



HAL
open science

Utilisation de nouvelles approches en intelligence artificielle pour le diagnostic de lésions de la muqueuse orale

Antoine Dubuc

► **To cite this version:**

Antoine Dubuc. Utilisation de nouvelles approches en intelligence artificielle pour le diagnostic de lésions de la muqueuse orale. Médecine humaine et pathologie. Université de Toulouse, 2024. Français. NNT : 2024TLSES058 . tel-04685280

HAL Id: tel-04685280

<https://theses.hal.science/tel-04685280v1>

Submitted on 3 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse III - Paul Sabatier

Utilisation de nouvelles approches en intelligence artificielle
pour le diagnostic de lésions de la muqueuse orale

Thèse présentée et soutenue, le 23 mai 2024 par

Antoine DUBUC

École doctorale

BSB - Biologie, Santé, Biotechnologies

Spécialité

ÉPIDÉMIOLOGIE

Unité de recherche

CERPOP - Centre d'Epidémiologie et de Recherche en santé des POPulations de Toulouse

Thèse dirigée par

Sara LAURENCIN-DALICIEUX et Sylvain CUSSAT-BLANC

Composition du jury

Mme Loredana RADOI, Rapporteur, AP-HP de Paris - Université Paris Cité

M. Adrian BASARAB, Rapporteur, Université de Lyon

M. Julien ALIGON, Examineur, Université Capitole I

Mme Anne-Gaëlle CHAUX, Examinatrice, Université de Nantes

Mme Sara LAURENCIN-DALICIEUX, Directrice de thèse, CHU de Toulouse - Université Toulouse III
Paul Sabatier

M. Sylvain CUSSAT-BLANC, Co-directeur de thèse, Université Toulouse Capitole

Doctorat de l'Université de Toulouse

préparé à l'Université Toulouse III - Paul Sabatier

Utilisation de nouvelles approches en intelligence artificielle
pour le diagnostic de lésions de la muqueuse orale

Thèse présentée et soutenue, le 23 mai 2024 par

Antoine DUBUC

École doctorale

BSB - Biologie, Santé, Biotechnologies

Spécialité

ÉPIDÉMIOLOGIE

Unité de recherche

CERPOP - Centre d'Epidémiologie et de Recherche en santé des POPulations de Toulouse

Thèse dirigée par

Sara LAURENCIN-DALICIEUX et Sylvain CUSSAT-BLANC

Composition du jury

Mme Loredana RADOI, Rapporteur, AP-HP de Paris - Université Paris Cité

M. Adrian BASARAB, Rapporteur, Université de Lyon

M. Julien ALIGON, Examineur, Université Capitole I

Mme Anne-Gaëlle CHAUX, Examinatrice, Université de Nantes

Mme Sara LAURENCIN-DALICIEUX, Directrice de thèse, CHU de Toulouse - Université Toulouse III
Paul Sabatier

M. Sylvain CUSSAT-BLANC, Co-directeur de thèse, Université Toulouse Capitole



**UMR 1295 Centre d'Épidémiologie et de Recherche en santé des
POPulations**



Service de Médecine Bucco-dentaire– Chirurgie Orale du CHU de Toulouse

Table des matières

Liste des abréviations	6
Introduction	7
Première partie : L'intelligence artificielle au service du diagnostic en dermatologie de la muqueuse orale.....	12
I- Préambule : Apport de l'intelligence artificielle en santé et dermatologie.....	12
II- <i>Deep Learning</i> , réseau de neurones et aide au diagnostic.....	22
A- Objectif	22
B- Evaluation du <i>Deep Learning</i>	23
1- Matériel et méthode.....	23
2- Résultats	35
3- Discussion	40
III- <i>Machine Learning</i> et apprentissage statistique, une solution alternative ⁸⁸	44
A- Objectif	44
B- Evaluation de l'apport du <i>Machine Learning</i>	45
1- Matériel et méthode.....	45
2- Résultats	49
3- Discussion	57
4- Conclusion.....	60
Deuxième partie : Développement d'un outil basé sur l'intelligence artificielle (IA) destiné aux professionnels de santé, ORAMEDAI.	61

I- Introduction	61
II- Acceptations de l'Intelligence Artificielle et Défis relatifs.....	63
III- Développement de l'outil	68
A- ORAMEDAI – Interface WEB	70
B- ORAMD.AI – Développement de l'application mobile	73
C- ORAMED.AI – Utilisation du serveur	76
D- ORAMED.AI – Développement du DL	77
IV- Conclusion	80
Discussion et Perspectives	82
Discussion	82
Perspectives	87
Conclusion.....	92
Références	95
Annexe	111

Liste des abréviations

AWS, *Amazon Web Services*

CNN, *Convolutional Neural Networks* (Réseau de neurones convolutifs)

DL, *Deep Learning* (Apprentissage profond)

DT, *Decision Tree* (Arbre décisionnel)

IA, *Intelligence Artificielle*

KNN, *K-Nearest Neighbors* (K plus proches voisin)

LightGBM, *Tree based Gradient Boosting Model* (méthode de gradient boosting basée sur les arbres)

LR, *Logistic Regression* (Regression Logistique)

ML, *Machine Learning* (Apprentissage automatique)

MLP, *Multilayer perceptron*

SHAP, *SHapley Additive exPlanations*

Introduction

La médecine dentaire joue un rôle crucial dans la santé globale des individus, en assurant le diagnostic et le traitement des pathologies orales. La santé bucco-dentaire fait partie de la santé générale et est reconnue comme une composante essentielle de la qualité de vie. Bien que la plupart des problèmes bucco-dentaires ne représentent pas un risque immédiat de décès, ils sont responsables de la diminution de la qualité de vie des individus. Ils peuvent également être l'expression d'une pathologie systémique, le témoin de son aggravation, et peuvent diminuer la réponse à un traitement¹. Plus localement, ils prolongent les états de douleur, de souffrance et provoquent des problèmes fonctionnels, esthétiques, nutritionnels et psychologiques².

Ces affections orales incluent, entre autres, les affections des tissus dentaires, parodontaux, osseux alvéolaires et basaux ainsi que les affections du revêtement muqueux. Parmi elles, les affections de la muqueuse orale constituent une entité clinique difficile à enseigner, appréhender et dont les spécialistes sont aujourd'hui peu nombreux. En France, les affections de la muqueuse orale pourraient concerner jusqu'à 61% de la population selon, Massereau et al³. Les cancers de la cavité orale sont classés 16ème en terme d'incidence et sont la 15^{ème} cause de mortalité par cancer dans le monde en 2022⁴. En 2022, l'incidence (/100000 sujets) des cancers oraux rapportée à l'âge est en France de 8,3 chez les hommes et 3,6 chez les femmes, ce qui place la France parmi les pays avec la plus forte incidence en Europe et dans le monde⁴. Cette incidence augmente avec l'âge⁵. Ceci est d'une importance capitale d'un point de vue de santé publique compte tenu du vieillissement de la population. Le carcinome épidermoïde représente le cancer le plus fréquent de la cavité orale et constitue 80 à 90 % des cancers de la tête et du cou, avec un taux de survie de 50 % à 5 ans⁶. En 2018, les cancers de la bouche et principalement les carcinomes épidermoïdes représentaient plus de la moitié des décès liés à un cancer de la tête et du cou⁷.

La connaissance des lésions de la muqueuse orale par les chirurgiens-dentistes et autres spécialistes de la sphère orale est pertinente pour diagnostiquer des lésions relativement anodines, des lésions précancéreuses, mais aussi des cancers oraux à un stade précoce. En cas de pathologie maligne, la détection précoce réduit la mortalité et la morbidité⁸. Les lésions de la muqueuse orale peuvent être malignes mais incluent surtout de très nombreuses lésions bénignes. Ces diverses maladies peuvent provoquer des symptômes tels qu'une sensation de brûlure, une tuméfaction, une irritation ou une douleur, mais souvent sont asymptomatiques^{9,10}. Dans le cas du carcinome épidermoïde, les symptômes sont souvent limités ou absents à un stade précoce de la maladie, ce qui amène les patients à consulter tardivement et à un stade avancé¹¹.

Le diagnostic des pathologies de la muqueuse orale peut parfois présenter des défis, et il est essentiel de mettre en place des méthodes de diagnostic précises et fiables. En effet, les pathologies de la muqueuse orale peuvent présenter une grande variété de manifestations et de caractéristiques cliniques ce qui complexifie leur diagnostic. Certaines lésions peuvent être évidentes et faciles à diagnostiquer, tandis que d'autres peuvent être plus subtiles et nécessiter des compétences spécialisées pour une évaluation précise et une prise en charge adéquate. Etant donné la complexité des pathologies de la muqueuse orale, il est crucial que les professionnels de la santé bucco-dentaire se tiennent au courant des dernières avancées diagnostiques et des recommandations de prise en charge. La formation continue et la participation à des programmes de perfectionnement professionnel sont essentielles pour améliorer les compétences diagnostiques^{12,13}.

Cependant, rares sont les études sur la perception et les connaissances des chirurgiens-dentistes concernant le diagnostic et la prise en charge des lésions précancéreuses et cancéreuses^{14,15}. Sont encore plus rares celles concernant les affections muqueuses de façon plus large^{16,17}. Une revue systématique récente sur le retard de diagnostic du cancer de la cavité orale souligne le manque de capacité à diagnostiquer correctement les lésions de la muqueuse orale par les professionnels de santé¹⁸. En 2019, Roume & al¹⁹ ont fait un état des lieux des connaissances en matière de pathologie de la muqueuse orale auprès des chirurgiens-dentistes. Ils concluent non seulement à un manque de formation mais également un manque d'expérience des praticiens.

Face à ces constats mais aussi face à la nécessité d'une prise en charge plus adaptée et précoce des patients, se pose l'importance de développer des outils d'aide au diagnostic destinés aux professionnels de santé. Quel rôle peut donc jouer l'intelligence artificielle (IA) dans ce domaine ?

L'intelligence artificielle (IA) fait partie des sciences informatiques et est apparue dans les années 1950²⁰. L'intérêt de l'IA en médecine et tout particulièrement en dermatologie connaît un essor majeur depuis quelques années^{21,22}. En effet de nombreuses applications telles que les classifications, la segmentation ou bien encore l'extraction de caractéristiques de lésions dermatologiques ont su démontrer leur intérêt²³. Ces applications sont toutefois aujourd'hui limitées à quelques lésions dermatologiques telles que les cancers cutanés (tout particulièrement le mélanome²⁴⁻²⁶, les ulcérations, le psoriasis ou les dermatites atopiques²⁷. Il a été démontré, ces dernières années, que la précision diagnostique de mélanomes reposant sur des algorithmes d'intelligence artificielle et notamment des réseaux neuronaux convolutifs était similaire²⁸ voire même meilleure^{29,30} que celle de dermatologues experts dans le domaine.

Le *Machine Learning* (ML) est défini comme étant une branche de l'IA. Les algorithmes de ML apprennent de façon autonome à effectuer des prédictions à partir de nombreuses données et sont capables d'améliorer leurs performances dans le temps²⁷. Les modèles de ML sont particulièrement bien adaptés pour mettre en évidence l'interaction entre différentes variables dans des modèles complexes. Ils présentent donc un intérêt majeur pour accompagner le praticien dans sa démarche diagnostique et le guider vers le traitement le plus approprié. De plus, l'explicabilité des interactions au sein du modèle est aujourd'hui facilement accessible et constitue un élément pédagogique intéressant pour les praticiens non-spécialistes.

Alors que les publications sur les lésions cutanées sont en nombre croissant (près de 300 à ce jour entre 1994 et 2021), l'apport de l'IA dans le domaine de la pathologie de la muqueuse orale reste rare ou anecdotique. Le diagnostic et la prise en charge des lésions de la muqueuse orale constituent un enjeu majeur, notamment dans le cadre de la prise en charge précoce des cancers oraux. Depuis 2019, certaines équipes s'intéressent à l'apport des algorithmes d'IA et notamment du ML dans la prise en charge des cancers oro-pharyngés principalement en termes de

prédiction du pronostic ³¹⁻³³, de classification d'images médicales ³⁴, mais aussi, pour optimiser le traitement de ces cancers ³⁵. Cependant, le manque de praticiens spécialisés en pathologie de la muqueuse orale ainsi que la variété des lésions pouvant être rencontrées font du diagnostic différentiel un objectif majeur afin d'améliorer le diagnostic et la prise en charge. En effet, en dermatologie cutanée, les modèles de ML ont le potentiel de transformer le dépistage du cancer de la peau en augmentant le nombre de patients dépistés ainsi que le nombre d'orientations vers les praticiens spécialistes. Les applications mobiles pourraient également être utilisées par les dermatologues pour l'aide à la décision clinique. Ces données pourraient aider les dermatologues à élargir ou à réduire la liste des diagnostics différentiels pour les patients présentant une pathologie particulière ^{36,37}.

L'objectif de ce travail est de contribuer à l'amélioration du diagnostic des pathologies de la muqueuse orale à l'aide de l'intelligence artificielle. Afin de répondre à cet objectif, nous avons organisé cette thèse en deux parties :

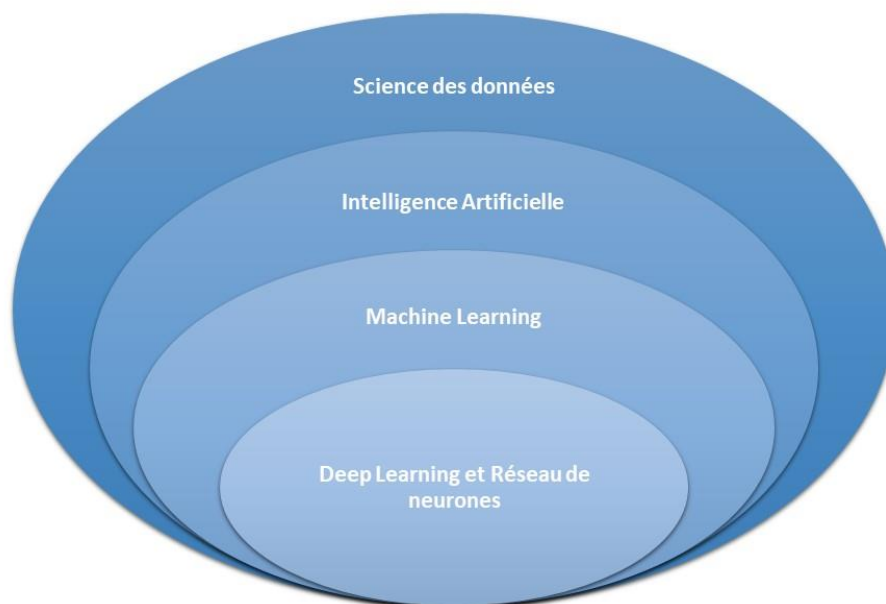
- 1- Dans un premier temps, nous avons étudié les capacités de prédictions de différents modèles de DL ou ML (*statistical learning*) dans le domaine de dermatologie orale.**
- 2- Dans un second temps, et une fois passée cette étape fondamentale d'évaluation d'algorithme, nous avons mis au point un outil d'orientation diagnostique destiné aux professionnels de santé.**

*Première partie : L'intelligence artificielle au service du diagnostic
en dermatologie de la muqueuse orale*

I- Préambule : Apport de l'intelligence artificielle en santé et dermatologie.

L'intelligence artificielle (IA) correspond au développement de systèmes informatiques capables d'effectuer des tâches qui requièrent normalement l'intelligence humaine telles que la perception visuelle, la reconnaissance vocale ou la prise de décision et la traduction entre langues. L'IA est devenue un élément indispensable de notre vie quotidienne, alors qu'elle pénètre de plus en plus d'activités humaines ³⁸

Figure I-0 : *Place des techniques d'IA dans la science des données.*



Il est possible de distinguer deux grandes entités de l'intelligence artificielle que sont le *Machine Learning* (ML) essentiellement représenté par des algorithmes statistiques et son évolution le *Deep Learning* (DL). Ce flux d'apprentissage profond a donné à de nombreux domaines médicaux, le potentiel d'évoluer grâce à l'IA³⁹. Actuellement, des avancées majeures ont été réalisées principalement dans les domaines de la radiologie et de la cardiologie. Les dispositifs médicaux validés par la *Food and Drug Administration* (FDA) ont été approuvés pour la première fois en 2016, donnant aux professionnels de la santé la possibilité d'améliorer la pratique médicale grâce à l'IA⁴⁰

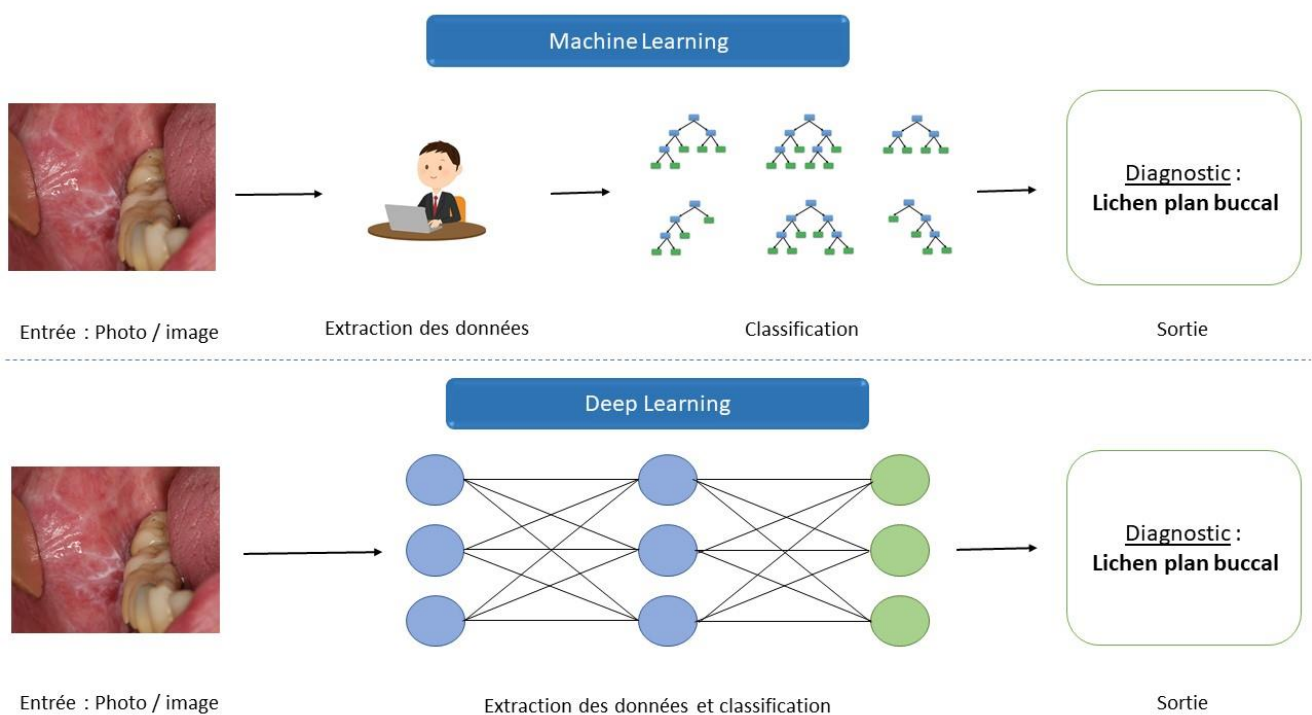


Figure I-I : Différence entre *Machine Learning* et *Deep Learning*

Le ML est un domaine qui se concentre sur l'aspect apprentissage de l'IA en développant des algorithmes qui représentent au mieux un ensemble de données. Contrairement à la programmation classique, dans laquelle un algorithme peut être explicitement codé à l'aide de caractéristiques connues, le ML intègre des sous-ensembles de données pour générer un « modèle » utilisant des combinaisons de caractéristiques et de poids différents⁴¹. En ML, il existe quatre méthodes d'apprentissage couramment utilisées, chacune utile pour résoudre des tâches différentes : l'apprentissage supervisé, non supervisé, semi-supervisé et l'apprentissage par renforcement⁴². L'apprentissage supervisé utilise des caractéristiques de l'ensemble des données d'apprentissage afin de les faire correspondre aux caractéristiques cibles et ainsi prédire dans les futurs ensembles de données. Les tâches d'apprentissage supervisé les plus courantes sont la régression et la classification. Contrairement à l'apprentissage supervisé, l'apprentissage non supervisé vise à détecter des tendances dans un ensemble de données et à classer les instances individuelles en catégories. Ces algorithmes sont non supervisés parce que les tendances qui peuvent ou non exister dans un ensemble de données sont laissées à l'appréciation de l'algorithme. Les tâches d'apprentissage non supervisé les plus courantes sont le regroupement, l'association, la détection d'anomalies et la réduction de dimensionalité. L'apprentissage semi-supervisé est particulièrement utile pour les ensembles de données qui contiennent à la fois des données labellisées et des données non labellisées. L'apprentissage semi-supervisé est souvent utilisé pour les images médicales, où un médecin peut labelliser un sous-ensemble d'images et les utiliser pour former un modèle. Enfin, l'apprentissage par renforcement est la technique qui consiste à former un algorithme à une tâche spécifique pour laquelle aucune réponse n'est correcte, mais un résultat global est souhaité. Il s'agit sans doute de la tentative la plus proche de modélisation de l'expérience d'apprentissage humaine. Le modèle apprend également à partir d'essais et d'erreurs plutôt qu'à partir de données

uniquement. Bien que l'apprentissage par renforcement soit une technique puissante, ses applications en médecine sont actuellement limitées.

En général, lors de la formation d'un modèle de ML, un sous-ensemble de données est "retenu" et est utilisé par la suite pour tester la précision du modèle formé. La précision du modèle est évaluée sur cet ensemble de données de test en fonction de sa capacité à faire correspondre correctement une image à son étiquette (par exemple une lésion maligne ou une lésion bénigne). Dans tout système de classification, il y aura un compromis entre la sensibilité et la spécificité. Par exemple, un système d'IA peut produire un score de probabilité de cancer compris entre 0 et 1, ce qui obligerait l'opérateur à fixer un seuil pour la limite de décision. Si le seuil est bas, une plus grande proportion de cancers sera détectée (sensibilité élevée), mais il y a un risque de classer des lésions bénignes comme cancer (faible spécificité)⁴³.

