique ainsi que les régions affectées.	8
Figure 1.6: Accident vasculaire cérébrale hémorragique ainsi que les régions affectées.	8
Figure 1.7: Blocage du débit sanguin.	9
Figure 1.8: Rupture d'un vaisseau sanguin.	10
Figure 1.9: Accidents vasculaires cérébraux dans le monde.	13
Figure 1.10: Analyse par scanner cérébral a été réalisée sur un patient présentant un AVC sylvien droit.	14
Figure 1.11: Lésion soudaine et intense provoquant un hématome. Une imagerie par résonance magnétique (IRM) a été effectuée deux heures après l'apparition d'une faiblesse du côté gauche du corps, accompagnée de maux de tête.	15
Figure 2.1: Matrice de cooccurrence pour l'extraction des caractéristiques de texture dans 4 directions et 4 distances.	28
Figure 2.2: La méthode avec trois voisins pour identifier à quelle classe appartient le point X.	33
Figure 2.3: Architecture d'un vecteur support machine.	35
Figure 2.4: Représentation matricielle d'une base de données.	37
Figure 2.5: A gauche, un échantillon correspondant à un problème de classification simple. À droite, la séparation optimale de ce problème.	37
Figure 2.6: Apprentissage d'un modèle à partir d'un ensemble d'apprentissage.	38
Figure 2.7: Prédiction de la classe d'une nouvelle instance.	38
Figure 2.8: Matrice de confusion.	39
Figure 3.1: AVC hémorragique.	45
Figure 3.2: AVC ischémique.	45
Figure 3.3: Image avant le prétraitement.	47
Figure 3.4: Image après le prétraitement.	47
Figure 3.5: Propriétés d'ordinateur utilisé.	50
Figure 3.6: DicomWorks 1.3.5.	50
Figure 3.7: Logo Matlab.	51

Table des tableaux

Tableau 3.1: Résultats du classifieur SVM avec les noyaux Linear,Polynom et RBF.	47
Tableau 3.2: Résultats obtenus avec le KNN pour la distance Minkowski et différentes valeurs de k.	50
Tableau 3.3: Les résultats obtenus avec le modèle KNN pour la distance Euclidienne et différentes valeurs de k.	53

Tables des matières

Introduction générale.....	1
Chapitre 1 : Contexte médicale	3
1. Introduction.....	4
2. Généralités	4
2.1 Définition du cerveau	4
2.2 Fonctionnement du cerveau.....	5
2.3 Anatomie du cerveau.....	5
2.4 Apport sanguin au cerveau	6
2.5 Définition de l'AVC.....	7
2.6 Type des AVC.....	7
2.6.1 Accident vasculaire cérébrale ischémique.....	7
2.6.2 Accident vasculaire cérébrale hémorragique.....	8
2.7 Physiopathologie de l'accident vasculaire cérébrale	9
2.7.1 Physiopathologie de l'ischémie.....	9
2.7.2 Physiopathologie de l'hémorragie.....	10
2.8 Causes et facteurs de risque de l'AVC	10
2.8.1 Causes d'AVC	10
2.8.2 Facteurs de risques	11
2.9 Signes précoces de l'AVC.....	12
2.10 Taux d'incidence et les conséquences de l'AVC dans le monde.....	12
2.11 Taux d'incidence et conséquences de l'AVC en Algérie	13
2.12 Modalités d'imagerie pour le diagnostic des AVC	13
2.12.1 Scanner cérébral	14
2.12.2 Imagerie par résonance magnétique	15
2.13 Classifications et fréquence des différents types d'AVC	16
2.14 Traitement.....	16
2.15 Conclusion.....	17
Chapitre 2 : Classification supervisée dans l'imagerie	18
1. Introduction.....	19
2. Notion d'image	19
3. Processus d'obtention d'une image	20
4. Traitement d'image.....	20

5.	Prétraitement d'images.....	21
6.	Opérateurs morphologiques	22
6.1	Dilatation Morphologique	22
6.2	L'érosion morphologique	22
6.3	Ouverture Morphologique.....	22
6.4	Fermeture Morphologique.....	23
7.	Notion de texture	23
8.	Définition de Grey level co-occurrence matrix (GLCM)	24
8.1	Energie.....	25
8.2	Contraste.....	25
8.3	Entropie	25
8.4	Homogénéité	25
8.5	Corrélation.....	26
8.6	Variance.....	26
8.7	Moyenne des sommes :.....	26
8.8	Variance des sommes :	26
8.9	Entropie des sommes :.....	27
8.10	Variance des différences :.....	27
8.11	Entropie des différences :	27
8.12	Information sur la corrélation 1 :	27
8.13	Information sur la corrélation 2 :	27
8.14	Coefficient de corrélation maximal (MCC) :.....	27
9.	Classification des images.....	28
10.	Évaluation des performances d'un classifieur.....	30
10.1	Mesures de classification	30
10.1.1	Taux de classification (Accuracy)	30
10.1.2	Précision	30
10.1.5	Sensibilité	31
10.1.6	Spécificité.....	31
11.	Apprentissage et reconnaissance.....	31
12.1	Phase d'apprentissage	33
12.2	Phase de test	33
12.	Etat de l'art.....	34
13.	Conclusion	36

Chapitre3 : Conception et Implémentation	37
1. Introduction.....	38
2. Base de données de notre système	38
2.1 Collecte et annotation de la base de données	38
2.2 Description de la base de données.....	39
3. Prétraitement des images	40
4. Extraction des caractéristiques	42
5. Classification des AVC avec SVM.....	43
5.1 Présentation du SVM (Support Vector Machine).....	43
5.2 Configuration des classifieurs SVM avec différents noyaux (linéaire, polynomiale, RBF) .	43
5.3 Entraînement des modèles SVM avec les caractéristiques extraites	44
6. Classification des AVC avec KNN	44
6.1 Présentation du KNN (K-Nearest Neighbors).....	44
6.2 Configuration des classifieurs KNN avec différentes distances (euclidienne, minkowski) ..	44
6.3 Entraînement des modèles KNN avec les caractéristiques extraites	44
7. Environnement de matériel.....	44
7. Application et outils d'implémentation	45
8.1 DicomWorks 1.3.5	45
8.2MATLAB.....	46
8. Résultats et discussion	46
9.1 Evaluation des performances des modèles SVM en termes de SE, SP, VP, VN, FP, FN, TC	46
9.2 Interprétation détaillée des résultats du classifieur SVM avec les noyaux Linear, Polynom et RBF.....	47
SVM avec noyau linéaire	47
SVM avec noyau polynomial	48
SVM avec noyau RBF	48
9. Evaluation des performances des modèles KNN en termes de SE, SP, VP, VN, FP, FN, TC.....	49
10. Limitations et considérations	57
11. Conclusion	58
Interface Graphique	59
Conclusion générale	62
12. Possibilités d'amélioration et de recherche future	58
13. Bibliographies	63

Introduction générale

Les AVC sont des conditions médicales graves qui nécessitent une intervention rapide pour minimiser les dommages et sauver des vies. Cependant, leur détection précoce peut être un défi pour les professionnels de la santé, ce qui souligne l'importance d'utiliser des outils de classification efficaces pour aider à identifier les AVC dès que possible.

Le présent projet de fin d'études vise à développer et à appliquer des techniques de classification pour la détection précoce et la classification des accidents vasculaires cérébraux. Ce travail a été réalisé en utilisant une base de données locale non annotée. Afin de pouvoir effectuer la notation des données, on a sollicité l'expertise d'un médecin spécialisé dans les accidents vasculaires cérébraux. Le médecin a évalué les échantillons de la base de données et a fourni les annotations nécessaires pour la classification des AVC.

Il est important de souligner que la collaboration avec un médecin expert et l'utilisation d'une base de données locale non annotée ont contribué à la réalisation d'une étude plus complète et fiable sur la détection et la classification des AVC. Ces éléments renforcent la pertinence et l'applicabilité des résultats obtenus et renforcent la validité des conclusions formulées.

Le projet est structuré en plusieurs chapitres qui couvrent différents aspects de la détection et la classification des AVC fournissant ainsi une compréhension approfondie du contexte médical des AVC, des techniques de classification utilisées et des résultats obtenus dans la partie pratique. Le chapitre 1, intitulé "Contexte médical", fournit une introduction générale à la maladie des AVC, en mettant l'accent sur sa définition, sa prévalence et ses conséquences. Il met également en évidence les enjeux liés à la détection précoce des AVC.

Le chapitre 2, intitulé "Techniques de classification", explore de manière générale les méthodes de classification utilisées dans le domaine médical. Il décrit les étapes typiques du processus de classification, telles que la collecte et la préparation des données, l'extraction des caractéristiques et le choix des classifieurs. Deux classifieurs spécifiques, à savoir le SVM et le KNN, sont présentés en détail en raison de leur pertinence pour la classification des AVC.

Le chapitre 3, intitulé "Partie pratique", présente la mise en œuvre concrète du projet. Il décrit les étapes spécifiques suivies pour préparer les données, extraire les caractéristiques à l'aide de la méthode GLCM et appliquer les modèles de classification SVM avec le noyau RBF et

KNN avec la distance euclidienne. Les résultats obtenus sont également présentés et discutés dans ce chapitre.

En définitive, nous allons mettre une conclusion dans laquelle on rappelle les principaux résultats et on annonce une extension du point de vue initiale.

Chapitre 1 : Contexte médicale

1. Introduction

Les AVC forment un groupe de maladies hétérogènes nécessitant des traitements qui doivent être adaptés au sous-type d'AVC concerné et à la cause présumée. Ce chapitre vise à clarifier nos connaissances sur les caractéristiques du cerveau, les AVC et les différentes techniques d'imagerie cérébrale. De plus, il résumera quelques recherches dans la littérature pour la détection de l'AVC ischémique et hémorragique.

2. Généralités

2.1 Définition du cerveau

Le cerveau est l'organe principal du système nerveux central situé dans la tête. Il est responsable de la coordination et du contrôle des fonctions et comportements corporels, tels que le mouvement, la sensation, la pensée et l'émotion. Le cerveau se compose de différentes régions, chacune ayant des fonctions spécialisées, comme le cerveau, le cervelet et le tronc cérébral. Il reçoit des informations des sens et traite ces informations pour créer une réponse appropriée. Le cerveau régule également le système nerveux autonome, qui contrôle les actions involontaires telles que la respiration et la fréquence cardiaque, et joue un rôle crucial dans la conscience et la formation de la mémoire.

Le cerveau humain est composé de différents éléments. Chez un adulte moyen, il pèse environ 3 livres. Le cerveau est constitué d'environ 60% de graisse, ainsi que d'une combinaison d'eau, de protéines, de glucides et de sels qui constituent les 40% restants. Contrairement à un muscle, le cerveau inclut des vaisseaux sanguins et des nerfs, qui se composent de neurones et de cellules gliales (Voir figure 1.1) [1].



Figure 1.1: Schéma d'un cerveau humain dans sa boîte crânienne.

2.2 Fonctionnement du cerveau

En substance, le cerveau fonctionne en transmettant des signaux chimiques et électriques dans tout le corps, qui sont responsables de la régulation de différents processus corporels. Le cerveau est capable d'interpréter ces signaux, tels que ceux qui vous font sentir fatigué ou causent de la douleur. Les messages peuvent être contenus dans le cerveau lui-même ou transmis à travers la moelle épinière et le réseau de nerfs dans tout le corps. Le système nerveux central dépend de milliards de neurones pour faciliter ce processus de communication (Voir figure 1.2) [2].

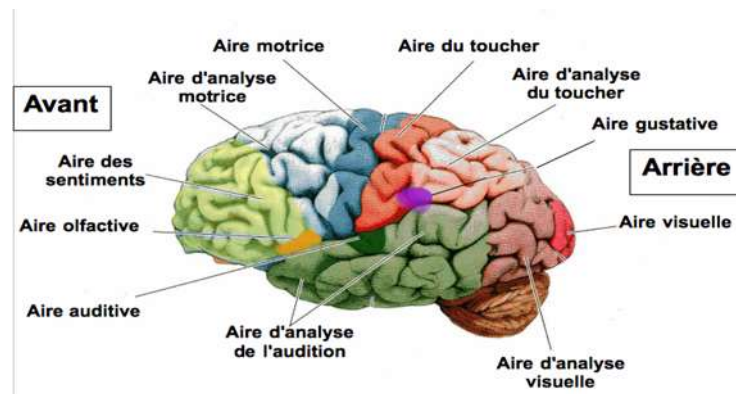


Figure 1.2: Communication nerveuse entre les organes.

2.3 Anatomie du cerveau

Le cerveau est divisé en deux hémisphères, chacun composé de quatre lobes : frontal, pariétal, temporal et occipital. Chaque lobe est responsable de fonctions spécifiques dans le cerveau. Le lobe frontal, situé à l'avant de la tête et le plus grand des lobes, joue un rôle crucial dans la prise de décisions, le mouvement et les caractéristiques de personnalité. De plus, le lobe frontal est associé à la capacité de parole et à la reconnaissance des odeurs. La partie médiane du cerveau, connue sous le nom de lobe pariétal, aide les individus à identifier les objets et à comprendre les relations spatiales. Il est également impliqué dans l'interprétation de la douleur et du toucher dans le corps. Le lobe pariétal contient l'aire de Wernicke, qui est responsable de la compréhension du langage. Situé à l'arrière du cerveau, le lobe occipital est principalement responsable de la vision. Enfin, les lobes temporaux, situés de chaque côté du cerveau, sont impliqués dans la mémoire à court terme, la parole, le rythme musical et, dans une certaine mesure, la reconnaissance des odeurs (Voir figure 1.3) [3].

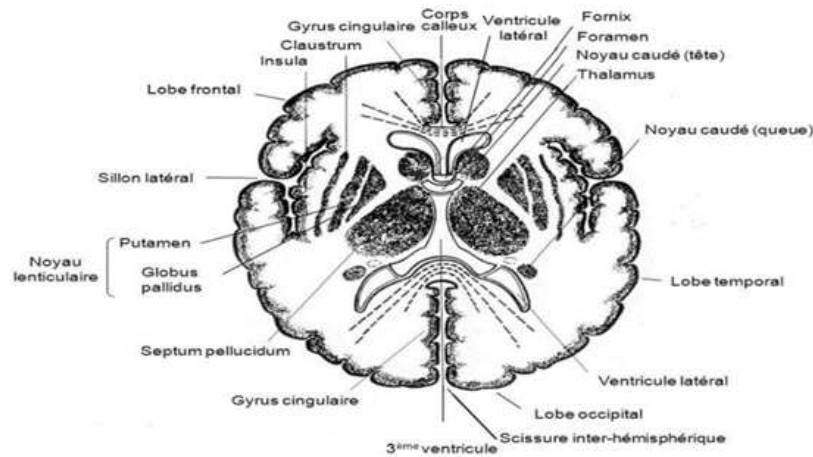


Figure 1.3: Représentation en coupe horizontale du cerveau présentant les diverses structures anatomiques.

2.4 Apport sanguin au cerveau

Le cerveau reçoit son approvisionnement sanguin à partir de deux ensembles distincts de vaisseaux sanguins, à savoir les artères carotides et les artères vertébrales. Les artères carotides externes se trouvent sur les côtés du cou et sont facilement palpables, permettant de sentir leur pouls en plaçant les doigts sur la zone. En revanche, les artères carotides internes se divisent dans le crâne et distribuent le sang à la partie antérieure du cerveau. Les artères vertébrales longent la colonne vertébrale, passent à travers le crâne pour fusionner au tronc cérébral et former l'artère basilaire. Cette artère fournit du sang aux régions postérieures du cerveau. Le cercle de Willis, qui est un groupe de vaisseaux sanguins situé près de la base du cerveau, relie les artères principales et aide à la circulation sanguine de l'avant vers l'arrière du cerveau. De plus, il favorise la communication inter artérielle entre les différents systèmes artériels (Voir figure 1.4) [3].

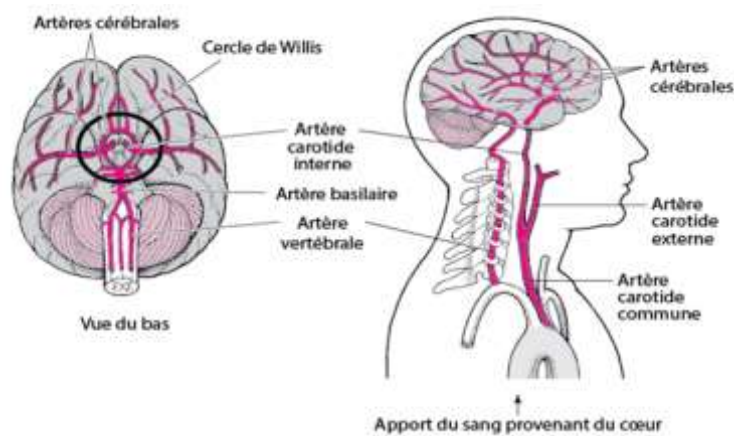


Figure 1.4: Irrigation sanguine du cerveau.

2.5 Définition de l'AVC

L'accident vasculaire cérébral (AVC), également connu sous le nom d'attaque cérébrale, se produit lorsque la circulation sanguine vers une partie du cerveau est interrompue ou réduite de manière significative. Cela peut être dû à l'obstruction d'un vaisseau sanguin par un caillot (AVC ischémique) ou à la rupture d'un vaisseau sanguin (AVC hémorragique).

2.6 Type des AVC

2.6.1 Accident vasculaire cérébrale ischémique

L'accident vasculaire cérébral ischémique survient lorsqu'il y a une réduction significative ou un arrêt de la circulation sanguine vers le cerveau en raison d'un blocage dans une artère cérébrale. Ce blocage est le plus souvent causé par une occlusion thrombotique ou embolique d'une artère, qui correspond au blocage de l'artère par un caillot de sang. Les artères plus petites ou les artères où il y a une accumulation de plaque artérioscléreuse sont plus sujettes à l'occlusion.

