



République Algérienne Démocratique et Populaire



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Akli Mohand Oulhadj de Bouira

Faculté des Sciences et des Sciences Appliquées

Département d'Informatique

Mémoire de Master

en Informatique

Spécialité : ISIL

Thème

Approche d'intelligence artificielle pour le diagnostic
des patients.

Encadré par

— ABBAS Akli

Réalisé par

— KOBBI Ahmed Abdessamed

— NEDJARI Mohamed Abdesslam

2023/2024

Remerciements

Nous remercions Dieu tout puissant pour son aide. Ce mémoire n'aurait pas été possible sans l'intervention d'un grand nombre de personnes, nous souhaitons les en remercier. Nous tenons à remercier notre promoteur Mr ABBAS Akli pour ses précieux conseils, sa confiance, sa patience qui ont constitué un apport considérable pour nous durant tous le déroulement de ce projet. Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils portent à notre travail en acceptant de l'examiner et de l'enrichir par leurs propositions.

Dédicaces

Nous dédions ce travail à :

Nos très chers parents.

Tous nos enseignants.

Nos frères et sœurs.

Tous nos adorables amis.

Ainsi qu'à toute personne qui nous a soutenus.

kobbi ahmed abdessamed.

Dédicaces

Nous dédions ce travail à :

Nos très chers parents.

Tous nos enseignants.

Nos frères et sœurs.

Tous nos adorables amis.

Ainsi qu'à toute personne qui nous a soutenus.

nedjari mohamed abdesslam.

ملخص

في المجال الطبي، يعد التشخيص الدقيق وفي الوقت المناسب أمرًا ضروريًا لضمان معاملة المرضى بشكل صحيح. ومع ذلك، فإن كمية البيانات الطبية التي يعالجها المهنيون الصحيون بشكل متكرر مهمة، مما يجعل إجراء التشخيص طويلًا وعرضة للأخطاء. يوفر التقدم في الذكاء الاصطناعي (اي) في مجال الطب فرصًا مثيرة للاهتمام لتعظيم نتائج الاختبار. استنادًا إلى الشبكات العصبية وخوارزميات التعلم الآلي، يستطيع الذكاء الاصطناعي تحليل العديد من البيانات الطبية بسرعة، وتحديد الأنماط الدقيقة وتقديم توصيات تشخيصية دقيقة.

تقيم هذه الدراسة فعالية الذكاء الاصطناعي لتشخيص المرضى في المجال الطبي باستخدام نموذج ذكاء وقاعدة بيانات بيبيث ٣ ، مع مقاييس التقييم غولجي سي ٤ ، بية يـ ٥ .
الكلمات الرئيسية : الذكاء الاصطناعي، تشخيص المرضى، الشبكات العصبية، التعلم

FLANT-5, bleu4, rouge-1, Meteor, MIMIC3.

Abstract

In the medical field, accurate and timely diagnosis is essential to ensure that patients are treated properly. However, the amount of medical data that health professionals frequently process is important, which makes the diagnostic procedure long and prone to errors. Advances in artificial intelligence (AI) in the field of medicine offer interesting opportunities to maximize test results. Based on neural networks and machine learning algorithms, artificial intelligence is able to analyze many medical data quickly, identify subtle patterns and make accurate diagnostic recommendations.

This study evaluates the effectiveness of artificial intelligence for the diagnosis of patients in the medical field using the flan-t5 model and the mimic3 database, with Blue-4, Red-1 and Meteor Evaluation Metrics.

Keywords : Artificial intelligence, Patient diagnosis, Neural networks, Learning, FLANT-5, blue-4, rouge-1, Meteor, MIMIC3.

...

Résumé

:

Dans le domaine médical, un diagnostic précis et rapide est essentiel pour s'assurer que les patients sont soignés correctement. Cependant, la quantité de données médicales que les professionnels de la santé traitent fréquemment est importante, ce qui rend la procédure de diagnostic longue et sujette aux erreurs. Les progrès de l'intelligence artificielle (IA) dans le domaine de la médecine offrent des possibilités intéressantes pour maximiser les résultats des tests. Basé sur les réseaux neuronaux et les algorithmes d'apprentissage automatique, l'intelligence artificielle est capable d'analyser de nombreuses données médicales rapidement, d'identifier des schémas subtils et de formuler des recommandations de diagnostic précis.

Cette étude évalue l'efficacité de l'intelligence artificielle pour le diagnostic des patients dans le domaine médical en utilisant le modèle flan-t5 et la base de données mimic3, avec les Métriques d'évaluation bleu-4, rouge-1 et Meteor.

Les mots clés : intelligence artificielle, diagnostic des patients, les réseaux neuronaux, l'apprentissage, FLANT-5, bleu-4, rouge-1, Meteor, MIMIC3.

...

Table des matières

Table des matières	i
Table des figures	iv
Liste des tableaux	v
Liste des abréviations	vi
Introduction générale	1
1 Revue de la littérature	3
1.1 Introduction	3
1.2 Histoire et évolution de l'IA en médecine	3
1.2.1 Utilisations précoces de l'IA dans le diagnostic médical	3
1.2.2 L'ère de l'apprentissage automatique (années 1980 à nos jours)	4
1.3 Évolution des techniques et des algorithmes	5
1.3.1 Le diagnostic médical	5
1.3.2 Recherche médicale	5
1.3.3 Développement de médicaments	6
1.3.4 Personnalisation des traitements	6
1.3.5 Surveillance de la santé	6
1.4 Techniques d'intelligence artificielle	6
1.4.1 Apprentissage supervisé et non supervisé	6
1.4.2 Réseaux de neurones, machine Learning, deep Learning	7
1.5 Applications actuelles de l'IA en diagnostic médical	8

1.5.1	Exemples de systèmes et de technologies utilisées	8
1.5.2	Études de cas et résultats notables	9
2	Méthodologie	13
2.1	Introduction	13
2.2	Description des données	13
2.2.1	Type et source des données utilisées	13
2.2.2	Méthodes de collecte et prétraitement des données	16
2.3	Modèles d'IA utilisés	17
2.3.1	Choix des modèles et justification	17
2.3.2	Description des algorithmes et des architectures	18
2.4	Validation et évaluation des modèles	20
2.4.1	Méthodes d'évaluation des performances des modèles	20
2.4.2	Indicateurs de performance	21
3	Résultats	26
3.1	Introduction	26
3.2	Environnement de développement	26
3.2.1	Outils et logiciels utilisés	26
3.2.2	Les caractéristiques de l'ordinateur	28
3.3	Analyse et interprétation des résultats	29
3.3.1	Analyse Résultats obtenus avec chaque métrique en fonction de re- cherche	29
3.3.2	Comparaison des performances des différents modèles	33
3.3.3	Discussion sur les points forts et les limites des modèles	34
3.4	Implications des résultats	37
3.4.1	Implications pratiques pour le diagnostic médical	37
3.4.2	Potentiel impact sur les pratiques cliniques	38
3.5	Comparaison avec les travaux antérieurs	39
3.5.1	ku-dmis-msra	39
3.5.2	WGSUM (Trans+TAG)	39
3.5.3	utsa-nlp-T5	39
3.5.4	BART+Oracle	40

3.6	Limites de l'étude et perspectives de recherche	40
3.6.1	Limites méthodologiques et techniques	40
3.6.2	Suggestions pour des recherches futures	41
	Conclusion générale	43
	Bibliographie	45

Table des figures

- 2.1 Structure de database MIMIC3 15
- 2.2 Architecture de modèle FLAT-5 19

- 3.1 NVIDIA Tesla P100 Performance. 28
- 3.2 Scores BLEU-p4 du modèle "flanT5-med-rad-reports" 29
- 3.3 Scores Meteor du modèle "flanT5-med-rad-reports" 30
- 3.4 Scores Rouge-L du modèle "flanT5-med-rad-reports" 31
- 3.5 Scores train eval loss du modèle "flanT5-med-rad-reports" 32

Liste des tableaux

3.1 COMPARAISON DES PERFERMANCES des métrique d'eval entre les modèles	34
---	----

Liste des abréviations

CNN	Convolutional Neural Networks
DL	Deep Learning
FDA	Food and Drug Administration
FPRC	Fuel Pressure Regulator
HIPAA	Health Insurance Portability and Accountability Act
IA	Intelligence Artificielle
IBM	International Business Machines
IDx-Dr	Index Dilation and Extraction
IEEE	Institute of Electrical and Electronics Engineers
ML	Machine Learning
MTMdr	Medication Therapy Management
NLP	Natural language processing

Introduction générale

À l'heure actuelle, l'intelligence artificielle (IA) apparaît clairement comme l'un des principaux moteurs de l'innovation et de la croissance.

L'intelligence artificielle est une branche de l'informatique qui vise à développer des systèmes qui ont la capacité d'imiter ou de reproduire les capacités mentales et les compétences d'un être humain - par exemple, l'apprentissage, la pensée, l'analyse, perception et adaptation — afin qu'un tel système soit capable de mener diverses activités complexes de façon autonome. Dans le secteur des soins de santé, un diagnostic précis en temps opportun est essentiel pour s'assurer qu'un patient reçoit un traitement approprié. Cependant, les professionnels de la santé sont toujours confrontés à d'énormes données médicales, ce qui rend le processus de diagnostic fastidieux tout en étant sujet aux erreurs.

Les applications médicales de l'intelligence artificielle conduisent de plus en plus à des résultats de test optimisés. Aujourd'hui, pour aller encore plus loin, l'intelligence artificielle est capable d'examiner des données étendues sur la médecine à l'aide d'algorithmes d'apprentissage automatique et de réseaux de neurones pour trouver des modèles subtils et des recommandations de cadre pour un diagnostic approprié. Le problème majeur de cette recherche est d'analyser l'efficacité de l'intelligence artificielle dans le diagnostic médical, principalement dans des spécialisations telles que la radiologie. Les hypothèses de travail comprennent le concept que l'intelligence artificielle peut améliorer la précision et la vitesse des diagnostics médicaux, réduire les erreurs humaines et élargir l'accessibilité des soins médicaux dans les zones mal couvertes. Le principal défi de ce travail est de produire l'approche conçue avec les techniques d'apprentissage automatique pour analyser les données et prendre des décisions.

Ce mémoire sera structuré en 4 chapitres.

Chapitre 1 : L'intelligence artificielle

L'utilisation de l'IA en médecine sera élaborée sur trois axes : l'historique de l'utilisation, les différentes techniques d'IA, l'apprentissage supervisé et non supervisé, les réseaux de neurones artificiels et les applications actuelles dans le cadre du diagnostic médical.

Chapitre 2 : Méthodologie

Méthodologie : Dans ce deuxième chapitre, nous décrivons en détail la méthodologie mise en œuvre dans notre étude. Pour ce faire, nous décrivons d'abord les données utilisées, qui sont MIMIC, en élaborant sur son type et en décrivant la méthode de collecte et de prétraitement. Cela est suivi d'une description des types de modèles d'IA appliqués et de la justification de leur sélection, ainsi que de brèves descriptions de leurs algorithmes et architectures. Nous décrivons également l'environnement de développement, y compris les outils et les logiciels utilisés, ainsi que l'infrastructure technique, tant matérielle que logicielle.

Enfin, nous expliquons les méthodes d'évaluation des performances des modèles, en précisant les indicateurs de performance utilisés.

Chapitre 3 : Résultats

Les résultats de notre étude sont exposés dans ce chapitre, en soulignant les résultats de chaque modèle utilisé et en comparant leurs performances. Les résultats sont également évalués en fonction des objectifs de la recherche, en évoquant à la fois les atouts et les limites des modèles. Par la suite, nous examinons les conséquences concrètes pour le diagnostic médical et l'incidence potentielle sur les pratiques cliniques, en comparant nos résultats avec ceux de la littérature existante et en identifiant les limites de notre étude ainsi que les voies de recherche à venir.

Nous terminons notre mémoire par une conclusion générale.

Revue de la littérature

1.1 Introduction

L'IA est apparue comme une idée depuis des décennies, mais le développement technologique rapide des dernières décennies a conduit à des progrès significatifs dans ce domaine.

L'IA est un domaine d'intérêt dans la création de systèmes technologiques capables d'effectuer des tâches qui nécessitent l'intelligence humaine. L'objectif principal de l'IA est de concevoir des logiciels et des appareils qui peuvent apprendre des données, acquérir des connaissances et prendre des décisions basées sur ces informations. Dans le domaine de la médecine, l'intelligence artificielle a joué un rôle de plus en plus important dans l'amélioration des soins aux patients, le diagnostic des maladies et le développement des traitements.

Dans ce chapitre, nous aborderons la définition de l'intelligence artificielle, son historique et son développement dans le domaine de la santé, ses types et de ses domaines.

1.2 Histoire et évolution de l'IA en médecine

1.2.1 Utilisations précoces de l'IA dans le diagnostic médical

L'étude scientifique de l'intelligence artificielle en informatique vise à construire des machines qui imiteraient et fonctionneraient intelligemment en tant qu'être humain. La façon d'apprendre l'histoire et de comprendre le contexte de l'IA est de revenir à son origine et de surveiller son développement jusqu'à présent.