Les professionnels de santé peuvent utiliser des algorithmes de ML pour gérer et analyser les immenses données présentes dans les dossiers médicaux des patients. Ceci permet d'en extraire des informations significatives, destinées à améliorer la santé et le bien-être de la population ⁴⁴. La médecine diagnostique intègre des technologies d'IA avancées pour prédire ou diagnostiquer les maladies et faciliter le processus de prise de décision clinique ⁴⁵⁻⁴⁷. Plusieurs études ont également montré que les systèmes de soins basés sur l'IA ont un grand potentiel pour améliorer les soins donnés aux patients, stimuler la recherche innovante et faciliter le partage d'informations ⁴⁸⁻⁵⁰. Il est tout particulièrement nécessaire de réaliser le prétraitement et le nettoyage des données présentes dans les dossiers médicaux électroniques. En effet, certaines caractéristiques ou contenus peuvent entraver l'analyse (par exemple, les erreurs grammaticales/orthographiques)⁵¹. Ce traitement des données est d'autant plus intéressant que certains auteurs ont montré que l'application du *text mining* (traitement automatique du langage

afin d'explorer des documents) aux dossiers médicaux peut accélérer l'extraction d'informations et améliorer les résultats en matière de santé⁵².

En dermatologie et en médecine en général, l'abondance des données contenues dans les dossiers cliniques, les informations démographiques des patients, les résultats des examens d'imagerie et les données recueillies à l'aide de questionnaires représentent une mine d'informations susceptibles de révolutionner la médecine personnalisée⁵³. L'apprentissage supervisé est le type d'apprentissage le plus couramment utilisé en dermatologie.

Parallèlement à ces algorithmes statistiques, on retrouve les réseaux de neurones. Ils font passer les données d'entrée par une série de nœuds interconnectés (analogues aux neurones biologiques). Chaque nœud réalise des opérations mathématiques très simples (addition, multiplication, etc.) permettant d'agrèger des signaux d'entrée pour produire un signal de sortie transmis à un autre neurone. Un groupe de nœuds interconnectés au sein du réseau est appelé "couche" d'un réseau. Les différentes couches, ensemble, constituent l'"architecture". Au cours de l'apprentissage, chaque nœud est ajusté et optimisé par un processus itératif appelé "rétropropagation", ce qui permet au réseau neuronal d'améliorer sa précision de classification par exemple.⁵⁴

Les réseaux neuronaux comportant plus de deux "couches cachées" sont appelés réseaux neuronaux "profonds" et effectuent un "apprentissage profond". Bien que le concept des réseaux neuronaux profonds ait été décrit il y a plusieurs dizaines d'années, il convient de noter que les réseaux neuronaux convolutifs (CNN) sont une forme spécifique d'architecture d'apprentissage profond qui s'est avérée efficace pour la classification des données issues d'image. Au cours des dernières années, l'utilisation des CNN dans les tâches de classification a explosé en raison d'une efficacité et d'une disponibilité constamment supérieures⁴³. De nouvelles architectures de CNN ont été développées, améliorées et mises à la disposition du public par des institutions

disposant d'un haut niveau d'expertise et de ressources informatiques. Une approche courante consiste à prendre une architecture de réseau de reconnaissance d'images pré-entraînée, et à spécialiser son application à un nouveau type d'images. Ce processus est appelé « *transfer learning* » (apprentissage par transfert)⁵⁵⁻⁵⁷.

La classification de lésions peut être réalisée à l'aide d'un CNN unique ou bien via l'association de différents CNN⁵⁸. Plusieurs architectures (ResNet, Inception, VGG, EfficientNet) pré-entraînées ont été étudiées (ImageNet, CIFAR, MNIST etc).^{59,60}. D'autre part l'utilisation d'une combinaison de réseaux a montré une meilleure prédiction en comparaison à un réseau utilisé seul^{61,62}.

La clé du développement d'un modèle d'apprentissage profond performant réside dans les données. Si les données sont de haute qualité, il est plus probable que les modèles apprennent à générer des prédictions plus précises. Plusieurs ensembles de données d'images de la peau accessibles au public sont fournis pour inciter les communautés de la dermatologie et de l'apprentissage profond à développer de nouveaux algorithmes ou à perfectionner ceux existants⁶³. Par exemple, les archives de l'International Skin Imaging Collaboration (ISIC) sont l'un des ensembles de données publiques d'images de cancer de la peau les plus connus. Les archives contiennent plus de 13 000 images dermoscopiques⁶⁴. Outre ces ensembles de données publiques, de nombreuses institutions cliniques ont collecté leurs propres ensembles de données pour des maladies telles que le psoriasis, la rosacée et les lésions labiales⁵⁸.

À l'heure de la quatrième révolution industrielle, les technologies et dispositifs numériques avancés sont largement utilisés pour l'innovation et la création de valeurs dans tous les secteurs d'activité. Le secteur des soins de santé ne fait pas exception. Les hôpitaux et les prestataires de soins du monde entier, en particulier dans les économies développées, déploient des technologies numériques, dont l'IA, pour améliorer la qualité des soins et l'efficacité opérationnelle⁶⁵. L'IA a été accueillie favorablement par les structures de santé, qui s'efforcent d'atteindre le "quadruple objectif" consistant à améliorer la santé et le bien-être de leurs patients, l'accès aux soins, la rentabilité et la vie des professionnels de santé⁶⁶.

L'approche du diagnostic et de la prise en charge des affections dermatologiques a considérablement changé avec l'avènement des nouvelles technologies. Les algorithmes informatiques ont trouvé leur application pour aider les dermatologues à diagnostiquer les maladies, en particulier les mélanomes malins. La dermatologie a pris la tête de la mise en œuvre de l'IA dans le domaine médical en raison de son importante base de données d'images cliniques, dermoscopiques et dermatopathologiques.^{67 68} En voici quelques exemples :

Classification :

Lésion maligne versus lésion bénigne :

Le principe de base de cette application est que les images dermoscopiques ou non dermoscopiques des lésions peuvent être décomposées. Ces applications sont généralement validées en comparant leur capacité à diagnostiquer correctement les lésions à celle d'un dermatologue certifié^{28,30}. Bien que la plupart des recherches sur l'IA soient basées sur des algorithmes de photo-reconnaissance, les valeurs numériques peuvent également être utilisées. Depuis quelques années, des applications pour smartphones sont disponibles et facilement accessibles pour le diagnostic du mélanome. Elles consistent en une évaluation du risque de

mélanome. Cependant la sensibilité et la spécificité varient respectivement de 7 à 73% et 37 à 94%⁶⁹.

Notons également que des réseaux de neurones convolutifs (CNN) sont également utilisés pour classifier les coupes histologiques et améliorer le diagnostic de mélanome⁷⁰.

L'évaluation des ulcération et escarres :

Les applications sont capables de mesurer les limites précises de la plaie et de différencier les différents types de tissus impliqués⁷¹. Des applications de prédiction du risque ont également été décrites avec une spécificité >80%.

Psoriasis et autres maladies inflammatoires de la peau :

Dans le domaine du psoriasis, l'IA se concentre principalement sur l'amélioration des méthodes de classification du psoriasis à l'aide de la reconnaissance d'images. Emam et al. ont démontré la capacité des systèmes à prédire la réponse biologique chez les patients atteints de psoriasis, en utilisant des paramètres évalués lors de la première visite, et ce afin d'optimiser le traitement⁷². L'IA peut également jouer un rôle diagnostique dans de nombreuses autres affections dermatologiques (l'acné, le lichen plan, le pityriasis lichénoïde, la dermatomyosite, et la dermatite séborrhéique.)

Télédermatologie

C'est l'un des domaines de la dermatologie qui se développe le plus rapidement, en particulier dans les régions où l'accès à un dermatologue est difficile. L'IA peut être intégrée dans des applications pour smartphones afin de photographier des lésions cutanées, recueillir des informations cliniques pertinentes et générer des rapports d'évaluation⁷³.

Enseignement et recherche

L'IA est une excellente ressource éducative pour la formation des dermatologues, des chercheurs, des étudiants et des spécialistes de la peau. Après avoir examiné des images visuellement similaires les praticiens sont en mesure de mieux comprendre leurs cas, ce qui les aide à poser un diagnostic plus précis⁷⁴.

Les applications concernant la muqueuse orale sont bien moins représentées. Toutefois l'engouement pour l'IA présent depuis quelques années touche aussi la médecine orale⁷⁵. Les études concernent essentiellement les lésions pré-cancéreuses et cancéreuses⁷⁶⁻⁸⁰. Fu & al⁷⁷ en 2020 ont développé un algorithme de détection des carcinomes épidermoïdes de la bouche avec une sensibilité 91% et une spécificité 93.5% à l'aide de 6176 photos. Des auteurs ont également mis en avant l'utilisation du DL pour segmenter des ulcérations aphtoïdes et des lésions herpétiques à l'aide d'un réseau U-Net⁸¹. Parmi les lésions bénignes on retrouve aussi la détection de leucoplasies^{82,83}. Jurczynszyn et al⁸³ ont utilisé un échantillon de 63 patients, divisés en leucoplasie, lichen plan et sains. L'étude a obtenu les meilleurs résultats en comparant les lésions *versus* la muqueuse saine. Toutefois le résultat est peu probant dans la différenciation entre leucoplasie et lichen plan.

Cependant, l'IA dans le domaine de la dermatologie présente des défis qu'il convient également de relever.⁶⁸ Actuellement, les données d'images des différentes maladies de la peau sont insuffisantes, le degré de partage des informations entre les sources est faible et la qualité des images de lésions cutanées n'est pas uniforme.

L'IA en dermatologie ne peut reconnaître que quelques maladies spécifiques, alors que le domaine dermatologique est extrêmement varié. Faire en sorte que l'IA identifie et classifie une longue et vaste liste de troubles dermatologiques aux présentations cliniques variées sera un véritable défi.

Le diagnostic par l'IA soulève également des questions juridiques, éthiques et de confidentialité des données, qui doivent être résolues ou tout du moins prises en compte.

Les études concernant la muqueuse orale étant presque inexistantes nous avons approché ces pathologies spécifiques avec une vision issue de la dermatologie cutanée et ceci tout au long de la thèse.

En effet, nous avons appliqué dans un premier temps des algorithmes de *Deep Learning* avant d'envisager d'autres approches plus statistiques.

II- *Deep Learning*, réseau de neurones et aide au diagnostic

A- Objectif

Du fait de l'utilisation prédominante du *Deep Learning* (DL) en dermatologie cutanée, nous avons fait le choix de l'appliquer à la muqueuse orale. Pour autant aucune base de données nationale voir même internationale n'étant disponible⁷⁵, nous avons constitué notre propre base de données. Cette dernière était issue de la consultation spécialisée de muqueuse orale du CHU de Toulouse. Cette base de données nous a donc servi à implémenter les différents modèles testés.

L'objectif de cette première phase fut de mettre au point un algorithme de DL capable d'apporter un diagnostic directement à partir des photographies. Pour ceci nous avons considéré les 5 lésions orales les plus rencontrées dans notre consultation de pathologie de la muqueuse orale au CHU de Toulouse. (Figure II-1)

Figure II-1 : Illustration des 5 différentes situations cliniques à prédire



A : Accroissement gingival, B : Maladie bulleuse (pemphigus), C : Ulcération aphte, D : Carcinome, E : Lichen plan buccal.

B- Evaluation du *Deep Learning*

1- Matériel et méthode

1-1 Critères de sélection

Les patients qui ont consulté à la consultation spécialisée de dermatologie orale du CHU de Toulouse entre janvier 2015 et décembre 2020 ont été pris en compte dans cette étude.

Cette étude a été réalisée dans le respect du code de la santé publique et des lois de bioéthique qui régissent la recherche en France⁸⁴. Cette étude est rétrospective et n'implique pas la personne humaine. Elle n'est donc pas soumise à un comité d'éthique mais respecte la méthodologie de référence (MR-004 de la Commission Nationale de l'Informatique et des Libertés (CNIL) (numéro : 2206723 v0)) pour laquelle le CHU de Toulouse a signé un engagement de conformité. La non-opposition du patient a été recueillie lors de la première consultation dans le service.

Tous les patients répondant aux critères suivants ont été inclus dans l'étude : (1) Patients consultant à la consultation de pathologie des muqueuses orales du CHU de Toulouse et présentant une lésion muqueuse, (2) Patients déjà suivis pour de telles lésions et nécessitant une réévaluation histologique, (3) Sujets âgés de plus de 18 ans au moment de la consultation, (4) Sujets capables de comprendre la langue française, de recevoir l'information sur l'étude et de comprendre le formulaire d'information pour participer à l'étude, (5) Sujets ne s'étant pas opposé à la réutilisation des données.

Les patients présentant les conditions suivantes n'ont pas pu être inclus dans cette étude : (1) Femmes enceintes et allaitantes, (2) Patients sous tutelle, curatelle ou sauvegarde de justice.

Tous les patients ont donné leur consentement pour la collecte des données et le traitement

informatique des données personnelles et médicales a été mis en œuvre pour les résultats de cette recherche.

1-2 Collecte et annotation des données

Pour chaque sujet, une série de photographies macroscopiques de haute résolution de la lésion (soit par appareil photo numérique, soit par smartphone) a été réalisée. Lorsque pour un même patient, plusieurs images de la même lésion ont été enregistrées, une seule d'entre elles a été prise en compte. Par ailleurs si les images du même patient concernent des lésions différentes et/ou une représentation différente d'une même lésion (par exemple, variation de la topographie ou la couleur) les images ont été traitées séparément. **Le Tableau II-1** détailles les caractéristiques de la base de données.

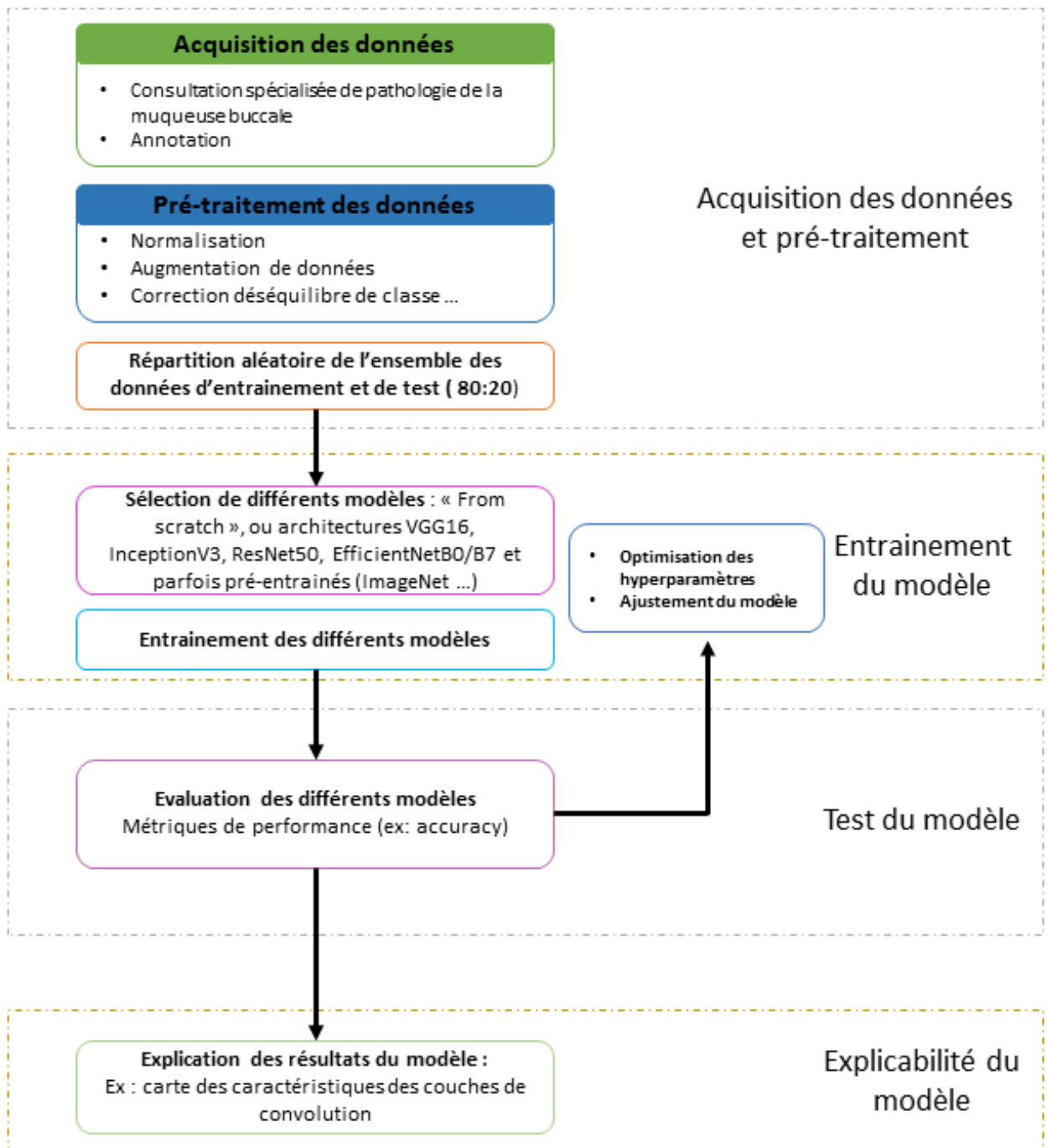
Diagnostics	Lichen plan buccal (n=77)	Leucokératose (n=41)	Ulcération aphtoïde (n=37)	Maladie Bulleuse (n=26)	Accroissement gingival (n=12)	Carcinome (n=11)	P-value^a
Genre:							
Homme	22 (28%)	20 (49%)	19 (51%)	9 (35%)	5 (42%)	5 (45%)	0.03
Femme	51 (66%)	16 (39%)	18 (49%)	15 (58%)	3 (25%)	4 (36%)	
Age moyen, (années)	65	65	50	65	43	54	<0.001
Alcool (n, %)	1 (1%)	3 (7%)	1 (3%)	-	-	2 (18%)	0.007
Tabac (n, %)	21 (27%)	18 (44%)	8 (22%)	3 (12%)	1 (8%)	5 (45%)	0.002
Localisation : Vermilion (n, %)	-	1 (2%)	-	1 (4%)	-	-	0.002
Commissure (n, %)		-	-	1 (4%)	-	-	
Muqueuse labiale (n, %)	-	2 (5%)	1 (3%)	1 (4%)	-	-	
Vestibule (n, %)	8 (10%)	3 (8%)	4 (10%)	2 (8%)	-	1 (9%)	
Gencive Attachée (n, %)	35 (45%)	5 (12%)	8 (22%)	14 (52%)	12 (100%)	4 (36%)	
Muqueuse juguale (n, %)	15 (19.5%)	9 (22%)	6 (16%)	2 (8%)	-	1 (9%)	
Palais Dur (n, %)	2 (3%)	4 (10%)	2 (5%)	1 (4%)	-	2 (18%)	

Palais mou (n, %)	-	2 (5%)	-	2 (8%)	-	-	
Pilier antérieur voile (n, %)	1 (1.5%)	1 (2%)	1 (3%)	1 (4%)	-	-	
Face dorsale de la langue (n, %)	9 (12%)	4 (10%)	5 (14%)	-	-	-	
Face ventrale de la langue (n, %)	1 (1.5%)	1 (2%)	3 (27%)	1 (4%)	-	1 (9%)	
Bord latérale de la langue (n, %)	5 (6%)	8 (20%)	5 (14%)	-	-	-	
Plancher buccal (n, %)	1 (1.5%)	1 (2%)	1 (3%)	-	-	2 (18%)	
Couleur de la lésion : Blanc (n, %)	31 (40%)	32 (78%)	10 (27%)	4 (16%)	-	1 (9%)	
Rouge (n, %)	8 (10%)	2 (15%)	4 (11%)	10 (38%)	2 (17%)	3 (27%)	
Mixte (n, %)	37 (49%)	7 (17%)	22 (60%)	12 (46%)	2 (17%)	7 (64%)	<0.001
Pigmenté (n, %)	1 (1%)	-	1(3%)	-	2 (17%)	-	
Sans changement de couleur (n, %)	-	-	-	-	6 (50%)	-	
Taille de la lésion : < 5mm (n, %)	9 (12%)	7 (17%)	10 (27%)	7 (27%)	2 (17%)	1 (9%)	
5 to 10 mm (n, %)	43 (56%)	13 (32%)	25 (68%)	18 (69%)	8 (66%)	2 (19%)	
10 to 50 mm (n, %)	25 (32%)	21 (51%)	2 (5%)	1 (4%)	2 (17%)	7 (63%)	<0.001
> 50 mm (n, %)	-	-	-	-	-	1 (9%)	

Lésion élémentaire : Hyperplasie gingivale (n, %)	-	-	-	-	9 (75%)	-	<0.001
Bulle (n, %)	-	-	-	4 (16%)	-	-	
Enduit cellulaire (n, %)	-	-	3 (8%)	2 (8%)	-	-	
Erosion (n, %)	3 (4%)	3 (8%)	1 (3%)	8 (30%)	-	1 (9%)	
Macule / Plaque (b, %)	30 (39%)	5 (12%)	5 (13%)	10 (38%)	3 (25%)	1 (9%)	
Nodule (n, %)	1 (1%)	-	1 (3%)	-	-	5 (46%)	
Papule / plaque (n, %)	37 (48%)	33 (80%)	3 (8%)	1 (4%)	-	2 (18%)	
Ulcération (n, %)	6 (8%)	-	24 (65%)	1 (4%)	-	2 (18%)	
Lésion homogène (n, %)	33 (43%)	24 (58%)	28 (76%)	10 (38%)	7 (58%)	1 (9%)	<0.001
Lésion bien délimitée (n, %)	33 (43%)	26 (63%)	31 (84%)	14 (54%)	6 (50%)	1 (9%)	<0.001
Bords surélevés de la lésion (n, %)	1 (1%)	-	4 (10%)	-	-	1 (9%)	0.29
Lésion réticulée (n, %)	34 (44%)	1 (2.5%)	-	1 (4%)	-	-	<0.001
Lésion bourgeonnante (n, %)	3 (4%)	-	-	-	-	7 (64%)	<0.001
Lésion papillomateuse (n, %)	-	2 (5%)	1(3%)	-	1 (8%)	1 (9%)	0.007
Nombre de dossiers cliniques correspondants (n)	1023	175	206	324	133	105	-

Figure II-2: Pipeline de Deep Learning

La bibliothèque Tensorflow[®] 2.5 sous Python[®] 3.7 (Python Software Foundation) a été utilisée comme cadre général pour le codage de l'algorithme et le traitement des données.



1-3 Développement du modèle et analyse des données

1-3-1 Étape de prétraitement :

La base de données précédemment décrite a été considérée pour chacun des modèles à tester. Toutefois, les images en doublon ont cette fois été conservées afin d'augmenter le nombre d'images. Ainsi l'entraînement et la validation du modèle ont été réalisés à l'aide de 2357 images. Les images ont été importées à l'aide de la fonction « *flow_from_dataframe* » de la bibliothèque Tensorflow[®]. Parallèlement à cette importation, les images ont toutes été redimensionnées à une taille de 256x256 pixels et normalisées par canaux (RGB). Chacune des photos importées a été associée à ses propres caractéristiques lésionnelles (couleur, taille, type de lésion élémentaire, forme ...). Une phase d'augmentation des données a été réalisée via des rotations, modifications de hauteurs, de largeur, de contraste, de la luminosité, de zoom ou bien encore en ajoutant des effets miroir. Enfin, la base de données obtenue a été partitionnée en deux groupes l'un d'entraînement et l'autre de validation selon un ratio de 80 :20.