L'athérosclérose est un processus au cours duquel des dépôts graisseux (plaques d'athérome) envahissent la paroi interne des vaisseaux sanguins, généralement les artères principales et les zones de turbulence accrue dans la vasculature. La plaque artérioscléreuse réduit progressivement la surface disponible pour la circulation sanguine, rendant le vaisseau plus susceptible à la thrombose ou à l'embolie. Le circuit carotidien et les petites artères cérébrales profondes sont des sites privilégiés pour la thrombose et l'embolie.

Le blocage de la circulation sanguine entraîne une anoxie dans la région alimentée par l'artère et la mort des neurones de cette région. Cela entraîne la perte de fonctions cérébrales localisées dans la zone normalement alimentée par le vaisseau bloqué. Des signes focaux de lésion cérébrale indiqueront le site de l'occlusion artérielle cérébrale. Par conséquent, la personne ressentira un déficit neurologique, tel qu'une perturbation des fonctions motrices, sensorielles, perceptuelles, cognitives, affectives, sociales ou de toute autre fonction contrôlée par la zone cérébrale affectée (Voir figure 1.5) [4].

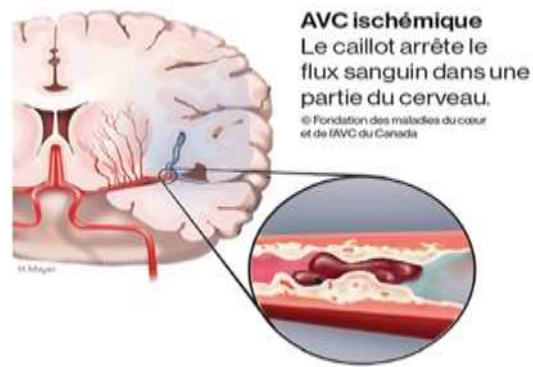


Figure 1.5:Accident vasculaire cérébrale ischémique ainsi que les régions affectées.

2.6.2 Accident vasculaire cérébrale hémorragique

Un AVC hémorragique est causé par une hémorragie dans le cerveau (hémorragie intracérébrale) ou autour du cerveau (hémorragie sous-arachnoïdienne) suite à la rupture d'un vaisseau sanguin. Les hémorragies cérébrales sont le plus souvent attribuées à une hypertension artérielle non contrôlée et, dans certains cas, à des anomalies liées à la structure des vaisseaux sanguins (comme des anévrismes ou des malformations vasculaires). Les AVC hémorragiques sont généralement trouvés dans de petites artères cérébrales athéroscléreuses affaiblies par l'hypertension chronique.

L'hémorragie entraîne une privation d'apport sanguin à la région du cerveau normalement alimentée par le vaisseau, et le sang qui s'accumule à l'extérieur du vaisseau exerce une pression supplémentaire sur les structures cérébrales, ce qui peut causer une anoxie pour les neurones dans les régions environnantes (Voir figure 1.6).

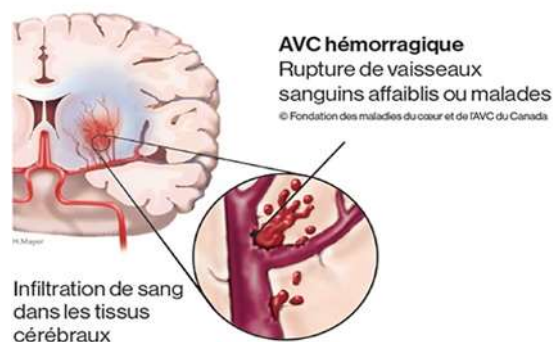


Figure 1.6:Accident vasculaire cérébrale hémorragique ainsi que les régions affectées.

Une personne subissant un AVC hémorragique ressentira une forte et soudaine douleur de tête, pourra avoir des nausées et des vomissements, perdre les fonctions cérébrales dans la région affectée, mais peut également présenter des signes de dommages à d'autres régions cérébrales comprimées. Par conséquent, les signes peuvent être plus diffus. Un AVC hémorragique important peut provoquer un coma dès le début [4].

2.7 Physiopathologie de l'accident vasculaire cérébrale

2.7.1 Physiopathologie de l'ischémie

Le cerveau humain assume la responsabilité de réguler et contrôler les fonctions corporelles. Même s'il ne représente que 2 % de notre poids total, il consomme plus de 20 % de l'oxygène que nous recevons. L'accident vasculaire cérébral ischémique se produit lorsque le flux sanguin vers le cerveau est insuffisant en raison d'un blocage ou d'une obstruction des artères qui y mènent, ce qui provoque une privation d'oxygène et de glucose nécessaires au maintien de l'équilibre cellulaire.

Cet événement déclenche divers processus qui conduisent à la mort cellulaire, tels que l'excitotoxicité, l'acidotoxicité, le déséquilibre des ions, le stress, l'inflammation et la dépolarisation péri-infarctus. La région touchée par l'ischémie subit une nécrose, caractérisée par une circulation sanguine cérébrale sévèrement restreinte ($DSC < 15 \text{ ml}/100 \text{ g}/\text{min}$) et un dysfonctionnement des mitochondries, ce qui empêche la production adéquate d'adénosine triphosphate (ATP) et entraîne la mort cellulaire en quelques minutes. La périphérie de la zone ischémique, connue sous le nom de pénombre ischémique, subit une apoptose, un processus de mort cellulaire programmée au niveau nucléaire, où une circulation sanguine collatérale ($DSC 15\text{-}20 \text{ ml}/100 \text{ g}/\text{min}$) peut atténuer les effets complets de l'accident vasculaire cérébral pendant plusieurs heures. Dans cette zone, la mort cellulaire survient moins rapidement grâce à des mécanismes de mort cellulaire active. On appelle cela le "tissu à risque", qui peut encore réagir à la reperfusion ou à une intervention thérapeutique. L'imagerie par résonance magnétique (IRM), en utilisant des techniques d'imagerie de perfusion, peut évaluer l'étendue relative de la zone à risque et le degré d'ischémie. (Voir figure 1.7) [5].

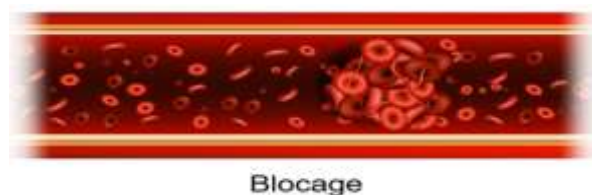


Figure 1.7:Blocage du débit sanguin.

2.7.2 Physiopathologie de l'hémorragie

Au cours d'un accident vasculaire hémorragique, l'entrée soudaine de sang dans le cerveau est causée par la rupture de petites artères à l'intérieur du cerveau. L'hypertension chronique, le tabagisme, les malformations artérielles ou la prédisposition génétique font partie des causes d'affaiblissement des artères qui peuvent conduire à une telle rupture. Dans une hémorragie sous-arachnoïdienne, la rupture spontanée d'un anévrisme dans les grandes artères à la surface du cerveau provoque des saignements dans les méninges. Cela peut endommager immédiatement le cerveau ou entraîner des complications dans les jours suivants. Dans le cas de la thrombose veineuse cérébrale, des caillots sanguins se forment à l'intérieur des veines cérébrales (pas des artères) en raison d'un épaissement du sang, par exemple à la suite de la prise de contraceptifs oraux, du tabagisme ou de facteurs génétiques. Cela peut conduire à une ischémie cérébrale ou à une hémorragie.

Les territoires vasculaires cérébraux sont responsables de l'approvisionnement en sang du cerveau et se composent de deux systèmes vasculaires : la circulation antérieure et la circulation postérieure, qui sont alimentées respectivement par l'artère carotide interne (ACI) et l'artère vertébrale (AV). Le cercle de Willis connecte l'ACI au système vertébrobasilaire. Les zones d'infarctus dépendent de divers facteurs, notamment de l'emplacement de l'obstruction dans le réseau artériel. L'artère communicante antérieure relie les deux artères cérébrales antérieures, tandis que l'artère communicante postérieure (ACoP) relie l'ACI à l'artère cérébrale postérieure (ACP). L'artère cérébrale antérieure (ACA) et l'artère cérébrale moyenne (ACM) sont les ramifications terminales de l'ACI (Voir figure 1.8) [4].



Figure 1.8:Rupture d'un vaisseau sanguin.

2.8 Causes et facteurs de risque de l'AVC

2.8.1 Causes d'AVC

L'AVC peut avoir plusieurs causes et des facteurs de risque multiples peuvent être présents chez un même patient. L'hypertension artérielle (HTA) est le principal facteur de risque

modifiable d'AVC, augmentant le risque d'infarctus cérébral jusqu'à 4 fois et d'hémorragie jusqu'à 10 fois si la pression artérielle systolique est supérieure à 160 mm Hg et la pression artérielle diastolique supérieure à 95 mm Hg (Fanny Pichard, 2014). L'HTA est prévalente chez 30 % des patients âgés de 20 à 75 ans et chez 90 % des personnes âgées de plus de 70 ans. L'hypercholestérolémie et le diabète sont également des facteurs de risque importants. Une dyslipidémie augmente le risque d'un AIC, avec un risque relatif de 1,3. Le diabète multiplie par deux ou trois l'incidence d'un AVC thrombo-embolique. Une consommation importante d'alcool augmente également le risque d'un AIC, avec un risque relatif de 1,8, tandis que la consommation de tabac favorise l'apparition de lésions athérosclérose carotidienne chez les sujets âgés.

Effectivement, l'âge et le sexe sont des facteurs de risque non-modifiables. En effet, le risque d'AVC augmente avec l'âge et les hommes sont plus susceptibles de développer un AVC que les femmes. Les antécédents familiaux d'AVC peuvent également augmenter le risque de survenue d'un AVC chez un individu. Il est important de souligner que la présence de plusieurs facteurs de risque, modifiables ou non, a un effet additif sur le risque d'AVC et nécessite une prise en charge globale et multifactorielle pour réduire ce risque.

2.8.2 Facteurs de risques

✓ Facteurs de risque inévitables

- **Âge** : l'AVC peut survenir à n'importe quel âge, même chez les enfants. Mais les personnes âgées sont plus à risque.
- **Sexe** : les AVC sont plus fréquents chez les hommes, mais les femmes en décèdent davantage.
- **Hérédité** : avoir des antécédents familiaux d'AVC.
- **Ethnie** : être d'origine africaine ou asiatique.
- AVC antérieur.

✓ Facteurs de risque contrôlables

- Tabagisme.
- Hypertension.
- Dyslipidémie : concentration très élevée de lipides dans le sang.
- Diabète.
- Maladies cardiaques.
- Troubles de la coagulation sanguine (thrombophilie).

- Maladie de l'artère carotide.
- Pilules contraceptives (à base d'oestrogènes, favorisant les thromboses).
- Accident ischémique transitoire (AIT).

2.9 Signes précoces de l'AVC

Il est important de noter que ces signes précoces d'ischémie ne sont pas spécifiques de l'AVC et peuvent être présents dans d'autres pathologies cérébrales. C'est pourquoi il est essentiel de réaliser une imagerie cérébrale pour confirmer le diagnostic d'AVC et en déterminer la nature (ischémique ou hémorragique) ainsi que l'étendue de la lésion cérébrale.

En ce qui concerne les signes cliniques précoces, il est important de reconnaître rapidement les signes de l'AVC pour une prise en charge urgente. Les signes les plus fréquents de l'AVC sont la faiblesse ou l'engourdissement d'un côté du corps, la difficulté à parler ou à comprendre la parole, la perte de vision soudaine d'un œil ou des deux, la perte d'équilibre ou de coordination, et une forte douleur de tête soudaine et inhabituelle. Il est important de noter que ces signes peuvent apparaître subitement et sans préavis. Si l'un ou plusieurs de ces signes sont présents, il est crucial de contacter immédiatement les services d'urgence pour une prise en charge rapide et efficace de l'AVC [5].

2.10 Taux d'incidence et les conséquences de l'AVC dans le monde

Chaque année, la France enregistre environ 130 000 cas d'AVC, tandis que ce chiffre s'élève à 1 400 000 en Europe et à 12 millions à l'échelle mondiale. Cette prévalence croissante est largement attribuée au vieillissement de la population. Bien que l'âge moyen des personnes touchées soit de 73 ans, il est important de noter que 25 % d'entre elles ont moins de 65 ans, révélant ainsi que cette maladie ne se limite pas uniquement aux personnes âgées.

Il est essentiel de réagir rapidement dès l'apparition des premiers symptômes, car en une minute, il est possible de perdre jusqu'à deux millions de neurones. Par conséquent, environ 15 % des personnes atteintes décèdent dans les jours suivant l'AVC, tandis que les survivants souffrent souvent de séquelles à long terme. Il convient également de souligner que les AVC sont la principale cause d'invalidité chez les adultes. (Voir figure 1.9) (PILBOX).

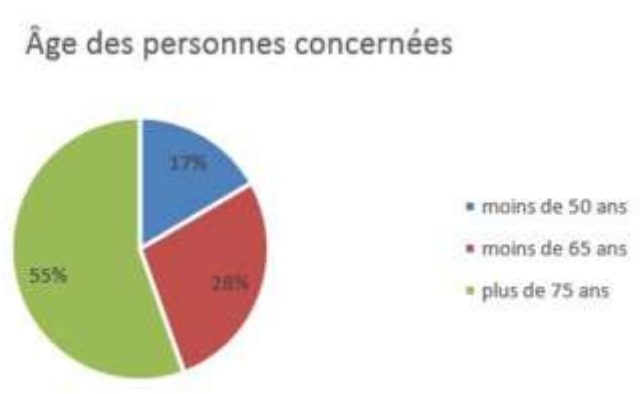


Figure 1.9:Accidents vasculaires cérébraux dans le monde.

2.11 Taux d'incidence et conséquences de l'AVC en Algérie

Selon l'Association des étudiants en pharmacie de l'université d'Alger (Asepa), en Algérie, l'accident vasculaire cérébral (AVC), connu également sous le nom d'attaque cérébrale, se positionne en tant que deuxième cause de décès chez les personnes âgées de plus de 60 ans et la cinquième cause chez les individus âgés de 15 à 59 ans. En 2018, 828 cas d'AVC ont été recensés, répartis en 85,1% d'AVC ischémiques et 13,5% d'AVC hémorragiques. Les nouveaux cas représentaient 70,5% de l'ensemble des AVC. Le taux global d'incidence des AVC était de 130,3 pour 100 000 habitants, avec un taux standardisé pour l'âge de 138,1 pour 100 000 habitants. Le taux d'incidence des hommes (143,6 pour 100 000 habitants) était significativement supérieur à celui des femmes (116,9 pour 100 000 habitants) : RR = 0,81 (IC à 95% : 0,71 ; 0,93). L'âge médian de survenue d'un AVC était de 70 ans. La létalité globale à 28 jours s'élevait à 22,1%. Le taux de mortalité global des AVC était de 28,8 pour 100 000 habitants, tandis que le taux standardisé correspondant était de 29,9 pour 100 000 habitants. Le taux de mortalité standardisé chez les hommes (34,5 pour 100 000 habitants) était significativement supérieur à celui des femmes (25,3 pour 100 000 habitants) :

SMR = 0,73 (IC à 95% : 0,54 ; 0,99). L'établissement d'un registre de population spécifique aux AVC dans la région de Blida a permis de quantifier le risque d'apparition et de décès liés à cette pathologie, et de souligner la charge de morbidité associée à ce problème de santé [6].

2.12 Modalités d'imagerie pour le diagnostic des AVC

L'accident vasculaire cérébral (AVC) représente une situation neurologique d'urgence majeure, et la perspective immédiate dépend en partie de la rapidité de la prise en charge appropriée. Il est essentiel d'acheminer rapidement le patient vers un établissement spécialisé afin de permettre un accès immédiat à l'imagerie cérébrale. Cette étape joue un rôle crucial

dans l'établissement d'un diagnostic précis et la mise en place d'une prise en charge diagnostique et thérapeutique adéquate. En effet, seule l'imagerie cérébrale permet de distinguer les deux types d'AVC.

2.12.1 Scanner cérébral

Lorsqu'il y a suspicion d'un AVC, le scanner cérébral est l'examen essentiel pour les premières étapes d'investigation. Il permet de confirmer le diagnostic d'AVC ischémique, de localiser la zone touchée, et d'éliminer d'autres causes possibles, y compris les AVC hémorragiques. Des changements dans l'image peuvent indiquer la présence de zones d'œdème cytotoxique qui témoignent d'une ischémie. Ces modifications sont généralement détectables par une altération du contraste dans la région où la matière grise et la matière blanche se rejoignent dans l'image. Le scanner cérébral doit être effectué en urgence, sans injection de produit de contraste. Repérer des altérations ischémiques subtiles sur une tomodensitométrie (TDM) sans produit de contraste représente un défi et nécessite l'expertise de radiologues expérimentés. Il est possible que les résultats soient normaux au cours des premières 48 heures (Voir figure 1.10).

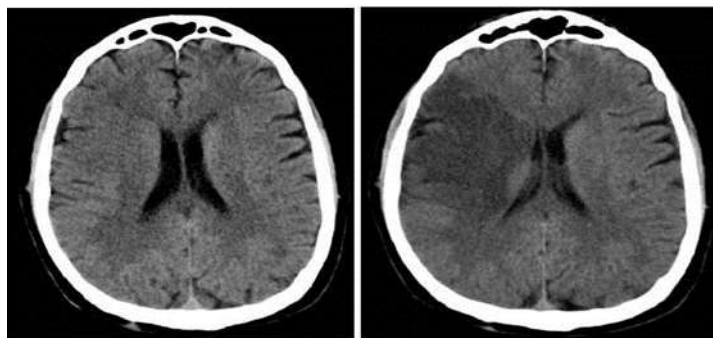


Figure 1.10: Analyse par scanner cérébral a été réalisée sur un patient présentant un AVC sylvien droit.