Les origines de l'IA (1940-1950)

Le fondement de la pensée de l'intelligence artificielle a été établi plus tôt par Alan Turing en 1950. L'hypothèse d'Alan Turing ne portait pas sur la médecine. Pourtant, son concept du test de Turing pour mesurer l'intelligence des machines a été une percée dans la mesure des fonctions cognitives des systèmes informatiques.

L'âge d'or de l'IA (1950-1960)

En 1956, John McCarthy organise une conférence au Dartmouth College. L'objectif général de cette conférence était de proposer le concept d'intelligence artificielle en tant que domaine scientifique. Les réalisations réussies de cette époque étaient plus généralisées et s'articulaient autour de la résolution logique de problèmes et de la mise en place de routes axées sur l'avenir vers ces techniques dans des secteurs particuliers tels que la médecine.

L'ère de l'IA symbolique (1960-1980)

Les systèmes experts développés pendant cette période ont été parmi les premières applications concrètes de l'IA dans le domaine médical. Ces systèmes utilisaient des bases de connaissances symboliques pour diagnostiquer des pathologies, recommander des traitements appropriés et fournir des interprétations du raisonnement clinique.

1.2.2 L'ère de l'apprentissage automatique (années 1980 à nos jours)

Les années 1980 ont apporté avec eux des réseaux de neurones, et les possibilités dans le domaine de la médecine ont été largement ouvertes. Aujourd'hui, les réseaux de neurones artificiels sont utilisés pour de nombreuses tâches, y compris l'analyse d'imagerie médicale, la prédiction des diagnostics ou la personnalisation de traitements spécifiques à effectuer sur des patients individuels.

Les progrès de l'informatique à haute performance, combinés à une explosion de données médicales, alimentent l'utilisation de techniques d'apprentissage automatique,

en particulier l'apprentissage profond dans des applications médicales avancées comme la détection de maladies par imagerie, la prédiction des résultats cliniques et la découverte de médicaments.

Cela signifie que chaque phase de l'évolution de l'IA a apporté des avancées importantes : des concepts de base à la formalisation, en passant par l'application de techniques avancées d'apprentissage automatique et profond pour résoudre des problèmes cliniques complexes en médecine [1].

1.3 Évolution des techniques et des algorithmes

L'IA révolutionne la médecine en fournissant des solutions innovantes pour le diagnostic, le traitement et la gestion des soins de santé. Grâce à l'intégration d'algorithmes sophistiqués et de techniques d'apprentissage automatique, l'IA peut analyser les données médicales avec une grande précision et fournir des recommandations cliniques précises, promettant un changement radical dans la prestation des soins de santé vers une approche plus personnalisée, médecine efficace et accessible. Des solutions innovantes que nous trouvons :

1.3.1 Le diagnostic médical

L'IA peut aider les médecins à poser des diagnostics plus précis en analysant des données médicales telles que les symptômes, les résultats d'examens médicaux, les images médicales, etc. Les algorithmes d'IA peuvent détecter des schémas et des signaux subtils qui pourraient échapper à l'œil humain, aidant ainsi à identifier des maladies plus tôt et à recommander des traitements appropriés [2].

1.3.2 Recherche médicale

L'IA peut accélérer la recherche médicale en analysant de vastes bases de données de recherche, en identifiant des modèles et des associations entre les facteurs de risque et les résultats des traitements. Cela peut aider à la découverte de nouvelles thérapies, à l'identification de nouvelles cibles de médicaments et à l'amélioration de la compréhension des maladies [2].

1.3.3 Développement de médicaments

Développer un médicament est un processus très long et réglementé. Il comprend globalement une phase de Drug Discovery, au cours de laquelle des candidats sont identifiés, puis une phase de recherche préclinique, pendant laquelle les candidats médicaments sont soumis à une validation de preuve de concept [3].

1.3.4 Personnalisation des traitements

L'IA, en passant au crible de vastes volumes de données médicales et génétiques, a la capacité de suggérer des procédures de traitement sur mesure qui prennent en compte les détails spécifiques du patient [3].

1.3.5 Surveillance de la santé

La fusion de gadgets portables et d'applications mobiles, pilotée par la technologie de l'IA, assure une surveillance continue de l'état de santé d'une personne; il est capable de signaler les anomalies sur le radar et d'agir préventivement contre les maladies avant qu'elles ne surviennent.

1.4 Techniques d'intelligence artificielle

1.4.1 Apprentissage supervisé et non supervisé

Apprentissage supervisé

L'apprentissage supervisé est une approche qui a pour objectif, la conception des modèles à partir des données annotées. Ainsi, chaque entrée est déjà associée à un label ou classe cible. Une fois les modèles entraînés, ils seront capables d'être utilisés dans différentes applications telles que, la prédiction ou la classification des données non annotées [4].

Apprentissage non supervisé

Dans ce cas, l'apprentissage se fait d'une manière autonome. Les données livrées à la machine ne sont pas annotées. Dans ce cas, c'est à l'algorithme de trouver les similarités et

les distinctions au sein de ces données, afin de regrouper celles qui ont des caractéristiques communes dans la même classe [4].

1.4.2 Réseaux de neurones, machine Learning, deep Learning

Réseaux de neurones

Formels se proposaient à l'origine de construire une intelligence artificielle s'inspirant des systèmes nerveux biologiques. Soixante ans après leur création, ces algorithmes inspirés du vivant occupent une position prépondérante parmi les techniques utilisées, tant dans les systèmes téléphoniques que dans les équipements industriels [5].

L'apprentissage automatique (Machine Learning)

Est un sous-domaine de l'intelligence artificielle faisant référence à l'utilisation d'algorithmes capables d'apprendre en reconnaissant des motifs et des tendances à partir de grands jeux de données. Par rapport aux approches statistiques plus « classiques » de recherche de facteurs ou de scores pronostiques employant par exemple des régressions linéaires ou logistiques simples, les approches plus « avancées » permettent l'analyse de données de natures très variées (données structurées et non structurées, textes et image-rie) et de très grandes dimensions « Big Data » pouvant avoir des associations complexes non-linéaires, grâce à des méthodes et des algorithmes de plus en plus sophistiqués [6].

Apprentissage profond : Deep Learning (DL)

En réalité, le terme « apprentissage profond » existait depuis les années 40. Néanmoins il a été utilisé officiellement pour la première fois en 2010, suite à son développement, dans divers domaines [7].

La puissance des ordinateurs qui ne cesse d'augmenter et l'accumulation des tonnes de données sont les causes majeures de l'apparition du DL. C'est dans ce cadre, que seules les approches basées sur cet apprentissage, permettent l'extraction automatique des caractéristiques, dans les différentes couches du réseau.

Le DL est aussi un logiciel informatique puissant, qui utilise diverses techniques pour entraîner des réseaux de neurones complexes (profonds). Ces réseaux composés de plusieurs couches cachés à partir d'un ensemble de données large, servent à prédire un résultat

exploitable, dont sa structure imite celle du cerveau humain.

Dès qu'un réseau de neurones profond reçoit une donnée ou une information, celle-ci excite l'ensemble des nœuds de la première couche, en changeant leurs états actuels afin de faire circuler l'information, jusqu'à la dernière couche et vice-versa. Ceci permet au système d'apprendre d'une manière autonome, en améliorant ses décisions et ses prévisions. La profondeur du modèle d'apprentissage, est mesurée par le nombre de couches qui le constituent. Un réseau de neurones contenant plus de 10 couches, est considéré comme un réseau de neurones très profond (very deep neural networks). De nos jours le DL a envahi divers domaines tels que : les assistants vocaux : Siri, Alexa, Google ; la rédaction des textes ; la cyber sécurité ; la finance ; la médecine ; le marketing et la vente ; l'agriculture ; l'exploration spatiale ; et le traitement d'images.

1.5 Applications actuelles de l'IA en diagnostic médical

1.5.1 Exemples de systèmes et de technologies utilisées

Watson for Oncology (IBM)

IBM Watson for Oncology is software as a service (SaaS) that delivers an advanced ability to analyze the meaning and context of structured and unstructured data in clinical notes and reports, easily assimilating key patient information written in plain English. By combining attributes from the patient's file with clinical expertise from Memorial Sloan Kettering, external research, and data, Watson for Oncology identifies and ranks potential treatment plans and options [8].

IDx-DR

IDx-DR est un logiciel innovant basé sur une intelligence artificielle I.A. qui détecte les signes précoces et diagnostiques automatiquement de la rétinopathie diabétique grâce à des images rétiniennes. Il est indiqué pour être utilisé par les ophtalmologues au cours de leur pratique clinique dans la détection automatique de plus de rétinopathie diabétique légère (mtmDR) chez les adultes (plus de 22 ans) diagnostiqués avec le diabète qui n'ont pas déjà été diagnostiqués avec la rétinopathie diabétique [9].

Le système d'IA autonome IDx-DR possède deux algorithmes de base, un algorithme

basé sur l'IA de qualité d'image et l'algorithme de diagnostic approprié [10].

Zebra Medical Vision

Zebra Medical Vision se concentre sur la gestion intelligente des soins de santé de la population, opérant dans les secteurs de la santé et de l'intelligence artificielle (IA). Il offre des solutions basées sur l'IA qui identifient les patients à risque de maladie chronique, permettant le lancement de voies de soins préventifs. Ces services profitent principalement aux réseaux de prestation intégrés et aux payeurs en améliorant l'exactitude de l'ajustement des risques, en permettant une intervention en temps opportun et en améliorant la qualité des services de santé. Elle a été fondée en 2014 et est basée en 2014 à Herzliya, en Israël. En août 2010, Zebra Medical Vision a été acquise par Nanox [11][12].

Butterfly iQ+

Butterfly transforme les soins aux patients, vous permettant d'intégrer l'échographie dans les évaluations quotidiennes, les diagnostics et les traitements. Nos sondes corps entier s'associent sans effort à une IA de pointe, à un logiciel de gestion des flux de travail et à une suite éducative 360° pour vous fournir la solution POCUS holistique qui répond à vos besoins [13].

1.5.2 Études de cas et résultats notables

Détection précoce du cancer du sein

Une étude menée par Google Health La mammographie de dépistage vise à identifier le cancer du sein aux stades précoces de la maladie, lorsque le traitement peut être plus successful. Malgré l'existence de programmes de dépistage dans le monde entier, l'interprétation des mammographies est affectée par des taux élevés de faux positifs et de faux négatifs². Nous présentons ici un système d'intelligence artificielle (IA) capable de surpasser les experts humains en prédiction du cancer du sein. Pour évaluer sa performance dans le cadre clinique, nous avons organisé un grand ensemble de données représentatif du Royaume-Uni et un grand ensemble de données enrichi des États-Unis. Nous montrons une réduction absolue de 5,7% et 1,2% (USA et UK) dans les faux positifs et 9,4% et 2,7% dans les faux négatifs. Nous démontrons la capacité du système à se généraliser

du Royaume-Uni aux États-Unis. Dans une étude indépendante menée auprès de six radiologues, le système d'IA a surpassé tous les lecteurs humains : la zone sous la courbe caractéristique de fonctionnement du récepteur (AUC-ROC) pour le système d'IA était supérieure de 11,5 % à celle du radiologue moyen. Nous avons effectué une simulation dans laquelle le système d'IA a participé au processus de double lecture utilisé au Royaume-Uni, et nous avons constaté que le système d'IA maintenait des performances non inférieures et réduisait la charge de travail du deuxième lecteur de 88%. Cette évaluation robuste du système d'IA ouvre la voie à des essais cliniques visant à améliorer la précision et l'efficacité du dépistage du cancer du sein [14].

Diagnostic dermatologique

Le cancer de la peau, la malignité humaine la plus fréquente, est principalement diagnostiqué visuellement, en commençant par un dépistage clinique initial et potentiellement suivi d'une analyse dermoscopique, d'une biopsie et d'un examen histopathologique. La classification automatisée des lésions cutanées à l'aide d'images est une tâche difficile en raison de la variabilité à grain fin de l'apparence des lésions cutanées. Les réseaux de neurones convolutifs profonds (CNNs) montrent un potentiel pour des tâches générales et très variables dans de nombreuses catégories d'objets à grain fin. Ici, nous démontrons la classification des lésions cutanées en utilisant un seul CNN, formé de bout en bout à partir d'images directement, en utilisant uniquement des pixels et des étiquettes de maladie comme entrées. Nous formons un CNN à l'aide d'un ensemble de données de 129450 images cliniques, soit deux ordres de grandeur de plus que les ensembles de données précédents, comprenant 2032 maladies différentes. Nous testons sa performance auprès de 21 dermatologues certifiés sur des images cliniques éprouvées par biopsie avec deux cas critiques d'utilisation de la classification binaire : les carcinomes kératinocytaires par rapport aux kératoses séborrhéiques bénignes et les mélanomes malins par rapport aux nævus bénins. Le premier cas représente l'identification des cancers les plus courants, le second représente l'identification du cancer de la peau le plus mortel. Le CNN réalise des performances équivalentes à celles de tous les experts testés pour les deux tâches, démontrant une intelligence artificielle capable de classer le cancer de la peau avec un niveau de compétence comparable aux dermatologues. Équipés de réseaux neuronaux profonds, les appareils mobiles peuvent potentiellement étendre la portée des dermatologues en dehors

de la clinique. Il est prévu que 6,3 milliards d'abonnements aux smartphones existeront d'ici 2021 et pourront donc potentiellement fournir un accès universel à faible coût aux soins diagnostiques vitaux [15].