1-3-2 Correction du déséquilibre de classe

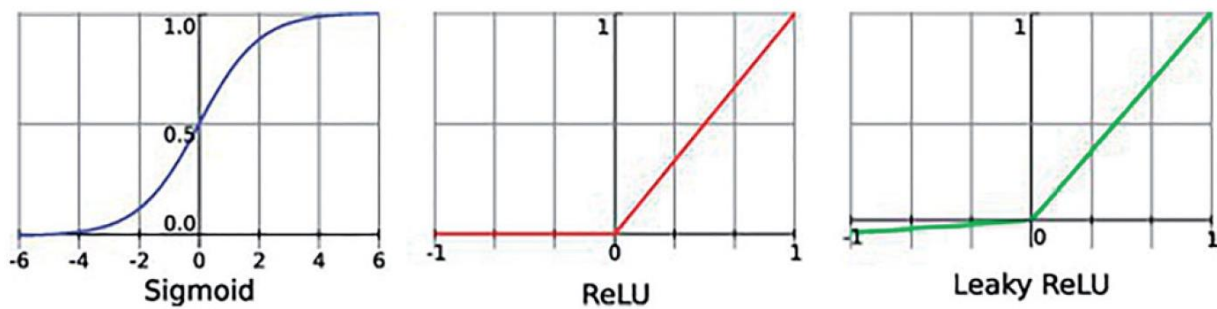
Compte tenu du déséquilibre existant entre les différentes classes à prédire, un calcul des poids relatifs à chacune des classes a été effectué puis imputé lors de l'entraînement du modèle.

1-3-3 Modélisation des données :

Différentes stratégies de *Deep Learning* ont été envisagées. Nous avons utilisé des réseaux neuronaux pré-entraînés types ResNet50, EfficientNet B0/B7, Inception V3. Nous avons également mis au point un réseau de neurones simplifié basé sur l'architecture VGG16. En effet, la profondeur du VGG16 était trop importante pour nos images et nous avons ainsi dû enlever certaines couches de convolution.

Notre architecture de réseau a utilisé un filtre de 3x3 pixels pour les couches de convolution et la fonction d'activation choisie était de type ReLU. Cette couche a comme caractéristique de ne s'activer qu'à partir de 0 et de transmettre au-delà les signaux de façon linéaire (**Figure II-3**).

Figure II-3 Illustration des différentes fonctions d'activation ⁸⁵



L'ajout d'une couche de « *pooling* » (c.-à-d. une couche de sous-échantillonnage) sous la couche de convolution permet de réduire la dimension de la caractéristique pour éviter le surentrainement et ainsi de condenser l'information. La couche « *maxpool* » (permet d'ajouter un filtre de taille définie sur l'image d'entrée afin de ne conserver que la valeur maximale pour chaque région de l'image couverte par le filtre.) est utilisée avec un filtre de 3x3 pixels. L'addition d'un « *dropout* » fixé à 0,5 (désactivation aléatoire et temporaire de certains neurones en définissant un taux de neurones à désactiver) concourt à diminuer le risque de sur-apprentissage.

Enfin, la structure du réseau est aplatie (« *flatten* », ce qui permet de transformer une information d'entrée en trois dimensions en une sortie en une seule dimension) et connectée à la couche dense de sortie, qui est suivie par une fonction sigmoïde (caractéristiques à prédire binaire) ou softmax (caractéristiques à prédire non binaire), afin d'effectuer la prédiction de l'étiquette de caractéristique lésionnelle voulue. Au total, le CNN comporte un total 33 928 147 paramètres (**Figure II-4**).

(A)

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 256, 256, 16)	448
conv2d_1 (Conv2D)	(None, 256, 256, 16)	2320
max_pooling2d (MaxPooling2D)	(None, 128, 128, 16)	0
dropout (Dropout)	(None, 128, 128, 16)	0
conv2d_2 (Conv2D)	(None, 128, 128, 32)	4640
conv2d_3 (Conv2D)	(None, 128, 128, 32)	9248
max_pooling2d_1 (MaxPooling2D)	(None, 64, 64, 32)	0
dropout_1 (Dropout)	(None, 64, 64, 32)	0
conv2d_4 (Conv2D)	(None, 64, 64, 64)	18496
conv2d_5 (Conv2D)	(None, 64, 64, 64)	36928
conv2d_6 (Conv2D)	(None, 64, 64, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 32, 32, 64)	0
dropout_2 (Dropout)	(None, 32, 32, 64)	0
flatten (Flatten)	(None, 65536)	0
dense (Dense)	(None, 512)	33554944
dense_1 (Dense)	(None, 512)	262656
dense_2 (Dense)	(None, 3)	1539

Total params: 33,928,147
Trainable params: 33,928,147
Non-trainable params: 0

(B)

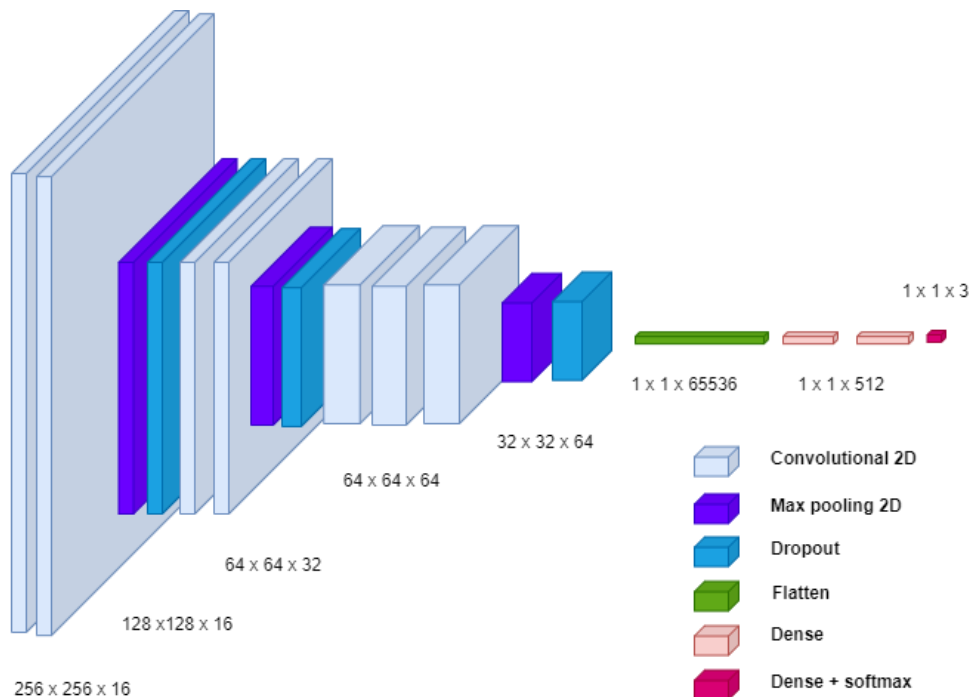


Figure II-4 Exemple des différentes couches composant un réseau de neurones (A) Listing des différentes couches composant le CNN. (B) Représentation schématique

Différentes fonctions de coût, (fonction mathématique permettant d'évaluer les performances du modèle en mesurant l'erreur entre les prédictions du modèle et les valeurs réelles) ont pu être utilisées selon les prédictions à effectuer. En effet, pour des prédictions binaires, la fonction de *binary-crossentropy* a été appliquée alors que pour des prédictions multi-classes, la fonction de *categorical-crossentropy* a été choisie. L'optimisation et donc la minimisation de cette fonction de coût repose sur un algorithme spécifique appelé descente de gradient. Il permet de calculer le minimum local d'une fonction (convexe) en changeant au fur et à mesure (itérations) les paramètres de cette fonction (*learning rate* notamment). Notre algorithme d'optimisation de la fonction de coût est *Adam optimizer*. Le *learning rate* pouvait varier de $1e-3$ à $1e-6$. Enfin, la métrique d'ajustement choisie était l'« *accuracy* », c'est-à-dire la précision.

1-3-4 Performance du modèle

L'entraînement des modèles a été réalisé sur 100 « epochs ». Le sur-apprentissage est défini par la détérioration des performances du modèle sur le set de test alors que les performances continuent de s'améliorer sur le test d'entraînement.⁸⁵

Afin de prévenir le sur-apprentissage (**Figure II-5**), une fonction d'arrêt précoce a été définie, elle était basée sur l'absence d'amélioration de la fonction de coût sur 3 epochs successifs. Pour chaque modèle, la matrice de confusion a été établie permettant d'évaluer les performances. Les courbes de précision et de *crossentropy* sur le set d'entraînement et de validation en fonction des epochs ont également été tracées.

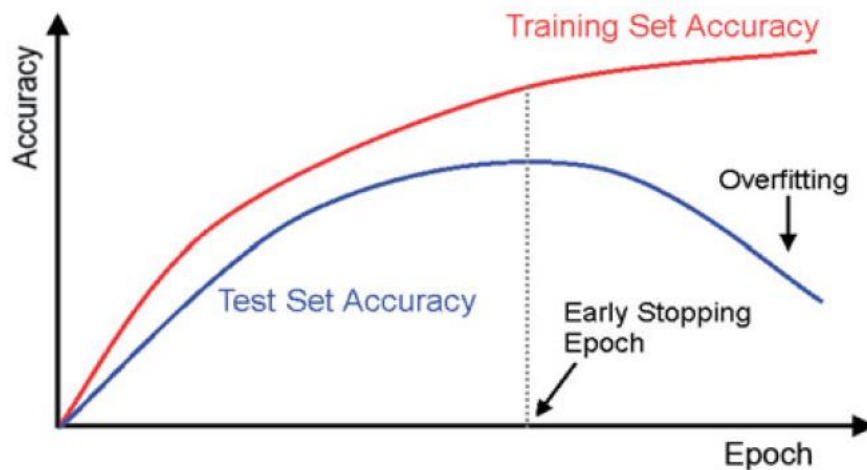


Figure II-5 Illustration du sur-apprentissage

Plusieurs métriques ont été utilisées pour évaluer le modèle. Pour rappel, l'«accuracy» ou exactitude est définie comme étant la proximité des valeurs par rapport à une mesure spécifique. Un ensemble de mesures répétées peut être exact si la moyenne de ces mesures est proche de la valeur réelle. Le F1-Score, tient compte de la valeur prédictive positive ainsi que de la sensibilité. Par conséquent, le nombre de vrais négatifs n'est pas pris en compte

1-3-4 Explicabilité du modèle

La principale composante utilisée pour construire l'architecture CNN est la couche de convolution. Plusieurs couches peuvent être empilées les unes sur les autres pour former la hiérarchie des caractéristiques ou motifs à extraire de l'image. Chaque couche peut être comprise comme fonctionnant en extrayant les caractéristiques de la couche précédente, puis en les utilisant comme données d'entrée pour la structure hiérarchique à laquelle elle est connectée. Du point de vue du réseau neuronal, la carte de caractéristiques correspond à l'unité cachée de la couche neuronale. Plus précisément, chaque coordonnée de la carte des caractéristiques correspond à un seul neurone dont le champ récepteur correspond à la taille du filtre. La valeur du filtre correspond au poids de la connexion entre les neurones de la couche

courante et les neurones de la couche supérieure. La pratique montre que le filtre est similaire à un détecteur de bords, et chaque filtre est ajusté à une fréquence spatiale, une échelle et une orientation différentes.

Afin d'expliquer notre modèle nous avons donc exploré le poids de chacune des couches de convolution et en particulier les plus superficielles permettant de mettre en évidence les zones de l'image participant à la prise de décision.

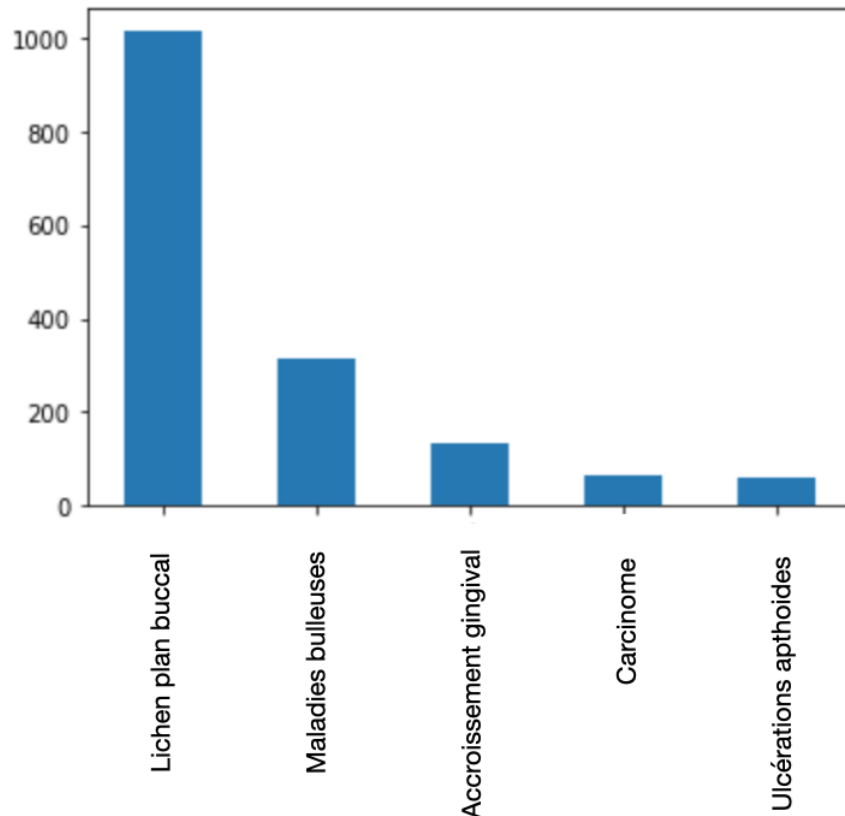
2- Résultats

2357 images ont été incluses initialement. Seules les photos de lésions associées aux diagnostics précédemment évoqués ont été conservées. C'est donc finalement 1591 images de lésions qui ont été utilisées pour les phases d'entraînement et de validation de notre modèle. La répartition des photos selon le diagnostic correspondant est exposée en **Figure II-6**. Il existe un déséquilibre de classe relativement important avec une sur-représentation du diagnostic de Lichen plan buccal. L'impact d'un set déséquilibré sur l'entraînement d'un CNN et les données obtenues peut être majeur.⁸⁶

Lorsqu'un ensemble de données est déséquilibré, le modèle affiche à la fois une baisse de la précision et un surajustement (« *overfitting* ») croissant pour les classes minoritaires. On observe également des difficultés à généraliser à la distribution de test des classes minoritaires. Le classificateur CNN fonde ses décisions sur une faible partie de la dimension de ses caractéristiques d'entrée. Dans le cas des classes minoritaires, le nombre de caractéristiques pertinentes est encore plus faible. En raison de la plus grande diversité des instances majoritaires, un plus grand nombre de caractéristiques est nécessaire pour prédire la classe dans sa totalité. Les stratégies de rééquilibrage des poids employées par certaines méthodes sensibles aux coûts, de suréchantillonnage ou de réentraînement du classificateur peuvent ne pas être

suffisantes pour remédier au problème du déséquilibre des classes. Enfin, un classificateur ré-entraîné avec des caractéristiques issues d'un extracteur déséquilibré ne peut pas retrouver les niveaux de précision d'un CNN entraîné sur des données équilibrées.

Figure II-6 : Répartition du nombre de photos de lésion selon le diagnostic associé.



Le **Tableau II-2** expose les valeurs d'«accuracy» et le F1-score pour chacune des caractéristiques lésionnelles à prédire. A l'exception de la prédiction du type de lésion élémentaire et de la couleur en 4 classes (Rouge, Blanc, Mixte, Pigmenté) l'exactitude de notre modèle varie de 0.62 à 1.00. Toutefois le F1-score reste faible allant de 0.1 à 0.53. Ainsi, il ressort que l'exactitude n'est pas une métrique correcte du fait du déséquilibre entre les différentes classes à prédire. Le modèle qui peut paraître efficace dans la prédiction effectuée en réalité des prédictions proches de l'aléatoire.

Concernant la prédiction du type de lésion élémentaire, l'exactitude est de 0,25, notre modèle n'est donc clairement pas adapté pour prédire ce type de caractéristique.

Deux approches ont été évaluées concernant la prédiction de la couleur des lésions. Une première approche consistait à effectuer une prédiction multi-classes. Ainsi on cherchait à prédire si la lésion appartenait à chacune des 4 classes : Blanc, Rouge, Mixte ou Pigmenté. Cette approche a donné des résultats relativement décevants avec une exactitude de 0.47 et un F1-score de 0.10. Une seconde approche dite « *multilabel* » a donc été envisagée, le but étant de prédire si la lésion était Rouge, Blanche ou bien à la fois Rouge et Blanche. Pour ce faire, une fonction de coût : *Binary cross-entropy* a été choisie et associée à une activation de type *sigmoid* et non *softmax*. Cette méthode pour annoter les images a permis une amélioration significative des performances avec une exactitude de 0.74 et un F1-score de 0.45. Toutefois ces résultats ne semblent pas satisfaisants à une utilisation clinique. En effet, un F1 score de 0,45 signifie que le modèle effectue plus de deux erreurs pour une prédiction positive correcte.

Les CNN sont souvent considérés comme des « boîtes noires », et il est difficile de comprendre les interactions entre les différents neurones amenant à la prédiction. Cependant, quelques moyens permettent d'essayer d'explorer les réseaux de neurones notamment la carte des caractéristiques ou « *Feature map* ». Cette carte des caractéristiques a pour but de mettre en évidence les informations présentes dans l'image et ayant servi de support à la prédiction. **La Figure II-7** est un exemple de carte des caractéristiques pour une lésion de lichen plan typique d'aspect réticulé. Notre algorithme a effectué une mauvaise prédiction en classant cette lésion comme étant non réticulée. L'observation du premier bloc de convolution confirme que le réseau identifie correctement la lésion par rapport à la muqueuse sous-jacente. Toutefois, l'information extraite dans le dernier bloc de convolution est sensiblement identique au premier bloc, signifiant que les informations les plus pertinentes n'ont pas été extraites.

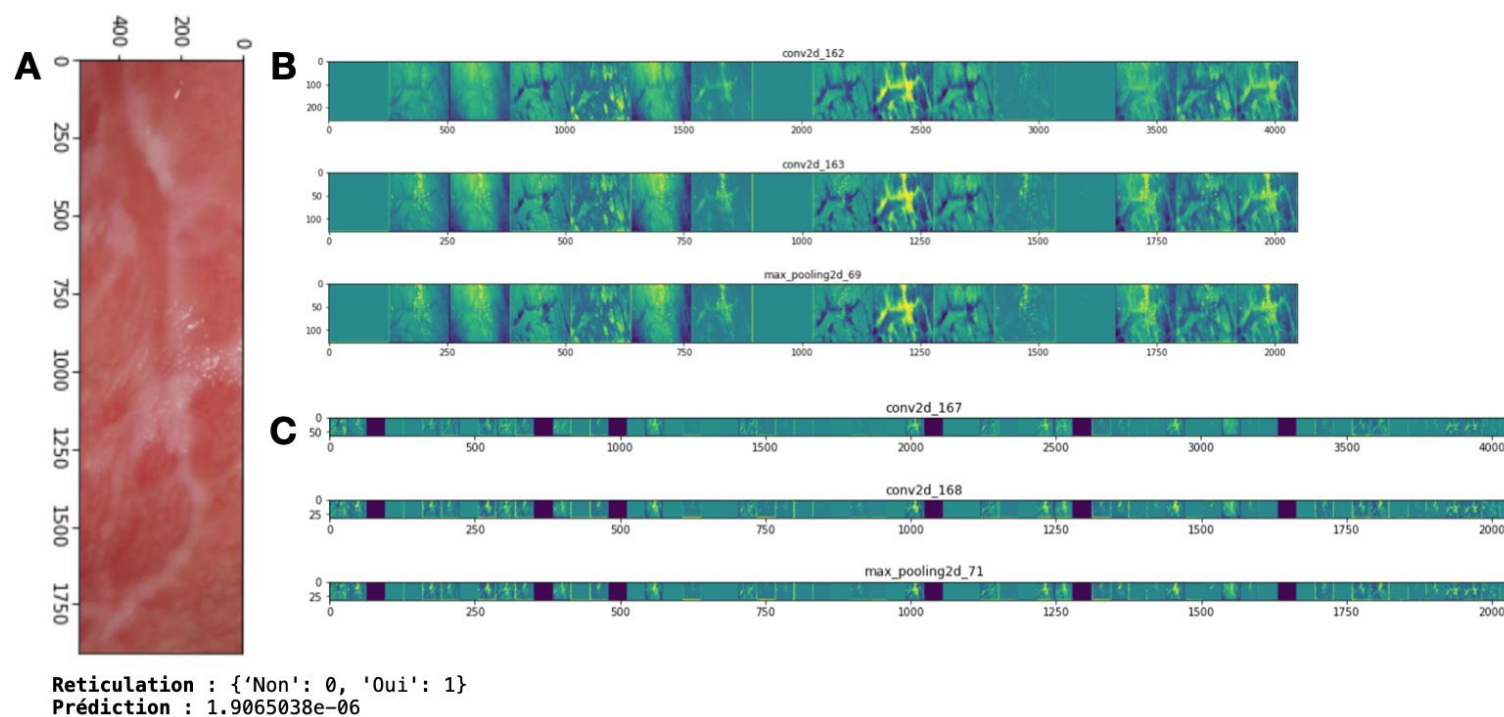
Tableau II-2 : Performances prédictives des différentes caractéristiques lésionnelles sur le set de test.

Caractéristiques	Couleur (4 classes)	Couleur Multilabel (2 classes)	Reticulation	Homogénéité	Lésion élémentaire	Contour bien délimité	Bourgeonnant	Papillomateux
Exactitude (« accuracy ») Données équilibrées	0.47	0.74	0.84	0.65	0.25	0.62	0.90	1.00
F1-Score Données équilibrées	0.10	0.45	0.3	0.53	0.1	0.57	0.17	0.00

Figure II-7 : Exploration des différentes couches de convolution

A : Image d'une lésion de lichen plan buccal avec son aspect typiquement réticulé. Le modèle prédit à tort que cette lésion est non réticulée.