Une diminution de densité précoce a été observée, avec une amélioration de la visualisation après 9 heures depuis l'apparition des premiers symptômes.

Pendant les six premières heures après l'apparition des symptômes, il est souvent difficile pour le scanner de détecter les signes précoces de modifications ischémiques. Ainsi, en cas d'urgence, le scanner est principalement utilisé pour exclure toute possibilité d'hémorragie intracérébrale. L'utilisation de l'angiographie par tomodensitométrie (CT) et de la perfusion par tomodensitométrie peut aider à analyser les AVC à l'aide du scanner, mais cela nécessite

l'administration d'un agent de contraste supplémentaire et entraîne une exposition accrue aux rayonnements.

2.12.2 Imagerie par résonance magnétique

Les séquences pondérées T2 classiques sont plus sensibles que le scanner lors de la phase aiguë d'un accident ischémique artériel. En effet, environ 80% des patients présentent une IRM anormale au cours des 24 premières heures, tandis que seulement 60% ont un scanner anormal pendant cette même période.

La séquence FLAIR, grâce à l'élimination de l'hypersignal du liquide cébrospinal, est maintenant considérée comme la référence pour la pathologie ischémique aiguë. Les séquences pondérées T2, en particulier la séquence FLAIR, sont très sensibles pour détecter l'augmentation de la teneur en eau des tissus, ce qui permet de mettre en évidence l'œdème intracellulaire ou l'œdème cytotoxique, se manifestant par un hypersignal. Les hypersignaux corticaux peuvent être observés dès la 8e heure sur la séquence FLAIR. En revanche, la séquence pondérée T1 est beaucoup moins sensible à un stade précoce que la séquence FLAIR. En situation d'urgence, l'injection de gadolinium n'est pas nécessaire, sauf en cas de doute sur une lésion non ischémique ou lorsqu'une IRM de perfusion est réalisée (AIDIS).

Bien que la séquence FLAIR présente une excellente sensibilité pour détecter les anomalies dès les stades très précoces de l'AIC (dans la première heure suivant l'apparition des symptômes), son utilisation dans le diagnostic et le suivi des AIC reste limitée en raison de son coût élevé et de sa disponibilité limitée dans certains centres (Voir figure 1.11).

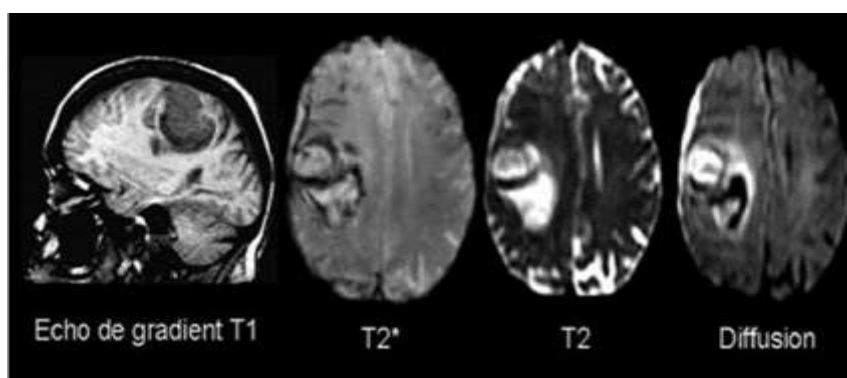


Figure 1.11: Lésion soudaine et intense provoquant un hématoème. Une imagerie par résonance magnétique (IRM) a été effectuée deux heures après l'apparition d'une faiblesse du côté gauche du corps, accompagnée de maux de tête.

Les examens mentionnés sont des examens complémentaires qui peuvent être utiles dans certains cas pour rechercher la cause de l'AVC ou pour évaluer les artères cervicales et intracrâniennes. Le Doppler transcrânien est une technique d'imagerie utilisée pour évaluer le flux sanguin dans les artères cérébrales. L'ARM (angiographie par résonance magnétique) permet de visualiser les artères intracrâniennes et cervicales sans utiliser de produit de contraste. L'angioscanner hélicoïdal est une technique d'imagerie utilisée pour évaluer les artères intracrâniennes et cervicales avec l'utilisation d'un produit de contraste. L'échocardiographie transœsophagienne est une technique d'imagerie cardiaque qui permet d'obtenir des images de l'oreillette et de l'auricule gauche ainsi que de la crosse de l'aorte pour rechercher d'éventuelles anomalies cardiaques qui pourraient être responsables de l'AVC.

En effet, la prise en charge des accidents vasculaires cérébraux (AVC) implique une recherche approfondie des causes potentielles afin de prévenir les récurrences et les complications. Les examens complémentaires peuvent varier en fonction du contexte clinique, de l'âge et des antécédents du patient.

2.13 Classifications et fréquence des différents types d'AVC

Le terme AVC est utilisé pour désigner toutes les pathologies cérébrales d'origine vasculaire, qu'elles soient artérielles ou veineuses. Il existe plusieurs types d'AVC, notamment :

- L'hémorragie cérébrale profonde, qui représente environ 15% des cas d'AVC.
- L'hémorragie sous-arachnoïdienne (HSA) et lombaire, qui sont plus rares.
- L'accident ischémique transitoire (AIT), qui constitue environ 20% des AVC.
- L'ischémie cérébrale constituée (AIC), également appelée infarctus cérébral, qui représente environ 65 à 70% des AVC.

Il est fréquent que ces différents types de pathologies soient associés. Par exemple, une HSA peut évoluer en AIC. De plus, un AIC peut parfois évoluer vers une transformation hémorragique au sein de la zone touchée par l'ischémie.

2.14 Traitement

Il est crucial que toute suspicion d'AVC soit prise en charge en urgence afin de réaliser un diagnostic précis à la fois radio-clinique et étiologique. Cela permettra de déterminer une stratégie thérapeutique spécifique en ce qui concerne l'ischémie cérébrale. Les options de traitement disponibles dépendent du délai dans lequel le patient est pris en charge.

La zone de pénombre est une zone critique pour les traitements d'urgence de l'ischémie cérébrale. Il existe deux grandes voies thérapeutiques théoriquement possibles :

- La désobstruction du vaisseau sanguin (thrombolyse).
- La prévention de la cascade de réactions biochimiques associées à l'ischémie cérébrale, qui aboutissent à la mort neuronale par afflux calcique intracellulaire (neuroprotection).

Des recherches ont prouvé que l'administration précoce du traitement par RTPA (Recombinant Tissue Plasminogen Activator) est associée à un pronostic fonctionnel favorable après 3 mois (217, 218). Certains auteurs suggèrent que les bienfaits du traitement pourraient s'étendre jusqu'à 4 heures et demie après l'apparition des symptômes (2009). Toutefois, la thrombolyse nécessite des techniques d'imagerie cérébrale spécifiques pour être effectuée de manière optimale. Le scanner, qui est l'appareil d'imagerie le plus couramment utilisé, ne permet pas un diagnostic précoce direct de l'ischémie cérébrale. Seule l'IRM, rarement disponible en situation d'urgence, peut clairement établir la nature ischémique d'un AVC dès les premières heures. De plus, la thrombolyse requiert une infrastructure adéquate pour accueillir les patients 24 heures sur 24, avec une équipe médicale permanente, une expertise en neuro-vasculaire, un accès urgent aux examens d'imagerie médicale et un radiologue spécialisé dans les pathologies neuro-vasculaires.

2.15 Conclusion

Dans ce chapitre, nous avons exploré les principes essentiels liés à la structure du cerveau, les deux types d'accidents vasculaires cérébraux (AVC) ainsi que l'importance des différentes techniques d'imagerie cérébrale, notamment la tomodensitométrie, pour diagnostiquer les AVC. Cependant, en raison de la complexité des lésions ischémiques et hémorragiques ainsi que du faible contraste des images obtenues par la tomodensitométrie, l'interprétation des images TDM représente un défi qui requiert une concentration soutenue sur une période prolongée. Dans cette perspective, notre thèse propose une méthode entièrement automatisée visant à classifier les AVC de types ischémiques et hémorragiques à partir des images obtenues par la tomodensitométrie, dans le but d'aider au diagnostic. Le deuxième chapitre présente une revue de la littérature sur les outils de traitement d'images et les recherches menées dans ce domaine d'étude.

Chapitre 2 : Classification supervisée dans l'imagerie

1. Introduction

L'analyse des accidents vasculaires cérébraux (AVC) est l'un des sujets les plus étudiés en imagerie médicale. Des méthodes automatiques et semi-automatiques sont développées pour la détection et la classification de ces AVC.

La classification automatique d'images trouve des applications dans divers domaines tels que l'analyse de documents, la médecine et le domaine militaire. Pour prendre des décisions éclairées, il est nécessaire de mettre en place une chaîne de traitement comprenant plusieurs étapes reposant sur les images disponibles. L'extraction de caractéristiques et l'apprentissage sont des méthodes indispensables à chaque étape du processus de classification.

Ce chapitre examine en détail les méthodes de classification supervisées et non supervisées, ainsi que la stratégie de détection des AVC à partir d'images. Nous commencerons par discuter de l'acquisition d'images et du prétraitement nécessaire pour améliorer la qualité des données. Ensuite, nous nous concentrerons sur la méthode GLCM (Matrice de Co-occurrence des Niveaux de Gris) et son utilisation pour capturer les relations spatiales entre les niveaux de gris dans une image. Enfin, nous aborderons les méthodes de classification les plus couramment utilisées, telles que le SVM (Machine à Vecteurs de Support) et le KNN (K Plus Proches Voisins), et expliquerons comment évaluer les performances des classifieurs. Ce chapitre fournit une compréhension approfondie des techniques et des outils nécessaires pour mener à bien la classification des AVC et permet de mieux appréhender les aspects pratiques de notre projet de détection des AVC.

2. Notion d'image

Une image peut être décrite comme une représentation visuelle d'un objet, d'une personne, d'un paysage ou d'une idée abstraite. Elle peut être créée à partir de divers médias tels que la peinture, la photographie, la sculpture, la gravure, la vidéo, et bien d'autres.

Dans le domaine informatique, une image est une représentation numérique d'une entité physique ou virtuelle, pouvant être stockée sous la forme d'un fichier image et modifiée à l'aide de logiciels de traitement d'images. En d'autres termes, une image numérique est constituée de pixels enregistrant la valeur moyenne de la zone capturée, ce qui implique l'absence d'objets continus au sens mathématique du terme.

Le concept de "image" peut également se référer à la représentation mentale d'un objet, d'un événement ou d'une idée, qui est formée à partir de l'expérience sensorielle ou de l'imagination. Cette représentation mentale peut influencer la perception, la réflexion et le comportement des personnes [7].

3. Processus d'obtention d'une image

L'objectif principal de l'acquisition d'images est de générer des représentations numériques à partir de sources physiques ou virtuelles afin de les stocker et de les traiter dans diverses applications telles que l'analyse scientifique, l'archivage, la production d'art numérique, la documentation d'éléments de preuve, la surveillance de sécurité, la reconnaissance d'images, la cartographie, la visualisation de données, et bien d'autres. Pour cela, différents dispositifs d'acquisition peuvent être utilisés, tels que des scanners ou des capteurs. Le processus d'acquisition varie en fonction du dispositif utilisé. Par exemple, lors de la numérisation d'un objet à l'aide d'un scanner, celui-ci est placé sur la vitre du scanner et balayé horizontalement et verticalement pour enregistrer l'image. Les images peuvent également être obtenues à partir de sources virtuelles, comme des modèles 3D créés par ordinateur ou des images satellite capturées à distance. Une fois l'acquisition effectuée, les images peuvent être modifiées et sauvegardées sous forme de fichiers numériques.

En créant des images numériques, il devient possible de reproduire fidèlement des objets et des scènes du monde réel afin de les étudier, de les manipuler et de les partager sous forme numérique. Les images numériques sont aisément stockables, transmissibles et modifiables sans altérer leur qualité originale.

En résumé, l'acquisition d'images permet de convertir des données visuelles en informations numériques exploitables pour une multitude d'applications dans différents domaines [8].

4. Traitement d'image

Le traitement d'images numériques englobe un ensemble de techniques utilisées pour manipuler, améliorer ou modifier des images, telles que l'amélioration de la qualité, en réduisant le bruit, en améliorant le contraste, la correction de défauts, l'extraction d'informations utiles et la création d'effets visuels. Les techniques principales comprennent le filtrage, qui permet de modifier les caractéristiques d'une image à l'aide de filtres flous, de netteté ou de détection de contours, la détection d'objets, qui permet de repérer des objets

spécifiques grâce à des algorithmes de reconnaissance faciale ou de texte, la transformation d'image, qui permet de modifier la forme, la taille ou l'orientation de l'image en utilisant des techniques de rotation, d'agrandissement ou de réduction, et enfin la reconnaissance de formes, qui permet de détecter des motifs spécifiques à l'aide d'algorithmes d'apprentissage automatique [9].

5. Prétraitement d'images

Le prétraitement d'images est une étape cruciale dans le traitement des images numériques, visant à préparer les images avant d'appliquer des techniques plus avancées. Il englobe diverses techniques pour améliorer la qualité, augmenter la netteté, améliorer la résolution et corriger les défauts tout en éliminant le bruit indésirable. Parmi les méthodes couramment utilisées, on retrouve la normalisation des couleurs et de la luminosité pour uniformiser les niveaux, le redimensionnement pour ajuster la taille, la suppression du bruit pour éliminer les interférences indésirables, ainsi que la correction de distorsion pour remédier aux déformations géométriques.

Pour le prétraitement spécifique des images de scanner cérébral, nous avons suivi les opérations suivantes : tout d'abord, l'application d'un filtre médian de noyau 3×3 aux images scanner afin de réduire le bruit. Le filtre médian est une technique de traitement d'image permettant d'éliminer les pixels de bruit en lissant l'image et en supprimant les valeurs aberrantes susceptibles de perturber le traitement ultérieur. Il fonctionne en remplaçant chaque pixel par la valeur médiane de ses pixels voisins, préservant ainsi les contours tout en supprimant le bruit. Cette méthode trouve de nombreuses applications, notamment le débruitage d'images, la détection de contours, la réduction de la taille de l'image et l'amélioration de sa qualité.

Le filtre médian est un filtre non linéaire, et son fonctionnement repose sur les étapes suivantes : la définition d'une fenêtre de voisinage autour de chaque pixel, le tri des pixels dans cette fenêtre, le calcul de la valeur médiane triée, et enfin, le remplacement de la valeur du pixel d'origine par la valeur médiane obtenue. Ces étapes sont répétées pour chaque pixel de l'image. Bien que le filtre médian puisse entraîner une perte de détails fins, il présente l'avantage de préserver les contours et les bords de l'image, le rendant ainsi plus efficace que d'autres filtres tels que le filtre moyen ou le filtre gaussien.

6. Opérateurs morphologiques

La morphologie fait référence à l'analyse de la forme et de la structure. En matière de traitement d'images, la morphologie mathématique est employée pour l'identification et l'extraction de descripteurs d'images pertinents, en se basant sur les caractéristiques de forme présentes dans l'image. Pour supprimer l'écriture des images, différents opérateurs morphologiques ont été utilisés :

6.1 Dilatation Morphologique

Pour réaliser une dilatation on translate l'élément structurant B centré au point x sur l'ensemble étudié, on se pose pour chaque point x la question : " B_x touche-t-il l'ensemble X ?", c'est à dire, y a-t-il une intersection non vide entre B_x et X ? L'ensemble des points de l'image correspondant aux réponses positives forme le nouvel ensemble Y des dilatés de X , noté comme suit :

$$\delta_B(X) = X \oplus B = \{x : B_x \cap X \neq \emptyset\} \quad (1)$$

6.2 L'érosion morphologique

L'érosion morphologique (notée $\varepsilon_B(X)$ ou $X \ominus B$) d'un ensemble X par un élément structurant B consiste à déplacer ce dernier de telle sorte que toutes les positions x de l'espace seront occupé son centre. Pour chaque position, on se demande si B est complètement inclus dans X ?

On garde un pixel si l'élément structurant centré sur ce pixel est complètement inclus dans l'ensemble de départ.

$$\varepsilon_B(X) = X \ominus B = \{x : B_x \subset X\} \quad (2)$$

6.3 Ouverture Morphologique

L'ouverture morphologique consiste à effectuer une érosion, puis une dilatation d'une image à l'aide du même élément structurant, Soit : $X \subseteq \mathbb{Z}^n$ et $B \subseteq \mathbb{Z}^n$, on définit l'ouverture de X par B comme :

$$\gamma_B(X) = X \circ B = (X \ominus B) \oplus B = \delta_B((\varepsilon_B(X))) \quad (3)$$

L'érosion permet de supprimer d'une image toutes les structures ne contenant pas l'élément structurant (les petits objets), les autres structures sont altérées (réduit la taille des objets restant).

Eviter cette altération, en dilatant le résultat de l'érosion par le même élément structurant :

Ainsi les structures éliminées par l'érosion auront disparu (les petits détails).

Les structures simplement altérées par l'érosion retrouveront de manière approximative leur forme originelle.

- La fonction imopen effectue l'érosion et la dilatation en une seule étape.
- L'ouverture morphologique est utile pour supprimer de petits objets dans une image tout en préservant la forme et la taille des objets plus gros, Elle Lisse les contours.

6.4 Fermeture Morphologique

C'est l'opération duale de l'ouverture. Elle permet la réalisation d'une dilatation suivie d'une érosion.