Diagnostic de la rétinopathie diabétique

L'intelligence artificielle (IA) a longtemps promis d'augmenter l'abordabilité, la qualité et l'accessibilité des soins de santé, mais la FDA, jusqu'à récemment, n'avait jamais autorisé un système de diagnostic d'IA autonome. Cet essai pivot d'un système d'IA pour détecter la rétinopathie diabétique (RD) chez les personnes atteintes de diabète a recruté 900 sujets, sans antécédents de RD dans les cliniques de soins primaires, en comparaison avec le Wisconsin Fundus Photograph Reading Center (FPRC) la photographie stéréoscopique à champ large et la tomographie par cohérence optique maculaire (TCO), par des photographes certifiés FPRC, et le classement FPRC de l'échelle de gravité de l'étude de rétinopathie diabétique de traitement précoce (ETDRS) et de l'œdème maculaire diabétique (DME). La RD plus que légère (mtmDR) a été définie comme étant de niveau ETDRS 35 ou supérieur, et/ou DME, dans au moins un œil. Les opérateurs de systèmes d'IA ont suivi un protocole de formation normalisé avant le début de l'étude. L'âge médian était de 59 ans (plage de 22 à 84 ans) ; parmi les participants, 47,5 % étaient des hommes ; 16,1 % étaient hispaniques, 83,3 % n'étaient pas hispaniques ; 28,6 % étaient afro-américains et 63,4 % ne l'étaient pas ; 198 (23,8 %) avaient un MTM. Le système d'IA a dépassé tous les paramètres de supériorité prédéterminés à une sensibilité de 87,2 % (IC à 95 %, de 81,8 à 91,2 %) (≥ 85 %), à une spécificité de 90,7 % (IC à 95 %, de 88,3 à 92,7 %) ($\geq 82,5$ %) et à un taux d'imageabilité de 96,1 % (IC à 95 %, de 94,6 à 97,3 %), ce qui démontre la capacité de l'IA à apporter des spécialités diagnostic de niveau aux établissements de soins primaires. Sur la base de ces résultats, la FDA a autorisé le système à utiliser par les fournisseurs de soins de santé pour détecter plus de DR léger et l'œdème maculaire diabétique, ce qui en fait, le premier système de diagnostic IA autonome autorisé par la FDA dans n'importe quel domaine de la médecine, avec le potentiel d'aider à prévenir la perte de vision chez des milliers de personnes diabétiques chaque année [16].

Conclusion

En bref, l'IA montre une influence énorme dans le domaine de la médecine, aidant à améliorer les soins aux patients, le diagnostic des maladies et le développement de traitements.

Dans ce chapitre, nous avons expliqué, historique de l'IA dans le domaine médicale les Techniques d'intelligence artificielle, et les Applications actuelles de l'IA en diagnostic médical.

Dans le prochain chapitre, nous présentons et nous concentrerons sur la technique de l'IA (l'apprentissage supervisée et non-supervisée, réseaux de neurone) et la donnée utilisé Et ses types et n'oublier pas les metriques d'évaluation.

Méthodologie

2.1 Introduction

De nos jours, l'analyse intelligente des données par l'IA, l'apprentissage automatique et les technologies d'apprentissage profond est un élément essentiel de divers domaines, y compris les technologies scientifiques

En appliquant les technologies d'IA au Big Data, des informations précieuses et une analyse précise peuvent être obtenues dans une variété de domaines scientifiques.

Dans ce chapitre, nous présenterons les technique d'intelligence artificielle qui comprennent (l'apprentissage et leur types, réseaux de neuron : CNN, RNN et TRANSFORMER (FLANT5) et les dataset (MIMIC) et les matrices d'évaluation.

2.2 Description des données

Dans cette section, nous expliquerons les type et sources de données utilisées.

2.2.1 Type et source des données utilisées

Les données utilisées

MIMIC est un Médical Information Mart pour les soins intensifs et se compose de plusieurs flux de données compréhensifs dans l'environnement des soins intensifs, avec des niveaux élevés de richesse et de détail, soutenir le traitement du signal complexe et les questions cliniques qui pourraient permettre la détection précoce de problèmes complexes,

fournir une orientation utile sur les interventions thérapeutiques et, en fin de compte, améliorer les résultats pour les patients.

Cet effort complexe a nécessité une collaboration engagée et coordonnée entre les établissements universitaires, industriels et cliniques pour fournir une plate-forme de données radicalement ouverte accessible aux chercheurs du monde entier. Nous traitons de l'information provenant de différentes versions du MIMIC publiées au fil des ans, notamment : MIMIC-IV, MIMIC-III, MIMIC-II [17].

Le type de données

MIMIC-III :

MIMIC-III (« Medical Information Mart for Intensive Care ») est une vaste base de données à centre unique qui contient des renseignements sur les patients admis dans les unités de soins intensifs d'un grand hôpital de soins tertiaires. Les données comprennent les signes vitaux, les médicaments, les mesures de laboratoire, les observations et les notes cartographiées par les fournisseurs de soins, le bilan hydrique, les codes de procédure, les codes de diagnostic, les rapports d'imagerie, la durée du séjour à l'hôpital, les données de survie, etc. La base de données prend en charge des applications, y compris la recherche universitaire et industrielle, les initiatives d'amélioration de la qualité et les cours de l'enseignement supérieur [18].

Contexte et résumé

Ces dernières années, il y a eu un mouvement concerté vers l'adoption de systèmes de dossiers de santé numériques dans les hôpitaux. Aux États-Unis, par exemple, le nombre d'hôpitaux de soins de courte durée non fédéraux dotés de systèmes numériques de base est passé de 9,4 % à 75,5 % entre 2008 et 2014. Malgré cette avancée, l'interopérabilité des systèmes numériques reste une question ouverte, ce qui pose des défis dans l'intégration des données. Par conséquent, le potentiel qu'offrent les données hospitalières en termes de compréhension des soins ciblés reste encore à réaliser. Parallèlement, la communauté de la recherche scientifique est de plus en plus critiquée pour le manque de reproductibilité des études.

Nous rapportons ici la publication de la base de données MIMIC-III, une mise à jour de la base de données MIMIC-II largement utilisée (Data Citation 1). MIMIC-III intègre des données cliniques complètes et désidentifiées des patients admis au Beth Israel

Deaconess Medical Center de Boston, au Massachusetts, et le rend largement accessible aux chercheurs du monde entier dans le cadre d'un accord d'utilisation des données (Fig. 1). La nature ouverte des données permet de reproduire et d'améliorer les études cliniques d'une manière qui ne serait pas possible autrement.

D'après notre expérience avec la version majeure précédente de MIMIC (MIMIC-II, publiée en 2010), nous prévoyons que MIMIC-III sera largement utilisé à l'échelle internationale dans des domaines tels que la recherche universitaire et industrielle, les initiatives d'amélioration de la qualité et les cours de l'enseignement supérieur.

Pour reconnaître l'utilisation de plus en plus large du MIMIC, nous avons renommé le titre complet de la base de données de « Multiparameter Intelligent Monitoring in Intensive Care » à « Medical Information Mart for Intensive Care ». III base de données sur les soins intensifs est unique et remarquable pour les raisons suivantes :

il s'agit de la seule base de données de soins intensifs en accès libre du genre ; l'ensemble de données s'étend sur plus d'une décennie et contient des renseignements détaillés sur les soins aux patients ; l'analyse est sans restriction une fois qu'un accord d'utilisation des données est accepté, ce qui permet la recherche clinique et l'éducation dans le monde entier [19].

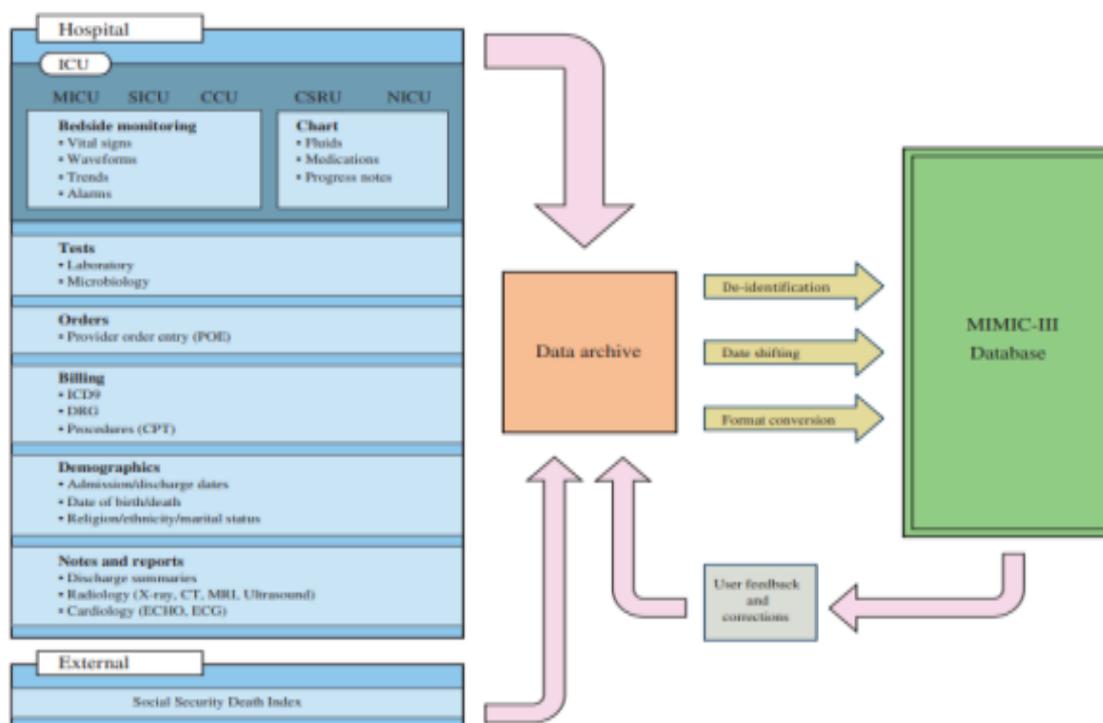


FIGURE 2.1 – Structure de database MIMIC3

2.2.2 Méthodes de collecte et prétraitement des données

Pour les rapports radiologiques et médicaux, les méthodes de collecte et de prétraitement des données MIMIC-III (Medical Information Mart for Intensive Care III) nécessitent également l'utilisation du traitement automatique du langage naturel (NLP) afin d'extraire des informations pertinentes. Voici la procédure à suivre :

Collecte des informations

* Les dossiers médicaux électroniques des patients admis en unité de soins intensifs sont utilisés pour extraire les rapports radiologiques et médicaux.

* Les données sont dé-identifiées afin de préserver la confidentialité des patients en respectant les règles HIPAA.

Prétraitement des données

- Le processus de nettoyage et de normalisation consiste à supprimer les bruits, les erreurs typographiques et autres incohérences des données textuelles. Par la suite, elles sont standardisées afin de standardiser la terminologie, les abréviations, etc.
- La tokenisation consiste à séparer les rapports en tokens (mots ou phrases) afin de simplifier leur traitement par les modèles NLP.
- La lemmatisation et la stemming consistent à réduire les mots à leur forme initiale (lemmes) ou à leur racine (stem) afin de regrouper les différentes formes de mots et d'améliorer la cohérence dans l'analyse.
- L'identification de la catégorie grammaticale (verbe, nom, adjectif, etc.) de chaque mot dans le rapport permet de saisir la structure grammaticale et le sens du texte.
- L'identification et la classification des entités nommées, comme les noms de maladies, les médicaments, les procédures médicales, etc., dans les rapports, sont réalisées par le Named Entity Recognition (NER).
- Analyse des émotions : Repérage des émotions exprimées dans les échanges, comme l'anxiété, la douleur, etc., afin de saisir l'état émotionnel du patient ou du médecin.
- Extraction de données essentielles : Afin de faciliter l'analyse ultérieure, il est important d'identifier des informations pertinentes telles que les diagnostics, les résultats des tests, les interventions médicales, etc.

- Automatisation du résumé : Élaboration de résumés courts des rapports afin de simplifier la visualisation et la compréhension rapide des informations.

Dans le cadre de la collecte et de la préparation des données, les chercheurs peuvent utiliser de manière efficace les rapports médicaux et radiologiques du MIMIC-III pour des analyses avancées, la modélisation prédictive et d'autres applications dans le domaine des soins de santé [20] .