B et C : Carte des caractéristiques ou *Feature map* pour le premier bloc de convolution (B) et le dernier bloc de convolution (C). Les blocs de convolution ont pour objectif d'extraire l'information à partir des pixels présents dans l'image. Plus on descend dans le réseau neuronal et plus l'information extraite doit être pertinente. La carte de caractéristiques correspond quant à elle à l'unité cachée de la couche neuronale. Plus précisément, chaque coordonnée de la carte des caractéristiques correspond à un seul neurone dont le champ récepteur correspond à la taille du filtre.



3- Discussion

Nous avons donc testé dans cette première partie plusieurs algorithmes de DL et ceci afin de prédire soit directement la lésion, soit une caractéristique sémiologique liée à cette dernière. Les résultats n'ont pas été probants. Plusieurs hypothèses, liées aux lésions elles-mêmes ou au DL en lui-même, peuvent expliquer ces résultats.

3-1 Hétérogénéité des lésions

En raison de l'hétérogénéité des lésions orales et de la similarité clinique entre certaines entités pathologiques^{77,83,87}, le processus diagnostique va bien au-delà de l'analyse des caractéristiques d'une image. Ce processus nécessite la recherche de stratégies qui permettent la combinaison d'informations issues de l'image et l'inclusion d'informations supplémentaires obtenues par l'examen clinique⁸⁸. Aucune étude sur les maladies bucco-dentaires ne décrit l'utilisation de bases de données publiques, ce qui indique l'inexistence de ce type de base de données et justifie le choix de Welikala. et al⁸⁷ de créer une base de données avec plusieurs entités pathologiques orales, associées aux données cliniques correspondantes. En raison du manque de bases de données publiques, ou même de l'absence de collecte, certaines études sur la médecine orale ont utilisé des images extraites d'internet pour le développement de modèles d'IA, obtenant ainsi un volume de données plus important mais non représentatif.^{78,80,82,87} Par ailleurs, les photos endo-buccales étant difficilement standardisables, il est nécessaire de normaliser au maximum les images pour les comparer⁸⁹

3-2 Rôle des informations cliniques

La plupart des modèles actuels ne sont entraînés qu'avec des images de lésions sans tenir compte d'autres informations cliniques liées aux patients. Étant donné que les médecins prennent généralement des décisions cliniques avec des informations supplémentaires autres que l'image, comme les examens de dossiers médicaux (antécédents, mode de vie, données biologiques), l'ajout de ces informations au modèle de DL pourrait conduire à de meilleures performances de classification. Haenssle et al,³⁰ ont montré qu'avec l'ajout des informations cliniques, il y a une augmentation de la sensibilité et de la spécificité de la performance diagnostique. Ainsi, il est prédit que les modèles de DL peuvent bénéficier de l'inclusion des métadonnées des patients.

3-3 Importance des véritables étiquettes et de la vérité terrain

Dans des ensembles de données plus importants (MNIST, CIFAR et ImageNet), les modèles de DL sont capables de généraliser à partir des données d'entraînement, lorsque les véritables étiquettes sont bien plus nombreuses que les étiquettes incorrectes⁹⁰. Cela souligne l'importance d'utiliser les rapports histopathologiques comme vérité de terrain pour des tâches importantes telles que la détection de lésions, étant donné que des lésions mal labellisées par des experts peuvent diluer l'ensemble de données, entraînant ainsi une performance inférieure du modèle.

3-4 Duplicité des données

Cassidy & al⁹¹ ont noté que plusieurs manuscrits utilisant l'ensemble de données ISIC avaient des images en double ou similaires dans les ensembles d'entraînement et de test, introduisant un biais dans le modèle CNN, et ont proposé une méthodologie pour supprimer les images en double. Les prédictions d'un modèle s'améliorent en précision en extrayant un plus grand nombre de caractéristiques uniques plutôt qu'en enrichissant simplement les données en se procurant un grand nombre d'images à partir d'un petit nombre de sources. Il convient de noter

qu'une riche source de données non dupliquées doit être utilisée pour la formation des CNN afin d'éviter les biais et la surestimation des performances du modèle.

3-5 Qualité et déséquilibre des données

La qualité des images peut être une source de préoccupation, en particulier avec les ensembles de données institutionnels non publics, car, la qualité des images cliniques peut varier en fonction de l'appareil et de l'opérateur qui capture les images. Pour les images dermoscopiques, les images sont prises avec un appareil désigné, et donc la qualité peut ne pas varier autant. Ce n'est pas le cas en pathologie de la muqueuse orale, les images ne sont pas standardisées tant en qualité, qu'en zoom ou en exposition. En effet, il est également à noter que les photos endobuccales contiennent des structures très différents, telles que de la muqueuse, mais aussi des dents, voir des restaurations prothétiques.

Le contrôle de la qualité des images dans de grands ensembles de données est une tâche difficile. L'augmentation de la diversité des images dans les ensembles de données et le développement d'outils pour évaluer la qualité des images et supprimer les données en double sont des solutions nécessaires pour améliorer le développement de modèles à l'avenir.

3-6 Validité externe des modèles

Bien que ces études montrent une utilisation potentielle des modèles d'IA en dermatologie et en médecine orale, il convient de noter que la majorité des articles sont en grande partie des preuves de concept, entraînées et testées sur des ensembles de données rétrospectives. Les limites à la généralisation peuvent être divisées en trois catégories : le manque d'ensembles de données en général, le manque de diversité dans les ensembles de données et le manque d'informations sur les patients.

Les obstacles à la validité externe seraient le déséquilibre des données entre l'âge, le sexe, l'origine ethnique, le type de maladie et la prévalence de la maladie, qui, s'il n'est pas suffisamment traité, pourrait entraîner une mauvaise performance lorsque ces modèles sont testés en dehors de leur population de d'entraînement et de test.

Une étude a rapporté que plusieurs algorithmes de ML peuvent être sous-performant selon le phototype des patients, car les ensembles de données utilisés pour former ces modèles, tels que les archives de défi ISIC, ont été largement collectés auprès de patients aux États-Unis, en Europe et en Australie ⁹².

3-7 Difficultés d'explicabilité des modèles de DL

La compréhension du processus de classification automatisée est une étape importante, en particulier dans le domaine de la santé. Bien que les connexions se produisent dans les couches intermédiaires des modèles DL et sont inconnues, l'analyse des erreurs peut donner des indices sur les données à l'origine de la confusion. Cette analyse peut améliorer la compréhension des paramètres que les modèles utilisent pour prendre leurs décisions. ^{93,94}

III- *Machine Learning* et apprentissage statistique, une solution alternative⁸⁸

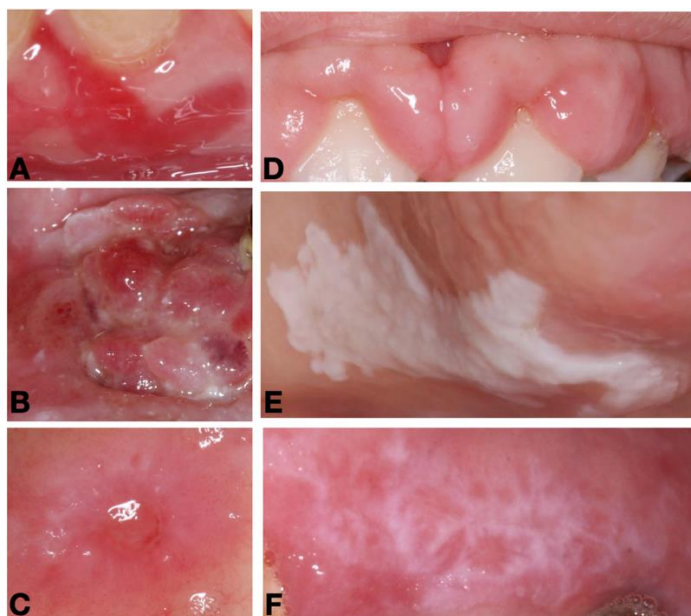
A- Objectif

Compte tenu des limites précédemment évoquées, nous avons choisi de tester une nouvelle approche, basée cette fois ci sur des modèles de *Machine Learning* (ML) reposant sur un apprentissage statistique.

L'objectif de cette deuxième phase fut de pallier le manque de précision diagnostic de notre modèle de DL. Pour cela nous avons opté pour un modèle de ML basé sur les caractéristiques sémiologiques des lésions. Cette approche nous a également permis de prendre en compte certaines données concernant l'état de santé général des patients.

Pour cela, 6 lésions orales les plus rencontrées dans notre consultation de pathologie de la muqueuse orale ont été considérées (**Figure III-1**).

Figure III-1 : Illustration des différentes situations cliniques à prédire



A : Maladie bulleuse (pemphigus),

B : Carcinome,

C : Ulcération aphtoïde,

D : Accroissement gingival,

E : Leucoplasie,

F : Lichen plan buccal.

B- Evaluation de l'apport du *Machine Learning*

1- Matériel et méthode

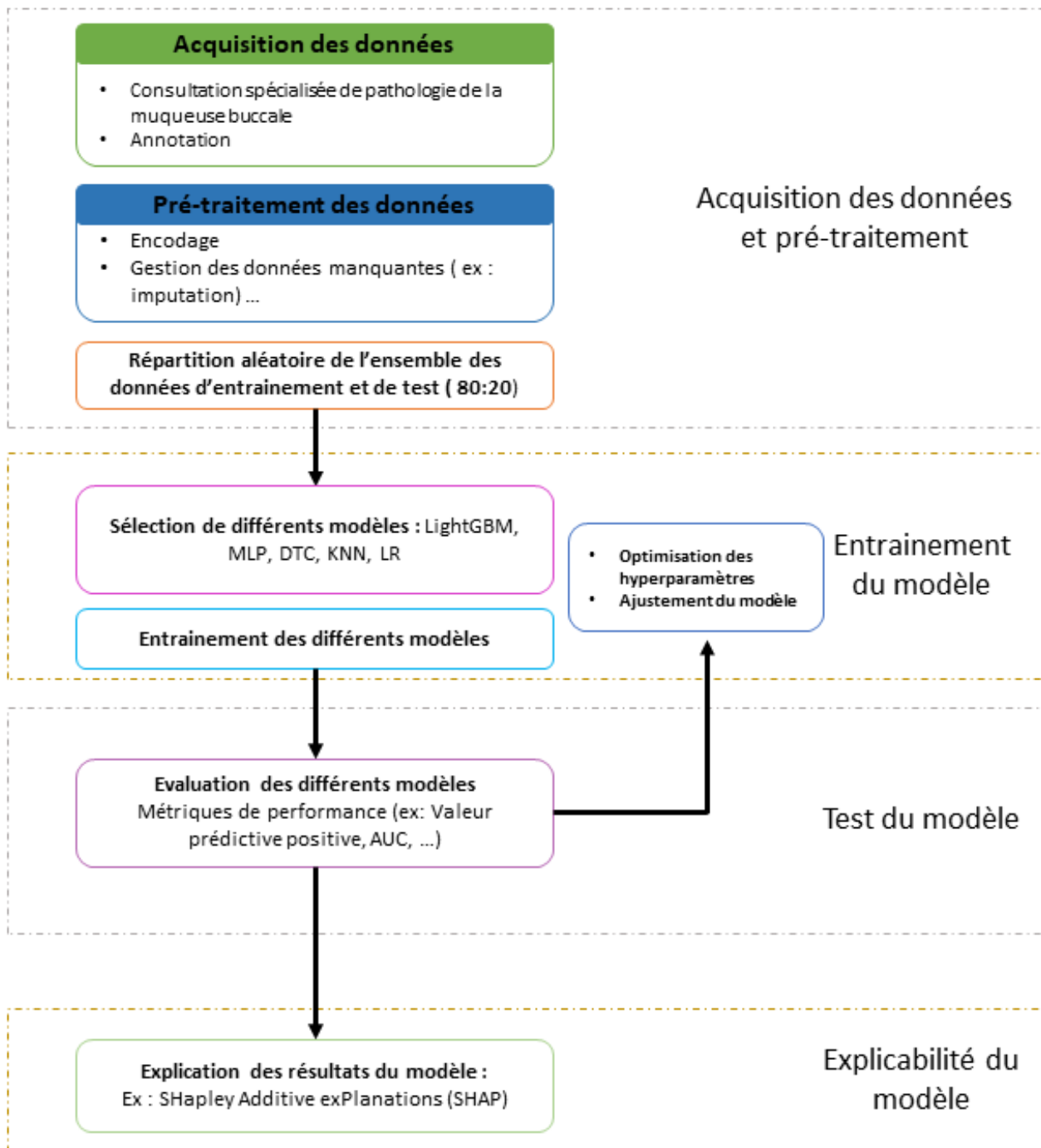
1-1 Critères de sélection

La base de données précédemment réalisée a été utilisée pour ce travail.

1-2 Collecte et annotation des données

Pour chacune des images incluses dans la base de données, chaque caractéristique de la lésion (lésion, couleur, topographie, taille, homogénéité et forme) a été annotée par un seul investigateur (A.D), préalablement calibré par deux experts en pathologie de la muqueuse orale (S.C, S.L) avant le lancement de l'étude en évaluant 50 images sélectionnées au hasard. Le kappa inter-examineur global a atteint 0,92. La fiabilité intra-examineur a été testée en réexaminant au hasard 30 lésions à un mois d'intervalle de l'évaluation initiale et a atteint 0,87. Les diagnostics associés à chacune des images ont été préalablement validés histologiquement. L'analyse anatomopathologique des lésions orales dans le cadre des soins de routine du patient a été réalisée dans le laboratoire anatomopathologique du CHU de Toulouse et de l'Institut Universitaire du Cancer de Toulouse. Enfin, les données démographiques (âge, sexe, consommation d'alcool, tabagisme), médicales et médicamenteuses ont également été enregistrées. Les antécédents médicaux et les médicaments ont été classés respectivement selon la classification internationale des maladies (CIM) et le système de classification anatomique thérapeutique chimique (ATC).

Figure III-2 : Pipeline de *Machine Learning*



1-3 Développement du modèle et analyse des données

1-3-1 Étape de prétraitement

Les données catégorielles binaires (oui/non ou homme/femme) ont d'abord été « label-encodées » par des nombres entiers. Pour les modèles de ML non basés sur les arbres décisionnels, le type de lésion élémentaire a été « one-hot encodées » (0 ou 1). Les données sociodémographiques et/ou médicales manquantes (pour l'âge (68, 7,1%), le sexe (61, 6,4%), le tabac (7, 0,7%), l'alcool (7, 0,7%) et les stupéfiants (7, 0,7%)) ont été considérées comme telles pour que l'approche algorithmique puisse les traiter (*tree based gradient boosting model* LightGBM) ou imputées à l'aide d'un modèle d'imputation unique multivarié⁹⁵.

1-3-2 Modélisation des données

Différentes stratégies de *Machine Learning* ont été comparées : LightGBM (une méthode de gradient boosting basée sur les arbres), *Logistic Regression with an elastic net penalization* (LR), *Decision tree* (DT), *K-Nearest Neighbors* (KNN) and *Multi-Layer Perceptron* (MLP).

Les modèles ont été utilisés à partir de *scikit-learn 1.0.1* et *lightgbm 3.3.1*. Pour optimiser les hyperparamètres de chaque modèle, une "grid search" a été réalisée en utilisant *optuna 2.10.0*, avec la moyenne ROC AUC un-vers-un après une validation croisée 5-fois considérée comme la métrique d'évaluation.

1-3-3 Performance du modèle :

Pour chaque modèle, nous avons calculé le taux de vrais positifs (proportion de données positives correctement considérées comme positives), le taux de vrais négatifs (proportion de données négatives correctement considérées comme négatives), la valeur prédictive positive (nombre de résultats positifs corrects divisé par le nombre de résultats positifs prédits par le classificateur), la valeur prédictive négative (nombre de résultats négatifs corrects divisé par le

nombre de résultats négatifs prédits par le classificateur), la précision globale (nombre de prédictions correctes divisé par le nombre total de prédictions) et le score F1 (moyenne harmonique entre les valeurs prédictives positives et le taux de vrais positifs).

1-3-4 Explicabilité du modèle :

Enfin, l'influence de chaque attribut de l'ensemble de données sur la prédiction du diagnostic a été explorée à l'aide de *SHapley Additive exPlanations* (SHAP)⁹⁶. Compte tenu de la grande dimensionnalité de l'ensemble de données (nombre élevé de variables) et de la nature du modèle final choisi (modèle arborescent lightGBM), le modèle a été expliqué grâce à l'optimisation TreeSHAP⁹⁷.

2- Résultats

L'ensemble des données comprenait un total de 1242 dossiers représentant 31 diagnostics différents pour 299 sujets (39,4 % d'hommes ; 60,6 % de femmes ; âge moyen 57,3 ans \pm 19,2). Compte tenu du nombre de cas par diagnostic, 7 catégories ont été considérées : lichen plan buccal (40,0 % de l'ensemble de données), maladies bulleuses (13,9 %), ulcères aphteux (8,9 %), leucoplasie (7,3 %), hypertrophie gingivale (3,7 %), carcinome épidermoïde buccal (3,3 %), et autres (22,9 %). Une description détaillée de l'ensemble des données finales (données sociodémographiques et médicales) est fournie dans le **tableau III-1**.

Seuls 8 (2,7 %) sujets étaient alcooliques et 72 (24,1 %) étaient fumeurs. Les lésions élémentaires les plus représentées étaient les plaques/papules (27 %), 36 % de ces lésions étant situées sur la gencive attachée. Une fois les modèles réglés pour l'optimisation des hyperparamètres et entraînés sur l'ensemble de données d'entraînement, les performances de chaque modèle ont été évaluées sur l'ensemble de données de test. Le modèle LightGBM a obtenu la meilleure performance avec une précision moyenne de 0,84. Les autres modèles de ML ont obtenu des performances plus faibles avec 0,54, 0,73, 0,49 et 0,62, respectivement, pour la *logistic regression* (LR), *K-nearest neighbors* (K-NN), *decision tree classifier* (DTC) et *multilayer perceptron models* (MLP). Les performances détaillées des modèles sont présentées dans le **Tableau III-2**. La matrice de confusion est présentée dans la **Figure III-3A** pour le modèle LightGBM.

Tableau III-1 : Caractéristiques des sujets utilisées pour la formation du modèle de machine learning en fonction de chaque diagnostic.

Number of Subjects	Lichen Planus (n = 77)	Leukoplakia (n = 41)	Aphthous Ulcer (n = 37)	Bullous Diseases (n = 26)	Gingival Enlargement (n = 12)	Carcinoma (n = 11)	Others (n = 95)	p-Value ^a
Gender (n,%)								
Male	22 (29%)	20 (49%)	19 (51%)	9 (35%)	5 (42%)	5 (45%)	42 (44%)	0.03
Female	51 (66%)	16 (39%)	14 (38%)	15 (58%)	3 (25%)	4 (36%)	53 (56%)	
Mean age	65	65	50	65	43	54	50	<0.001
Alcohol (n,%)	2 (3%)	4 (10%)	1 (3%)	-	-	2 (18%)	2 (2%)	0.007
Smoking (n,%)	21 (27%)	18 (44%)	8 (22%)	3 (12%)	1 (8%)	5 (45%)	16 (17%)	0.002
Number of lesions (n)	497	91	110	173	46	41	252	
Localization (n,%)								
Vermilion	19 (3.8 %)	6 (6.5%)	5 (4.5%)	3 (1.7%)	-	-	16 (6.3%)	
Commissure	1 (0.2%)	-	-	1 (9.6%)	-	-	8 (3%)	
Labial mucosa	8 (1.7%)	3 (3%)	15 (13.6%)	16 (9.2%)	-	1 (2.4%)	21 (8.3%)	
Muco-buccal fold	136 (27.3%)	25 (27.5%)	18 (16.4%)	16 (9.2%)	-	5 (12%)	21 (8.3%)	
Attached gingiva	172 (34.6%)	12 (13.1%)	12 (10.9%)	91 (52.6%)	45 (98%)	16 (39%)	85 (33.7%)	
Buccal mucosa	37 (7.4%)	3 (3.3%)	10 (9.0%)	14 (8.0%)	1 (2.0%)	2 (4.8%)	11 (4.4%)	0.002
Hard palate	22 (4.4%)	13 (14.3%)	5 (4.5%)	10 (5.7%)	-	11 (26.8%)	20 (7.9%)	
Soft palate	1 (0.2%)	4 (4.4%)	5 (4.5%)	19 (10.9%)	-	-	2 (0.7%)	
Tonsillar pillar	3 (0.6%)	1 (1.1%)	1 (0.9%)	1 (0.5%)	-	-	2 (0.7%)	
Dorsal tongue	60 (12.0%)	7 (1.8%)	3 (2.7%)	2 (1.2%)	-	2 (4.8%)	51 (20.2%)	
Ventral tongue	9 (1.8%)	1 (1.0%)	10 (9.0%)	6 (3.5%)	-	1 (2.4%)	2 (0.7%)	
Lateral tongue	24 (4.8%)	14 (15.4%)	13 (11.8%)	-	-	1 (2.4%)	11 (4.4%)	
Floor of mouth	5 (1.0%)	2 (2.2%)	2 (1.8%)	-	-	2 (4.8%)	2 (0.7%)	
Lesion color (n,%)								
White	160 (32.0%)	70 (7.8%)	33 (30.0%)	46 (26.6%)	-	13 (31.7%)	50 (18.8%)	
Red	55 (11.0%)	2 (2.2%)	9 (8.1%)	55 (31.8%)	15 (32.6%)	8 (19.5%)	51 (20.2%)	<0.001
Mixt (red and white)	268 (54%)	19 (20.8%)	66 (60.0%)	69 (39.9%)	5 (11.8%)	16 (39.0%)	79 (31.3%)	
Pigmented	10 (2.0%)	-	1 (0.9%)	-	5 (11.8%)	3 (2.3%)	40 (15.8%)	
No color changes	4 (0.8%)	-	1 (0.9%)	-	21 (46.0%)	1 (2.4%)	32 (12.7%)	
Lesion size (n,%)								
<5 mm	39 (7.8%)	13 (14.3%)	27 (24.5%)	32 (18.5%)	27 (58.6%)	1 (2.4%)	41 (16.2%)	
5 to 10 mm	268 (54.0%)	41 (45.1%)	66 (60.0%)	116 (67.0%)	38 (82.6%)	9 (21.9%)	135 (53.1%)	<0.001
10 to 50 mm	188 (38.0%)	37 (40.7%)	17 (15.5%)	25 (14.5%)	6 (13.0%)	27 (65.8%)	75 (29.7%)	
>50 mm	2 (0.4%)	-	-	-	-	4 (9.7%)	1 (0.4%)	