Soit : $X \subseteq \mathbb{Z}^n$ et $B \subseteq \mathbb{Z}^n$, on définit la *fermeture* de X par B comme :

$$\varphi_B(X) = X \bullet B = \varepsilon_B \left((\delta_B) \right) = (I \oplus E) \ominus E \quad (4)$$

A l'opposé de l'érosion, la dilatation augmente les objets et ferme les trous et fissures. L'élargissement des objets peut être évité par l'application d'une érosion avec le même élément structurant.

La fermeture morphologique est utile pour combler les petits trous d'une image tout en préservant la forme et la taille des objets de l'image [10].

7. Notion de texture

La compréhension de la texture joue un rôle crucial dans le traitement d'images. L'étape d'extraction des caractéristiques repose sur la représentation de l'ensemble de l'image par un pixel et ses attributs correspondants. Les définitions mathématiques de la texture se basent sur des propriétés statistiques et, par observation empirique, elles décrivent la texture comme un domaine homogène et cohérent de l'image. Les trois principaux types de texture sont les textures aléatoires, directionnelles et périodiques. Dans notre étude sur la détection des accidents vasculaires cérébraux (AVC) à partir d'images médicales, nous utilisons une méthode avancée appelée la Matrice de Co-occurrence des Niveaux de Gris (GLCM - Grey-Level Co-occurrence Matrix). Cette méthode est largement utilisée dans l'analyse d'images pour capturer les relations spatiales entre les niveaux de gris d'une image.

8. Définition de Grey level co-occurrence matrix (GLCM)

La matrice de cooccurrence est un outil utilisé pour estimer la probabilité d'occurrence de paires de valeurs de pixels à une distance spécifique dans une image. Son fonctionnement repose sur le calcul de la probabilité $P(i, j, \delta, \theta)$, qui représente le nombre d'occurrences où un pixel de niveau de couleur i est situé à une distance relative δ d'un pixel de niveau de couleur j , dans une orientation θ donnée. L'objectif est d'analyser les relations spatiales entre les pixels et d'extraire des informations pertinentes sur la structure de l'image (Voir figure 2.1).

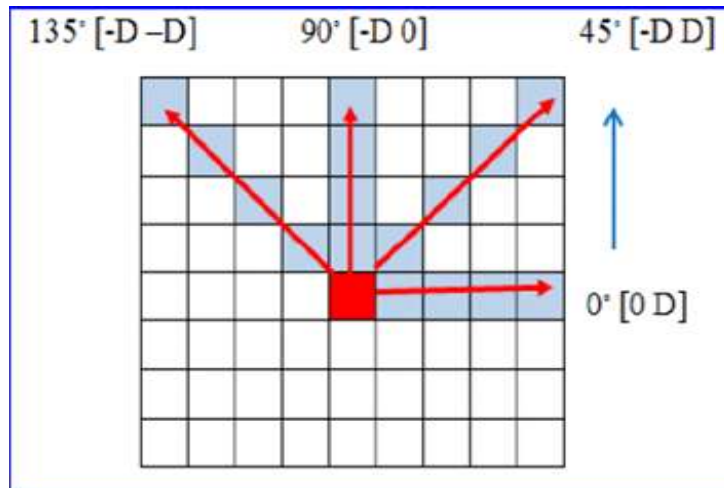


Figure 2.1: Matrice de cooccurrence pour l'extraction des caractéristiques de texture dans 4 directions et 4 distances.

Les directions angulaires couramment utilisées sont 0, 45, 90 et 135 degrés. Les relations de voisinage entre pixels, nécessaires au calcul des matrices, sont illustrées dans la figure (2).

Les caractéristiques extraites à partir de ces matrices fournissent des informations précieuses sur des aspects tels que l'homogénéité, les dépendances linéaires entre les niveaux de gris, le contraste, etc.

Les paramètres de caractérisation reposent principalement sur l'utilisation de la matrice de cooccurrence, qui est l'outil le plus populaire parmi les méthodes statiques de caractérisation des textures.

Dans ce travail, notre contribution se concentre sur la technique de calcul de la matrice de cooccurrence appliquée à nos images cérébrales. Cette approche nous permettra d'obtenir les paramètres de textures les plus pertinents. La matrice de cooccurrence contient une quantité considérable d'informations, exploitées à travers des mesures appelées attributs, paramètres ou indices de texture. Harralick a proposé des indices définis par les expressions suivantes :

8.1 Energie

L'énergie mesure le caractère régulier de la texture. De manière générale, une image très régulière est caractérisée par une énergie élevée. Dans ce cas, les valeurs élevées de la GLCM sont concentrées à quelques endroits de la matrice. C'est le cas des images dont la distribution des niveaux de gris présente soit un aspect constant, soit un aspect périodique. La GLCM d'une image aléatoire ou fortement bruitée est distribuée de manière plus uniforme et présente une énergie faible.

$$f_1 = \sum_i \sum_j P_{i,j}^2 \quad (5)$$

(i,j) désignent les coordonnées d'un élément de la matrice

P (i,j) désigne la valeur normalisée (divisée par le nb de pixels)

8.2 Contraste

Pour des GLCM présentant des valeurs plus larges en dehors de la diagonale, le contraste est plus élevé, c'est le cas des images affichant des changements locaux d'intensité [Oli,2014].

$$f_2 = \sum_i \sum_j P_{i,j} (i - j)^2 \quad (6)$$

8.3 Entropie

Une diagonale de la GLCM étalée indique une entropie très élevée, le cas extrême étant une GLCM uniforme. L'entropie est l'inverse de l'énergie et caractérise l'aspect irrégulier de l'image, d'où une corrélation forte entre ces deux attributs [Oli, 2014].

$$f_3 = \sum_i \sum_j P_{i,j} (-\ln P_{i,j}) \quad (7)$$

8.4 Homogénéité

L'homogénéité est inversement proportionnelle au contraste et est maximale lorsque les différences entre les paires de pixels analysées sont faibles. Celle-ci est donc plus sensible aux éléments diagonaux de la GLCM, contrairement au contraste qui dépend plus des éléments éloignés de la diagonale [Oli, 2014].

$$f_4 = \sum_i \sum_j \frac{P_{i,j}}{1 + (i-j)^2} \quad (8)$$

8.5 Corrélation

Elle peut s'apparenter à une mesure de la dépendance linéaire des niveaux de gris dans l'image. [Oli, 2014].

$$f_5 = \sum_i \sum_j \left[\frac{(i - \mu_i)(j - \mu_j)}{\sqrt{\sigma_i^2 \sigma_j^2}} \right] \quad (9)$$

Où :

μ_i et μ_j la moyenne des entrées normalisées pour le pixel de référence de valeur i et j

σ_i et σ_j : l'écart-type des entrées normalisées pour le pixel de référence de valeur i et j

8.6 Variance

C'est une mesure de l'hétérogénéité et est fortement corrélée à la variable statistique de premier ordre telle que l'écart-type. La variance augmente lorsque les valeurs de niveau de gris sont différents de leur moyenne [Dha, 2000].

$$f_6 = \sum_i \sum_j P_{i,j} (i - \mu_i)^2 \quad (10)$$

μ : La moyenne

8.7 Moyenne des sommes :

Il s'agit de la moyenne des valeurs de la somme des éléments de la matrice de cooccurrence des niveaux de gris. Elle représente la tendance centrale des relations de cooccurrence dans l'image.

$$f_7 = \sum_i \sum_j i P_{i,j} \quad (11)$$

8.8 Variance des sommes :

C'est la mesure de la dispersion des valeurs de la somme des éléments de la matrice de cooccurrence. Elle indique la variation des relations de cooccurrence dans l'image.

$$f_8 = \sum_i (i - f_7)^2 P_{x+y}(i) \quad (12)$$

8.9 Entropie des sommes :

L'entropie des sommes quantifie le degré d'incertitude ou de désordre dans les valeurs de la somme des éléments de la matrice de cooccurrence. Elle est utilisée pour évaluer la complexité de la texture de l'image.

$$f_9 = -\sum_i P_{x+y}(i) \log P_{x+y}(i) \quad (13)$$

8.10 Variance des différences :

L'entropie des sommes quantifie le degré d'incertitude ou de désordre dans les valeurs de la somme des éléments de la matrice de cooccurrence. Elle est utilisée pour évaluer la complexité de la texture de l'image.

$$f_{10} = \sum_i (i - f_7)^2 P_{x-y}(i) \quad (14)$$

8.11 Entropie des différences :

Cette mesure évalue l'irrégularité et le désordre des différences entre les niveaux de gris des pixels voisins dans la matrice de cooccurrence.

$$f_{11} = -\sum_i P_{x-y}(i) \log P_{x-y}(i) \quad (15)$$

8.12 Information sur la corrélation 1 :

Elle renseigne sur la corrélation entre les niveaux de gris des pixels dans la direction horizontale.

$$IMC_1 = \frac{HXY - HXY1}{\max\{HX, HY\}} \quad (16)$$

$$HXY1 = -\sum_i \sum_j P(i, j) \log(P_x(i)P_y(i)) \quad (17)$$

Où HX et HY sont respectivement les entropies de P_x et P_y

8.13 Information sur la corrélation 2 :

C'est une mesure de la corrélation entre les niveaux de gris des pixels dans la direction verticale.

$$IMC2 = \sqrt{1 - \exp[-2(HXY2 - HXY)]} \quad (18)$$

8.14 Coefficient de corrélation maximal (MCC) :

Le MCC est une mesure de la corrélation maximale entre les niveaux de gris des pixels dans la matrice de cooccurrence. Il indique le degré de relation linéaire entre les pixels voisins.

$$HXY1 = -\sum_i \sum_j P_x(i)P_x(j) \log(P_x(i)P_y(i)) \quad (19)$$

Où:

$$MCC = \sqrt{\text{la deuxième valeur propre de } Q} \quad (20)$$

$$Q_{i,j} = \sum_k \frac{g(i,k)g(j,k)}{g_x(i)g_y(k)} \quad (21)$$

9. Classification des images

La classification d'images est un processus qui vise à regrouper des images numériques dans des catégories prédéfinies, que ce soit par des moyens manuels ou à l'aide d'algorithmes informatiques comme les réseaux de neurones convolutifs. Son objectif est de sélectionner des caractéristiques ou des descripteurs de l'image qui la représentent de manière quantitative. Ces informations sont ensuite utilisées par les algorithmes de classification pour identifier et classer les images en fonction de leur contenu. La classification d'images trouve des applications dans de nombreux domaines tels que la reconnaissance faciale, la reconnaissance de caractères, la surveillance de la sécurité et la reconnaissance de formes.

9.1 Méthode des K plus proches voisins (KNN)

La méthode des K plus proches voisins, également connue sous le nom de K-nearest neighbors (KNN), est un algorithme d'apprentissage supervisé utilisé principalement pour la classification, bien qu'il puisse également résoudre des problèmes de régression. Cet algorithme s'appuie sur des données étiquetées pour entraîner un modèle, qui est ensuite utilisé pour prédire la classe associée à de nouvelles observations en calculant leur distance par rapport aux points d'entraînement. Les points d'entraînement sont enregistrés et les nouveaux points sont classifiés en fonction de leur similarité avec les données enregistrées. Voici un exemple illustrant cette méthode avec trois voisins. L'objectif est d'identifier à quelle classe appartient le point X. (Voir figure 2.2) [11].

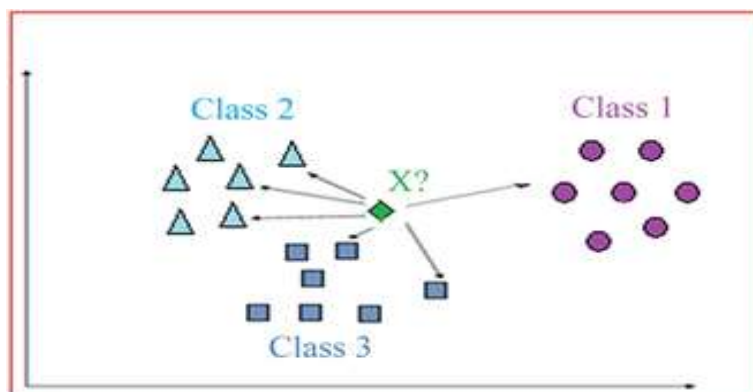


Figure 2.2: La méthode avec trois voisins pour identifier à quelle classe appartient le point X.

9.2 Machines à vecteurs de soutien (SVM)

Les machines à vecteurs de support (SVM) sont des algorithmes d'apprentissage supervisé qui ont été développés par Vapnik en 1999. Ils sont couramment utilisés pour résoudre des problèmes de classification et de régression. Les SVM sont réputés pour leur capacité à traiter efficacement des données de grande dimension, leur faible consommation de mémoire, ainsi que leur garantie théorique de trouver la meilleure frontière de décision.

Le principe des SVM repose sur la séparation des données en différentes classes à l'aide d'un hyperplan qui maximise la marge, c'est-à-dire la distance entre l'hyperplan et les points de données les plus proches de chaque classe. Ces points de données, appelés vecteurs de support, sont essentiels dans le fonctionnement des SVM.

De plus, les SVM peuvent exploiter des fonctions de noyau pour projeter les données d'entrée dans un espace de dimension supérieure, où il devient plus facile de trouver un hyperplan séparateur. Cette technique, connue sous le nom de "kernel trick", permet aux SVM de traiter des données qui ne sont pas linéairement séparables en les transformant dans un espace de dimension supérieure, où elles peuvent être séparées de manière linéaire (Voir figure 2.3).

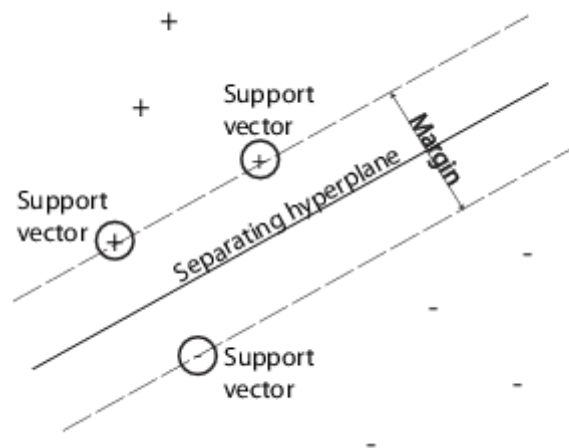


Figure 2.3: Architecture d'un vecteur support machine.

Les SVM (Support Vector Machines) sont largement employés dans divers domaines, notamment la classification d'images, la classification de texte, la bio-informatique et la finance. Leur utilité se démarque notamment lorsque le nombre de caractéristiques est élevé par rapport au nombre d'échantillons, car ils sont capables de gérer efficacement des données en haute dimension.

10.Évaluation des performances d'un classifieur

10.1 Mesures de classification

La matrice de confusion peut être utilisée pour des mesures plus approfondies afin d'obtenir une meilleure évaluation de la qualité du modèle. Plusieurs mesures peuvent être utilisées pour évaluer les performances d'un classifieur, notamment : l'accuracy, la précision, le rappel, le score F1 et la spécificité ,la sensibilité .

10.1.1 Taux de classification (Accuracy)

Il s'agit du rapport entre le nombre d'exemples correctement classés et le nombre total d'exemples. C'est une mesure globale qui donne une indication générale de la performance du classifieur. via la formule suivante :

$$T_{classification} = \frac{TP + TN}{TP + FP + FN + TN}$$

Cette mesure est utilisée lorsque le nombre de True Positive et de True Négative sont les plus important.

10.1.2 Précision

La précision correspond au nombre d'éléments corrects rendus par le modèle. Cela correspond au rapport entre le nombre de classifications positives correctes et le nombre total de prédictions positives. Elle est calculée selon la formule suivante :

$$\text{Précision} = \frac{TP}{TP + FP}$$

Cette mesure est utilisée lorsque le nombre de False Positive est le plus élevé.

10.1.3 rappel

Un rappel (recall) est une mesure de performance utilisée en apprentissage automatique et en traitement du langage naturel pour évaluer la capacité d'un modèle à identifier tous les exemples pertinents d'une classe donnée. Il est calculé comme le rapport entre le nombre d'exemples pertinents correctement identifiés et le nombre total d'exemples pertinents réels.

10.1.4 score F1

Le score F1 est une mesure de performance globale qui combine à la fois la précision (precision) et le rappel (recall) d'un modèle. Il est calculé en prenant en compte la moyenne harmonique de la précision et du rappel, donnant ainsi une mesure équilibrée entre ces deux métriques. Le score F1 est souvent utilisé lorsque la classe positive est déséquilibrée par rapport à la classe négative.

Formellement, le score F1 est calculé en utilisant la formule suivante :

$$F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$$

Précision, rappel et score F1 : Ces mesures sont souvent utilisées pour évaluer les performances d'un classifieur binaire. La précision représente la proportion de prédictions positives qui sont correctes, le rappel mesure la proportion d'exemples positifs correctement prédits, et le score F1 est une moyenne harmonique entre la précision et le rappel.

10.1.5 Sensibilité

La sensibilité détermine la proportion des valeurs positives qui ont été prédites avec précision. Elle correspond donc au rapport de nombre de prédictions positives correctes et le nombre total de classifications de classe positive. Elle est donnée par la formule suivante :

$$Se = \frac{TP}{FN + TP}$$

Cette mesure est utilisée lorsque le nombre de False Négative est le plus important.