2.3 Modèles d'IA utilisés

2.3.1 Choix des modèles et justification

Transformer

Transformer, une architecture de modèle évitant la récurrence et reposant entièrement sur un mécanisme d'attention pour attirer les dépendances globales entre l'entrée et la sortie. Le transformateur permet une parallélisation nettement plus importante et peut atteindre un nouvel état de l'art en matière de qualité de traduction après avoir été formé pendant aussi peu que douze heures sur huit GPU P100 [21] .

Justification de choix

La caractéristique de FLAN T5 réside dans sa capacité à prendre en compte de manière efficace de nouvelles tâches, même si elles ne sont pas connues, ce qui les rend pratiques pour diverses applications.

En comparaison avec d'autres modèles réglés sur des instructions telles que FLAN, Super-Natural Educations dans la série pensée, MMLU et BBH, FLAN T5 se distingue par ses performances exceptionnelles, ce qui en fait une option privilégiée dans de nombreux scénarios.

Il s'agit d'un instrument puissant dans le domaine des soins de santé, où il peut être employé pour analyser les dossiers médicaux électroniques et extraire des informations essentielles telles que les diagnostics, les médicaments et les résultats de laboratoire. De plus, il peut être utilisé pour fournir des recommandations personnelles concernant les médicaments, les rendez-vous et l'analyse de données volumineuses afin de comprendre les tendances et les tendances dans le domaine des soins de santé. Cela permet aux profes-

sionnels de la santé de prendre des décisions rapides et efficaces et d'améliorer la qualité des soins prodigués.

Grâce à l'utilisation de FLAN T5, les utilisateurs peuvent profiter de sa capacité à prendre en compte de nouvelles tâches, de ses performances spécialisées et de son efficacité dans l'apprentissage à faible débit. Cela en fait un outil puissant et efficace dans diverses applications dans le domaine de l'analyse du langage naturel [22] .

2.3.2 Description des algorithmes et des architectures

FLAN-T5

Définition de modèle (FLANT-5)

”Si vous connaissez déjà le T5, le FLAN-T5 est tout simplement meilleur. Pour le même nombre de paramètres, ces modèles ont été affinés sur plus de 1000 tâches supplémentaires couvrant également plus de langues”

Il fonctionne bien sur un large éventail de tâches de traitement du langage naturel, y compris la traduction de la langue, la classification du texte et la réponse aux questions. Le modèle est connu pour sa vitesse et son efficacité, ce qui en fait une option attrayante pour les applications en temps réel. De plus, FLAN-T5 est conçu pour être hautement personnalisable, permettant aux développeurs de l'affiner pour répondre à leurs besoins spécifiques. Avec ses fonctionnalités avancées et ses hautes performances, FLAN-T5 est sur le point de devenir un acteur majeur dans le domaine du traitement du langage naturel [23][24] .

Architecture de modèle FLAT-5

Le modèle FLAN-T5 est un modèle encodeur-décodeur qui a été pré-entraîné sur un mélange multi-tâches de tâches non supervisées et supervisées et pour lequel chaque tâche est convertie en un format texte-texte.

Au cours de la phase de formation, FLAN-T5 a reçu un grand corpus de données textuelles et a été formé pour prédire les mots manquants dans un texte d'entrée via un remplissage dans l'objectif de style vide. Ce processus est répété plusieurs fois jusqu'à ce que le modèle ait appris à générer du texte similaire aux données d'entrée.

Une fois formé, FLAN-T5 peut être utilisé pour effectuer une variété de tâches NLP, telles que la génération de texte, la traduction de la langue, l'analyse des sentiments et la classification du texte.

Le transformateur suit cette architecture globale en utilisant des couches superposées d'auto-attention et de point-sage, entièrement connectées pour l'encodeur et le décodeur, montrés dans les moitiés gauche et droite de la figure 1, respectivement [25].

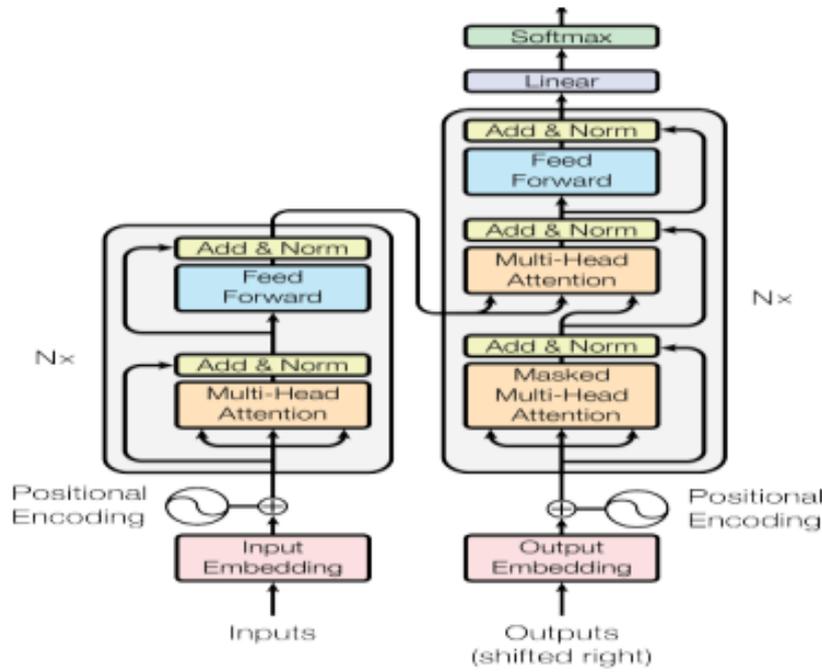


FIGURE 2.2 – Architecture de modèle FLAT-5

Piles d'encodeurs et de décodeurs :

Encodeur :

L'encodeur est composé d'une pile de $N = 6$ couches identiques. Chaque couche a deux sous-couches. Le premier est un mécanisme d'auto-assistance multi-tête, et le second est un réseau d'alimentation en avant simple, position Wise entièrement connecté. Nous utilisons une connexion résiduelle [11] autour de chacune des deux sous-couches, suivie d'une normalisation de la couche [1]. C'est-à-dire que la sortie de chaque sous-couche est $\text{LayerNorm}(x + \text{Sublayer}(x))$, où $\text{Sublayer}(x)$ est la fonction implémentée par la sous-couche Lui-même. Pour faciliter ces connexions résiduelles, toutes les sous-couches du modèle, ainsi que les couches d'intégration, produisent des sorties de dimension de model = 512.

Décodeur :

Le décodeur est également composé d'une pile de $N = 6$ couches identiques. En plus des deux sous-couches dans chaque couche d'encodeur, le décodeur insère une troisième sous-couche, qui effectue une attention multi-tête sur la sortie de la pile d'encodeur. Comme pour l'encodeur, nous utilisons des connexions résiduelles autour de chacune des sous-couches, suivies d'une normalisation des couches. Nous modifions également l'attention personnelle sous-couche dans la pile du décodeur pour empêcher les positions de s'occuper des positions suivantes. Ce masquage, combiné au fait que les embeddings de sortie sont compensés par une position, garantit que les prédictions pour la position i ne peuvent dépendre que des sorties connues à des positions inférieures à i [26].

2.4 Validation et évaluation des modèles

2.4.1 Méthodes d'évaluation des performances des modèles

En bref, les mesures prises pour diagnostiquer le patient et analyser les données. Grâce à l'organisation détaillée du processus :

Diagnostic du patient :

- Recueillir des informations médicales et cliniques sur le patient.

- Effectuer les tests nécessaires et analyser les résultats.
- Les données de la dataset Mimic contiennent une partie représentant 10% du pourcentage de données.

Compilation de l'information provenant de la radio :

- Extraire les informations disponibles de la radio.

Traitement :

- Cette partie des données que nous lui offrons les résultats recueillis à partir de la radio (informations patient).
- Traitement et analyse des fichiers.

Affichage des résultats :

- Présentation des conclusions(sorties).
- Database contient des sorties et des entrées.

Comparaison des résultats :

Comparer les résultats obtenus à partir du modèle avec les résultats trouvés dans la dataset Mimic pour identifier les similitudes et les différences.

- Ce règlement permet de s'assurer que les étapes sont effectuées de manière ordonnée et efficace, conduisant à une analyse approfondie et complète de l'état de santé du patient.

2.4.2 Indicateurs de performance

Bilingual Evaluation Understudy (BLEU) :

L'UEBL évalue la qualité du texte traduit par une machine d'une langue naturelle à une autre. BLEU Score est une mesure de performance pour mesurer la performance des modèles de traduction automatique. Il évalue la façon dont un modèle se traduit d'une langue à l'autre. Le TM comparera l'unigramme, le bigramme ou le trigramme en sortie avec la vérité du sol. Il ne tient pas compte du sens, de la structure de la phrase et du langage morphologiquement riche [30].

La formule de BLEU Score est :

1. Précision des n-grammes de BLEU (n-gram precision) :

$$P_n = \frac{\sum_{n-gram} Count_{clip}(n-gram)}{\sum_{n-gram} Count(n-gram)}$$

où $Count_{clip}(n-gram)$ est le nombre d'occurrences de l'n-gram dans la traduction générée, limité par son nombre maximal d'occurrences dans une référence quelconque.

et $Count(n-gram)$ nombre total d'occurrences de l'n-gramme dans la traduction générée.

2. Score BLEU :

$$BLEU = BP \cdot \exp \left(\sum_{n=1}^N w_n \log P_n \right)$$

- N est le nombre maximal de grammes.
- w_n est le poids de chaque précision des n-grammes, souvent $\frac{1}{N}$ pour tous les n-grammes.
- BP est le Brevity Penalty (pénalité de brièveté) qui pénalise les phrases générées trop courtes :

$$BP = \begin{cases} 1 & sic > r \\ e^{(1-\frac{r}{c})} & sic \leq r \end{cases}$$

- c est la longueur de la traduction générée.
- r est la longueur de la référence.

METEOR

The Metric for Evaluation of Translation with Explicit ORdering (METEOR) est une mesure basée sur la précision pour l'évaluation de la sortie de traduction automatique. Il surmonte certains des pièges du score UEBL, tels que l'appariement exact des mots tout en calculant la précision. Le score METEOR permet de faire correspondre les synonymes et les mots à tige avec un mot de référence.

Les n-grammes peuvent être appariés en fonction des mots et des significations. METEOR utilise la précision et le rappel d'un unigramme pour calculer un score [31].

1. Rappel

Le rappel mesure la proportion des mots de la référence qui sont trouvés dans la traduction

générée. Un rappel élevé indique que la plupart des mots importants de la référence sont présents dans la traduction générée.

$$R = \frac{\text{Nombre d'unigrammes correspondants}}{\text{Nombre total d'unigrammes dans la référence}}$$

- **Nombre d'unigrammes correspondants** : C'est le nombre d'unigrammes (mots individuels) dans la traduction générée qui sont également présents dans la référence.
- **Nombre total d'unigrammes dans la référence** : C'est le nombre total d'unigrammes dans le texte de référence.

2. Précision)

La précision mesure la proportion des mots dans la traduction générée qui sont également dans la référence. Une précision élevée signifie que peu de mots dans la traduction générée sont superflus ou incorrects par rapport à la référence.

$$P = \frac{\text{Nombre d'unigrammes correspondants}}{\text{Nombre total d'unigrammes dans l'hypothèse}}$$

- **Nombre d'unigrammes correspondants** : C'est le nombre d'unigrammes dans la traduction générée qui sont également présents dans la référence.
- **Nombre total d'unigrammes dans l'hypothèse** : C'est le nombre total d'unigrammes dans la traduction générée.

3. F-mesure harmonisée

La F-mesure harmonisée est une moyenne pondérée de la précision et du rappel. Dans le cas de METEOR, la précision est pondérée 9 fois plus que le rappel, ce qui signifie que l'on accorde plus d'importance à la précision qu'au rappel. Cela aide à s'assurer que la traduction générée n'ajoute pas trop de mots incorrects, tout en gardant un certain niveau de rappel.

$$F_{mean} = \frac{10 \cdot P \cdot R}{9 \cdot P + R}$$

- P : Précision.
- R : Rappel.

4. Pénalité de fragmentation

Cette pénalité réduit le score METEOR si les correspondances sont trop fragmentées, encourageant des correspondances plus continues et ordonnées.

$$Pen = 0.5 \left(\frac{ch}{m} \right)^3$$

- ch : C'est le nombre de groupes de mots (chunks) correspondants qui sont contigus dans l'ordre correct.
- m : C'est le nombre total de mots (Nombre total de correspondances d'unigrammes) dans la traduction générée qui correspondent à des mots dans la référence.

5. Score METEOR

Le score METEOR combine la F-mesure harmonisée et la pénalité de fragmentation. En multipliant la F-mesure par $(1 - Pen)$, on réduit le score proportionnellement à la fragmentation. Ainsi, une traduction avec une bonne précision et rappel mais avec une forte fragmentation aura un score plus bas.

$$METEOR = F_{mean} \cdot (1 - Pen)$$

- F_{mean} : F-mesure harmonisée.
- Pen : Pénalité de fragmentation.