Number of Subjects	Lichen Planus (n = 77)	Leukoplakia (n = 41)	Apthous Ulcer (n = 37)	Bullous Diseases (n = 26)	Gingival Enlargement (n = 12)	Carcinoma (n = 11)	Others (n = 95)	<i>p</i> -Value ^a
Elementary Lesion (n,%)								
Gingival hyperplasia	3 (0.6%)	-	-	1 (0.5%)	40 (86.9%)	-	9 (3.5%)	
Bulla	7 (1.4%)	2 (2.2%)	2 (1.8%)	31 (18.0%)	-	-	16 (6.3%)	
Cellular coating	11 (2.2%)	1 (1.1%)	13 (11.8%)	16 (9.2%)	-	2 (4.8%)	12 (4.7%)	
Erosion	18 (3.6%)	3 (3.3%)	6 (5.5%)	30 (18.3%)	-	1 (2.4%)	8 (3.1%)	<0.001
Macula/patch	184 (37.0%)	12 (13.2%)	11 (10.0%)	79 (45.7%)	6 (13.0%)	4 (9.7%)	78 (30.9%)	
Nodule	13 (2.6%)	2 (2.2%)	2 (1.8%)	-	-	18 (43.9%)	100 (39.6%)	
Papula/plaque	220 (44.3%)	70 (77.0%)	8 (7.3%)	7 (4.0%)	-	13 (31.7%)	23 (9.1%)	
Ulcer	40 (8.0%)	1 (1.1%)	68 (61.8%)	9 (5.2%)	-	3 (7.3%)	4 (1.5%)	
Lesion homogeneity (n,%)	187 (37.6%)	52 (75.1%)	77 (70.0%)	71 (41.0%)	32 (69.5%)	5 (12.1%)	142 (56.4%)	<0.001
Well circumscribed lesion (n,%)	230 (46.3%)	55 (60.4%)	93 (84.5%)	81 (46.8%)	19 (41.3%)	7 (17.0%)	166 (65.8%)	<0.001
Elevated edges of the lesion (n,%)	5 (1.0%)	-	13 (11.8%)	2 (1.2%)	1 (2.0%)	1 (2.4%)	4 (1.6%)	0.29
Reticular lesion (n,%)	175 (35.2%)	3 (3.3%)	1 (0.9%)	5 (2.9%)	-	-	5 (1.9%)	<0.001
Growing lesion (n,%)	13 (2.6%)	2 (2.2%)	-	1 (0.5%)	3 (6.5%)	24 (58.5%)	25 (9.9%)	<0.001
Papillomatus lesion (n,%)	3 (0.6%)	8 (8.8%)	1 (0.9%)	1 (0.5%)	1 (2.0%)	8 (19.5%)	41 (16.2%)	0.007

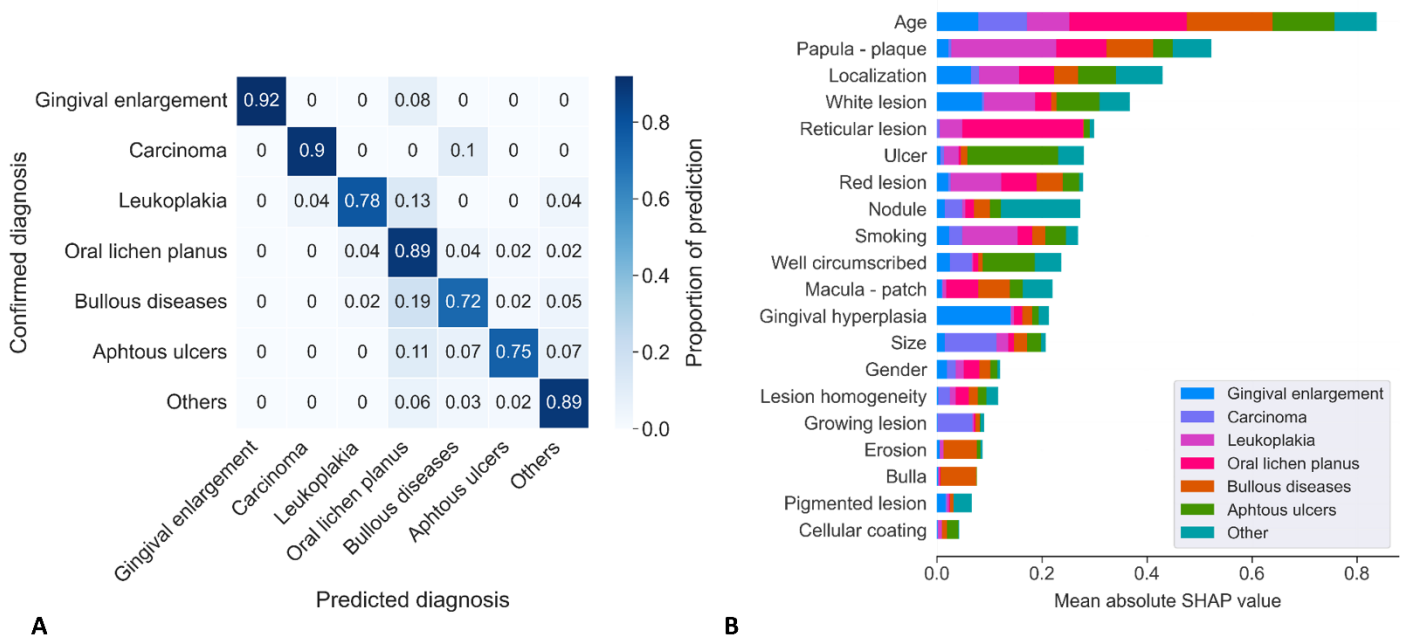
^a for categorical values, the *p*-value resulted from a χ^2 test; for numerical values, the *p*-value resulted from an ANOVA test.

Tableau III-2 : Performances des différents modèles testés pour chacun des diagnostics

Diagnostics	LightGBM					Elastic Net Regression					K-Nearest Neighbors					Decision Tree					Multilayer Perceptron				
	TPR	TNR	PPV	NPV	F1	TPR	TNR	PPV	NPV	F1	TPR	TNR	PPV	NPV	F1	TPR	TNR	PPV	NPV	F1	TPR	TNR	PPV	NPV	F1
Gingival enlargement	0.92	1.00	1.00	1.00	0.96	0.92	0.99	0.85	1.00	0.88	0.83	0.99	0.71	0.99	0.77	0.67	0.97	0.47	0.99	0.55	0.83	1.00	1.00	0.99	0.91
Carcinoma	0.90	1.00	0.90	1.00	0.90	0.60	0.92	0.21	0.99	0.32	0.70	0.99	0.64	0.99	0.67	0.80	0.96	0.40	0.99	0.53	0.50	0.98	0.50	0.98	0.50
Leukoplakia	0.78	0.98	0.75	0.98	0.77	0.74	0.91	0.40	0.98	0.52	0.57	0.94	0.42	0.96	0.48	0.74	0.89	0.36	0.98	0.49	0.35	0.98	0.62	0.95	0.44
Oral lichen planus	0.89	0.89	0.85	0.92	0.87	0.38	0.92	0.77	0.68	0.51	0.85	0.88	0.83	0.89	0.84	0.27	0.98	0.92	0.66	0.42	0.84	0.70	0.66	0.86	0.74
Bullous diseases	0.72	0.96	0.76	0.95	0.74	0.70	0.81	0.38	0.94	0.49	0.70	0.96	0.75	0.95	0.72	0.74	0.76	0.34	0.95	0.46	0.35	0.95	0.56	0.90	0.43
Apthous ulcers	0.75	0.99	0.84	0.97	0.79	0.54	0.96	0.56	0.95	0.55	0.46	0.97	0.59	0.95	0.52	0.75	0.92	0.50	0.97	0.60	0.50	0.98	0.74	0.95	0.60
Other lesions	0.89	0.97	0.89	0.97	0.89	0.60	0.94	0.72	0.90	0.66	0.67	0.93	0.72	0.91	0.69	0.46	0.93	0.64	0.87	0.54	0.73	0.92	0.70	0.93	0.71
Mean accuracy	0.84					0.54					0.73					0.49					0.67				

Abbreviations : TPR, true positive rate; TNR, true negative rate; PPV, positive predictive value; NPV, negative predictive value. True positive rate: proportion of positive data that are correctly considered as positive. True negative rate: proportion of negative data that are correctly considered as negative. Positive predictive value: number of correct positive results divided by the number of positive results predicted by the classifier. Negative predictive value: number of correct negative results divided by the number of negative results predicted by the classifier. F1-score: the harmonic mean between predictive positive values and true positive rate.

Figure III-3 : Matrice de confusion pour le modèle LightGBM et importance des caractéristiques pour les prédictions globales.



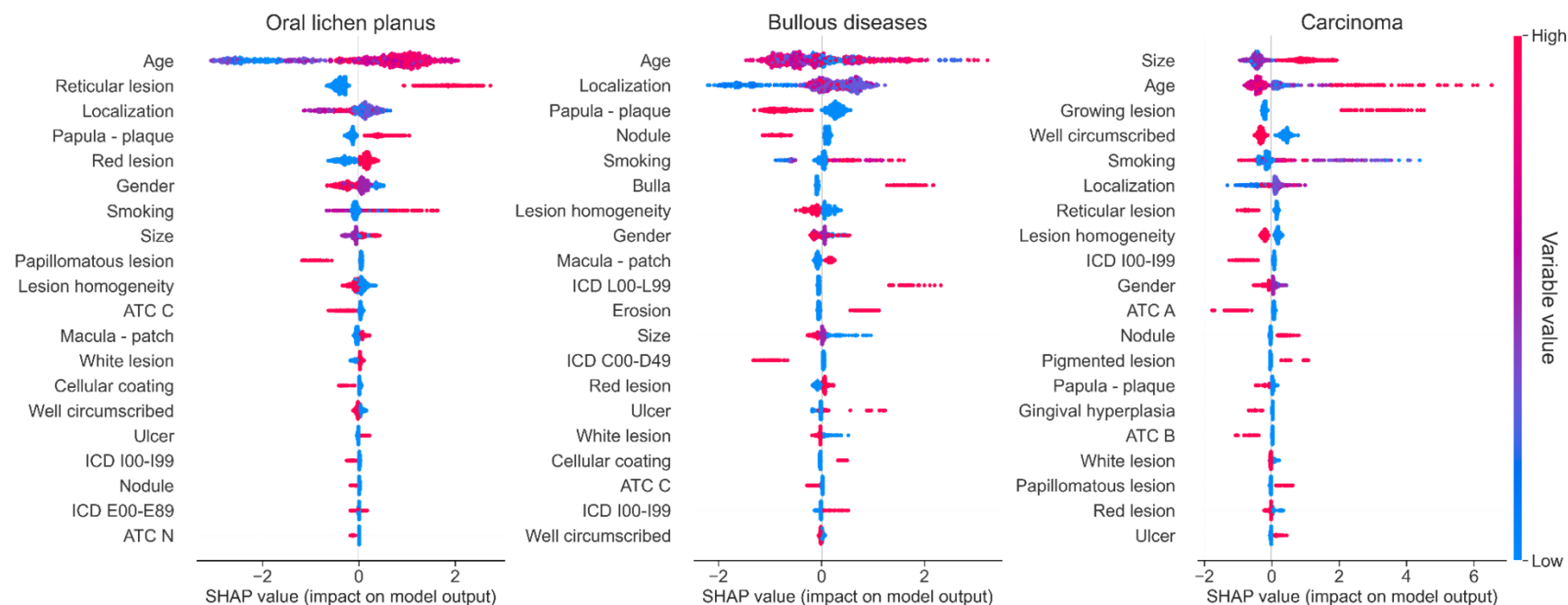
(A) La matrice de confusion décrit les performances du modèle de classification LightGBM sur l'ensemble de données de test. Le nombre de lésions correctement classées varie de 92 % à 72 % en fonction des différents diagnostics évalués.

(B) L'importance des caractéristiques SHAP pour le modèle LightGBM entraîné à prédire les lésions de la muqueuse orale est mesurée par les valeurs absolues moyennes de SHAP.

Quel que soit le diagnostic à prédire, le modèle LightGBM a obtenu les meilleures performances. La méthode TreeSHAP a été utilisée pour donner un aperçu de l'explicabilité des prédictions. La **Figure III-3B** montre l'importance des différentes caractéristiques dans la prise de décision globale. Les facteurs influençant la prise de décision globale sont essentiellement les caractéristiques des lésions (type de lésion, couleur, réticulation.). Bien que l'âge reste le facteur le plus important dans la prise de décision, d'autres données sociodémographiques sont prises en compte mais avec moins d'importance. Le modèle ne prend pas en compte les antécédents médicaux et les traitements dans l'explication globale du comportement du modèle. **La Figure III-4** montre l'importance des différentes caractéristiques pour la prédiction des diagnostics de lichen plan buccal, maladie bulleuse et carcinome. Si l'âge du patient semble influencer toutes les lésions, pour le lichen plan buccal, des caractéristiques telles que la forme réticulaire de la lésion, le sexe, le type de lésion élémentaire et la localisation ont plus d'influence sur le diagnostic que les antécédents médicaux du patient. Au contraire, pour le carcinome, la taille de la lésion, sa localisation et le statut tabagique du patient sont plus importants pour prédire le diagnostic que la couleur de la lésion ou le sexe. Les antécédents médicaux ont souvent une influence mineure sur la prédiction individuelle, à l'exception des antécédents de pathologie dermatologique, qui influencent fortement le diagnostic de la maladie bulleuse.

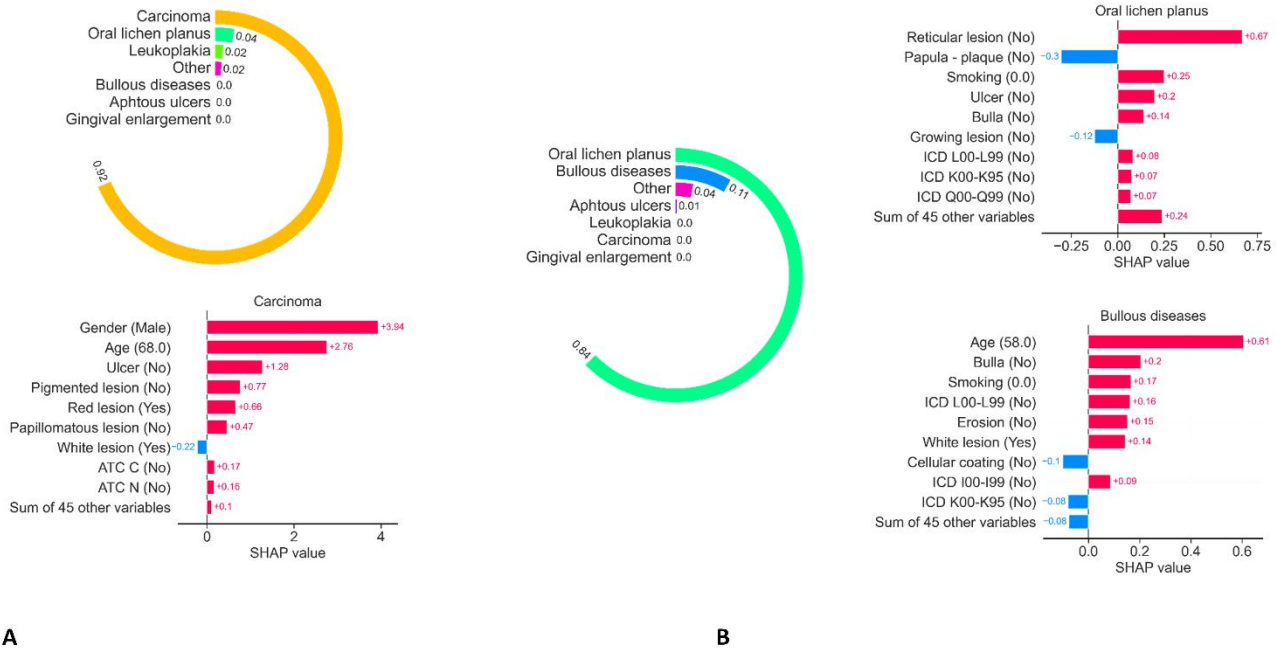
Enfin, la **Figure III-5** illustre deux exemples de prédiction, respectivement pour un carcinome (**Figure III-5A**) et une lésion de lichen plan buccal (**Figure III-5B**). L'algorithme SHAP permet de comprendre la prise de décision en soulignant les facteurs qui influencent positivement ou négativement les différents diagnostics.

Figure III-4 : Graphique récapitulatif de l'importance des caractéristiques pour chaque prédiction diagnostique : le lichen plan buccal, les maladies bulleuses et le carcinome.



Chaque point du graphique récapitulatif est une valeur SHAP pour une caractéristique et une instance donnée. La position sur l'axe des y est déterminée par la caractéristique et sur l'axe des x par la valeur SHAP. La couleur représente la valeur de la caractéristique, de la plus faible à la plus élevée. Les points qui se chevauchent sont divisés dans le sens de l'axe des ordonnées, ce qui nous donne une idée de la distribution des valeurs SHAP par caractéristique. Les caractéristiques sont classées en fonction de leur importance.

Figure III-5 : Illustration de deux exemples de prédiction, respectivement pour un carcinome (A) et une lésion de lichen plan buccal (B).



L'algorithm SHAP permet de comprendre la prise de décision en soulignant les facteurs qui influencent positivement ou négativement les différents diagnostics

Abréviations : ATC A, Classification anatomique thérapeutique chimique : Tube digestif et métabolisme ; ATC C, Classification anatomique thérapeutique chimique : Système cardiovasculaire ; ATC H, Classification anatomique thérapeutique chimique : Préparations hormonales systémiques ; ATC M, Classification anatomique thérapeutique chimique : Système musculo-squelettique ; ATC N, Classification anatomique thérapeutique chimique : Système nerveux ; ATC R, Classification anatomique thérapeutique chimique : Système respiratoire ; ICD C00-D49, classification internationale des maladies : Tumeurs ; ICD E00-E89, classification internationale des maladies : Maladies endocriniennes, nutritionnelles et métaboliques ; ICD F00-F99, classification internationale des maladies : Troubles mentaux et du comportement ; ICD G00-G99, classification internationale des maladies : Maladies du système nerveux ; ICD I00 I99, classification internationale des maladies : Maladies de l'appareil circulatoire ; ICD K00-K95, classification internationale des maladies : Maladies de l'appareil digestif ; ICD L00-L99, classification internationale des maladies : Maladies de la peau et du tissu sous-cutané ; ICD M00-M99, classification internationale des maladies : Maladies du système digestif ; ICD L00-L99, classification internationale des maladies : Maladies de la peau et du tissu sous-cutané. ICD M00-M99, classification internationale des maladies : Maladies de l'appareil locomoteur et du tissu conjonctif ; ICD N00-N99, classification internationale des maladies : Maladies de l'appareil génito-urinaire

3- Discussion

Sur la base de la topographie des lésions et des caractéristiques médicales et sociodémographiques des patients, le modèle LightGBM a permis de prédire les 6 lésions de la muqueuse orale sélectionnées dans cette étude avec une précision globale d'environ 0,84. La spécificité (taux de vrais négatifs) et la sensibilité (taux de vrais positifs) restent très élevées, allant respectivement de 0,89 à 1,00 et de 0,72 à 0,92, illustrant ainsi la capacité de discrimination de notre modèle. Par conséquent, cet outil semble adapté aux praticiens non spécialistes pour les guider dans leur processus diagnostique et thérapeutique. Le diagnostic précoce des lésions orales représente un défi pour les traiter rapidement et réduire les pénalités thérapeutiques, en particulier dans le cas du cancer. D'autant plus que les progrès thérapeutiques ont permis d'augmenter l'espérance de vie de ces patients. Bien que la dermatologie orale soit au carrefour de plusieurs spécialités médicales (dentistes, oto-rhino-laryngologistes, chirurgiens oraux et maxillo-faciaux, dermatologues), les lésions orales restent sous-diagnostiquées, en particulier pour les lésions précancéreuses et cancéreuses^{15,98}. En effet, les médecins généralistes manquent souvent de formation spécifique et les patients ne sont pas orientés vers des spécialistes dans ce domaine^{19,99}.

D'autre part, alors que l'apport de l'IA est en plein essor en médecine et notamment en dermatologie, seuls quelques modèles ont été développés en dermatologie orale^{77-80,83,100-102}. Pour répondre à ce problème, nous avons démontré que nous pouvions prédire avec précision le diagnostic d'une lésion de la muqueuse orale à l'aide d'une stratégie algorithmique de ML basée sur une consultation de routine en médecine orale. De même, d'autres modèles créés pour la détection automatique des cancers de la peau ont montré l'importance d'inclure les informations cliniques du patient pour optimiser ces outils¹⁰³. Les algorithmes de ML sont considérés comme des "boîtes noires". L'explicabilité vise à mettre en évidence l'importance de

chaque donnée dans le processus de décision afin d'améliorer la transparence, le questionnement et la compréhension. Cette explicabilité augmente la confiance des modèles en assurant que la prise de décision est cohérente avec les données sémiologiques classiquement admises dans la littérature. De plus, l'identification de nouvelles corrélations doit être étayée par des preuves afin d'éliminer les décisions trompeuses. Enfin, la compréhension des modèles n'est pas seulement un objectif de recherche ; elle permet au praticien d'être guidé dans sa réflexion clinique, d'identifier les variables de grande importance et, par conséquent, d'optimiser le diagnostic des lésions. Par ailleurs, si l'explicabilité d'un modèle permet de mettre en évidence des interactions importantes entre certaines données, elle ne permet pas de conclure à l'existence d'un lien entre ces données et le modèle. Le développement d'applications accessibles au grand public à des fins de prévention et de diagnostic est un enjeu important¹⁰⁴. En 2018, la revue systématique de Chuchu et al⁶⁹ a évalué la sensibilité et la spécificité des applications smartphone utilisées pour le diagnostic du mélanome. Bien que prometteurs, les résultats ont montré un manque de précision qui a conduit à des erreurs de diagnostic par rapport au diagnostic posé par des dermatologues experts. Deux ans plus tard, Udréa et al. en 2020¹⁰⁵ ont amélioré le système en développant une application pour smartphone basée sur un algorithme de ML qui permet de détecter les lésions cutanées cancéreuses ou précancéreuses avec une sensibilité de 95,1 % et une spécificité de 77,3 %. Dans le diagnostic des lésions orales, en permettant la détection des "patients à haut risque de cancer" à partir des données sociodémographiques et comportementales du patient associées à la description élémentaire de la lésion, le présent algorithme peut ainsi constituer un outil simple à utiliser pour les spécialistes et les non-spécialistes, avec une précision élevée. Aujourd'hui, les modèles d'IA les plus puissants utilisés pour le diagnostic en dermatologie sont désormais basés sur des réseaux neuronaux convolutifs profonds. Ils permettent d'utiliser l'image de la lésion en considérant les pixels qui la constituent comme des caractéristiques à part entière¹⁰⁶⁻¹⁰⁸, voire en y associant

des informations sur le patient¹⁰⁹. L'utilisation de la photographie elle-même pour déterminer le diagnostic est la suite logique de notre travail, notamment pour affiner la prédiction. Sur la base des résultats de cette étude, notre algorithme pourrait être utilisé en pratique courante pour aider au diagnostic et effectuer un premier triage. En effet, il pourrait faciliter la détection des patients présentant des lésions à risque de développement malin. Bien entendu, il ne permet pas à ce jour de remplacer le spécialiste en raison de certaines limitations. Une anamnèse appropriée, suivie d'un examen permettant de déceler des signes spécifiques, complétés par des examens complémentaires ou une biopsie est la méthode standard de diagnostic en dermatologie. Cette approche holistique ne peut être remplacée par aucune méthode, y compris l'IA.