10.1.6 Spécificité

La spécificité représente le nombre de classes négatives prédites par le modèle. Cette mesure est déterminée par le ratio entre le nombre de prédictions négatives correctes et le nombre total de prédictions négatives. Elle est donnée par la formule suivante :

$$SP = \frac{TN}{TN + FP}$$

11. Apprentissage et reconnaissance

Dans le domaine de l'apprentissage et de la reconnaissance, l'apprentissage est considéré comme supervisé lorsque les différentes étiquettes ou classes sont préalablement connues et attribuées à chaque élément de la base de données. Dans le cadre de l'apprentissage supervisé, les données sont constituées d'une collection de N objets (variables d'entrée) qui sont décrits par M caractéristiques (attributs). Ces caractéristiques fournissent des informations sur l'étiquette de chaque objet (variable de sortie ou cible). On peut représenter un ensemble de données sous forme de matrice (voir Figure 1.12), où X représente l'espace des vecteurs d'entrée avec M variables et N individus. De même, Y représente l'espace de sortie (voir Figure 2.4).

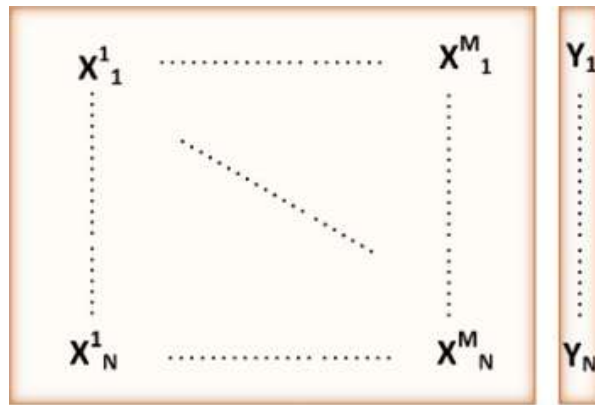


Figure 2.4: Représentation matricielle d'une base de données.

Reprenons l'exemple du problème bidimensionnel illustré dans la figure 1.13. dans lequel chaque point de cet espace représente trois patients décrits par deux variables, M1 et M2. Supposons que l'objectif de l'apprentissage soit de trouver une fonction qui permet de séparer de manière optimale les données en deux classes distinctes : "malade" ou "en bonne santé". Dans ce contexte, l'apprentissage supervisé a pour but de fournir une règle permettant à un médecin de prédire la classe d'un nouveau patient (voir figure 2.5).

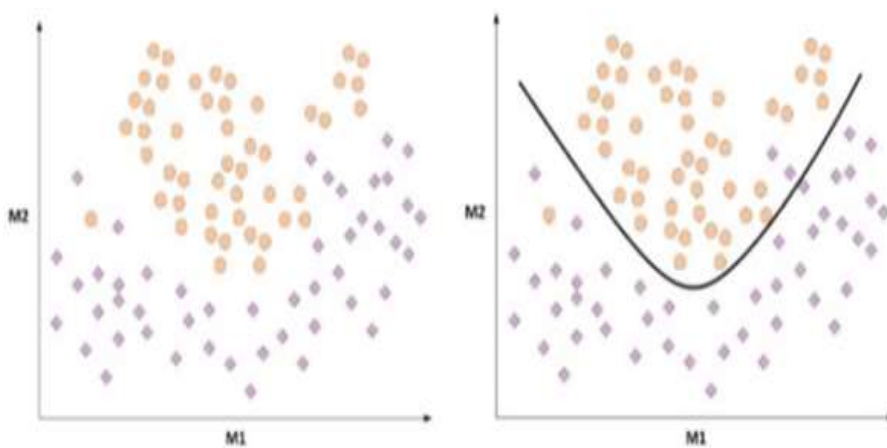


Figure 2.5: A gauche, un échantillon correspondant à un problème de classification simple. À droite, la séparation optimale de ce problème.

L'algorithme d'apprentissage prend en entrée un ensemble de données contenant les informations nécessaires pour caractériser un problème donné et renvoie un modèle qui représente les concepts caractérisant ces données et qui doit être capable de prédire l'étiquette de nouveaux objets en fonction de leurs valeurs d'entrée. En suivant ce processus

d'apprentissage et de test, il est possible de développer et de valider des modèles d'apprentissage automatique efficaces [12].

12.1 Phase d'apprentissage

Dans cette phase, un algorithme d'apprentissage automatique est entraîné sur une base de données étiquetées, au cours de cette étape, le modèle peut améliorer ses performances en se guidant avec une sous partie de la base d'apprentissage appelée base de validation (Voir figure 2.6).

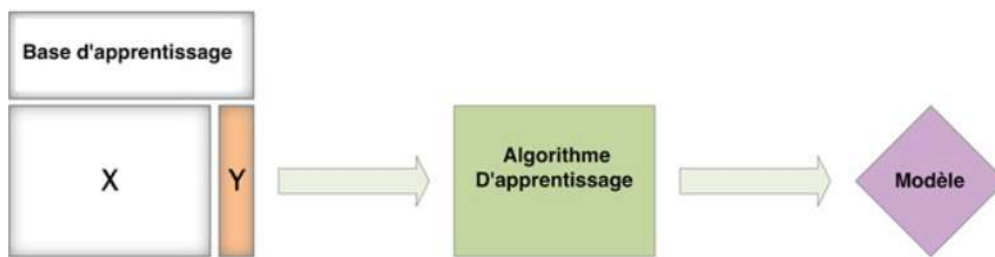


Figure 2.6: Apprentissage d'un modèle à partir d'un ensemble d'apprentissage.

12.2 Phase de test

Au moment du test, dans le cadre des problèmes de classification, le modèle est censé attribuer une classe à une instance de test en se basant sur sa valeur d'entrée, en respectant les règles qu'il a apprises lors de l'étape d'apprentissage (Voir figure 2.7).

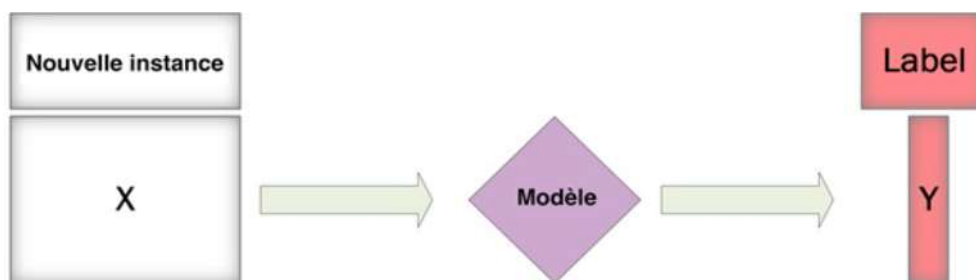


Figure 2.7: Prédiction de la classe d'une nouvelle instance.

La matrice de confusion est un outil qui établit une relation entre les décisions prises par un classifieur et les étiquettes attribuées aux échantillons. Elle constitue une mesure de la qualité d'un système de classification. Comme illustré dans la figure 2.1, la diagonale de la matrice de confusion représente les exemples correctement classés, tandis que les autres éléments

correspondent aux exemples mal classés. La matrice de confusion est un paramètre d'évaluation qui prend en compte à la fois la précision de la classification et la répartition des différentes classes (Voir figure 2.8).

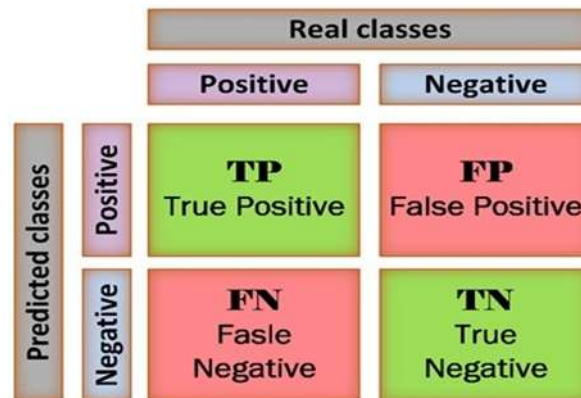


Figure 2.8:Matrice de confusion.

Sachant que, par exemple, dans le problème de diagnostic médical :

VP (TP en anglais) : représente le nombre de personnes malades classifiées malades.

FP (FP en anglais) : représente le nombre d'individus en bonne santé classés comme malades.

FN (FN en anglais) : représente le nombre d'individus malades classés comme patients en bonne santé.

VN (TN en anglais) : représente le nombre d'individus en bonne santé classés en bonne santé.

Les matrices de confusion sont conçues pour donner plus de détails sur la classification des échantillons d'une classe donnée. A partir d'une matrice de confusion, nous pouvons calculer des mesures statistiques telles que le Taux de Classification (TC), c'est une mesure très utilisée pour évaluer les performances des classifieurs.

$$TC = \frac{TP + TN}{TP + FP + FN + TN}$$

12. Etat de l'art

La classification des AVC en types ischémiques et hémorragiques est un domaine de recherche actif depuis des décennies, car elle a des implications importantes pour la prise en charge des patients atteints d'AVC. Des chercheurs du monde entier ont travaillé sur le

développement des algorithmes de classification précis pour différencier les AVC ischémiques et les AVC hémorragiques. Parmi ces méthodes:

1- Yahiaoui Amina a utilisé une nouvelle méthode automatique de détection des lésions ischémiques cérébrales précoces sur des images de tomodensitométrie, en se basant sur la symétrie entre les hémisphères cérébraux.

Les caractéristiques obtenues à partir de dix régions d'intérêt (ROI) ont été utilisées pour sélectionner les régions anormales et calculer le score ASPECTS correspondant, La méthode est satisfaisante avec une sensibilité de 98.8% [13].

2- Aymen BOUGUERRA a Appliqué la technique émergente de classification d'images à base d'apprentissage profond « réseau neuronal convolutionnel » à une large base de données d'images d'AVC segmentées à la main « ATLAS », avec une précision de validation de 95% [14].

3- Anjali Gautam a présenté une nouvelle méthode pour classer les images de tomodensitométrie (CT) cérébrales en images normales, en cas d'accident vasculaire cérébral ischémique ou hémorragique. Pour ce faire, un extracteur de caractéristiques est utilisé pour extraire des informations de l'image et générer un vecteur de caractéristiques appelé "descripteur de motif de voisinage local (LNP)" en utilisant les valeurs binaires de l'image. Cette méthode est comparée à d'autres descripteurs de caractéristiques tels que le "descripteur de modèle binaire local (LBP)", le "descripteur de modèle ternaire local (LTP)", le "descripteur de modèle d'ondelettes local (LWP)" et le "descripteur de motifs d'extrêmes diagonaux locaux (LDEP)" pour déterminer sa précision de classification. Les résultats des expériences montrent que le descripteur de caractéristiques proposé est plus efficace que les autres descripteurs de caractéristiques à la pointe de la technologie pour la classification des images de tomodensitométrie cérébrales, La performance de cette nouvelle méthode est tout à fait satisfaisante avec une sensibilité de 97.6% [15].

4- Asit Subudhi a principalement classé les AVC en syndrome de circulation antérieure partielle (PACS), syndrome lacunaire (LACS) et l'AVC total de la circulation antérieure (TAC). La partie du cerveau affectée par un AVC était segmentée à l'aide de l'algorithme de maximisation des attentes (EM) et puis la région segmentée a été traitée plus avant avec l'optimisation de l'essaim de particules darwiniennes d'ordre fractionnaire (FODPSO) afin d'améliorer la précision de détection. Un total de 192 scans de L'IRM a été considérée pour l'évaluation. Différentes caractéristiques morphologiques et statistiques ont été extraites des lésions segmentées pour former un ensemble de caractéristiques qui a ensuite été classé avec

des classificateurs de machine à vecteurs de support (SVM) et de forêt aléatoire (RF) avec une précision de validation de 93.4% [16].

5- U Rajendra Acharya a proposé un système basé sur l'extraction de l'entropie bispectre d'ordre supérieur et de ses caractéristiques de phase à partir de l'IRM cérébrale (Imagerie Magnétique par résonance). Pour la classification, qui sert à établir la gravité de l'AVC, un support machine vectorielle a été incorporé dans la conception. La technique développée a détecté efficacement l'accident vasculaire cérébral et atteint une sensibilité, une spécificité, une précision et une valeur prédictive positive égales à 96,4 %, 100 %, 97,6 % et 100 %, respectivement. Les résultats ont été obtenus sans avoir besoin de manuel d'intervention. Cette conception est avantageuse par rapport à la détection automatisée de la gravité de l'AVC par des systèmes à la technologie de pointe qui nécessitent que le neuroradiologue lecteur détermine manuellement la région d'intérêt. Par conséquent, le procédé est efficace pour fournir une aide à la décision dans le diagnostic de l'AVC ischémique aidant ainsi le neuroradiologue dans les procédures de dépistage de routine [17].

13. Conclusion

Ce chapitre a essentiellement été consacré à l'exposé de la méthodologie mise en œuvre dans cette étude pour la détection des différents types d'AVC sur les images de scanner cérébral et sur la classification de ces images. nous avons établi les fondements nécessaires pour la mise en œuvre pratique du projet et la réalisation d'une détection précise et efficace des AVC. Les connaissances acquises dans ce chapitre seront appliquées dans le chapitre suivant, où nous mettrons en pratique ces techniques pour analyser les données et obtenir des résultats concrets.

Chapitre3 : Conception et Implémentation

1. Introduction

Ce chapitre présente la partie pratique de notre projet de fin d'études axé sur la classification des AVC en utilisant la méthode GLCM avec les algorithmes SVM et KNN. L'objectif est d'améliorer la détection précoce des AVC en utilisant des caractéristiques extraites à partir d'images médicales. La méthode GLCM est utilisée pour capturer les informations sur la distribution spatiale des niveaux de gris des pixels. L'algorithme SVM vise à trouver un hyperplan optimal pour séparer les données en différentes classes, tandis que le KNN attribue une classe basée sur les échantillons voisins les plus proches. Ce chapitre détaille, également, les étapes de prétraitement des données, l'extraction des caractéristiques avec GLCM, et l'application de SVM et KNN pour la classification des AVC. Les performances des deux approches sont discutées pour objectif de contribuer aux progrès de la détection précoce et de la classification des AVC.

2. Base de données de notre système

2.1 Collecte et annotation de la base de données

La collecte de données et l'annotation sont des étapes essentielles dans tout projet de classification, et dans le contexte de notre projet de fin d'études sur la classification des AVC, elles revêtent une importance particulière. Cette partie du projet vise à décrire en détail la méthode utilisée pour collecter les données ainsi que le processus d'annotation qui a été suivi.

Pour commencer, nous avons effectué une collecte rigoureuse de données médicales spécifiques aux AVC. Nous avons consulté diverses sources, telles que des bases de données médicales en ligne, des hôpitaux et des cliniques spécialisées dans le traitement des AVC. Cette phase a nécessité plusieurs mois de recherche intensive et des efforts soutenus pour rassembler un ensemble de données suffisamment représentatif.

Cependant, la collecte de données n'a pas été sans obstacles. Nous avons rencontré divers obstacles administratifs lors du dépôt d'un dossier spécial afin d'obtenir l'accès aux données médicales nécessaires. Ces obstacles ont été surmontés grâce à une correspondance approfondie entre le département de génie biomédical et l'hôpital Meslem Al-Tayeb de Mascara, qui nous a accordé les autorisations nécessaires pour accéder aux images médicales des patients atteints d'AVC. Tout au long de la collecte de données et de l'annotation, nous avons strictement respecté les protocoles et les réglementations éthiques en matière de

recherche médicale. La confidentialité des patients a été préservée à tout moment, et les données ont été utilisées de manière responsable et sécurisée.

Une fois que les images médicales nécessaires soient obtenues, nous avons procédé à l'annotation de ces images en collaboration avec un médecin expert « Manseri Nabil ». L'expertise du médecin dans le domaine des AVC a été d'une valeur inestimable pour garantir l'exactitude et la fiabilité de l'annotation. Le médecin a examiné attentivement chaque image et a identifié la présence ou l'absence d'une lésion due à un AVC. Cette étape d'annotation a permis de créer un ensemble de données annotées qui servira de base pour l'entraînement et la validation de nos modèles de classification.

En utilisant des données réelles provenant de patients atteints d'AVC et en collaborant avec un médecin expert, nous nous assurons que nos modèles de classification sont basés sur des situations cliniques réelles. Cela améliore la validité et la pertinence de nos résultats et renforce la fiabilité de notre approche.

En résumé, la collecte de données et l'annotation sont des étapes essentielles de notre projet de classification des AVC. Elles nous ont permis d'obtenir un ensemble de données annotées précieuses, basées sur des images médicales réelles, et d'établir une collaboration fructueuse avec un médecin expert. Ces éléments contribuent à la qualité, à la validité et à la pertinence de notre approche de classification des AVC, renforçant ainsi la valeur ajoutée de notre projet.

2.2 Description de la base de données

La base de données utilisée dans notre projet de classification des AVC est composée de 360 images médicales. Après avoir procédé à l'annotation de la base de données, nous avons effectué un processus de lissage pour ne conserver que les images où la lésion était clairement visible. Cela a réduit le nombre d'images à 266, comprenant à la fois des AVC hémorragiques et ischémiques.

Parmi les 266 images, il y avait 133 cas d'AVC hémorragiques et 133 cas d'AVC ischémiques. Cela a permis d'avoir un échantillon équilibré pour chaque type d'AVC, ce qui est essentiel pour une classification précise et fiable.

Les images médicales étaient initialement au format DICOM, qui est un format standard utilisé pour les images médicales. Cependant, nous avons rencontré des difficultés lors de la conversion des images DICOM vers d'autres formats. Nous avons essayé d'utiliser un site web

et un programme Matlab pour effectuer cette conversion, mais malheureusement, nous avons perdu une partie des informations pendant le processus.

Pour résoudre ce problème, nous avons utilisé un logiciel spécifique, appelé logiciel DICOM, qui nous a permis de convertir les images DICOM en format JPG. Cette conversion image par image nous a permis de conserver toutes les informations essentielles de chaque image, y compris les niveaux de gris et les détails anatomiques pertinents.