ROUGE

Recall-Oriented Understudy for Gisting Evaluation (Lin, 2004) est une autre mesure d'évaluation lexicale plus axée sur le rappel. Il a été conçu à l'origine pour mesurer la synthèse de texte, principalement en comptant le nombre d'unités de chevauchement (n-gramme, séquences de mots et paires de mots) entre la prédiction et les références. ROUGE-L le plus commun dans nos expériences[32]. Les variantes les plus courantes du ROUGE sont les suivantes :

ROUGE-N : Se concentre sur les n-grammes (phrases de N-mot).

ROUGE-1 et **ROUGE-2** (unigrammes et bigrammes, respectivement) sont les plus communs.

ROUGE-L : Basé sur la plus longue sous-séquence commune (LCS), qui prend en compte la similitude de structure au niveau de la phrase naturellement et identifie automatiquement les n-grammes co-occurents les plus longs dans la séquence [33].

1. Rappel

Le rappel mesure la proportion de la LCS par rapport à la longueur totale du texte de

référence. Un rappel élevé indique qu'une grande partie du texte de référence est capturée dans le texte généré.

$$R_{LCS} = \frac{LCS(X, Y)}{|Y|}$$

- $LCS(X, Y)$: Longueur de la plus longue sous-séquence commune entre le texte généré X et le texte de référence Y .
- $|Y|$: Longueur du texte de référence Y (nombre de mots).

2. Précision

La précision mesure la proportion de la LCS par rapport à la longueur totale du texte généré. Une précision élevée indique que la plupart des mots du texte généré sont pertinents par rapport au texte de référence.

$$P_{LCS} = \frac{LCS(X, Y)}{|X|}$$

- $LCS(X, Y)$: Longueur de la plus longue sous-séquence commune entre le texte généré X et le texte de référence Y .
- $|X|$: Longueur du texte de référence X (nombre de mots).

3. F-mesure

La F-mesure combine la précision et le rappel en une seule mesure. Elle est la moyenne harmonique pondérée de la précision et du rappel. Lorsque $\beta = 1$, la F-mesure donne une importance égale au rappel et à la précision.

$$F_{LCS} = \frac{(1 + \beta^2) \cdot R_{LCS} \cdot P_{LCS}}{R_{LCS} + \beta^2 \cdot P_{LCS}}$$

- R_{LCS} : Rappel basé sur LCS.
- P_{LCS} : Précision basé sur LCS.
- β : Coefficient qui contrôle l'importance relative du rappel par rapport à la précision (souvent fixé à 1, ce qui donne une importance égale au rappel et à la précision).

Le **ROUGE-L** fournit une évaluation équilibrée qui prend en compte à la fois la quantité de contenu pertinent capturé par la traduction et la concision de cette traduction.

Résultats

3.1 Introduction

Dans cette chapitre Les résultats de notre étude sont exposés dans ce chapitre, en soulignant les résultats de chaque modèle utilisé et en comparant leurs performances. Les résultats sont également évalués en fonction des objectifs de la recherche, en évoquant à la fois les atouts et les limites des modèles. Par la suite, nous examinons les conséquences concrètes pour le diagnostic médical et l'incidence potentielle sur les pratiques cliniques, en comparant nos résultats avec ceux de la littérature existante et en identifiant les limites de notre étude ainsi que les voies de recherche à venir.

3.2 Environnement de développement

3.2.1 Outils et logiciels utilisés

Kaggle

Kaggle est une plate-forme pour les concours de science des données, où les data scientistes et les ingénieurs en apprentissage automatique peuvent se concurrencer pour créer les meilleurs modèles pour résoudre des problèmes spécifiques ou analyser certains ensembles de données. La plateforme fournit également une communauté où les utilisateurs peuvent collaborer sur des projets, partager du code et des ensembles de données, et apprendre du travail de chacun. Fondée en 2010, Google a acquis Kaggle en 2017 et la plateforme fait désormais partie de Google Cloud.

Kaggle organise une variété de concours parrainés par des organisations, allant de la prévision des résultats médicaux à la classification des images ou à l'identification des transactions frauduleuses. Les participants peuvent soumettre leurs modèles et voir comment ils performant sur un classement public, ainsi que recevoir des commentaires d'autres concurrents et de la communauté.

En plus des compétitions, Kaggle propose également des ensembles de données publiques, des ordinateurs portables d'apprentissage automatique et des tutoriels pour aider les utilisateurs à apprendre et à pratiquer leurs compétences en science des données et en apprentissage automatique. Il est devenu une plate-forme populaire pour les scientifiques de données novices et expérimentés pour améliorer leurs compétences, construire leurs portefeuilles et se connecter avec d'autres dans l'industrie [27].

Python

Python est un langage de programmation interprété, orienté objet et de haut niveau avec une sémantique dynamique. Ses structures de données intégrées de haut niveau, combinées au typage dynamique et à la liaison dynamique, le rendent très attrayant pour le développement d'applications rapides, ainsi que pour une utilisation comme langage de script ou de collage pour connecter des composants existants. La syntaxe simple et facile à apprendre de Python met l'accent sur la lisibilité et réduit donc le coût de la maintenance du programme. Python prend en charge les modules et les paquets, ce qui encourage la modularité du programme et la réutilisation du code. L'interpréteur Python et la vaste bibliothèque standard sont disponibles sous forme source ou binaire sans frais pour toutes les principales plateformes, et peuvent être distribués librement [28].

Carte graphique (GPU NVIDIA Tesla P100)

Les accélérateurs GPU NVIDIA Tesla P100 sont les plus avancés jamais construits, alimentés par l'architecture révolutionnaire NVIDIA Pascal™ et conçus pour augmenter le débit et économiser de l'argent pour les centres de données HPC et hyperscale. Dernier né de cette famille, le Tesla P100 pour PCIe permet à un seul nœud de remplacer un demi-rack de nœuds CPU de base en offrant des performances ultra-rapides dans un large éventail d'applications HPC [29].

NVIDIA Tesla P100 Performance

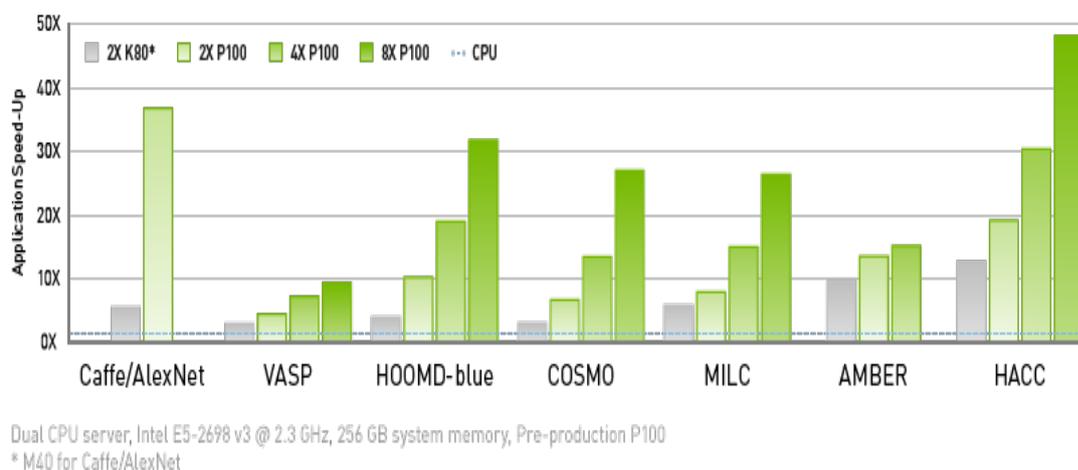


FIGURE 3.1 – NVIDIA Tesla P100 Performance.

Hugging Face

Hugging Face, Inc. est une entreprise franco-américaine basée à New York qui développe des outils informatiques pour la création d'applications utilisant l'apprentissage automatique. Leur offre la plus remarquable est la bibliothèque Transformers, qui propose des implémentations open-source de modèles de type transformer pour des tâches de traitement de texte, d'image et de son. Cette bibliothèque prend en charge des modèles populaires tels que BERT et GPT-2 et est compatible avec PyTorch, TensorFlow et JAX. De plus, Hugging Face propose une plateforme où les utilisateurs peuvent partager des modèles d'apprentissage automatique, des ensembles de données et présenter leur travail. Si vous souhaitez explorer des modèles et des ressources d'intelligence artificielle, vous pouvez accéder au Hugging Face Hub depuis leur site web, leur interface en ligne de commande (CLI) ou leur bibliothèque client Python.

3.2.2 Les caractéristiques de l'ordinateur

- Ordinateur : TOSHIBA.
- Processeur : Intel(R) Core(TM) i3-5005U CPU @ 2.00GHz.
- RAM : 12,00 Go

3.3 Analyse et interprétation des résultats

3.3.1 Analyse Résultats obtenus avec chaque métrique en fonction de recherche

Bleu-p4

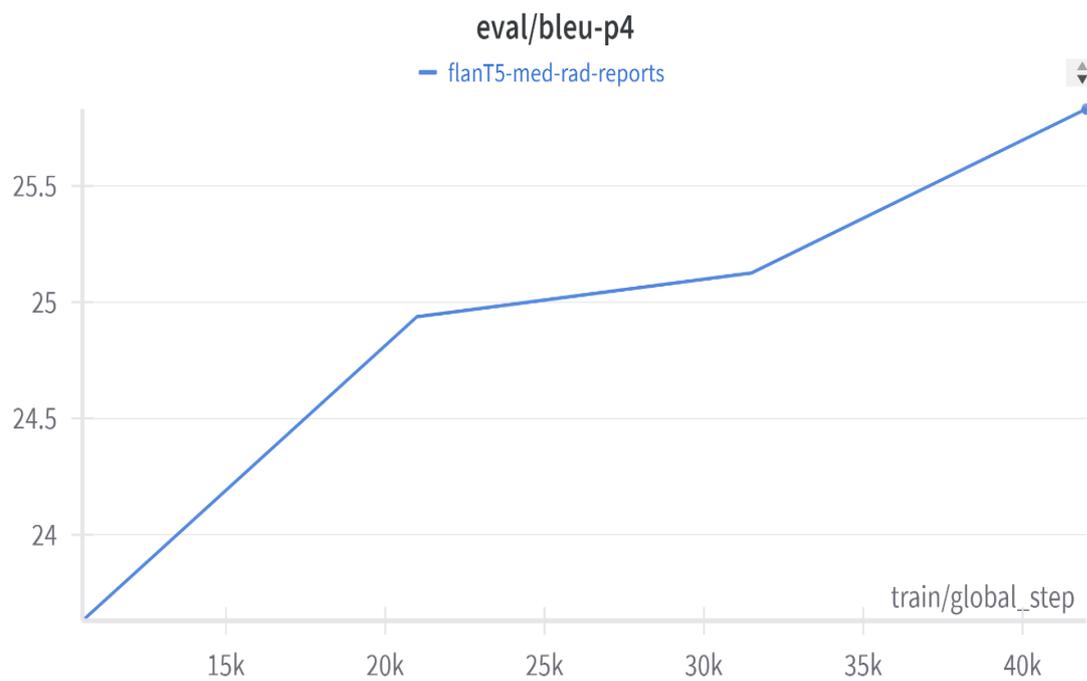


FIGURE 3.2 – Scores BLEU-p4 du modèle "flanT5-med-rad-reports"

Le graphe ci dessus montre l'évaluation des scores BLEU-p4 du modèle "flanT5-med-rad-reports" en fonction du nombre d'étapes d'entraînement.

Analyse de résultat (Bleu4)

Le graphe montre l'évolution des scores BLEU-4 du modèle "flanT5-med-rad-reports" par rapport au nombre d'étapes de formation.

Le graphe ci-dessus montre que les scores BLEU-4 du modèle "flanT5-med-rad-reports" augmentent linéairement par rapport aux étapes, passant d'environ 24 à 25,5. Ce type de croissance droite sans grande oscillation suggère un apprentissage stable et approprié. L'absence de saturation implique directement que ce modèle bénéficiera encore d'une formation supplémentaire. Une telle amélioration de la qualité du texte généré est prometteuse en ce qui concerne les applications pratiques dans des domaines tels que la génération de rapports de radiologie médicale. Pour une efficacité maximale, une validation croisée, une analyse qualitative des textes et une optimisation des hyperparamètres pourraient être nécessaires.