4- Conclusion

Cette première partie a donc mis en lumière deux approches d'IA pouvant être utilisées dans le diagnostic des lésions de la muqueuse orale. Se basant sur la littérature scientifique et les approches actuellement utilisées dans le domaine de la dermatologie cutanée, nous avons tout d'abord envisagé une approche par *Deep Learning*. Mais face à des résultats peu probants, nous avons secondairement envisagé l'utilisation de modèles statistiques. En effet, l'utilisation de modèles de ML en médecine est prometteuse, notamment en raison de leur capacité à calculer une très grande quantité de données pour faire des prédictions. Notre base de données présente certaines limites (taille de la base, standardisation des images), et il n'existe actuellement aucune base de données publique pour les lésions orales^{110,111}. Pourtant, les résultats de notre étude sont prometteurs car ils montrent une grande précision pour prédire les lésions cancéreuses (précision de 0,90) et une bonne précision globale de 0,84 pour cinq des autres lésions les plus fréquentes de la muqueuse orale (lichen plan, leucoplasie, maladies bulleuses, ulcérations aphtoïdes, hypertrophie gingivale). D'autres études avec des bases de données beaucoup plus importantes devraient être menées pour valider la précision de notre modèle. Cette alternative ne permet pourtant pas de se passer de la description de la lésion par un utilisateur et s'adresse donc à des praticiens ayant des connaissances minimales. Ainsi, l'optimisation de modèles d'apprentissage profond pourrait être particulièrement intéressante et devrait être étudiée pour prédire directement à partir de l'image de la lésion.

Des outils basés sur de tels algorithmes créent de nouvelles opportunités de démocratisation et d'amélioration des soins aux patients. En prenant une photo avec l'appareil photo de leur smartphone¹¹². Nous avons donc dans la suite de cette thèse transposé les recherches précédentes dans une applications mobile de santé d'aide au diagnostic des lésions de la muqueuse orale.

Deuxième partie : Développement d'un outil basé sur l'intelligence artificielle (IA) destiné aux professionnels de santé, ORAMEDAI.

I- Introduction

Depuis peu, les algorithmes d'IA ont été mis en œuvre dans des applications de santé mobiles grand public. On dénombre à ce jour, de très nombreuses applications de santé mobiles (*mHealth*) : Soins infirmiers et assistance virtuelle¹¹³⁻¹¹⁵, médecine de précision^{116,117}, amélioration du traitement des maladies, amélioration de la réduction des erreurs de diagnostic aide à la décision¹¹⁸, analyse des données des patients^{52,119}, robotique médicale¹²⁰, assistance au flux de travail administratif, amélioration de l'efficacité opérationnelle et de la rentabilité¹²¹ ou bien encore l'amélioration de la recherche biomédicale^{46,122}. Certaines applications permettent aux utilisateurs de recevoir instantanément une évaluation du risque d'une lésion cutanée.

Nous nous sommes intéressés aux applications destinées à l'aide au diagnostic et notamment dans le domaine de la dermatologie générale qui reste notre référence. Les applications mobiles peuvent faciliter la détection des cancers de la peau, car la population générale peut dépister une lésion cutanée n'importe où et à n'importe quel moment. En outre, un outil de diagnostic facile et rapide peut permettre de détecter des cancers de la peau tels que le mélanome à un stade plus précoce et avec un meilleur pronostic¹²³. Ces applications semblent être relativement bien perçues par la population¹²⁴. Plusieurs auteurs^{123,125,126} ont défini des barrières et des éléments facilitateurs concernant l'utilisation de telles applications.

Figure V-I : Barrières et éléments facilitateurs liés à l'utilisation d'application mobile diagnostique.



Il nous a semblé ainsi important de pouvoir imaginer le développement d'une application mobile destinée au diagnostic des lésions de la muqueuse orale. En réutilisant les travaux précédents et notre base de données une version « bêta » a été mise au point.

Toutefois et préalablement à ce développement, nous nous sommes questionnés quant à l'acceptation de l'IA en santé. Il s'agit d'un défi majeur conditionnant l'outil en lui-même.

II- Acceptations de l'Intelligence Artificielle et Défis relatifs

Dans le contexte plus général de la technologie, l'acceptation est définie comme la volonté, l'intention et la motivation interne d'utiliser une technologie en raison d'attitudes positives à l'égard de la technologie ou du système¹²⁷. Cette notion d'acceptation est différente de la notion d'acceptabilité qui fait, elle, référence à la compréhensibilité ou la transparence.

L'acceptation de la technologie par l'utilisateur est fondamentale pour l'adoption réussie des dispositifs¹²⁸. L'IA pouvant profiter à de nombreuses personnes, les utilisateurs doivent accepter cette technologie pour l'adopter et l'utiliser de manière adéquate. Une faible acceptation peut diminuer l'adoption de l'IA par les utilisateurs, entraînant une désaffectation des ressources, un excès de dispositifs d'IA et un déclin potentiel de l'innovation technologique au détriment des consommateurs. Plusieurs modèles ont été développés pour évaluer l'acceptation de l'IA par les utilisateurs, notamment le modèle d'acceptation de la technologie (TAM¹²⁸), la théorie unifiée de l'acceptation et de l'utilisation de la technologie (UTAUT¹²⁹) et, plus récemment, le modèle d'acceptation de l'utilisation de l'appareil d'IA (AIDUA^{130,131}.)

Le modèle TAM (*Technology acceptance model*) postule que des variables externes, telles que les médias et les références sociales, influencent l'utilité perçue et la facilité d'utilisation perçue, qui contribuent à l'intention d'utiliser la technologie et donc, à l'utilisation effective du système. L'utilité perçue est souvent la notion la plus importante dans l'acceptation de la technologie. Le modèle TAM est également souvent complété par des variables supplémentaires, telles que la confiance et la connaissance, afin d'améliorer son pouvoir prédictif. Ainsi, une évolution est le modèle TRAM (*Technology Readiness and Acceptance Model*) qui intègre la notion de préparation à la technologie dans le domaine de l'adoption d'innovations par les consommateurs¹³².

A défaut d'utilité perçue, la confiance dans l'IA et dans le fournisseur de technologie est un facteur déterminant de son acceptation¹³³.

La méthodologie UTAU est définie comme *l'Unified Theory of Acceptance of Use of Technology*. Elle permet d'évaluer l'adoption d'une nouvelle technologie¹²⁹. L'UTAUT identifie quatre facteurs clés (performance espérée, effort attendu, influence sociale et conditions facilitantes) et quatre modérateurs (âge, sexe, expérience et volontariat) liés à la prédiction de l'intention comportementale d'utiliser une technologie et de l'utilisation réelle de la technologie. Cette théorie a été adaptée aux systèmes utilisant l'IA¹³⁴, car certains points y sont spécifiques.

Certains modèles d'IA sous-jacents sont souvent considérés comme une « boîte noire ». L'utilisateur n'a que peu ou pas de visibilité sur l'algorithme ou le processus sous-jacent de prise de décision¹³⁵. Il est peu probable que les utilisateurs acceptent toujours cette situation, surtout s'ils doivent rendre compte des conséquences. Outre les doutes des cliniciens, le manque de confiance des patients à l'égard des technologies basées sur l'IA entrave également le développement. S'ils acceptent la réalité des erreurs humaines, ils ne tolèrent guère les erreurs des machines¹³⁶. Afin de pallier cette problématique, certains algorithmes spécifiques ont été développés pour tenter d'expliquer les prises de décisions (ex : *SHapley Additive exPlanations* (SHAP)¹³⁷). Les approches permettant d'ouvrir la « boîte noire » peuvent être différenciées en deux catégories : agnostiques et spécifiques à un modèle. L'analyse de sensibilité et les modèles locaux interprétables sont des exemples d'analyses spécifiques à un modèle. Dans les deux cas, il s'agit de modifier les données d'entrée et d'observer les effets qui en résultent, ce qui permet de mieux comprendre la pertinence des caractéristiques. L'amélioration de l'interprétabilité des algorithmes d'IA est une étape cruciale dans l'adoption de l'IA en médecine¹³⁸. En effet, l'explicabilité des modèles peut permettre d'anticiper une erreur diagnostique et y pallier le cas

échéant. Très récemment Lambert & al¹³⁹ ont étudié l'acceptation de l'IA en santé dans le monde hospitalier en se basant sur la méthodologie UTAU¹²⁹.

Selon Gursoy et al¹³⁰ les modèles traditionnels d'acceptation des technologies (TAM et UTAUT) ne devraient être utilisés que pour étudier les technologies non intelligentes, car leurs prédicteurs ne sont pas pertinents pour l'utilisation de l'IA. Ils ont donc développé un nouveau modèle : *AI Device Use Acceptance* (AIDUA). L'évaluation repose sur l'expérience des utilisateurs en trois étapes (c'est-à-dire l'évaluation primaire, l'évaluation secondaire et l'étape des résultats). Au stade de l'évaluation primaire, les consommateurs évaluent l'importance de l'utilisation de l'appareil d'IA en fonction de l'influence sociale, de la motivation hédonique (plaisir perçu) et de l'anthropomorphisme (qualités humaines que le dispositif peut reproduire). Sur la base de cette évaluation, les consommateurs détermineront ensuite les avantages et les coûts du dispositif d'IA en fonction de son espérance de performance perçue et de son espérance d'effort, ce qui entraînera le développement d'une émotion à l'égard de l'IA. Ces évaluations aboutissent à l'étape du résultat qui traduit la volonté des utilisateurs d'accepter la technologie ou de s'y opposer. A ce jour, aucune publication n'a utilisé ce modèle afin d'évaluer l'acceptation des outils d'IA dans le domaine de la santé.

Quel que soit le modèle d'acceptation, il ressort le fait que pour être accepté dans une utilisation en pratique courante, les performances de l'outil d'IA ont un rôle important à jouer. En effet, tant pour les patients que les praticiens ces performances doivent être satisfaisantes. Les algorithmes sont aujourd'hui de plus en plus performants. En dermatologie, nombreuses sont les études comparant l'IA et les praticiens dans des tâches diagnostiques qui montrent des résultats de classification diagnostique comparables entre l'IA et des experts humains. Cependant, plusieurs algorithmes souffrent d'une faible généralisation en raison des

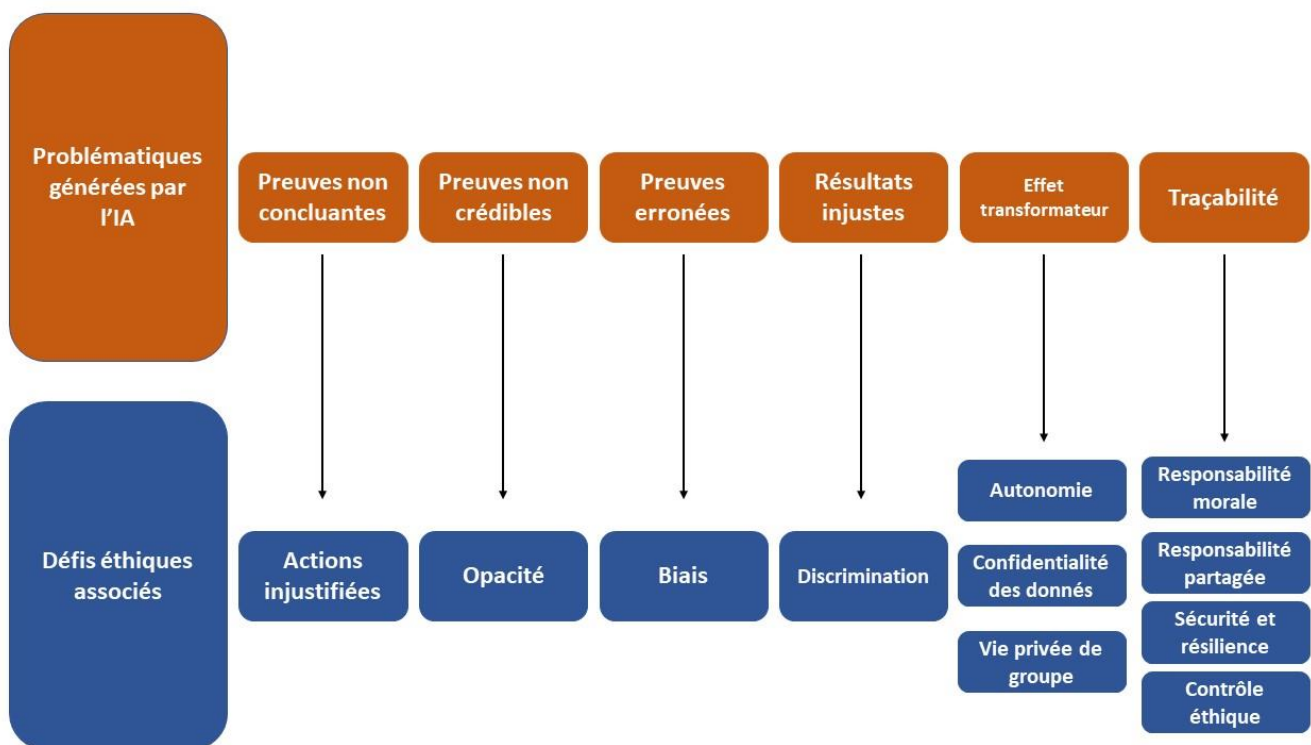
performances variables des modèles lorsqu'ils sont testés en dehors de ses conditions d'entraînement^{43,140}. Plus récemment encore, certains auteurs ont montré les limites de certains algorithmes de DL avec une diminution des performances dès lors que l'on modifie certains paramètres d'images (rotation/translation, balance de couleur)¹⁴¹. Cela conduit à des cas d'IA défectueuse, ce qui peut avoir un impact néfaste sur la confiance et la promesse que les chercheurs et les cliniciens ont envers l'IA dans le domaine de la dermatologie. Bien que les systèmes de classification basés sur l'IA ne puissent pas remplacer les experts humains, ils peuvent coopérer avec des experts et leur donner les moyens de faire des diagnostics cutanés précis^{70,142}. Hekler et al., en 2019⁷⁰ ont formé un modèle CNN en utilisant plus de 11 000 images dermoscopiques pour effectuer une classification multi classe de cinq affections cutanées et ont constaté que la précision moyenne combinée de l'IA et de l'humain était de 83 %, soit 1,4 % de plus que l'IA seule et 40 % de plus que les experts seuls.

Certains facteurs tels que le sexe, l'âge, l'expérience des utilisateurs sont considérés comme essentiels dans l'acceptation d'une nouvelle technologie. Mais qu'en est-il concernant l'utilisation de l'IA en santé ? Une étude récente¹⁴³, a exploré l'acceptation de l'IA durant la crise du coronavirus par les patients et concluent que la sensibilisation à la santé, les besoins d'empathie et les facteurs sociodémographiques des patients décrits dans le modèle TRAM, tels que l'âge, le sexe et le niveau d'éducation, ne sont pas des facteurs prédictifs significatifs de la volonté d'utiliser la technologie basée sur l'IA.

Du point de vue des cliniciens, le genre ne semble pas non plus avoir un impact sur l'acceptation¹⁴⁴⁻¹⁴⁶. Concernant l'âge, il apparaît pour certains auteurs¹⁴⁷ que les jeunes praticiens accepteraient plus facilement de nouveaux outils tandis que pour d'autres auteurs,^{144,146} l'âge n'a pas d'influence. Enfin, concernant l'expérience des praticiens, ici encore, les résultats diffèrent d'une étude à une autre^{121,145}.

L'acceptation de l'IA par les professionnels de santé mais plus largement par les gouvernements impose le respect de considérations éthiques afin de protéger les patients et respecter leurs droits¹⁴⁸. Or, Les algorithmes ne sont pas neutres d'un point de vue éthique¹⁴⁹. Mittelstadt & al¹⁴⁹ ont défini en 2016, six problématiques clés liées à l'utilisation des algorithmes. La commission des droits de l'Homme et de la biomédecine du Conseil Européen a repris ces problématiques en les liant aux problèmes éthiques pouvant être soulevés (**Figure VI-1**)¹⁵⁰.

Figure VI-I : Problématiques soulevées par l'IA et défis éthiques associés^{151,152}



l'IA fait donc face à de nombreux challenges¹⁴⁸ éthiques, tel que : la protection des données, le consentement éclairé, les disparités sociales ou bien encore la relation d'empathie liée à la consultation. La prise en compte de ces défis ou tout du moins leur connaissance a constitué un socle pour le développement de notre outil d'aide au diagnostic

III- Développement de l'outil

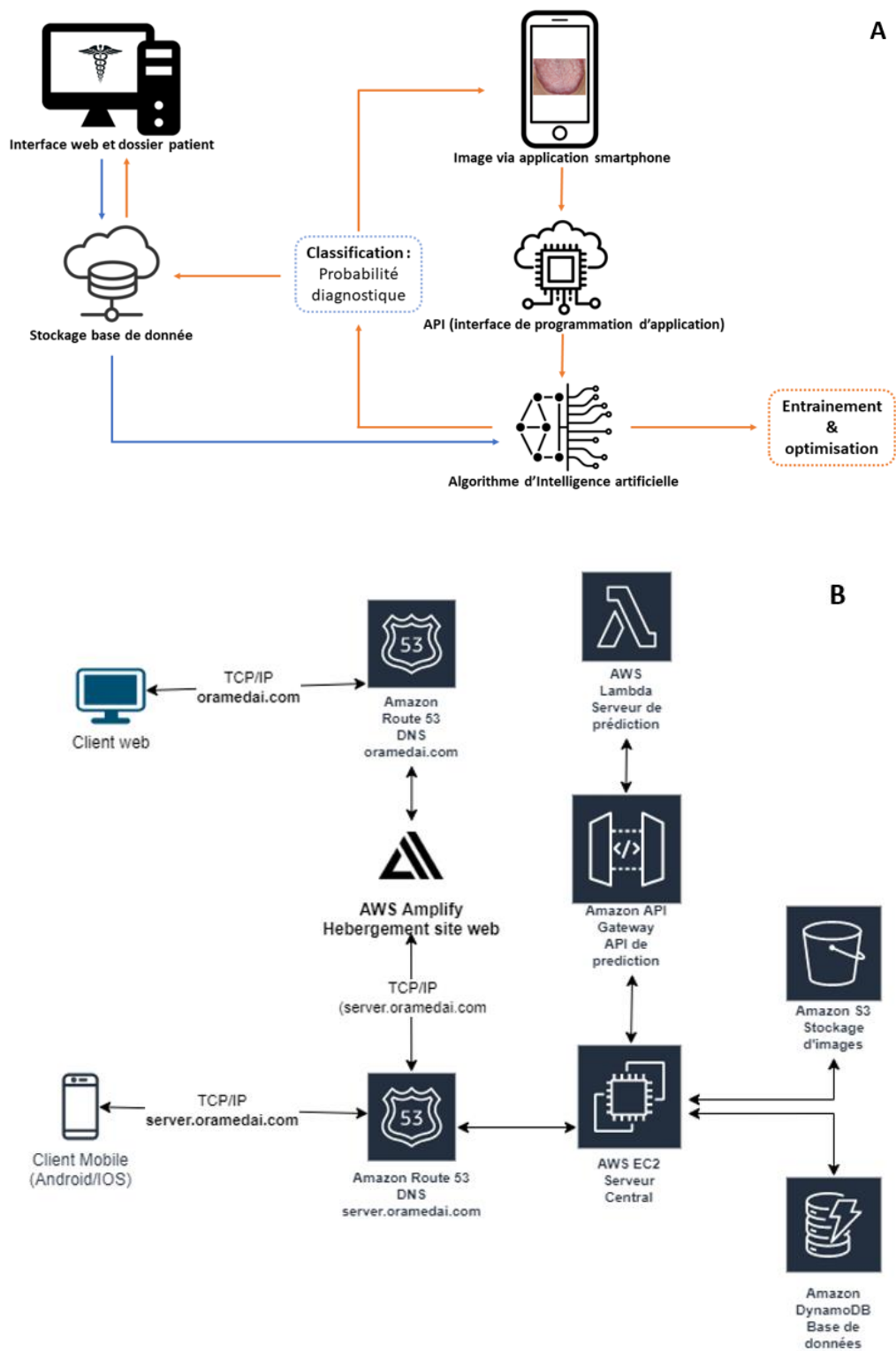


Figure V-II : Architecture de l'application ORAMEDAI (A) Objectif - (B) Finalisée

C'est avec l'aide de l'École Polytechnique de Montréal que cet outil nommé ORAMEDAI a été développé.

ORAMED.AI est une application mobile de santé ayant pour objectif principal l'aide au diagnostic en pathologie de la muqueuse orale. Elle n'est pas destinée au patient directement mais à des professionnels de santé non spécialistes en la matière (pharmaciens, médecins généralistes, dentistes omnipraticiens). Ce public ciblé nous a semblé indispensable car ayant des connaissances minima pour pouvoir répondre à certains volets de l'application et afin de pouvoir garder un certain contrôle sur les données enregistrées et diagnostics obtenus.

Notre application est donc composée d'une interface internet et d'une interface mobile pour smartphone. Ainsi les données ne sont enregistrées que sur le serveur, les smartphones ne servant que d'outil afin de prendre les photographies des lésions à diagnostiquer. La plateforme internet permet quant à elle de recueillir les données socio-démographiques et éléments sémiologiques en lien avec l'affection à diagnostiquer. ORAMED AI est un outil qui n'a pas vocation à donner un diagnostic de certitude mais plus raisonnablement à orienter les praticiens et éventuellement favoriser l'adressage des patients de façon non urgente, semi-urgente ou urgente vers les spécialistes adéquats.

Parallèlement à cet objectif d'aide au diagnostic, il existe un objectif secondaire qui est la constitution d'une base de données conséquente en vue de l'amélioration permanente des performances des modèles d'IA utilisés mais pouvant également servir de support en termes d'évaluation des besoins de santé public.

Enfin, un troisième volet de cette application non développé à ce jour aurait pour vocation l'éducation thérapeutique en fournissant des fiches cliniques et guides thérapeutiques à destinée du grand publique.