En résumé, la base de données utilisée dans notre projet est composée de 266 images médicales d'AVC hémorragiques et ischémiques. Les images ont été lissées pour ne conserver que celles où la lésion était clairement visible. Initialement au format DICOM, les images ont été converties en format JPG en utilisant le logiciel DICOM. Cette base de données préparée et traitée constitue un ensemble d'images représentatives des différents types d'AVC, et elle servira de base pour notre étude de classification (Voir figure 3.1, 3.2).

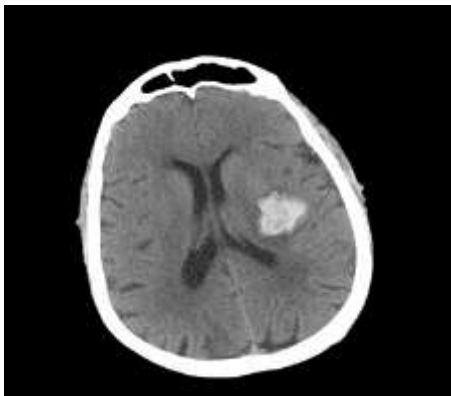


Figure 3.1:AVC hémorragique

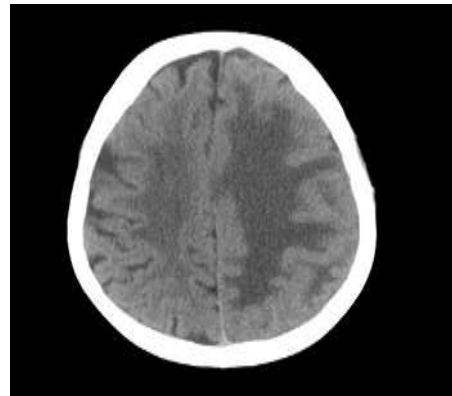


Figure 3.2:AVC ischémique

3. Prétraitement des images

Dans cette partie de notre projet, nous nous concentrons sur le prétraitement des images avant de procéder à l'extraction des caractéristiques et à la classification des AVC. Le prétraitement joue un rôle essentiel en améliorant la qualité de l'image et en préparant les données pour une analyse plus approfondie.

Voici les étapes détaillées du prétraitement des images que nous avons effectuées :

1. Chargement des images CT : Nous commençons par charger les images CT à partir de notre base de données. Ces images sont des données médicales spécifiques aux AVC que nous utilisons dans notre projet.

2. Conversion en niveaux de gris : Pour simplifier les opérations de traitement, nous convertissons les images en niveaux de gris. Cela nous permet de travailler avec une seule composante de couleur et de simplifier les calculs ultérieurs.

3. Binarisation de l'image : Nous appliquons un seuil à l'image pour la binariser. Cette étape consiste à convertir l'image en une image binaire, où les pixels d'intérêt (représentant les tissus cérébraux et les lésions) sont séparés des pixels de fond. Cette binarisation nous permet de segmenter l'image et de focaliser notre attention sur les régions d'intérêt.

4. Remplissage des trous : Une fois que l'image soit binarisée, nous procédons au remplissage des trous dans l'image. Cela consiste à identifier les régions connexes non remplies et à les remplir pour obtenir une segmentation plus complète et précise de l'image.

5. Opérations morphologiques : Pour éliminer les artefacts indésirables et améliorer la qualité de l'image binaire, nous appliquons des opérations morphologiques telles que l'érosion, la dilatation, l'ouverture et la fermeture. Ces opérations permettent de modifier la forme et la taille des objets dans l'image, ce qui nous aide à éliminer les détails indésirables et à améliorer la clarté de l'image.

6. Suppression des petits objets : Nous supprimons également les petits objets basés sur leur aire dans l'image binaire. Cela nous permet d'éliminer les petites zones indésirables qui peuvent être le résultat du bruit ou d'autres artefacts.

7. Sélection des tissus cérébraux : À l'aide du masque binaire obtenu après le prétraitement, nous sélectionnons les tissus cérébraux dans l'image d'origine. Cela nous permet de nous concentrer uniquement sur la région d'intérêt, c'est-à-dire les tissus cérébraux et les éventuelles lésions.

8. Affichage de l'image nettoyée : Enfin, nous affichons l'image nettoyée résultant du prétraitement. Cela nous permet de visualiser le résultat du prétraitement et de vérifier si les étapes précédentes ont été effectuées correctement.

Le prétraitement des images joue un rôle crucial en réduisant le bruit, en éliminant les artefacts indésirables et en mettant en évidence la région d'intérêt. Cela facilite grandement l'extraction des caractéristiques pertinentes pour la classification des AVC (Voir figure 3.3, 3.4).



Figure 3.3:Image avant le prétraitement.

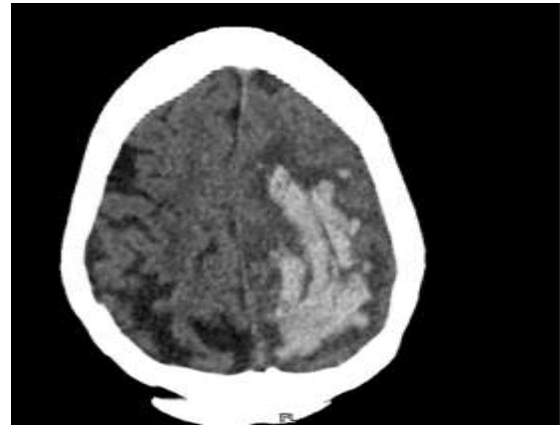


Figure 3.4:Image après le prétraitement.

4. Extraction des caractéristiques

Dans cette partie de notre travail, nous avons utilisé la méthode GLCM pour extraire des caractéristiques significatives à partir des images AVC prétraitées. La GLCM est une technique couramment utilisée dans l'analyse d'images pour capturer les relations spatiales entre les niveaux de gris des pixels.

Le processus d'extraction des caractéristiques avec la GLCM se déroule comme suit :

1. Matrice de co-occurrence des niveaux de gris : À partir des images prétraitées, nous calculons la matrice de co-occurrence des niveaux de gris. Cette matrice représente la fréquence d'apparition des paires de niveaux de gris adjacents dans une image.

2. Calcul des caractéristiques : À partir de la matrice de co-occurrence normalisée, nous extrayons les cinq paramètres caractéristiques suivants :

- Énergie : Il mesure la distribution de l'énergie des niveaux de gris dans l'image. Une valeur élevée indique une texture plus homogène et régulière.

- Entropie : Elle mesure la complexité et l'irrégularité de la texture de l'image. Une entropie élevée indique une texture plus complexe et variée.

- Variance : Elle mesure la dispersion des niveaux de gris dans l'image. Une variance plus élevée indique une texture plus hétérogène.

- Homogénéité : Elle mesure la proximité des niveaux de gris dans l'image. Une valeur élevée indique une texture plus uniforme.

- Contraste : Il mesure la différence entre les niveaux de gris adjacents dans l'image. Un contraste élevé indique une texture plus distincte et contrastée.

Ces paramètres fournissent des informations sur la texture de l'image et sont utilisés comme descripteurs pour la classification des AVC.

3. Calcul des caractéristiques pour chaque image : Nous répétons le processus de calcul des caractéristiques pour chaque image de la base de données. Ainsi, nous obtenons un ensemble de valeurs de caractéristiques pour chaque image.

L'extraction des caractéristiques avec la méthode GLCM nous permet de quantifier les propriétés de texture des images AVC. Ces caractéristiques jouent un rôle crucial dans la distinction entre les AVC hémorragiques et ischémiques lors de l'étape de classification. En capturant les informations texturales importantes, elles améliorent la précision et la performance globale du système de classification.

5. Classification des AVC avec SVM

5.1 Présentation du SVM (Support Vector Machine)

Le SVM (Support Vector Machine) est un algorithme d'apprentissage supervisé utilisé pour la classification et la régression. Il cherche à trouver un hyperplan optimal qui sépare les données en classes distinctes. Dans notre cas, nous avons utilisé le SVM pour la classification des AVC hémorragiques et ischémiques.

5.2 Configuration des classifieurs SVM avec différents noyaux (linéaire, polynomiale, RBF)

Nous avons configuré les classifieurs SVM en utilisant trois différents noyaux : linéaire, polynomiale et RBF (Radial Basis Function).

- Le noyau linéaire crée un hyperplan linéaire pour séparer les données en classes. Il est généralement utilisé lorsque les données sont linéairement séparables.
- Le noyau polynomiale utilise des fonctions polynomiales pour créer des frontières non linéaires entre les classes. Il est plus adapté aux données présentant des relations non linéaires.
- Le noyau RBF utilise des fonctions de base radiale pour créer des frontières non linéaires plus complexes. Il est capable de capturer des relations spatiales plus complexes dans les données.

5.3 Entraînement des modèles SVM avec les caractéristiques extraites

Nous avons entraîné des modèles SVM en utilisant les caractéristiques extraites à partir des images AVC, telles que l'énergie, le contraste, l'entropie, la variance et l'homogénéité. Ces caractéristiques ont été utilisées comme descripteurs pour entraîner les modèles SVM.

6. Classification des AVC avec KNN

6.1 Présentation du KNN (K-Nearest Neighbors)

Le KNN est un algorithme de classification basé sur l'idée que les échantillons similaires ont tendance à appartenir à la même classe. Il fonctionne en recherchant les k échantillons les plus proches dans l'espace des caractéristiques et en attribuant une classe à l'échantillon cible en fonction de la classe majoritaire parmi ses k voisins les plus proches.

6.2 Configuration des classifieurs KNN avec différentes distances (euclidienne, minkowski)

Dans cette étude, nous avons utilisé différentes distances pour mesurer la similarité entre les échantillons. Nous avons évalué les performances du KNN en utilisant les distances Minkowski et Euclidienne.

6.3 Entraînement des modèles KNN avec les caractéristiques extraites

Nous avons entraîné les modèles KNN en utilisant les caractéristiques extraites de notre ensemble de données spécifique aux AVC. Les caractéristiques sélectionnées étaient pertinentes pour la classification des types d'AVC telles que l'énergie, le contraste, l'entropie, la variance et l'homogénéité.

7. Environnement de matériel

On a utilisé dans notre travail un ordinateur (Voir figure 3.5) ayant les propriétés suivantes :

- processeur Intel® Core™ i7-3770CPU3 .40GHZ .
- Mémoire installée(RAM) de taille 8.00 Go.
- Système d'exploitation (64 bits).

- Windows 7 professionnel.

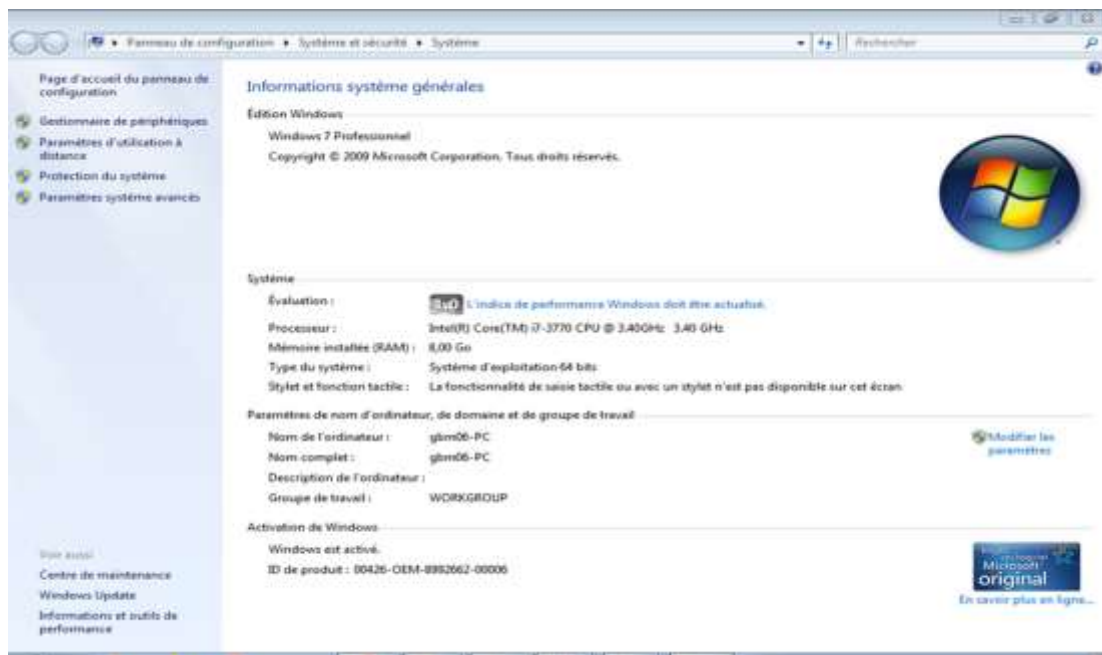


Figure 3.5: Propriétés d'ordinateur utilisé.

7. Application et outils d'implémentation

7.1 DicomWorks 1.3.5

DicomWorks 1.3.5 est une application logicielle spécialisée dans la visualisation, la manipulation et la gestion des images médicales au format DICOM. Elle offre une interface conviviale pour afficher et manipuler les images DICOM, avec des fonctionnalités avancées telles que le zoom, la rotation, l'ajustement de contraste, les annotations et les mesures. L'application permet également la conversion des images en d'autres formats courants, la reconstruction 3D, la fusion d'images, la comparaison côte à côte et la création de rapports d'imagerie. Elle facilite également l'organisation et la gestion des données DICOM, y compris la recherche et l'archivage sécurisé des images. En résumé, DicomWorks 1.3.5 est un outil puissant et convivial pour les professionnels de l'imagerie médicale, offrant des fonctionnalités avancées et une gestion efficace des données (Voir figure 3.6).



Figure 3.6: DicomWorks 1.3.5.

7.2 MATLAB

MATLAB est un environnement de programmation et un langage de programmation largement utilisé pour effectuer des calculs numériques, des analyses de données, de la modélisation et de la simulation dans divers domaines scientifiques et d'ingénierie. Il offre une syntaxe intuitive et des fonctionnalités intégrées pour manipuler des matrices, effectuer des opérations mathématiques avancées, visualiser des données, traiter des signaux et des images, développer des algorithmes, et plus encore. MATLAB 2015 est une version spécifique du logiciel qui apporte des améliorations de performance, de nouvelles fonctionnalités et des boîtes à outils spécialisées. Nous avons utilisé MATLAB pour développer notre programme en raison de sa polyvalence et de ses fonctionnalités adaptées à nos besoins (Voir figure 3.7).

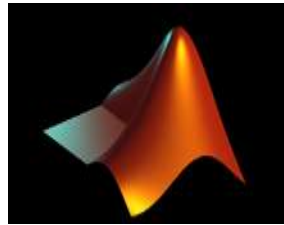


Figure 3.7: Logo Matlab.

8. Résultats et discussion

Evaluation des performances des modèles SVM en termes de SE, SP, VP, VN, FP, FN, TC

Nous avons évalué les performances des modèles SVM en utilisant plusieurs métriques pour mesurer la qualité de la classification des AVC.

- SE (Sensitivity) : Il représente le taux de vrais positifs, c'est-à-dire la proportion d'AVC ischémiques réellement détectés par le modèle.
- SP (Specificity) : Il représente le taux de vrais négatifs, c'est-à-dire la proportion d'AVC hémorragiques réellement identifiés comme tels par le modèle.
- VP (True Positive) : Il représente le nombre d'AVC ischémiques correctement classifiés.
- VN (True Negative) : Il représente le nombre d'AVC hémorragiques correctement classifiés.

- FP (False Positive) : Il représente le nombre d'AVC hémorragiques incorrectement classifiés comme ischémiques.

- FN (False Negative) : Il représente le nombre d'AVC ischémiques incorrectement classifiés comme hémorragiques.

- TC (Taux de classification) : Il représente le taux global de classification correcte des AVC.

Après l'évaluation des performances des modèles SVM avec les différents noyaux, voici les résultats obtenus :

	SE	SP	VP	VN	FP	FN	TC
Linear	1	0.7627	31	45	14	0	0.8444
Polynom	0.9535	0.9149	41	43	4	2	0.9333
RBF	0.9556	0.9556	43	43	2	2	0.9556

Tableau 3.1: Résultats du classifieur SVM avec les noyaux Linear, Polynom et RBF.

Interprétation détaillée des résultats du classifieur SVM avec les noyaux Linear, Polynom et RBF

SVM avec noyau linéaire

Sensibilité (SE) : 1

Spécificité (SP) : 0.7627

Taux de vrais positifs (VP) : 31 (AVC ischémiques correctement détectés)

Taux de vrais négatifs (VN) : 45 (AVC hémorragiques correctement détectés)

Taux de faux positifs (FP) : 14 (échantillons hémorragiques classés comme ischémiques)

Taux de classification (TC) : 0.8444

Le noyau linéaire a une sensibilité parfaite de 1, ce qui signifie qu'il a correctement détecté tous les AVC ischémiques de l'ensemble de données. Cependant, la spécificité est relativement faible à 0.7627, ce qui indique que le modèle a classé incorrectement certains échantillons hémorragiques comme étant ischémiques. Le taux de classification global est de 0.8444, ce qui montre une performance modérée du modèle.

SVM avec noyau polynomial

Sensibilité (SE) : 0.9535

Spécificité (SP) : 0.9149

Taux de vrais positifs (VP) : 41 (AVC ischémiques correctement détectés)

Taux de vrais négatifs (VN) : 43 (AVC hémorragiques correctement détectés)

Taux de faux positifs (FP) : 4 (échantillons hémorragiques classés comme ischémiques)

Taux de faux négatifs (FN) : 2 (AVC hémorragiques classés comme ischémiques)

Taux de classification (TC) : 0.9333

Le noyau polynomial présente une sensibilité élevée de 0.9535, indiquant qu'il a correctement détecté la plupart des AVC ischémiques. La spécificité est également élevée à 0.9149, ce qui suggère que le modèle a correctement classé un grand nombre d'échantillons comme étant hémorragiques. Cependant, il y a eu quelques erreurs de classification avec des échantillons AVC hémorragiques classés comme ischémiques. Le taux de classification global est de 0.9333, montrant une performance prometteuse du modèle avec ce noyau.