Meteor

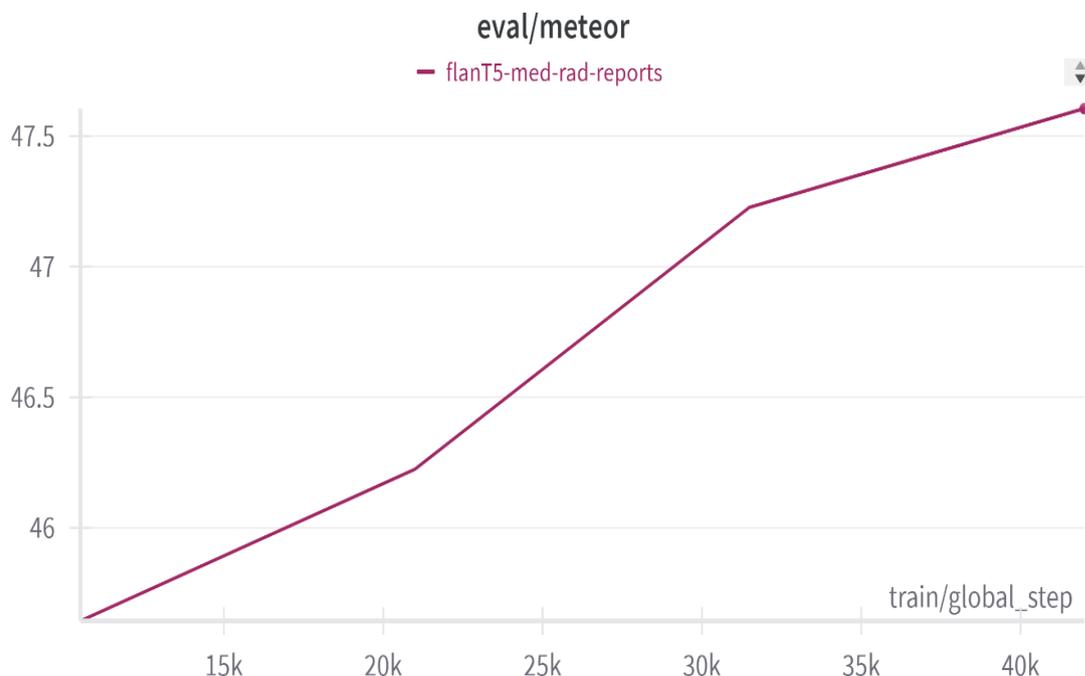


FIGURE 3.3 – Scores Meteor du modèle "flanT5-med-rad-reports"

Le graphe ci dessus montre l'évaluation des scores METEOR pour le modèle "flanT5-med-rad-reports" au fur et à mesure des étapes d'entraînement.

Analyse de résultat Meteor

Par conséquent, le graphe montre une augmentation continue des scores METEOR pour le modèle "flanT5-med-rad-reports." En d'autres termes, il passe d'un score initial d'environ 46 à 47,5 à chaque étape du processus de formation. Cette augmentation constante exige que le modèle produise des rapports de radiologie précis et linguistiquement cohérents avec les diagnostics médicaux. Grâce à leur rapprochement avec les normes humaines, ces rapports améliorés rendent les diagnostics plus précis, réduisent au minimum les risques d'erreurs et accélèrent le processus de prise de décision clinique, ce qui rend la manière de faciliter les soins aux patients plus efficace et fiable.

Rouge-L

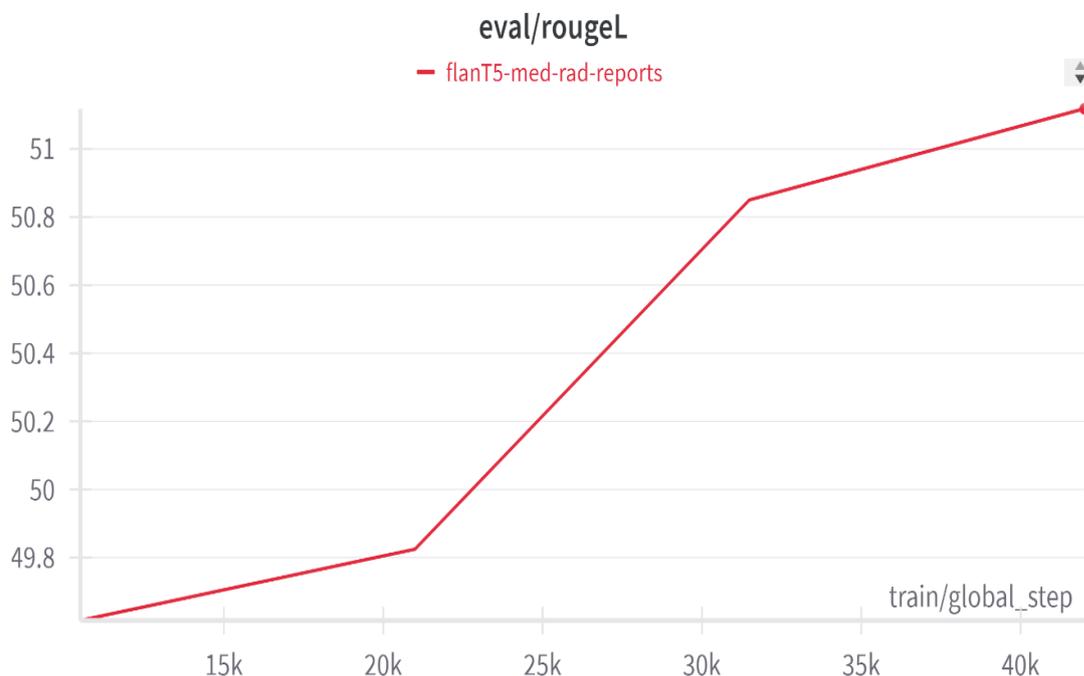


FIGURE 3.4 – Scores Rouge-L du modèle "flanT5-med-rad-reports"

Le graphe ci dessus montre l'évaluation des scores ROUGE-L pour le modèle "flanT5-med-rad-reports" au fur et à mesure des étapes d'entraînement.

Analyse de résultat Rouge-L

Ceci est un graphique des scores ROUGE-L ou Recall-Oriented Understudy for Gisting Evaluation résultats de l'évaluation pour un modèle nommé "flanT5-med-rad-reports". Des améliorations progressives des scores sont affichées avec des étapes de formation ou des étapes globales. À partir de 14k pas, au début, le score était d'environ 49,8, puis en augmentation constante jusqu'à environ 51 à 40k pas. Cette augmentation montre que le modèle devient lentement mais sûrement plus apte à générer des résumés plus proches des rapports radiologiques médicaux de référence. On pourrait en conclure que la formation aide effectivement le modèle à améliorer ses résumés diagnostiques générés et ce qui est très important pour des informations claires et précises pour les professionnels de la santé dans le diagnostic des patients.

Train Eval loss

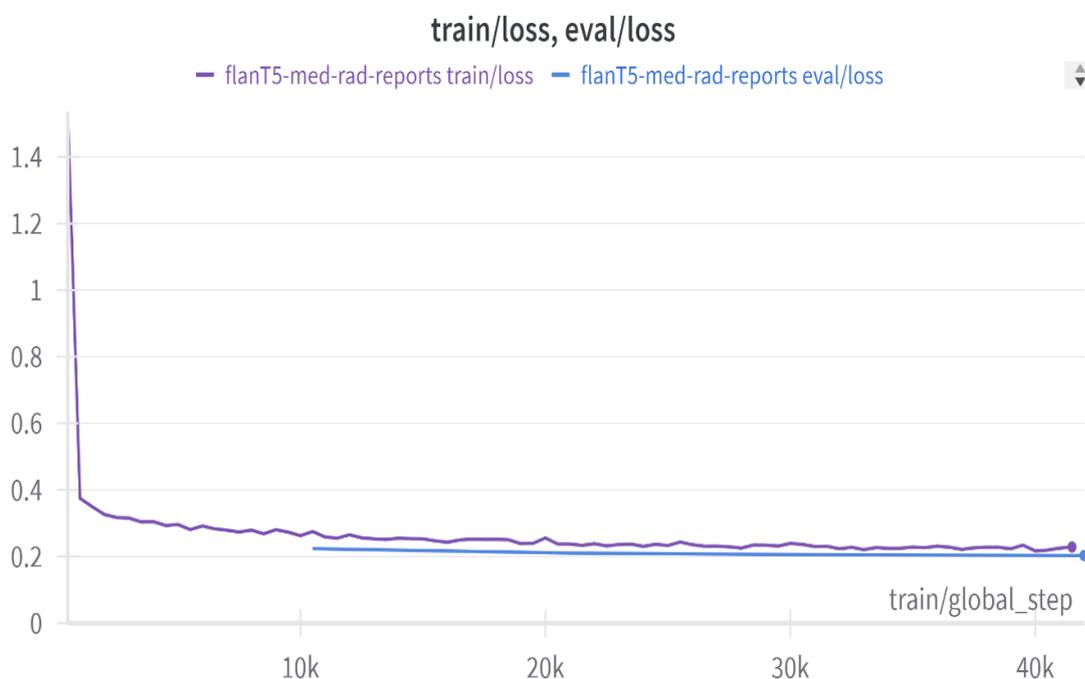


FIGURE 3.5 – Scores train eval loss du modèle "flanT5-med-rad-reports"

Le graphe ci dessus montre l'évaluation des scores (TRAIN/LOSS, EVAL/LOSS) pour le modèle "flanT5-med-rad-reports" au fur et à mesure des étapes d'entraînement.

Analyse de résultat Train/loss, Eval/loss

Il s'agit d'un graphique des pertes de formation et d'évaluation pour le modèle "flanT5-med-rad-reports" au cours de 40000 étapes de formation. Les pertes commencent à un niveau élevé puis chutent très rapidement de valeurs initiales autour de 1,4 à environ 0,4 à 10000 pas. De là, ils se stabilisent autour de 0,2 pour la plupart des étapes restantes. Cette baisse rapide initiale indique donc que le modèle apprend correctement à identifier les structures et les modèles dans les données de formation, comme prévu au début de la formation où les changements de poids sont les plus significatifs.

les pertes se stabilisent à un faible niveau après 10 000 pas ; c'est-à-dire qu'il a optimisé ses paramètres pour minimiser les erreurs de prévision sur les données de formation, tout en montrant une bonne généralisation sur les données de validation. La proximité des courbes de formation, les pertes d'évaluation prouvent qu'il n'y a pas eu de surajustement lorsqu'un modèle s'adaptait trop bien aux données de formation, mais qu'il ne fonctionnait pas aussi bien sur les nouvelles données.

Ceci est important dans les objectifs de diagnostic des patients. Un modèle contenant une perte minimale et stable a la capacité de faire des résumés de rapports de radiologie qui sont précis et cohérents, diminuant ainsi la possibilité d'un mauvais diagnostic. Puisque le modèle garantit que tout est bien généralisé, il est stable et fiable ; par conséquent, il fournit des informations critiques aux cliniciens de manière systématique et précise, améliorant la qualité et la sécurité des soins aux patients.

3.3.2 Comparaison des performances des différents modèles

Model	ROUGE-1	ROUGE-2	ROUGE-L	METEOR	BLEU-p4
ku-dmis -msra	/	/	47.75	/	25.58
WGSUM(Trans+GAT)	48.37	33.34	46.68	/	/
utsa-nlp-T5	/	/	47.86	/	25.87
BART+Oracle Guidance	51.46	31.71	49.99	/	/
flanT5-mimic-med- rad-reports(ours)	52.87	42.42	51.11	47.60	25.83

TABLE 3.1 – COMPARAISON DES PERFORMANCES des métrique d’éval entre les modèles

3.3.3 Discussion sur les points forts et les limites des modèles

MODELE FLANT-5

Les points forts :

- **Polyvalence** : La possibilité d’effectuer plusieurs tâches NLP avec une grande précision en changeant le préfixe d’entrée est une force significative du modèle Flan T5.
- **Évolutivité** : Avec des variantes allant de petit à xxl, le modèle peut gérer des tâches de complexité et de taille variables [47].
- **Personnalisation** : Le modèle peut être affiné pour des ensembles de données spécifiques, améliorant son applicabilité sur différents domaines [48].
- Se concentre sur la création d’informations en comprenant le langage et en créant de nouveaux contenus. Bon pour raccourcir les informations médicales complexes.

Les limites

- Rae et al. (2021) ont averti que le Flan-T5 pourrait avoir des applications nocives malgré son potentiel de génération linguistique. Par conséquent, les auteurs recommandent que le Flan-T5 ne soit pas utilisé sans une évaluation approfondie des préoccupations en matière d’innocuité et d’équité propres à la demande en question [49].
- **Considérations et risques éthiques** : Étant donné que Flan-T5 est formé sur une grande quantité de données textuelles qui n’ont pas été examinées pour un contenu explicite ou évaluées pour des biais potentiels, le modèle peut avoir le potentiel de générer un contenu inapproprié similaire ou de reproduire les biais existants présents dans les données originales. Par conséquent, il faut faire preuve

de prudence avant d'utiliser Flan-T5 dans toute application, et des mesures appropriées doivent être prises pour régler les problèmes liés à la sécurité et à l'équité [49].

- **temps de formation** : La formation des modèles FLAN-T5 nécessite une grande quantité de ressources informatiques et prend un temps considérable. Cela peut rendre difficile la mise en œuvre rapide de nouveaux modèles et le test de différentes configurations [50].