A- ORAMEDAI – Interface WEB

- Description du site internet

L'interface internet est destinée à compiler et gérer les dossiers patients. En effet, chaque prédiction diagnostique doit être rattachée à un dossier patient. Ce dossier patient est créé à partir de l'interface internet, il contient des informations socio-démographiques et médicales, telles qu'utilisées dans notre modèle de ML précédemment décrit. Le recueil de ces données est essentiel pour augmenter notre base de données et pouvoir optimiser l'algorithme. Par ailleurs, il n'est accessible qu'à des professionnels de santé qui doivent s'inscrire via leur numéro RPPS.

- Outils utilisés et environnement

Le site web <https://oramedai.com> est construit à l'aide du *framework* Next.js

- Structure du projet

Tout le code de développement est localisé dans le dossier source. On distingue :

actions : contient toutes les actions pour mettre à jour le contenu du stockage de session de l'utilisateur en utilisant *little-state-machine*.

assets : contient les images utilisées pour ce fichier README et le tutoriel utilisateur.

components : contient plusieurs composants d'interface utilisateur avec leurs styles qui sont utilisés dans le code.

functions : contient plusieurs fonctions réutilisables qui peuvent être importées et réutilisées dans le code.

helpers : contient des classes utiles au fonctionnement du site internet et au cryptage des données de stockage de la session qui peuvent être importées et réutilisées dans tout le code.

interfaces : contient les définitions des objets qui sont envoyés et reçus du serveur.

pages : Comme Nextjs utilise ce dossier pour effectuer le routage au lieu d'utiliser *react-router*,

services : contient les définitions des classes qui font des requêtes API (interfaces de programmation d'applications) au serveur en utilisant Axios.

styles : contient les feuilles de style pour tous les composants à l'intérieur de /pages.

- Déploiement

L'application est déployée sur Amazon Web Services (AWS) à l'aide d'Amplify. Amplify qui synchronise automatiquement ce Git avec le nom de domaine oramedai.com. Le processus de construction et de déploiement de l'application est géré par Amplify et aucune maintenance manuelle n'est nécessaire.

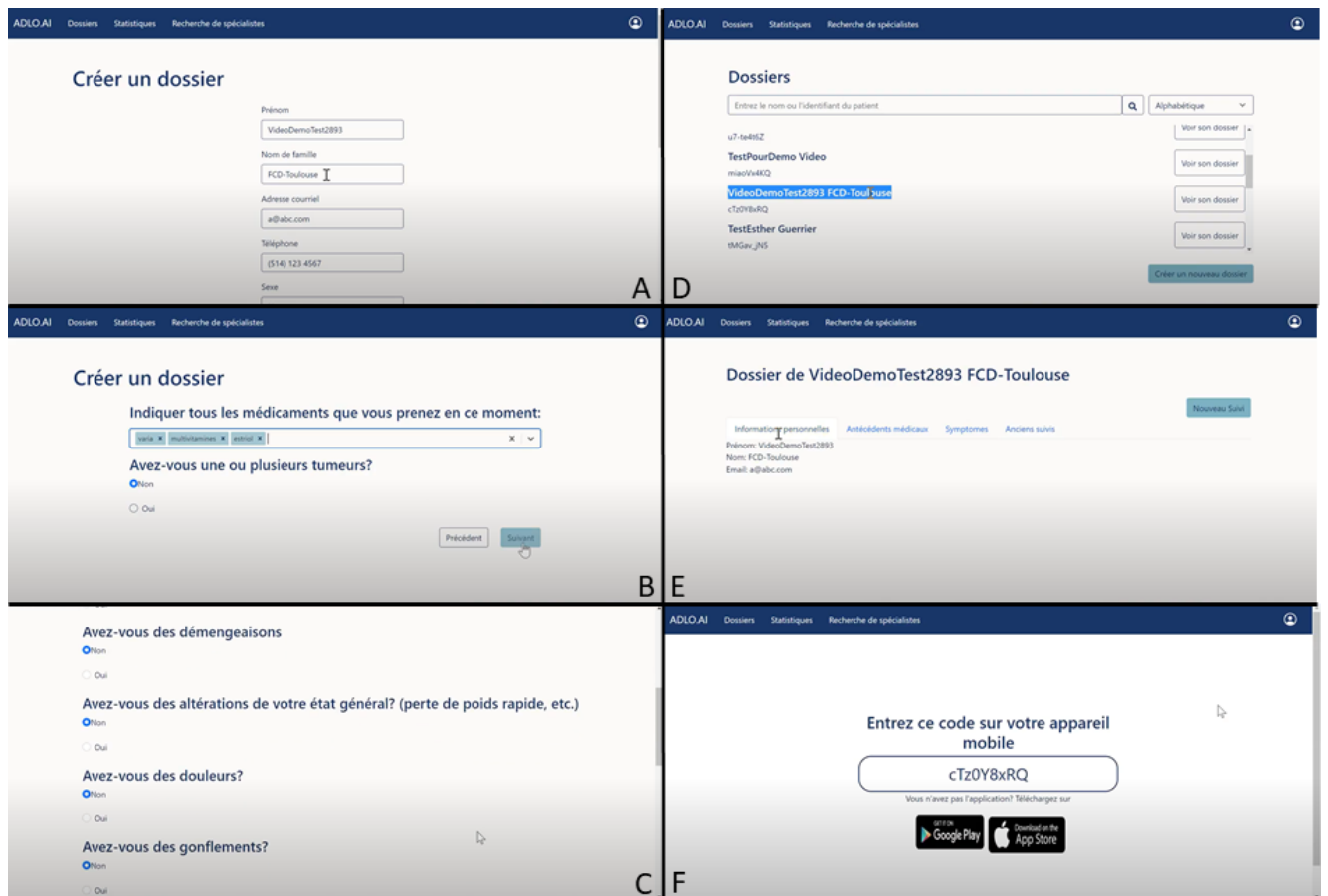


Figure V-IV : Interface ORAMED.AI, Différentes pages internet liées à la création d'un nouveau dossier patient. (A) Eléments socio-démographiques, (B) Recueil des traitements médicamenteux et antécédents médicaux, (C), Symptômes et éléments cliniques en rapport avec la lésion à diagnostiquer, (D) Récapitulatif des dossiers en cours, (E) Synthèse du dossier ouvert, (F) Lien permettant d'ajouter la photographie de la lésion à partir d'un smartphone (importation ou prise de photo).

B- ORAMD.AI – Développement de l'application mobile

- Description de l'application

L'intérêt de l'application est de pouvoir prendre une photo de la lésion directement à partir de son smartphone ou bien d'importer une image déjà présente sur son smartphone afin de l'envoyer vers la plateforme internet et l'algorithme de DL.

- Outils utilisés et environnement

L'application mobile multiplateforme d'ORAMED.AI est construite à l'aide de *React Native* et *Typescript*. L'environnement de développement repose sur un éditeur de code tel que Visual Studio Code, NodeJs installé, Android Studio pour le développement Android, et Xcode pour le développement iOS (uniquement disponible sur macOS).

- Structure du projet

Tout le code de développement est localisé dans le dossier source. On distingue :

components : Il s'agit des multiples composants réutilisables de l'interface utilisateur, tels que les boutons, qui sont utilisés dans l'ensemble du code.

hooks : Il s'agit des crochets avec état utilisés dans l'application.

navigation : il s'agit de tous les composants utilisés pour la navigation React Native.

screens : Voici les composants les plus importants. Tous les écrans de l'application mobile se trouvent dans ce dossier.

styles: Ce dossier contient toutes les conventions de style utilisées dans l'application, telles que les boutons, les couleurs, les formulaires, l'index, les contours, la taille et la typographie.

- Publication de l'application

Publier l'application sur l'App Store et le Google Play Store est une tâche délicate. L'application est actuellement déployée sur le *Google Play Store*. Concernant l'application *IOs.*, cette dernière n'a pu être testée qu'en version bêta. Son déploiement n'étant pas possible avec un compte de particulier car les données traitées sont considérées comme sensibles par Apple.

Google Play Store – Android

La version 1.0.0 de cette application mobile a été entièrement réalisée à partir des données de la base de données du CHU de Toulouse en collaboration avec un étudiant en master de l'équipe Polytechnique de Montréal et la faculté de Santé de l'université Paul Sabatier de Toulouse.

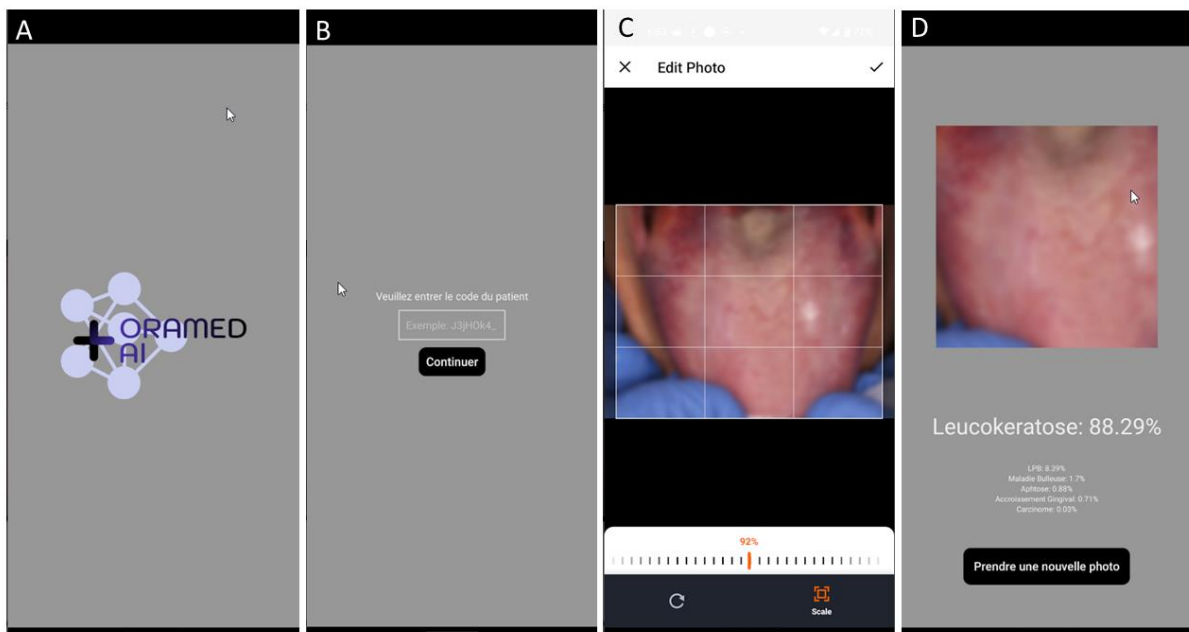


Figure V-V : Application ORAMED.AI (version Android)

Cette application est conçue de façon intuitive afin de généraliser son accessibilité. Elle est constituée d'un écran d'accueil (**Figure V-VA**). Dans un second temps, le numéro du dossier patient préalablement créé sur le site internet est demandé (**Figure V-VB**) afin d'associer l'image et le diagnostic au dossier clinique du patient. Ceci permet de ne stocker aucune donnée sensible sur le smartphone du praticien. Une fois le dossier correspondant chargé, il est possible de prendre une photo directement depuis son smartphone ou bien d'importer une photo déjà réalisée. A ce stade il est possible d'éditer la photo (Zoom, Longueur, centrage) (**Figure V-VC**). Enfin, une fois la photographie validée l'image est traitée en ligne par l'algorithme d'IA et une probabilité diagnostique est donnée au praticien (**Figure V-VD**).

C- ORAMED.AI – Utilisation du serveur

- Description

Le serveur coordonne les communications entre les clients (internet, mobile) et le reste des composants sur AWS tels que la base de données et l'API permettant la prédiction des classes d'images. Les interfaces API fournissent essentiellement un moyen de communication normalisé entre deux applications logicielles qui ne sont pas nécessairement du même type. Une API permet la communication entre une page internet ou une application et une application d'IA.

- Outils et Environnement

Il s'agit d'un serveur *NodeJS* basé sur le *framework* *express*

- Déploiement

Le serveur est déployé sur le nuage AWS sur une machine EC2, cette machine est liée au nom de domaine *oramedai.com* pour permettre des connexions sécurisées HTTPS.

- Connexion à la machine

La machine est liée à un Git, dans le cas où des changements sont effectués sur le serveur, ou que des mises à jour sont nécessaires voir à des fins de maintenance. Une connexion à la machine via SSH en utilisant une clé privée est possible.

D- ORAMED.AI – Développement du DL

- Description

L'objectif du modèle est de classer les images de lésions orales dans la catégorie de diagnostic appropriée.

Afin d'éliminer autant de bruit que possible et d'augmenter les performances, nous avons entraîné nos modèles sur des images recadrées des lésions orales. C'est-à-dire que les images de lésions orales ont été recadrées de manière à ne conserver que la partie de l'image qui contient la lésion.

- Outils utilisés et environnements

○ Google colab

Google colab permet à quiconque d'écrire et d'exécuter du code python [®] via le navigateur. Il est particulièrement adapté à l'apprentissage automatique en raison de sa facilité d'utilisation et des ressources qu'il fournit. Google colab met à la disposition de ses utilisateurs des ressources en processeur, en mémoire vive et même en GPU qui peuvent être utilisées pour le développement. Pour ce projet, nous avons utilisé le service d'abonnement Google colab pro afin d'accéder à du matériel plus puissant (notamment de meilleurs gpus).

○ Wandb

Lorsque l'on travaille avec plusieurs modèles simultanément, il devient rapidement évident qu'un outil est nécessaire pour garder une trace des métriques et des configurations des différents modèles développés. Nous avons choisi d'utiliser « *Weights and Biases* » en raison de sa popularité et de sa facilité d'utilisation.

- Tensorflow[®]

Tensorflow[®] est une bibliothèque populaire d'apprentissage automatique en python[®] qui facilite le développement de modèles d'apprentissage automatique.

- *Workflow*

Vous trouverez ci-dessous une vue d'ensemble du processus de développement de DL utilisé pour ce projet :

- Nettoyage de la base de données :

Une étape préliminaire très importante est le nettoyage des données. Dans notre cas, puisque nous travaillons avec des données d'images, nous avons parcouru manuellement la base de données et supprimé les images jugées inappropriées (images floues, mal cadrées, non pertinentes, etc.).

- Fractionnement de l'ensemble de données

Dans le cadre du développement de l'apprentissage automatique, il est courant de diviser l'ensemble de données en trois catégories distinctes : formation, validation et test. L'ensemble d'entraînement est utilisé (comme son nom l'indique) pour entraîner le modèle. L'ensemble de validation sert de référence pour sélectionner les meilleurs hyperparamètres du modèle et/ou les meilleures architectures de modèle. L'ensemble de validation est nécessaire pour éviter un surajustement de l'ensemble de test. Enfin, l'ensemble de test est généralement utilisé comme référence finale. Dans notre cas, en raison du manque de données, nous avons opté pour une validation croisée 5 fois au lieu d'un ensemble de validation afin d'obtenir une évaluation plus "robuste" des performances du modèle pour une configuration donnée.

- Augmentation des données

Lorsqu'il s'agit d'un petit ensemble de données, l'augmentation des données est une technique utilisée pour "augmenter" artificiellement la taille de l'ensemble d'apprentissage.

Essentiellement, des transformations telles que la rotation ou le retournement sont appliquées aux images de l'ensemble d'apprentissage afin de les modifier légèrement. L'apprentissage par transfert est un outil puissant utilisé en vision artificielle pour obtenir des performances élevées sans nécessairement avoir accès à une grande quantité de données ou de ressources informatiques. Essentiellement, l'apprentissage par transfert exploite les "connaissances" de modèles d'apprentissage automatique puissants et de grande taille et les utilise comme point de départ pour une tâche différente.

https://www.tensorflow.org/tutorials/images/transfer_learning?hl=fr

IV- Conclusion

L'utilisation d'applications mobiles de santé a récemment augmenté considérablement pour fournir des soins de santé indépendants en temps et en lieu. Entre 2014 et 2017, 235 nouvelles applications dans le domaine de la dermatologie ont été enregistrées, ce qui représente une croissance de 80,8 %¹⁵³. Outre leur fonction principale d'information ou de sensibilisation sur le mélanome, la prévention et les stratégies d'autosurveillance de la peau, les applications offrent de plus en plus de possibilités de diagnostic et de suivi des lésions cutanées pigmentées¹⁵⁴. Jahn & al¹⁵⁵ ont monté en 2022 une sur-détection de mélanome par une application smartphone contrairement au diagnostic fait par un dermatologue ou celui fait par l'IA sur des images dermoscopiques. Cela souligne l'intérêt que peut avoir une telle application pour sensibiliser les patients. Toutefois le diagnostic par un spécialiste reste de rigueur.

Dans le domaine de la dermatologie orale, cette version « bêta » constitue une preuve de concept, en effet elle nous a permis de valider une architecture d'application compatible avec son utilisation à l'échelle d'un service hospitalier. Toutefois, des optimisations sont à envisager avant une éventuelle généralisation. En effet, les probabilités diagnostiques reposent sur un algorithme de DL qui, comme nous l'avons montré dans notre première partie s'avère peu performant. Ainsi, une alternative pourrait être de le combiner avec notre second algorithme de ML. Notons également que seuls 6 diagnostics ont été définis à ce jour, or les pathologies de la muqueuse orale sont des affections nombreuses. Il s'agira donc de prendre en compte bien d'autres diagnostics pour en faire une application plus complète.

D'autre part, cette application peut permettre également de recueillir un grand nombre de photos et de données socio-démographiques associées. En effet, cela est un point important pour étoffer notre base de données et donc entraîner et améliorer perpétuellement les algorithmes.

Dans notre spécialité, la dermatologie orale, alors que les données sont plus limitées (faible nombre mais aussi images peu standardisées) la comparaison des performances de prédiction de l'IA versus des spécialistes semble être comparable^{156,157,93}. Bien que le diagnostic direct soit un objectif final de l'IA, l'étape consistant à s'assurer que l'utilisateur confirme une lésion primaire correctement identifiée présente deux avantages distincts : Un contrôle de leur qualité et l'éducation simultanée du non-spécialiste à l'examen clinique. Ces avantages sont synonymes de transparence pour les patients et garantissent que la décision clinique finale revient à un médecin et non à un modèle¹⁵⁸.

Discussion et Perspectives

Discussion

Nous avons au cours de cette thèse adapté différents algorithmes à notre domaine spécifique qu'est la muqueuse orale pour enfin envisager leur intégration dans une application mobile de santé.

Nous devons tenir compte du fait qu'un pourcentage élevé de patients n'aura pas accès à des soins spécialisés, ce qui peut entraîner une évolution défavorable de la maladie¹⁵⁹. De nombreuses études ont mis en évidence l'utilité de la télédermatologie ainsi que des applications automatisées pour smartphone¹⁶⁰. Dans une étude récente, les patients ont été invités à utiliser une application téléphonique pour envoyer des photos de leurs lésions cutanées suspectes. Environ 70 % des participants ont déclaré qu'ils n'auraient pas consulté un dermatologue sans le programme, ce qui indique que la télédermatologie pourrait contribuer à sensibiliser les patients¹⁶¹.

Dans notre première partie, nous sommes basés sur les modèles d'IA classiquement utilisés en dermatologie cutanée. Les modèles d'apprentissage automatique sont de plus en plus appliqués à la dermatologie en tant qu'outil d'aide à la décision diagnostique utilisant l'analyse d'images, en particulier pour la détection du cancer de la peau et la classification des maladies. Dans certains cas, les performances sont comparables à celles des médecins spécialistes. Dans leur revue systématique, Ping Choi & al¹⁶² ont fourni une évaluation complète des études d'analyse d'images DL sur les maladies de la peau (l'acné, le psoriasis, l'eczéma, la rosacée, le vitiligo et l'urticaire). La précision diagnostique rapportée des algorithmes de DL est globalement encourageante pour les maladies inflammatoires courantes de la peau, contrairement aux infections cutanées, pour lesquelles la précision diagnostique des algorithmes de DL

validés/testés en externe est généralement faible. Bien que les algorithmes de diagnostic de DL soient pour la plupart spécifiques, leur sensibilité varie. Elle est notamment plus faible dans les études moins nombreuses, mais plus robustes, qui ont procédé à une validation ou à un test externe. Les algorithmes d'apprentissage profond avaient une spécificité élevée et une sensibilité variable pour diagnostiquer ces maladies. La précision des algorithmes dans le diagnostic est respectivement de 94%, 94%, 93% et 89% pour l'acné, la rosacée, l'eczéma et le psoriasis. Concernant le versant le plus étudié en dermatologie, la détection des cancers cutanés, une récente méta analyse¹⁶³ retrouve une sensibilité, une spécificité et une aire sous la courbe respectivement de 0,78, 0,98 et 0,97.

Nous avons appliqué à notre recherche des méthodologies similaires et pourtant n'avons pas obtenus les résultats espérés. Nous avons fait face à la complexité d'entraînement des réseaux profonds. Ceci est à mon sens, essentiellement lié à notre base de données qui est limitée en termes de quantité d'images mais également peu homogène dans la distribution des différents diagnostics. La taille de la base de données utilisée pour former et évaluer un modèle d'IA peut avoir un impact significatif sur ses performances et sa généralisation. Les modèles d'IA formés sur de petits ensembles de données peuvent ne pas être aussi performants lorsqu'ils sont appliqués à de nouvelles données, ce qui limite leur utilité dans les environnements cliniques du monde réel. En outre, la qualité, la disponibilité et la cohérence des données utilisées dans ces études posent des problèmes importants. En effet, il n'existe pas à ce jour de banque d'images cliniques de lésions orales accessibles pour l'entraînement. Pour autant ces bases de données commencent à se développer dans le domaine concernant la muqueuse orale et l'on commence à retrouver des banques d'images histopathologiques, permettant des taux de détection de cancers excellents.¹⁶⁴ Les défis ne sont cependant pas tout à fait les mêmes avec des images cliniques car il existe une grande hétérogénéité des atteintes tant entre les lésions elles-mêmes que pour une même entité clinique. La littérature concernant l'IA et la pathologie

orale s'intéresse aussi principalement aux cancers de la cavité orale et ne considère que très peu les autres affections muqueuses¹⁶⁵. Ceci s'expliquant très certainement par le risque vital de telles affections. Une des principales différences avec la dermatologie cutanée repose sur la qualité des images qui est bien meilleure et inclue une certaine standardisation. La dermoscopie est une technique d'imagerie non invasive qui utilise la lumière polarisée et non polarisée pour améliorer la sensibilité et la spécificité du diagnostic du cancer de la peau¹⁶⁶. De plus, la plupart des dermatoscopes disponibles dans le commerce ont un grossissement standardisé de 10×, tout en préservant les identifiants des patients, ce qui en fait un cas test idéal pour l'apprentissage de l'intelligence artificielle. En conséquence, une grande partie de la recherche en IA s'est concentrée sur les images dermoscopiques pour la détection précoce du cancer de la peau¹⁶⁷.