SVM avec noyau RBF

Sensibilité (SE) : 0.9556

Spécificité (SP) : 0.9556

Taux de vrais positifs (VP) : 43 (AVC ischémiques correctement détectés)

Taux de vrais négatifs (VN) : 43 (AVC hémorragiques correctement détectés)

Taux de faux positifs (FP) : 2 (échantillons hémorragiques classés comme AVC ischémiques)

Taux de faux négatifs (FN) : 2 (AVC ischémiques classés comme hémorragiques)

Taux de classification (TC) : 0.9556

Le noyau RBF a obtenu des performances élevées avec une sensibilité et une spécificité de 0.9556. Cela indique que le modèle a correctement détecté la plupart des AVC ischémiques et hémorragiques. Le taux de classification global est également de 0.9556, démontrant une performance élevée du modèle avec ce noyau.

En comparant les performances des modèles SVM avec les différents noyaux, nous pouvons conclure que le noyau RBF donne les meilleurs résultats. Il présente à la fois une haute sensibilité (SE), une haute spécificité (SP) et un taux de classification élevé (TC). Cela signifie qu'il est capable de détecter avec précision à la fois les AVC hémorragiques et les AVC ischémiques.

Le noyau RBF offre une meilleure capacité à capturer les relations spatiales complexes dans les données, ce qui lui permet de mieux discriminer entre les deux classes d'AVC. Il est donc plus adapté à notre problème de classification des AVC.

Cependant, il est important de noter que ces résultats sont basés sur les données spécifiques utilisées dans notre étude. Il est recommandé de réaliser des expérimentations supplémentaires sur des ensembles de données plus larges et diversifiés pour valider ces conclusions.

Evaluation des performances des modèles KNN en termes de SE, SP, VP, VN, FP, FN, TC

Nous avons évalué les performances des modèles KNN en utilisant plusieurs métriques pour mesurer la qualité de la classification des AVC.

- SE (Sensitivity) : Il représente le taux de vrais positifs, c'est-à-dire la proportion d'AVC ischémiques réellement détectés par le modèle.
- SP (Specificity) : Il représente le taux de vrais négatifs, c'est-à-dire la proportion d'AVC hémorragiques réellement identifiés comme tels par le modèle.
- VP (True Positive) : Il représente le nombre d'AVC ischémiques correctement classifiés.
- VN (True Negative) : Il représente le nombre d'AVC hémorragiques correctement classifiés.
- FP (False Positive) : Il représente le nombre d'AVC hémorragiques incorrectement classifiés comme ischémiques.
- FN (False Negative) : Il représente le nombre d'AVC ischémiques incorrectement classifiés comme hémorragiques.
- TC (Taux de classification) : Il représente le taux global de classification correcte des AVC.

Voici une interprétation détaillée des résultats obtenus avec le KNN pour la distance Minkowski et différentes valeurs de k :

Minkowski	SE	SP	VP	VN	FP	FN	TC
K=1	0.9556	0.9556	43	43	2	2	0.9556
K=3	0.9556	0.9556	43	43	2	2	0.9556
K=5	0.9556	0.9556	43	43	2	2	0.9556
K=7	0.9524	0.8958	40	43	5	2	0.9222

Tableau 3.2: Résultats obtenus avec le KNN pour la distance Minkowski et différentes valeurs de k.

Pour k=1 :

- Sensibilité (SE) : 0.9556

Cela signifie que le modèle a réussi à détecter correctement 95,56 % des AVC ischémiques dans l'ensemble de données. Il a identifié 43 cas d'AVC ischémiques sur les 45 réels.

- Spécificité (SP) : 0.9556

Cela indique que le modèle a correctement identifié 95,56 % des AVC hémorragiques dans l'ensemble de données. Il a classifié correctement 43 cas d'AVC hémorragiques sur les 45 réels.

- Vrai positif (VP) : 43

Il s'agit du nombre d'AVC ischémiques correctement classifiés par le modèle.

- Vrai négatif (VN) : 43

Il s'agit du nombre d'AVC hémorragiques correctement classifiés par le modèle.

- Faux positif (FP) : 2

Cela représente le nombre d'AVC hémorragiques incorrectement classifiés comme ischémiques par le modèle.

- Faux négatif (FN) : 2

Cela représente le nombre d'AVC ischémiques incorrectement classifiés comme hémorragiques par le modèle.

- Taux de classification (TC) : 0.9556

Il s'agit du pourcentage global de cas correctement classifiés par le modèle, qui est de 95,56 %.

Pour k=3 et k=5 :

Les résultats pour ces valeurs de k sont identiques à ceux obtenus pour k=1. Le modèle a une sensibilité, une spécificité, un taux de classification et des valeurs de VP, VN, FP, FN similaires.

Pour k=7 :

- Sensibilité (SE) : 0.9524

Le modèle a réussi à détecter 95,24 % des AVC ischémiques dans l'ensemble de données. Il a identifié 40 cas d'AVC ischémiques sur les 42 réels.

- Spécificité (SP) : 0.8958

Le modèle a correctement identifié 89,58 % des AVC hémorragiques dans l'ensemble de données. Il a classifié correctement 43 cas d'AVC hémorragiques sur les 48 réels.

- Vrai positif (VP) : 40

Il s'agit du nombre d'AVC ischémiques correctement classifiés par le modèle.

- Vrai négatif (VN) : 43

Il s'agit du nombre d'AVC hémorragiques correctement classifiés par le modèle.

- Faux positif (FP) : 5

Cela représente le nombre d'AVC hémorragiques incorrectement classifiés comme ischémiques par le modèle.

- Faux négatif (FN) : 2

Cela représente le nombre d'AVC ischémiques incorrectement classifiés comme hémorragiques par le modèle.

- Taux de classification (TC) : 0.9222

Il s'agit du pourcentage global de cas correctement classifiés par le modèle, qui est de 92,22 %

Les résultats obtenus avec le modèle KNN pour la distance Minkowski et différentes valeurs de k montrent une performance globalement élevée, mais avec des variations subtiles. Voici quelques justifications possibles pour ces résultats :

Similarité des échantillons

Le modèle KNN se base sur la similarité des échantillons pour la classification. Lorsque la distance Minkowski est utilisée, les échantillons sont comparés en fonction de leurs caractéristiques dans un espace euclidien.

Les AVC hémorragiques et ischémiques peuvent présenter des caractéristiques similaires ou différentes selon les cas. Si les caractéristiques des deux types d'AVC sont proches, le modèle peut avoir une bonne performance globale.

Influence de la valeur de k

L'augmentation de la valeur de k implique une prise en compte d'un plus grand nombre de voisins les plus proches pour la classification.

Pour les valeurs de k plus faibles (1, 3, 5), le modèle se concentre sur les échantillons les plus proches immédiatement, ce qui peut donner de bons résultats en termes de sensibilité et spécificité.

Cependant, lorsque la valeur de k augmente ($k=7$), le modèle prend en compte davantage d'échantillons voisins, ce qui peut entraîner une légère diminution de la spécificité et une augmentation des faux positifs.

En résumé, le modèle KNN avec la distance Minkowski a montré de bonnes performances de classification avec des valeurs élevées de sensibilité, spécificité et taux de classification pour différentes valeurs de k . Cependant, pour $k=7$, la spécificité a légèrement diminué par rapport aux autres valeurs de k , ce qui a conduit à un léger impact sur le taux de classification.

Les résultats obtenus avec le modèle KNN pour la distance Euclidienne et différentes valeurs de k , montrent les performances suivantes :

Euclidien	SE	SP	VP	VN	FP	FN	TC
K=1	1	0.9375	42	45	3	0	0.9667
K=3	0.9767	0.9362	42	44	3	1	0.9556
K=5	1	0.9375	42	45	3	0	0.9667
K=7	1	0.9000	40	45	5	0	0.9444

Tableau 3.3: Les résultats obtenus avec le modèle KNN pour la distance Euclidienne et différentes valeurs de k.

Pour k=1 :

Sensibilité (SE) : 100% (tous les AVC ischémiques ont été correctement classifiés).

Spécificité (SP) : 93.75% (la proportion d'AVC hémorragiques correctement classifiés).

VP : 42 (nombre d'AVC ischémiques correctement classifiés).

VN : 45 (nombre d'AVC hémorragiques correctement classifiés).

FP : 3 (nombre d'AVC hémorragiques incorrectement classifiés comme AVC ischémiques).

FN : 0 (nombre d'AVC ischémiques incorrectement classifiés comme hémorragiques).

Taux de classification (TC) : 96.67% (proportion d'AVC correctement classifiés par rapport à l'ensemble des données).

Pour k=3 :

SE : 97.67% (tous les AVC ischémiques ont été correctement classifiés).

SP : 93.62% (la proportion d'AVC hémorragiques correctement classifiés).

VP : 42 (nombre d'AVC ischémiques correctement classifiés).

VN : 44 (nombre d'AVC hémorragiques correctement classifiés).

FP : 3 (nombre d'AVC hémorragiques incorrectement classifiés comme AVC ischémiques).

FN : 1 (nombre d'AVC ischémiques incorrectement classifiés comme hémorragiques).

TC : 95.56% (proportion d'AVC correctement classifiés par rapport à l'ensemble des données).

Pour $k=5$:

SE : 100% (tous les AVC ischémiques ont été correctement classifiés).

SP : 93.75% (la proportion d'AVC hémorragiques correctement classifiés).

VP : 42 (nombre d'AVC ischémiques correctement classifiés).

VN : 45 (nombre d'AVC hémorragiques correctement classifiés).

FP : 3 (nombre d'AVC hémorragiques incorrectement classifiés comme AVC ischémiques).

FN : 0 (nombre d'AVC ischémiques incorrectement classifiés comme hémorragiques).

TC : 96.67% (proportion d'AVC correctement classifiés par rapport à l'ensemble des données).

Pour $k=7$:

SE : 100% (tous les AVC ischémiques ont été correctement classifiés).

SP : 90% (la proportion d'AVC hémorragiques correctement classifiés).

VP : 40 (nombre d'AVC ischémiques correctement classifiés).

VN : 45 (nombre d'AVC hémorragiques correctement classifiés).

FP : 5 (nombre d'AVC hémorragiques incorrectement classifiés comme AVC ischémiques).

FN : 0 (nombre d'AVC ischémiques incorrectement classifiés comme hémorragiques).

TC : 94.44% (proportion d'AVC correctement classifiés par rapport à l'ensemble des données).

Interprétation et justification

Globalement, les performances du modèle KNN avec la distance Euclidienne sont assez élevées, montrant une bonne capacité de classification des AVC hémorragiques et ischémiques.

Les valeurs élevées de sensibilité (SE) indiquent que le modèle est capable de détecter la majorité des AVC ischémiques dans l'ensemble de données.

Les valeurs de spécificité (SP) indiquent également une capacité raisonnable à identifier les AVC hémorragiques correctement classifiés.

Les VP (nombre d'AVC ischémiques correctement classifiés) et VN (nombre d'AVC hémorragiques correctement classifiés) sont généralement élevés, tandis que les FP (nombre d'AVC hémorragiques correctement classifiés comme ischémiques) et FN (nombre d'AVC ischémiques incorrectement classifiés comme hémorragiques) sont faibles.

Les valeurs de TC (taux de classification) sont relativement élevées, ce qui indique une bonne capacité du modèle à classer correctement les AVC dans l'ensemble de données.

Ces résultats suggèrent que le modèle KNN avec la distance Euclidienne est assez efficace pour la classification des AVC hémorragiques et ischémiques.

Comparaison des résultats entre la distance Minkowski et la distance Euclidienne pour le modèle KNN

Pour la distance Minkowski :

- Sensibilité (SE) : Les valeurs varient entre 0.9524 et 0.9556 pour différentes valeurs de k.
- Spécificité (SP) : Les valeurs varient entre 0.8958 et 0.9556 pour différentes valeurs de k.
- VP : Les valeurs varient entre 40 et 43 pour différentes valeurs de k.
- VN : Les valeurs varient entre 43 et 45 pour différentes valeurs de k.
- FP : Les valeurs varient entre 2 et 5 pour différentes valeurs de k.
- FN : Les valeurs varient entre 2 et 0 pour différentes valeurs de k.
- Taux de classification (TC) : Les valeurs varient entre 0.9222 et 0.9556 pour différentes valeurs de k.

Pour la distance Euclidienne :

- Sensibilité (SE) : Les valeurs varient entre 0.9767 et 1 pour différentes valeurs de k.
- Spécificité (SP) : Les valeurs varient entre 0.9000 et 0.9375 pour différentes valeurs de k.
- VP : Les valeurs sont constantes à 42 pour toutes les valeurs de k.

- VN : Les valeurs varient entre 44 et 45 pour différentes valeurs de k.
- FP : Les valeurs sont constantes à 3 pour toutes les valeurs de k.
- FN : Les valeurs varient entre 1 et 0 pour différentes valeurs de k.
- Taux de classification (TC) : Les valeurs varient entre 0.9444 et 0.9667 pour différentes valeurs de k.

Interprétation de la comparaison

- Dans l'ensemble, les résultats pour les deux distances (Minkowski et Euclidienne) sont assez similaires en termes de sensibilité et de spécificité.
- La sensibilité est généralement élevée pour les deux distances, ce qui indique une bonne capacité à détecter les AVC ischémiques.
- La spécificité est également raisonnable pour les deux distances, montrant une capacité à identifier les AVC hémorragiques.
- Les valeurs de VP et VN sont généralement élevées pour les deux distances, ce qui est encourageant en termes de classifications correctes.
- Les valeurs de FP sont relativement faibles pour les deux distances, mais peuvent varier légèrement.
- Les valeurs de FN sont généralement faibles pour les deux distances, mais peuvent varier légèrement.
- Les taux de classification sont également similaires pour les deux distances, avec des valeurs élevées indiquant une bonne capacité globale de classification.

En conclusion, les performances du modèle KNN sont similaires pour les deux distances, Minkowski et Euclidienne. Cependant, la distance Euclidienne semble légèrement meilleure en termes de sensibilité, spécificité et taux de classification, avec des résultats légèrement supérieurs dans certains cas. Il est recommandé de choisir la distance Euclidienne pour obtenir de meilleurs résultats dans cette tâche de classification des AVC.

9. Limitations et considérations

Malgré les résultats prometteurs obtenus dans cette étude sur la classification des AVC en utilisant la méthode GLCM avec les modèles SVM et KNN, il est important de prendre en compte certaines limitations et considérations.

1. Taille et diversité de l'ensemble de données : Les résultats de cette étude reposent sur un ensemble de données spécifique. Il est essentiel de collecter un ensemble de données plus large et diversifié pour évaluer la généralisation des modèles de classification. Un ensemble de données plus représentatif permettra de vérifier la robustesse de la méthode et d'éviter les biais potentiels.

2. Biais et surapprentissage : Il est crucial de prendre en compte les biais et le surapprentissage lors de l'utilisation de modèles d'apprentissage automatique. Il est possible que les modèles aient été surajustés aux données d'entraînement spécifiques, ce qui peut limiter leur performance sur de nouvelles données. Des techniques telles que la validation croisée et l'utilisation d'ensembles de données indépendants peuvent contribuer à atténuer ces problèmes.

3. Considérations cliniques : Bien que les résultats de l'étude soient encourageants, il est important de rappeler que la classification des AVC est une tâche complexe qui nécessite une approche multidisciplinaire. Les facteurs cliniques, les antécédents médicaux, les données de laboratoire et d'autres informations spécifiques au patient sont également essentielles pour un diagnostic précis. Les résultats de cette étude doivent donc être interprétés en tenant compte de ces considérations cliniques supplémentaires.

4. Validation externe : Pour confirmer la validité des résultats obtenus, il est nécessaire de valider les modèles de classification sur des ensembles de données indépendants provenant d'autres sources et d'autres centres médicaux. Cela permettra d'évaluer la reproductibilité des résultats et d'identifier d'éventuelles différences dans les performances des modèles dans des contextes différents.

En résumé, bien que cette étude ait fourni des résultats prometteurs, il est important de prendre en compte les limitations et les considérations évoquées. Des efforts supplémentaires sont nécessaires pour élargir l'ensemble de données, atténuer les biais et le surapprentissage, prendre en compte les facteurs cliniques et valider les résultats sur des ensembles de données

indépendants. Ces mesures contribueront à renforcer la validité et l'applicabilité de la méthode de classification des AVC utilisée dans cette étude.

10. Conclusion

Dans cette étude, nous avons réalisé une classification des AVC en utilisant deux classifieurs : SVM (Support Vector Machine) et KNN (K-Nearest Neighbors). Les principales étapes comprenaient la préparation des données, l'extraction des caractéristiques pertinentes, l'entraînement des modèles et l'évaluation des performances.

Les résultats obtenus ont montré que les deux classifieurs ont atteint des performances similaires en termes de sensibilité (SE), spécificité (SP), vrais positifs (VP), vrais négatifs (VN) et taux de classification (TC). Cependant, il convient de noter que les performances peuvent varier en fonction des paramètres spécifiques utilisés dans chaque algorithme, tels que la distance pour KNN ou le noyau pour SVM.