MODELE BART

Les points forts :

- **Cohérence et fluidité** : BART est particulièrement fort pour générer des résumés cohérents et pertinents sur le plan contextuel en capturant les dépendances à long terme dans le texte [51].
- **Connaissances pré-enseignées** : Sa formation préalable à grande échelle sur divers corpus lui permet de comprendre divers modèles linguistiques, ce qui est bénéfique pour résumer différents types de sources de contenu [64] .
- BART+Oracle Guidance s'est montré très prometteur dans l'amélioration des tâches de synthèse de texte en intégrant des conseils oracle de type humain dans le processus de réglage fin du modèle. Cette approche améliore la qualité et la précision des résumés générés, en particulier dans les domaines où des orientations ou des contraintes spécifiques sont critiques [53] .

Les Limites :

- **Complexité de calcul** : BART est intensif en calcul et nécessite des ressources substantielles pour le réglage fin et l'inférence, ce qui peut être une limitation pour les utilisateurs avec une source de capacités de calcul restreinte.
- **Variation des performances** : Les performances de BART peuvent se dégrader si l'ensemble de données personnalisé diverge significativement de la distribution des données de préformation, ce qui conduit à des résumés sous-optimaux dans de tels cas [52] .

MODELE L'UTSA-NLP-T5

Les points fort :

Le modèle T5 se distingue dans les tâches d'extraction de l'étendue du sentiment en raison de ses capacités génératives avancées et de sa capacité à produire des résultats divers et pertinents sur le plan contextuel. Par rapport à d'autres modèles, tels que les modèles GPT-3, BART et Seq2seq, T5 a gagné en importance pour son efficacité dans l'extraction de l'étendue de sentiment [54] .

Les limites :

- **Ressources informatiques** : Les exigences informatiques de l'UTSA-NLP-T5, en particulier lors des étapes de réglage et d'inférence, peuvent poser des défis aux chercheurs ayant un accès limité aux ressources informatiques haute performance. Des modèles de grande taille et des architectures complexes peuvent nécessiter une infrastructure de calcul importante, limitant l'accessibilité et l'évolutivité du modèle [55] .
- **Interprétabilité et explicabilité** : Comme avec de nombreux modèles d'apprentissage profond, les processus décisionnels de l'UTSA-NLP-T5 peuvent être opaques et difficiles à interpréter. Le manque de transparence dans les prédictions des modèles peut entraver la confiance et l'adoption, en particulier dans les applications où l'interprétabilité est cruciale, comme les domaines juridiques ou de santé [56] .

MODELE WGSUM

Les Points forts :

Cette combinaison permet au modèle de générer des résumés plus cohérents et pertinents sur le plan contextuel. Les résultats expérimentaux ont validé les fortes performances du modèle, notamment dans le cadre de la génération automatique d'empreintes pour les résultats radiologiques. En utilisant des graphiques de mots construits à partir des résultats, le modèle WGSUM excelle à préserver la signification originale du texte d'entrée tout en le condensant efficacement en résumés concis. Cette approche a démontré son efficacité sur des ensembles de données comme OpenI et MIMIC-CXR[57]

Les limites :

- **Besoin de guide faible** : Comme son nom l'indique, le WGSUM utilise le guidage léger pour faire du texte. Cela signifie qu'il peut être freiné par la qualité et la justesse de l'information du guide. De mauvaises informations de guide peuvent

conduire à de mauvais résultats de texte.

MODELE KU-DMIS-MSRA

Les Points forts :

- **Excellentes performances** : Le modèle Ku-DMIS-MSRA fonctionne exceptionnellement bien dans une variété de tâches de traitement du langage naturel (TAL), telles que la classification du texte, la génération et la traduction. Son architecture sophistiquée lui permet de reconnaître des relations complexes dans les données textuelles.
- **Polyvalence** : La conception modulaire et les mécanismes d'attention du Ku-DMIS-MSRA lui permettent de s'adapter facilement à diverses tâches PNL sans nécessiter de changements importants. Cette polyvalence le rend pertinent pour un large éventail de scénarios d'utilisation.

Les limites :

- **Puissance de calcul** : Comme de nombreux modèles de langage avancés, le Ku-DMIS-MSRA peut avoir besoin de beaucoup de puissance informatique pour s'entraîner et faire des prédictions, ce qui peut le rendre coûteux en termes de temps et d'équipement.
- **Comprendre** : Les modèles de langage profond comme Ku-DMIS-MSRA peuvent être difficiles à comprendre car ils sont complexes. Savoir comment le modèle fait ses choix peut être difficile, ce qui rend difficile de repérer les erreurs ou les biais possibles [58] .

3.4 Implications des résultats

3.4.1 Implications pratiques pour le diagnostic médical

Les applications pratiques pour le diagnostic en médecine, comme l'utilisation de réseaux de neurones, l'application de transformateurs comme FLANT5, l'application de l'apprentissage automatique et des mesures d'évaluation comme ROUGE-L, BLEU-4 et METEOR, sont nombreuses dans le MIMIC-Jeu de données III lorsque la plateforme Kaggle est appliquée :

- **Amélioration de la précision du diagnostic** : Ces technologies permettent une analyse plus approfondie des données médicales, ce qui permet d'identifier plus précisément les symptômes, de poser un diagnostic potentiel et de recommander le traitement.
- **Optimisation du flux de travail clinique** : Des modèles pré-entraînés sont intégrés dans les systèmes de gestion de l'information médicale ; ceux-ci peuvent effectuer certaines des tâches courantes de ces systèmes et sur lesquelles les médecins ont pris du retard dans les aspects cliniques et décisionnels.
- **Personnalisation du traitement** : De tels modèles peuvent aider davantage à personnaliser le traitement en analysant les dossiers médicaux des patients en fonction des caractéristiques individuelles des patients et des antécédents médicaux.
- **Reconnaissance et réduction des erreurs médicales** : Ils aideront les professionnels de la santé à assurer le suivi de tous les processus, du diagnostic au traitement, soit en surveillant l'identification des renseignements pertinents, soit en suggérant diverses formes de traitement [59] .

3.4.2 Potentiel impact sur les pratiques cliniques

Puisque les mesures d'évaluation telles que BLEU, METEOR et ROUGE sont couramment utilisées en TALN pour évaluer la qualité des traductions automatiques ou résumés automatiques de textes, l'impact immédiat sur les Pratiques Cliniques peut ne pas être discernable. Cependant, il y a un impact indirect manifeste de ces mesures dans la perspective des Pratiques Cliniques . Ceux-ci peuvent se manifester en :

- **Évaluation à base de la performance du système** : il peut être effectué en utilisant des métriques d'évaluation infrastructures telles que le ROUGE pour noter la qualité du résumé des enregistrements médicaux. Améliorer la qualité du résumé facilite la compréhension rapide des aspects critiques de l'information clinique, un facteur qui est d'une grande importance pour les praticiens.
- **Évaluation de la traduction automatique au-dessus du système** : par exemple, ces évaluations peuvent être faites à partir de ces métriques BLEU et METEOR pour noter la qualité de la traduction. Dans des environnements multilingues, la qualité des traductions automatiques a un rôle important dans la prestation du service de qualité aux patients.

- **Recherche clinique** : Les mesures d'évaluation de la qualité de la production de texte peuvent être utilisées dans le domaine de la recherche clinique pour évaluer la qualité de la littérature médicale automatiquement abrégée. Cela permettrait une identification rapide et précise de la littérature pertinente à utiliser par le praticien. Cela peut avoir une incidence sur les décisions cliniques prises.
- **Création et formation médicale** : Le résumé automatique des dossiers médicaux en création médicale et en formation continue peut être facilité par : la façon dont l'évaluation de la qualité des sommaires créés automatiquement sur les dossiers médicaux par les mesures d'évaluation, comme le ROUGE, peut avoir une incidence sur le transfert des connaissances médicales aux praticiens et sur leur assimilation [60] .

3.5 Comparaison avec les travaux antérieurs

3.5.1 ku-dmis-msra

En 2023, Ku-DMiS-MSRA a été introduit, formé sur l'ensemble de données MIMIC-CXR (Johnson et al., 2019) et évalué à la fois sur l'ensemble de tests MIMIC-CXR et un ensemble de tests cachés nouvellement assemblés. MIMIC-CXR comprend 128 032 paires d'images de rapport et 227 835 images multi-vues, accessibles au public. Ce modèle est affiné en utilisant la perte d'entropie croisée vers des séquences de vérité du sol, fonctionnant sur les entrées (x) et les sorties (y) [61] .

3.5.2 WGSUM (Trans+TAG)

WGSUM (Trans+GAT), dévoilé en 2021, a suivi une formation sur l'ensemble de données MIMIC-III. Utilisant l'approche standard de séquence à séquence pour la génération automatisée d'images (AIG), ce modèle comprend trois composants principaux : l'encodeur FINDINGS, l'encodeur graphique et le décodeur guidé par graphique [62] .

3.5.3 utsa-nlp-T5

UTSA-NLP-T5, dévoilé en 2023, a suivi une formation sur les ensembles de données MIMIC-III et MIMIC-CXR. L'approche du modèle comporte deux étapes : Initialement,

il récupère le rapport de radiologie le plus similaire à l'aide d'un encodeur multimodal pré-entraîné basé sur des images radiographiques thoraciques et des résultats d'un corpus médical. Ensuite, il augmente les résultats de l'exemple d'entrée en préparant les principales conclusions et impressions du rapport le plus similaire.

Dans l'étape suivante, un modèle textuel génère l'impression en fonction des entrées augmentées. Cette méthodologie s'inspire d'études antérieures menées par Endo et al. (2021) et Jeong et al. (2023), traitant la génération de rapports médicaux comme une tâche de récupération de texte d'image et tirant parti d'un grand corpus de récupération pour une couverture complète des diagnostics potentiels d'un thorax d'entrée X-ray image (Yu et al., 2022) [63] .

3.5.4 BART+Oracle

Guidance BART+Oracle Guidance, introduit en 2023 et formé sur le jeu de données MIMIC-III, s'appuie sur le modèle BART. Il utilise des résumés échantillonnés de l'ensemble de formation comme guide des connaissances préalables. Cette approche implique l'encodage de représentations contextuelles supplémentaires du guidage avec l'encodeur et l'amélioration des représentations de décodage dans le décodeur [64] .

3.6 Limites de l'étude et perspectives de recherche

3.6.1 Limites méthodologiques et techniques

L'IA dans le diagnostic médical n'est pas exempte de problèmes, parmi lesquels des problèmes méthodologiques et techniques. Bien sûr, parmi les plus importants est l'interprétabilité des algorithmes utilisés. Dans la plupart des cas, en particulier avec les réseaux de neurones d'apprentissage profond, les modèles ont tendance à être assez complexes, agissant ainsi comme des boîtes noires dans l'explication de leurs décisions. Une telle situation nuit à l'acceptation et à la confiance des systèmes d'IA parmi les professionnels de la santé [38] .

Un autre problème fondamental concerne la qualité et la quantité des données. Des ensembles de données de haute qualité et non biaisés sont nécessaires en grande quantité pour développer des modèles d'IA qui peuvent faire des inférences vraiment fiables. En

fait, la plupart des données médicales présentent certains problèmes, comme des lacunes ou des biais, qui peuvent entraîner une erreur dans les résultats (MDPI, 2020) [39].

De plus, l'absence de formats de données normalisés et le manque d'interopérabilité entre les différents systèmes de soins de santé se sont avérés très dissuasifs pour l'intégration d'applications d'intelligence artificielle dans les services de santé [40].

Les considérations éthiques sont également essentielles. La confidentialité des patients et la sécurité des données doivent être garanties, mais c'est très difficile à faire. Il y a également un niveau de considération éthique à prendre en compte en termes d'équité et de biais qui peuvent être construits dans les modèles d'IA, affectant ainsi certains groupes de patients [44].

Un autre obstacle important à l'application de l'IA dans les diagnostics médicaux est ses approbations réglementaires. Étant donné que le secteur de la santé est fortement réglementé, les systèmes d'IA doivent respecter des normes strictes avant que les systèmes soient approuvés pour une utilisation clinique. Cela peut retarder le déploiement dans des secteurs comme la santé, qui sont très réglementés [41].

3.6.2 Suggestions pour des recherches futures

Pour que l'IA travaille plus largement dans le diagnostic médical, nous avons besoin :

- **Explicabilité et transparence** : Nous avons besoin de modèles d'IA plus explicables afin que les fournisseurs de soins de santé et les patients puissent leur faire confiance. XAI (IA explicable) tente de rendre la prise de décision du modèle d'IA plus transparente [41] .
- **Améliorer les données et l'interopérabilité** : Nous devons améliorer la collecte et la gestion des données. Cela signifie des formats de données standardisés et une meilleure interopérabilité entre les différents systèmes de santé [39] .
- **Techniques de préservation de l'apprentissage fédéré et de la confidentialité** : La recherche en apprentissage fédéré peut permettre des modèles d'IA collaboratifs sans compromettre les données des patients. Les techniques de préservation de la confidentialité telles que la confidentialité différentielle peuvent garder les données des patients privées [41] .
- **Des cadres de validation robustes** : Des cadres de validation robustes et des normes réglementaires pour l'IA dans les soins de santé peuvent garantir que les

outils de diagnostic de l'IA sont sûrs et efficaces [40] .