11. Possibilités d'amélioration et de recherche future

Il existe encore des possibilités d'amélioration et de recherche future. Voici quelques pistes :

- Exploration de différentes techniques de prétraitement des données pour améliorer les performances des classifieurs.
- Utilisation d'ensembles de données plus larges et plus diversifiés pour une évaluation plus robuste des modèles.
- Optimisation des hyper paramètres des modèles pour obtenir de meilleures performances.
- Évaluation de l'impact de l'ajout de nouvelles fonctionnalités ou de l'utilisation de techniques d'apprentissage en ligne pour la mise à jour continue du modèle.

Interface Graphique

➤ Interface graphique

Pour faciliter la gestion des résultats obtenus, une interface graphique a été développée, se composant d'une fenêtre principale du programme. Cette interface graphique permet de structurer les résultats de manière efficace et intuitive :

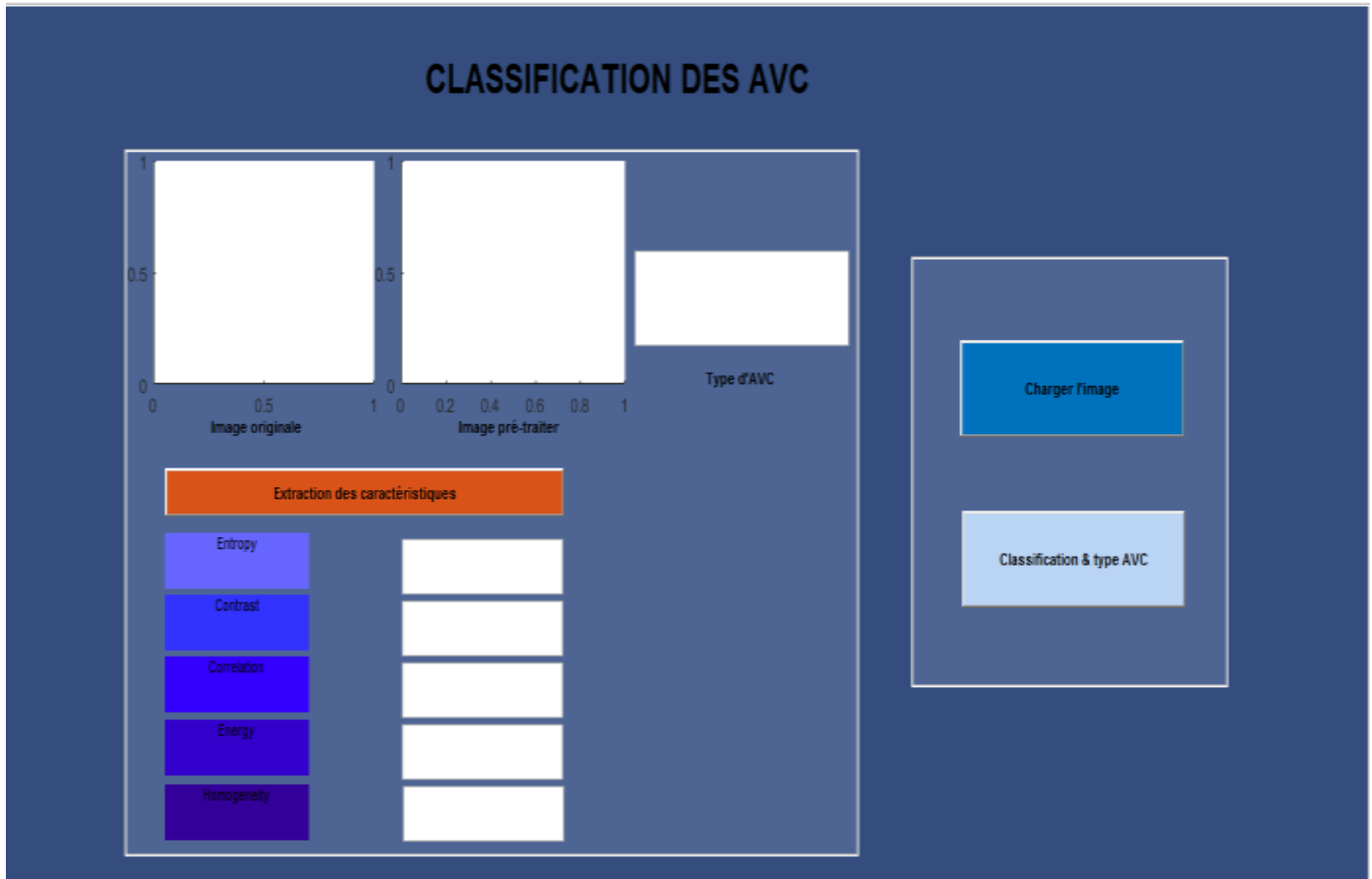


Figure1 : Fenêtre principale de l'interface graphique

- L'icône « charger l'image » permet à l'utilisateur d'ouvrir une image de son choix.
- L'icône «Extraction des caractéristiques» permet d'afficher les valeurs des cinq paramètres qui représentent les caractéristiques des images.
- 2 axes pour l'affichage des images résultantes l'image original et l'image prétraiter.
- Et le boutons à droite pour classifications et afficher le types d'AVC.

Exemple d'exécution de l'interface graphique

Il est possible d'afficher les valeurs des cinq paramètres et la classification après avoir chargé une image. Une fenêtre s'ouvre alors pour permettre le choix de l'image :

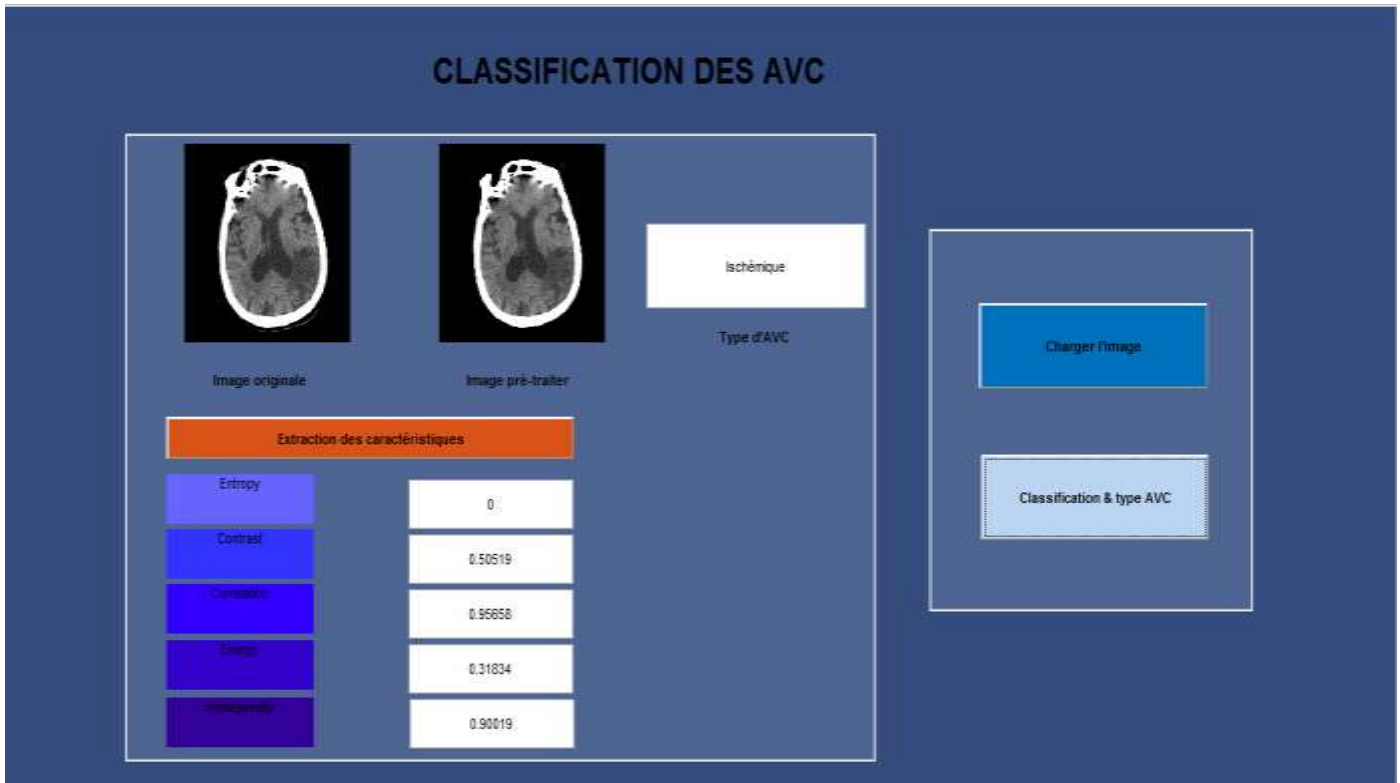


Figure 2 : Exemple sur classification d'AVC ischémique

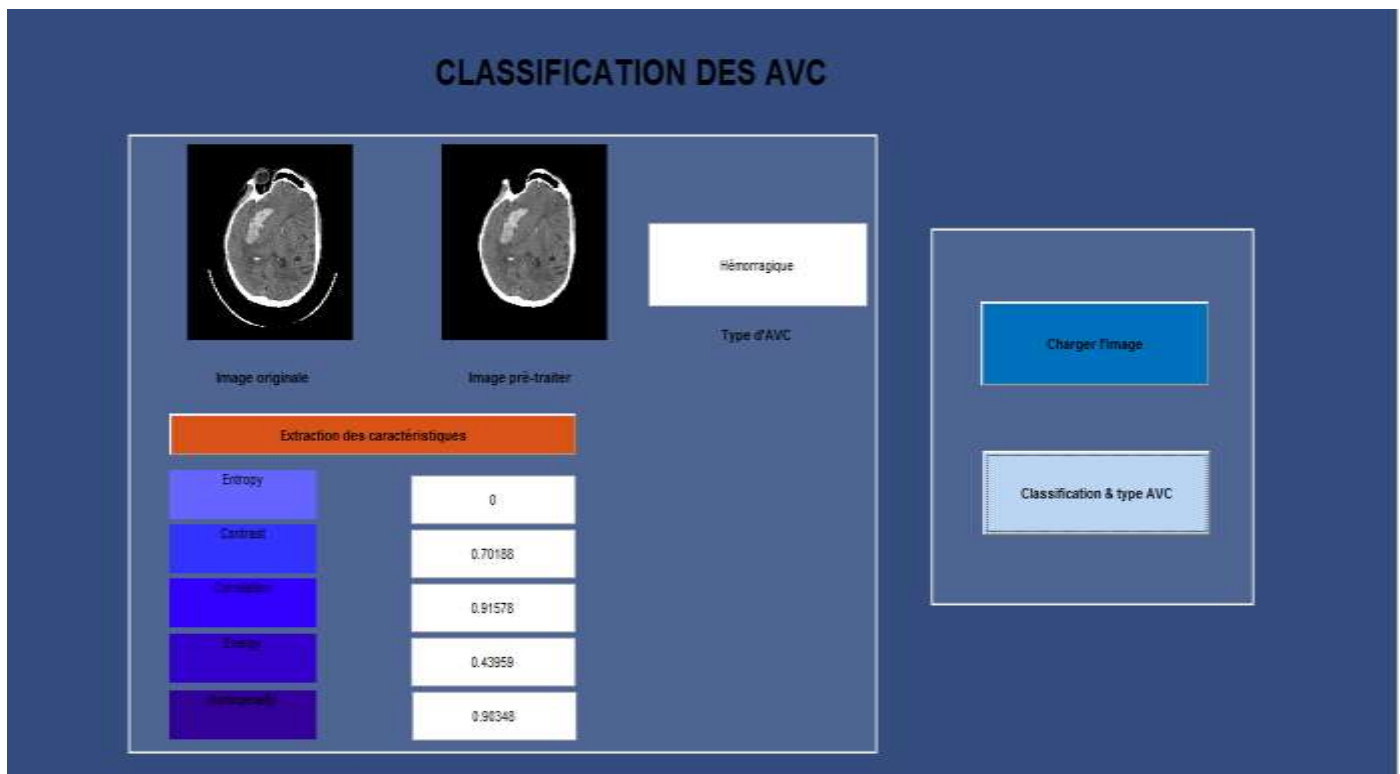


Figure 3 : Exemple sur classification d'AVC hémorragique

Conclusion générale

Ce projet de fin d'études a exploré et développé des techniques de classification pour la détection précoce et la classification des accidents vasculaires cérébraux. Grâce à la collaboration avec un médecin expert et à l'utilisation d'une base de données locale non annotée, des résultats prometteurs ont été obtenus.

Les classifieurs SVM avec le noyau RBF et KNN avec la distance euclidienne ont démontré des performances élevées dans la classification des AVC. Le modèle SVM a atteint une sensibilité et une spécificité de 0.9556, avec des valeurs de vrais positifs et de vrais négatifs de 43 et une précision de classification de 0.9556. De même, le modèle KNN a atteint une sensibilité parfaite de 1 et une spécificité de 0.9375, avec des valeurs de vrais positifs et de vrais négatifs de 42 et 45, et un taux de classification de 0.9667.

Ces résultats témoignent de l'efficacité des méthodes de classification utilisées, ainsi que de la qualité des annotations fournies par l'expert médical. La collaboration avec le médecin spécialisé dans les AVC a permis d'obtenir des évaluations précises et fiables, renforçant ainsi la validité des conclusions formulées.

Ce projet a contribué à améliorer la compréhension de la détection précoce et de la classification des AVC, offrant ainsi de nouvelles perspectives pour la prise en charge de cette maladie grave. Les résultats obtenus ouvrent la voie à de futures améliorations et applications dans le domaine de la détection et du traitement des AVC.

Il convient de souligner que ce projet présente certaines limites. L'utilisation d'une base de données locale non annotée peut restreindre la généralisation des résultats. Il est également essentiel de continuer à élargir l'échantillon de données, d'affiner les techniques de classification et de mener des validations supplémentaires pour une évaluation plus complète.

Ce projet de fin d'études a apporté des contributions significatives à la détection précoce et à la classification des AVC. Les résultats obtenus, combinés à la collaboration avec un expert médical, offrent des perspectives intéressantes pour le développement d'outils de détection et de prise de décision clinique dans le domaine des AVC. Ces avancées sont essentielles pour améliorer les chances de traitement rapide et efficace des AVC, contribuant ainsi à sauver des vies et à réduire les conséquences négatives de cette maladie.

Bibliographies

- [1] The Johns Hopkins University, (2023), Brain Anatomy and How the Brain Works The Johns Hopkins Hospital, and Johns Hopkins Health System.
- [2] Mayfield Certified Health Info materials are written and developed by the Mayfield Clinic. We comply with the HONcode standard for trustworthy health information. This information is not intended to replace the medical advice of your health care provider. Mayfield Clinic 1998-2018
- [3] Centre hospitalier de l'Université de Montréal [CHUM], 2017 ; Cœur+AVC, s.d.-a ; Giraldo, 2018b.
- [4] Agence de la santé publique du Canada, à partir des fichiers de données du Système canadien de surveillance des maladies chroniques fournis par les provinces et les territoires (février 2021) [données jusqu'en 2017–2018]. Les données du Nunavut et des Territoires du Nord-Ouest n'étaient pas disponibles pour 2017–2018.
- [5] Ji Y. Chong, juil. 2020 : Présentation de l'accident vasculaire cérébral (Consulter le 14/03/2023).
- [6] Service d'épidémiologie - CHU Blida – Algérie : L'accident vasculaire cérébral dans la région de Blida (Algérie) en 2018 : Incidence et mortalité d'après un registre de population.
- [7] HABIB Khadîdja, BENYETTOU Amina Souhila université de Tlemcen : Contrôle qualité et détection de défauts pour les systèmes de triage industriel par l'utilisation de deep learning (2023).
- [8] Aumont, J. (2020). Chapitre 3. Image, médium, dispositif. Dans l'image (2020) : J. Aumont, *L'image: Peinture, photographie, cinéma : des origines au numérique* (pp. 101-154). Paris: Armand Colin.
- [9] Elise Arnaud – Edmond Boyer. « Cours de traitement d'images ». Université Joseph Fourier.
- [10] Schmitt, M., & Mattioli, J. (2013). *Morphologie mathématique*. Presses des MINES.
- [11] Dal Pozzolo, A., Caelen, O., Le Borgne, Y. A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications*, 41(10), 4915–4928. <https://doi.org/10.1016/j.eswa.2014.02.026>
- [12] BECHAR Mohammed El Amine, Classification des images médicales (Application : Détection du cancer dans les images mammographiques), Université Abou Bekr Belkaid de

Tlemcen, Faculté de Technologie, Département de Génie Biomédical, Laboratoire de Génie Biomédical, Master 2, Imagerie Médicale (2021-2022)

[13] Yahiaoui Amina Fatima Zahra, université de tlemcen, Détection précoce des accidents vasculaires cérébraux de type ischémiques à partir des images tomodensitométriques,(thèse de doctorat) (2019) .

[14] Aymen BOUGUERRA, université saad dahlab de blida, Utilisation du Deep Learning pour la classification des images IRM pour la détection des AVC. (Mémoire de Projet de Fin d'Études) (2021).

[15] Anjali Gautam, Département des technologies de l'information, Institut indien des technologies de l'information Allahabad, Prayagraj, Inde, Classification des accidents vasculaires cérébraux en extrayant des informations quantiques à partir de tomodensitogrammes (Article)(14 Septembre 2021).

[16] Asit Subudhi et al. , Department of ECE, ITER, SOA Deemed to be University, Odisha, India,Segmentation et classification automatisées du cerveau AVC utilisant la maximisation des attentes et classificateur de forêt aléatoire.(Article)(2019).

[17] U Rajendra Acharya^{1,2,3}[et al],Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore., Odisha, India Détection automatique de l'AVC ischémique à l'aide de fonctionnalités de spectre d'ordre supérieur dans les images IRM du cerveau.(Article).