- **Équipes interdisciplinaires** : Les experts en IA, les professionnels de la santé et les décideurs politiques qui travaillent ensemble peuvent stimuler l'innovation et résoudre les problèmes uniques de l'IA dans les soins de santé [43] .

Conclusion générale

Les résultats du graphique montrent une amélioration constante des performances par le modèle "flanT5-med-rad-reports" dans les étapes de la formation. C'est-à-dire qu'il y a eu des augmentations linéaires dans les scores UEBL 4, METEOR et ROUGE-L, indiquant ainsi de plus en plus de capacités à générer des rapports de rayonnement précis et cohérents.

La courbe d'apprentissage, Train / perte, Eval / perte : il a une rétention rapide dans la perte de formation et d'évaluation et se stabilise à un faible niveau, donc un apprentissage efficace sans haute précision. À partir des résultats, il est bien adapté pour une application pratique. Il produit ainsi des résumés diagnostiques fiables et de qualité, ce qui est essentiel pour améliorer la précision du diagnostic médical et la qualité des soins qui en résultent.

La mise en œuvre de modèles d'IA dans le secteur des soins de santé exige le respect des pratiques exemplaires pour assurer une intégration et une utilisation efficaces. Les principaux facteurs à prendre en considération sont les suivants :

- **Identification des problèmes spécifiques** : L'identification des problèmes spécifiques doit être effectuée en premier, ou défi que le modèle d'IA cherche à aborder dans les milieux de soins de santé. Des objectifs clairs doivent être fixés pour démontrer l'efficacité de la solution d'IA [1] .
- **Qualité et sécurité des données** : La qualité et la sécurité des données doivent être garanties. Cela comprend un nettoyage approfondi des données, un prétraitement et un stockage conforme aux réglementations [2] .
- **Participation des cliniciens et des intervenants** : La participation des cliniciens et des intervenants dès le début est importante afin que l'adhésion et l'adaptation du système d'IA coche toutes les cases pour les besoins en soins de santé [3]

- **Sélectionner la technologie d’IA appropriée** : Une sélection appropriée de la technologie d’IA, de l’apprentissage automatique, du traitement du langage naturel ou de la robotique doit être adoptée en fonction des exigences de l’organisation et du problème spécifique à résoudre [4] .
- **Validation et essais** : La validation et l’essai de base des algorithmes d’IA parmi de multiples ensembles de données représentatifs des populations cibles sont très importants. L’observation de protocoles de validation robustes assure l’exactitude et la généralisation [5] .

Éthique et explication IA : La considération éthique la transparence et l’explication des modèles d’IA sont importantes pour renforcer la confiance des professionnels de la santé et du grand public [6] .

L’avenir de l’IA dans le diagnostic des soins de santé regorge de possibilités et de nombreuses pistes de recherche et de développement. Conformément à l’IEEE, les futures orientations de recherche comprennent l’amélioration de la précision et de l’efficacité du diagnostic, la mise en place de plans de traitement personnalisés basés sur des données spécifiques aux patients, l’amélioration de la surveillance de la santé en temps réel, et le développement de systèmes d’IA capables d’interpréter l’imagerie médicale complexe avec une grande précision. De plus, MDPI souligne l’évolution vers la médecine personnalisée, en tirant parti de l’analyse prédictive basée sur l’IA pour révolutionner les soins aux patients en prédisant les résultats pour la santé et en identifiant des stratégies de traitement optimales. D’autres recommandations incluent l’intégration de l’IA dans les infrastructures techniques existantes, comme l’a souligné Springer, et l’amélioration de l’interprétabilité et de l’explicabilité des conclusions dérivées de l’IA pour instaurer la confiance parmi les professionnels de la santé. L’article de ScienceDirect souligne la nécessité d’améliorer les solutions d’IA pour l’imagerie médicale en combinant des procédures de diagnostic visuel avec des données cliniques, des risques liés au mode de vie et des données démographiques. De plus, l’intégration de l’IA avec d’autres technologies telles que l’IoT pour la surveillance en temps réel et l’analyse des données est une autre direction prometteuse [8] .

Bibliographie

a. Bibliographie :

[4] EL AGOURI, H. (2022). Évaluation d’algorithmes d’intelligence artificielle pour prédire le diagnostic histologique du cancer du sein : à propos d’une série de 328 images numérisées.

[5] Jean-Philippe Rennard. Réseaux neuronaux : une introduction accompagnée d’un modèle Java. Vuibert, 2006.

[6] R Sun, M Lerousseau, T Henry, A Carré, A Leroy, T Estienne, S Niyoteka, S Bockel, A Rouyar, É Alvarez Andres, et al. Intelligence artificielle en radiothérapie : radiomique, pathomique, et prédiction de la survie et de la réponse aux traitements. *Cancer/Radiothérapie*, 25(6-7) :630–637, 2021.

[7] LeCun, Y., Bengio, Y. and Hinton, G., “Deep learning,” *Nature* 521(7553), 436–444 (2015).

[9] <https://www.healthvisors.com/en/idx-dr/>

[10] Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ digital medicine*, 1(1), 39.

[11] Article scientifique : Yaniv, Z., Amit, H., Barak, A., Elad, B., Elad, S., Ohad, Z. (2018). Deep Learning Algorithm for Detection of Thoracic Pathologies in Chest Radiographs. *Clinical Imaging*, 52, 239-248.

[14] McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashraffian, H., ... Reicher, J. J. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89-94.

[15] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., Thrun,

S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.

[16] Abramoff, M. D., Lavin, P. T., Birch, M., Shah, N., Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *npj Digital Medicine*, 1(1), 39.

[17] Data, M. C., Mark, R. (2016). The story of MIMIC. *Secondary Analysis of Electronic Health Records*, 43-49.

[18] Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., ... Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1), 1-9.

[19] <https://people.cs.pitt.edu/jlee/note/Notes-NLP-Clinical-Notes-on-MIMIC-3-Database.html>

[20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

[22] Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., ... Wei, J. (2024). Scaling instruction-finetuned language models. *Journal of Machine Learning Research*, 25(70), 1-53.

[23] <https://exemplary.ai/blog/flan-t5>

[24] Alex Graves. Generating sequences with recurrent neural networks. *arXiv preprint arXiv :1308.0850*, 2013.

[25] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N. Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

[29] Kishore Papineni, Salim Roukos, Todd Ward, and WeiJing Zhu. 2002. Bleu : a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

[30] Yashaswini, S., Shylaja, S. S. (2021). Metrics for Automatic Evaluation of Text from NLP Models for Text to Scene Generation. *European Journal of Electrical Engineering and Computer Science*, 5(4), 20-25.

[31] Wang, Z., Zhou, S., Fried, D., Neubig, G. (2022). Execution-based evaluation for open-domain code generation. *arXiv preprint arXiv :2212.10481*

- [34] Kim, G., Kim, H., Ji, L., Bae, S., Kim, C., Sung, M., ... Kang, J. (2023). KU-DMIS-MSRA at RadSum23 : Pre-trained Vision-Language Model for Radiology Report Summarization. arXiv preprint arXiv :2307.07409.
- [35] Hu, J., Li, J., Chen, Z., Shen, Y., Song, Y., Wan, X., Chang, T. H. (2021). Word graph guided summarization for radiology findings. arXiv preprint arXiv :2112.09925.
- [36] Wang, T., Zhao, X., Rios, A. (2023, July). UTSA-NLP at RadSum23 : Multi-modal Retrieval-Based Chest X-Ray Report Summarization. In The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks (pp. 557-566).
- [37] Zhu, Y., Yang, X., Wu, Y., Zhang, W. (2023). Leveraging summary guidance on medical report summarization. IEEE Journal of Biomedical and Health Informatics.
- [38] Namikawa, K., Hirasawa, T., Nakano, K., Ikenoyama, Y., Ishioka, M., Shiroma, S., ... Tada, T. (2020). Artificial intelligence-based diagnostic system classifying gastric cancers and ulcers : comparison between the original and newly developed systems. *Endoscopy*, 52(12), 1077-1083.
- [39] Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., Fanos, V. (2020). Comparison of conventional statistical methods with machine learning in medicine : diagnosis, drug development, and treatment. *Medicina*, 56(9), 455.
- [40] Kononenko, I. (2001). Machine learning for medical diagnosis : history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1), 89-109.
- [41] Kumar, Y., Koul, A., Singla, R., Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis : a systematic literature review, synthesizing framework and future research agenda. *Journal of ambient intelligence and humanized computing*, 14(7), 8459-8486.
- [43] Peloso, A., Naesens, M., Thauinat, O. (2023). The Dawn of a New Era in Kidney Transplantation : Promises and Limitations of Artificial Intelligence for Precision Diagnostics. *Transplant International*, 36, 12010.
- [44] Panayides, A. S., Amini, A., Filipovic, N. D., Sharma, A., Tsaftaris, S. A., Young, A., ... Pattichis, C. S. (2020). AI in medical imaging informatics : current challenges and future directions. *IEEE journal of biomedical and health informatics*, 24(7), 1837-1857
- [51] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... Zettlemoyer, L. (2019). Bart : Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv :1910.13461.
- [53] Das, S., Samuel, V., Noroozizadeh, S. (2024). TLDR at SemEval-2024 Task 2 :

T5-generated clinical-Language summaries for DeBERTa Report Analysis. arXiv preprint arXiv :2404.09136.

[54] Yang, S., Cho, W. I. (2024). Evaluating Span Extraction in Generative Paradigm : A Reflection on Aspect-Based Sentiment Analysis. arXiv preprint arXiv :2404.11539.

[55] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv :1910.10683.[67] [56] Fine-tuning pre-trained transformer language models to distantly supervised relation extraction tasks. arXiv preprint arXiv :2006.06432[68]

[57] Hu, J., Li, J., Chen, Z., Shen, Y., Song, Y., Wan, X., Chang, T. H. (2021). Word graph guided summarization for radiology findings. arXiv preprint arXiv :2112.09925.

[61] Kim, G., Kim, H., Ji, L., Bae, S., Kim, C., Sung, M., ... Kang, J. (2023). KUDMIS-MSRA at RadSum23 : Pre-trained Vision-Language Model for Radiology Report Summarization. arXiv preprint arXiv :2307.07409.

[62] Hu, J., Li, J., Chen, Z., Shen, Y., Song, Y., Wan, X., Chang, T. H. (2021). Word graph guided summarization for radiology findings. arXiv preprint arXiv :2112.09925.

[63] Wang, T., Zhao, X., Rios, A. (2023, July). UTSA-NLP at RadSum23 : Multi-modal Retrieval-Based Chest X-Ray Report Summarization. In The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks (pp. 557-566).

[64] Zhu, Y., Yang, X., Wu, Y., Zhang, W. (2023). Leveraging summary guidance on medical report summarization. IEEE Journal of Biomedical and Health Informatics.

b. Webographie :

[1]<https://fr.linkedin.com/pulse/historique-et-contexte-de-lintelligence-artificielle-ia-kouame> ... consulter le 10/06/2024

[2] <https://aleia.com/solutions-metiers/cas-dusage/sante/> ... consulter le 2024 19 :00 13/06/2024

[3]<https://www.labtoo.com/fr/page/developpement-de-medicaments-drug-discovery-et-recherche-preclinique> ... consulter le 18/06/2024

[8]<https://www.ibm.com/docs/en/announcements/watson-oncology?region=CAN> ... consulter le 09/06/2024

[12] <https://www.cbinsights.com/company/zebra-medical-vision> ... consulter le 06/06/2024

[13] <https://www.butterflynetwork.com/> ... consulter le 11/06/2024

- [21] <https://swngui.medium.com/what-is-flan-t5-db87f077d857> ... consulter le 13/06/2024
- [26] <https://www.coursera.org/articles/kaggle> ... consulter le 14/06/2024
- [27] <https://www.python.org/doc/essays/blurb/> ... consulter le 02/06/2024
- [28] <https://www.nvidia.com/en-us/data-center/tesla-p100/> ... consulter le 07/06/2024
- [32] voir <https://medium.com/@kbdhunga/nlp-model-evaluation-understanding-bleu-rouge-meteor-and-bertscore-9bad7db71170>. ... consulter le 09/06/2024
- [47] https://huggingface.co/docs/transformers/en/model_doc/flan-t5...consulter le 15/06/2024
- [48] <https://stackoverflow.com/questions/75474086/how-to-train-flan-t5-to-summarization-task-with-a-custom-dataset-of-legal-docume> ... consulter le 10/06/2024
- [49] <https://accubits.com/large-language-models-leaderboard/flan-t5> ... consulter le 11/06/2024
- [50] <https://exemplary.ai/blog/flan-t5> ... consulter le 15/06/2024
- [52] <https://www.width.ai/post/bart-text-summarization> ... consulter le 08/06/2024
-