Et pourtant, malgré cela, il existe de nombreux facteurs de confusion entraînant des mauvaises classifications par les algorithmes, tels que les bulles d'air, les poils, les maladies cutanées de fond, la peau endommagée par le soleil et les sites anatomiques particuliers qui ne peuvent pas être éliminés par les humains¹⁶⁸. Enfin et de la même façon que pour la muqueuse orale, les lésions bénignes d'apparence banale sont le plus souvent sous-représentées ou absentes des ensembles d'entraînement des études, ce qui cause une sous-performance des algorithmes¹⁶⁹. L'inclusion de lésions bénignes typiques évite le biais de vérification et élimine donc ces limitations, comme le montre une étude de Tschandl et al¹⁷⁰.

Face aux difficultés de diagnostic directement à partir de l'image, la suite de notre travail a donc été de détourner la problématique. Pour cela nous avons considéré non plus les images en elles-mêmes mais leurs caractéristiques sémiologiques en y associant également certaines données socio-démographiques. En effet, à ce jour et dans notre domaine d'intérêt, les technologies d'IA ont été principalement appliquées pour différencier les conditions normales, précancéreuses et malignes, prédire la probabilité d'incidence du cancer de la bouche¹⁷¹, le

pronostic, la détection précoce des lésions précancéreuses et cancéreuses, prédire le risque de récurrence, prédire la possibilité de développement de la maladie à partir d'une lésion maligne potentielle, et prédire la survie des patients¹⁷².

Ces nouvelles données ont permis une prédiction via des modèles de ML statistiques qui se sont avérés bien plus performants. Toutefois alors que cette approche est entièrement satisfaisante elle implique l'intervention de l'Homme et des connaissances soles en matière de description de lésions dermatologiques. Ceci n'est en réalité pas un frein et permet de garder un contrôle accru sur la prédiction. Pour autant, la prédiction directement à partir de l'image est tout particulièrement intéressante du fait du manque de professionnels de santé capables d'apporter ces connaissances de bases. Les recherches futures pourraient porter sur le développement d'algorithmes plus sophistiqués pour l'interprétation des images, ainsi que sur l'amélioration des techniques d'acquisition et de traitement des images afin de garantir des données d'entrée de haute qualité pour les systèmes d'intelligence artificielle. Par ailleurs, les systèmes d'IA fonctionnent souvent comme des "boîtes noires", de sorte qu'il est difficile pour les cliniciens de comprendre comment ils parviennent à un résultat particulier. Cela peut être problématique dans un contexte clinique où il est crucial de comprendre le raisonnement qui sous-tend un diagnostic ou un pronostic. Des efforts devraient être faits pour développer des modèles d'IA interprétables, ou au moins pour fournir une certaine forme d'information sur la prise de décision. Si un système d'IA commet une erreur entraînant un préjudice, la responsabilité n'est pas clairement établie : le clinicien, les développeurs du système d'IA ou l'institution qui l'a mis en œuvre. Des lignes directrices et des réglementations claires sont nécessaires pour résoudre ces problèmes.

Une fois un modèle optimisé afin d'obtenir des résultats satisfaisants en termes de sensibilité et spécificité, celui-ci pourra être inclus dans notre application mobile de santé. En effet, à ce stade il s'agit plus d'une preuve de concept que d'une application prête au lancement sur le marché du soin. A ce jour, cette application constitue un socle pour l'aide au diagnostic d'une part mais également pour le recueil de données cliniques et sémiologiques d'autre part. Ce dernier point étant indispensable à l'optimisation de futurs algorithmes.

L'intégration d'algorithmes CNN dans les applications pour smartphones permettra de disposer d'un système automatisé et fiable qui classera les lésions du patient et fournira des instructions en conséquence. La télédermatologie orale est donc susceptible de contribuer à l'amélioration du dépistage et à la sensibilisation des patients, ainsi qu'à la réduction de la charge de travail des dermatologues et à l'octroi de l'attention nécessaire aux patients à haut risque.

Perspectives

Concernant l'amélioration des performances de prédiction, deux voies de recherches sont à développer : La constitution d'une base de données mutualisées et standardisées ainsi que l'application à notre problématique de muqueuse orale d'approches algorithmiques différentes.

Une des principales limites de ce travail est en lien avec le manque de données disponibles. Il n'existe en effet aucune banque d'images de lésions orales comparables aux données disponibles en dermatologie cutanée. La constitution d'une base de données spécifique à notre problématique pourrait être une solution à l'entraînement des algorithmes classiquement utilisés en dermatologie. Un élément essentiel pour garantir la sécurité algorithmique est de garantir que les ensembles de données soient suffisamment diversifiés et représentatifs de la population à laquelle ils sont destinés¹⁷³.

Ainsi il s'agira de développer un outil permettant :

- un stockage de nombreuses images
- l'association avec diagnostic validé histologiquement,
- une annotation standardisée de l'image,
- le recueil de données socio-démographiques.

Afin d'éviter les biais liés à la constitution d'une base de données, il existe un besoin de transparence dans la préparation des ensembles de données, notamment en ce qui concerne les personnes incluses ou exclues, le traitement des données manquantes et la manière dont les données sont étiquetées¹⁷⁴. L'importance de la collecte d'informations sur les données socio-démographiques ne réside pas nécessairement dans le fait qu'elles soient intégrées en tant que

variables d'entrée ou prédicteurs dans les modèles, mais surtout dans le fait qu'elles soient collectées pour vérifier les performances des modèles au sein de sous-groupes désagrégés, en tant que méthode de découverte des biais^{175,176}. Par ailleurs une autre limitation repose sur la collecte d'images de différentes résolutions issues de smartphones ou divers appareils photos numériques. Cette problématique sera difficile à solutionner. L'utilisation d'un protocole standardisé pour une utilisation de qualité de l'imagerie est essentielle. Digital Imaging and Communications in Medicine (DICOM) est la norme internationale pour l'imagerie médicale. Elle définit les formats des images médicales qui peuvent être échangées avec les données et la qualité nécessaires à l'utilisation clinique. Bien que l'utilisation de DICOM soit devenue la méthode standard de traitement des images dans d'autres spécialités telles que la radiologie et la cardiologie, des améliorations sont encore possibles en dermatologie. Les images sont collectées par des méthodes non standardisées via des smartphones et des appareils photos sans la capacité d'inclure du matériel supplémentaire. Des objets tels que des images redimensionnées ou sous-échantillonnées, des images de segmentation et le résultat de la classification des lésions par l'algorithme, ainsi que des métadonnées, peuvent être joints à un fichier DICOM¹⁷⁷.

Aussi, la poursuite de ce travail pourrait être l'utilisation d'un autre type de modèle d'IA : Le *cartesian genetic programming (CGP)*, décrit dans les années 2000 par Miller & al¹⁷⁸.

Il s'agit d'une technique d'optimisation et de recherche automatique basée sur l'idée de l'évolution darwinienne et utilisée pour résoudre des problèmes complexes, notamment dans les domaines de l'ingénierie, de l'informatique et de l'apprentissage automatique. Cortacero & al¹⁷⁹ ont tout récemment décrit son utilité dans le traitement d'images biomédicales et mettent en avant l'intérêt de nouvelles approches plus explicables dites « boîtes blanches » (« *White boxes* »). Leur méthode nommée Kartesio est entraînable sur de petits ensembles de données, transparente et interprétable par les humains, et capable d'intégrer les connaissances humaines.

Le CGP est une méthode de représentation, d'évolution et d'optimisation de programmes informatiques sous forme de graphes acycliques dirigés dont voici les principales caractéristiques :

Représentation sous forme de graphe : Les programmes sont représentés sous forme de graphes, où les nœuds du graphe représentent des opérations ou des fonctions, et les arêtes définissent les connexions entre ces nœuds. Cette représentation facilite la modularité et la réutilisation des composants.

Évolution génétique : Le CGP utilise des techniques d'évolution génétique pour optimiser les graphes représentant les programmes. Cela signifie qu'une population de graphes est générée, évaluée en fonction de leurs performances dans la résolution du problème donné, puis soumise à des opérations d'évolution telles que la mutation et la recombinaison pour créer de nouvelles générations de graphes.

Flexibilité : Le CGP est flexible en termes de types de programmes qu'il peut représenter. Il peut être utilisé pour évaluer des programmes simples ou complexes, y compris des algorithmes, des fonctions mathématiques, des circuits électroniques, etc.

Optimisation automatique : L'objectif du CGP est généralement de trouver un programme qui maximise ou minimise une fonction objective donnée. Cette fonction objective est déterminée par le problème spécifique que l'on souhaite résoudre.

Paramètres ajustables : Le CGP offre de nombreux paramètres ajustables, tels que la taille des graphes, les opérations de mutation, les opérations de recombinaison, etc. Ces paramètres peuvent être adaptés au problème particulier que l'on cherche à résoudre.

C'est une technique puissante pour explorer de vastes espaces de recherche et trouver des solutions innovantes à des problèmes complexes. L'approche Kartésio ajoute au CGP un volet de traitement d'image basé sur la conception évolutive de pipelines assemblés à partir de filtres préexistants conçus par l'Homme. Ceci permet d'extraire des informations quantitatives des images numériques.

Face aux problématiques rencontrées lors de l'utilisation du DL dans notre travail et notamment en lien avec le polymorphisme des lésions ou l'absence de standardisation, cette nouvelle approche pourrait avoir un réel intérêt. En effet, Kartezio nécessite un ensemble de données d'apprentissage minimales par comparaison au DL tout en générant des pipelines d'une précision comparable¹⁷⁹. Par ailleurs, cette méthode est explicite et rend le processus d'extraction d'informations à partir d'images médicales transparent et interprétable afin de favoriser la confiance, d'assurer le respect des normes éthiques professionnelles.

Concernant notre application de santé, il semble également important d'aborder l'apport d'une telle application dans l'éducation tant des professionnels de santé que des patients. Un module

de recommandations et de conseils thérapeutiques et préventifs pourra être ajouté afin de participer aux efforts en matière de santé publique. De nombreuses applications ont des objectifs éducatifs et préventifs pour les patients¹⁸⁰. Masters & al¹⁸¹ ont même souligné l'intérêt d'enseigner le développement de telles applications aux étudiants en santé afin de favoriser le partage d'informations médicales de haute qualité. L'enseignement ne concerne pas que les patients, et notre application pourrait au travers d'un cheminement en sens inverse constituer une très grande base de données de cas cliniques pour les étudiants en santé. Ainsi il est possible d'imaginer un module d'autoformation qui consisterait en des « quizz diagnostiques ». Ces derniers sont importants pour solliciter la démarche diagnostique mais également confronter les étudiants à des pathologies parfois très polymorphes mais aussi parfois rarement rencontrées pendant leurs études médicales.

Conclusion

Pour finir et de façon plus générale, tout au long de mon travail de thèse de nombreuses interrogations se sont posées sur la place de l'IA dans la société et dans le soin. L'IA et les technologies connexes peuvent être des auxiliaires très utiles pour les professionnels de santé. Elles ne doivent pas être considérées comme un substitut au personnel médical, mais plus comme des outils d'optimisation. L'IA surpasse l'Homme sur deux aspects importants : la connectivité et la possibilité de mise à jour. En raison de sa nature transformatrice, le secteur des soins de santé est particulièrement concerné par le potentiel des applications de l'IA. Les applications ont le potentiel de changer non seulement les processus de traitement et de diagnostic, mais aussi les modes de vie des patients et des praticiens.

Il faut cependant garder à l'esprit que ces technologies par leur bouleversement des pratiques courantes peuvent également entraîner des problématiques éthiques majeures. En effet, l'application de l'IA en santé pose des exigences plus élevées en raison de la réglementation et de son impact éventuel. Jusqu'à présent, l'IA a été introduite dans le secteur médical dans le but de proposer des solutions efficaces et axées sur la clinique tant pour les patients que les professionnels de santé. Dans ce domaine, l'IA est développée pour aider les professionnels à prendre des décisions, à établir des diagnostics, des pronostics et des traitements, et à se soulager des tâches physiquement exigeantes¹⁸². Les questions éthiques, l'absence de normalisation et le manque de clarté en matière de responsabilité juridique sont autant de défis auxquels est confrontée la généralisation de l'IA dans le secteur de la santé aujourd'hui¹⁸³. Les êtres humains sont connus pour résister au changement, accepter le changement n'est donc pas un processus simple. Pourtant l'acceptation est une étape indispensable au développement des outils basés sur l'IA

L'exigence d'un consentement éclairé récurrent facilité par la technologie pour les nouvelles utilisations de données, lorsque cela est possible, contribuerait à respecter la vie privée et l'autonomie des patients. De même, le droit de retirer des données pourrait être clairement communiqué et surtout facile à exercer. Dans la mesure du possible, des données génératives pourraient être utilisées pour combler les lacunes en matière de données créées par ces retraits afin d'éviter de rendre les systèmes d'IA inopérants¹⁸⁴.

Les nouvelles technologies telles que l'intelligence artificielle, l'apprentissage automatique, la robotique, le big data et les réseaux devraient révolutionner les processus de production, mais elles pourraient également avoir un impact majeur sur les économies en développement. Les nouvelles technologies risquent de creuser le fossé entre les pays riches et les pays pauvres en déplaçant davantage d'investissements vers les économies avancées où l'automatisation est déjà bien établie. Cela pourrait à son tour avoir des conséquences négatives sur l'emploi¹⁸⁵. La montée en puissance des robots chirurgicaux et des robots infirmiers dans le secteur des soins de santé, menace leurs futures perspectives d'emploi.

D'un point de vue moins global, les formes dominantes d'IA ou les algorithmes de ML d'aujourd'hui, sont entraînés à l'aide d'ensembles de données qui reflètent les jugements, les priorités et les catégories conceptuelles de l'Homme. Lorsque ces données sont biaisées ou, présentent une inégalité, cette dernière peut être codée et reproduite dans les algorithmes basés sur l'apprentissage automatique^{186,187}. Mais lorsque la vérité est que la société est structurellement injuste et inégale, et que les technologies font partie de ces structures, la question est de savoir si nos algorithmes doivent reproduire fidèlement l'inégalité ou œuvrer pour la changer.¹⁸⁸

Enfin, Les soins centrés sur la personne sont considérés comme la norme d'excellence pour les relations médecin-patient¹⁸⁹. Les nouveaux outils d'IA vont ainsi prendre en compte les données propres de chaque patient et pourraient répondre à cette approche centrée sur le patient. Pourtant l'empathie reste un prérequis indispensable à l'établissement de cette relation patient-praticien¹⁹⁰. Le risque de déshumanisation est donc à connaître. D'un autre côté, il est possible de considérer que l'apport de tels outils peut simplifier les tâches, libérer du temps aux praticiens et donc favoriser les relations humaines. Un des points clés afin de préserver l'autonomie du patient est de l'impliquer dans la prise de décision¹⁹¹.

Il est donc possible d'intégrer des outils d'IA dans la relation médecin-patient mais cela implique nécessairement un accord de confiance tant du patient que du médecin envers cette nouvelle entité. Or, cette confiance sera liée à la transparence des systèmes d'IA. L'explicabilité de tels systèmes permettra tant au praticien, qu'au patient, de comprendre quels facteurs ont joué un rôle dans la prise de décision partagée et permettra de renforcer la relation tripartite : Médecin - Outil d'IA - Patient.

Il semblerait donc à ce stade que tant les patients¹⁹² que les praticiens¹⁹³ considèrent positivement l'IA comme des outils d'aide ne devant pas totalement remplacer les médecins. Cette hypothèse est d'autant plus vraie lorsque les patients nécessitent des qualités humaines plus prononcées comme c'est le cas dans les pathologies chroniques ou en fin de vie. Ceci, suggérant une utilisation de l'IA dépendante du type de maladie et du niveau d'empathie requis^{194,195}. Il est donc aujourd'hui nécessaire de réfléchir à la place et aux rôles des outils d'IA développés en santé afin de s'assurer que leur déploiement reste bénéfique à la relation médecin-patient.

Références

1. Sampaio-Maia B, Caldas IM, Pereira ML, Pérez-Mongiovi D, Araujo R. Chapter Four - The Oral Microbiome in Health and Its Implication in Oral and Systemic Diseases. In: Sariaslani S, Michael Gadd G, eds. *Advances in Applied Microbiology*. Vol 97. Academic Press; 2016:171-210. doi:10.1016/bs.aambs.2016.08.002
2. Spanemberg JC, Cardoso JA, Slob EMGB, López-López J. Quality of life related to oral health and its impact in adults. *J Stomatol Oral Maxillofac Surg*. 2019;120(3):234-239. doi:10.1016/j.jormas.2019.02.004
3. Massereau E, Laurans R, Foletti JM, Gallucci A, Chossegras C. [Epidemiology of oral diseases in a university hospital specialized consultation]. *Rev Stomatol Chir Maxillo-Faciale Chir Orale*. 2016;117(3):136-141. doi:10.1016/j.revsto.2016.01.013
4. Cancer Today. Accessed April 22, 2024. <https://gco.iarc.who.int/today/>
5. Diz P, Meleti M, Diniz-Freitas M, et al. Oral and pharyngeal cancer in Europe: Incidence, mortality and trends as presented to the Global Oral Cancer Forum. *Transl Res Oral Oncol*. 2017;2:2057178X17701517. doi:10.1177/2057178X17701517
6. Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and aetiology - Johnson - 2011 - Periodontology 2000 - Wiley Online Library. Accessed July 29, 2021. <https://onlinelibrary.wiley.com/doi/10.1111/j.1600-0757.2011.00401.x>
7. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394-424. doi:10.3322/caac.21492
8. Mignogna MD, Fedele S, Lo Russo L. The World Cancer Report and the burden of oral cancer. *Eur J Cancer Prev Off J Eur Cancer Prev Organ ECP*. 2004;13(2):139-142. doi:10.1097/00008469-200404000-00008
9. Villanueva-Vilchis M del C, López-Ríos P, García IM, Gaitán-Cepeda LA. Impact of oral mucosa lesions on the quality of life related to oral health. An etiopathogenic study. *Med Oral Patol Oral Cirurgia Bucal*. 2016;21(2):e178-184. doi:10.4317/medoral.20866
10. Warnakulasuriya S, Kerr AR. Oral Cancer Screening: Past, Present, and Future. *J Dent Res*. 2021;100(12):1313-1320. doi:10.1177/00220345211014795
11. Dhull AK, Atri R, Dhankhar R, Chauhan AK, Kaushal V. Major Risk Factors in Head and Neck Cancer: A Retrospective Analysis of 12-Year Experiences. *World J Oncol*. 2018;9(3):80-84. doi:10.14740/wjon1104w
12. Lloyd JS, Abrahamson S. Effectiveness of Continuing Medical Education: A Review of the Evidence. *Eval Health Prof*. 1979;2(3):251-280. doi:10.1177/016327877900200301
13. Jenkins B, Lester K, Noble A, Such H, Yawn B, Scott A. Evaluating the Impact of Continuing Medical Education in the Interdisciplinary Team: A Novel, Targeted Approach. *J CME*. 2023;12(1):2161730. doi:10.1080/28338073.2022.2161730

14. Clovis JB, Horowitz AM, Poel DH. Oral and pharyngeal cancer: knowledge and opinions of dentists in British Columbia and Nova Scotia. *J Can Dent Assoc.* 2002;68(7):415-420.
15. Seoane J, Warnakulasuriya S, Varela-Centelles P, Esparza G, Dios PD. Oral cancer: experiences and diagnostic abilities elicited by dentists in North-western Spain. *Oral Dis.* 2006;12(5):487-492. doi:10.1111/j.1601-0825.2005.01225.x
16. Sardella A, Demarosi F, Lodi G, Canegallo L, Rimondini L, Carrassi A. Accuracy of referrals to a specialist oral medicine unit by general medical and dental practitioners and the educational implications. *J Dent Educ.* 2007;71(4):487-491.
17. Ergun S, Ozel S, Koray M, Kürklü E, Ak G, Tanyeri H. Dentists' knowledge and opinions about oral mucosal lesions. *Int J Oral Maxillofac Surg.* 2009;38(12):1283-1288. doi:10.1016/j.ijom.2009.07.004
18. Lima AM, Meira IA, Soares MS, Bonan PR, Mélo CB, Piagge CS. Delay in diagnosis of oral cancer: a systematic review. *Med Oral Patol Oral Cirurgia Bucal.* 2021;26(6):e815-e824. doi:10.4317/medoral.24808
19. Roume M, Azogui-Levy S, Lescaille G, Descroix V, Rochefort J. Knowledge and practices of dentists in France regarding oral mucosal diseases: a national survey. *J Oral Med Oral Surg.* 2019;25(1):10. doi:10.1051/mbcb/2018036
20. McCorduck P. *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence.* AK Peters Ltd; 2004. Accessed May 10, 2021. <https://www.routledge.com/Machines-Who-Think-A-Personal-Inquiry-into-the-History-and-Prospects-of/McCorduck/p/book/9781568812052>
21. Olsen TG, Jackson BH, Feeser TA, et al. Diagnostic Performance of Deep Learning Algorithms Applied to Three Common Diagnoses in Dermatopathology. *J Pathol Inform.* 2018;9:32. doi:10.4103/jpi.jpi_31_18
22. Hogarty DT, Su JC, Phan K, et al. Artificial Intelligence in Dermatology—Where We Are and the Way to the Future: A Review. *Am J Clin Dermatol.* 2020;21(1):41-47. doi:10.1007/s40257-019-00462-6
23. Zahia S, Garcia Zapirain MB, Sevillano X, González A, Kim PJ, Elmaghraby A. Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods. *Artif Intell Med.* 2020;102:101742. doi:10.1016/j.artmed.2019.101742
24. Day GR, Barbour RH. Automated skin lesion screening – a new approach: *Melanoma Res.* 2001;11(1):31-35. doi:10.1097/00008390-200102000-00004
25. Han SS, Lim W, Kim MS, Park I, Park GH, Chang SE. Interpretation of the Outputs of a Deep Learning Model Trained with a Skin Cancer Dataset. *J Invest Dermatol.* 2018;138(10):2275-2277. doi:10.1016/j.jid.2018.05.014
26. Burki TK. Machine learning model in melanoma. *Lancet Oncol.* 2018;19(7):e340. doi:10.1016/S1470-2045(18)30432-7

27. Thomsen K, Iversen L, Titlestad TL, Winther O. Systematic review of machine learning for diagnosis and prognosis in dermatology. *J Dermatol Treat.* 2020;31(5):496-510. doi:10.1080/09546634.2019.1682500
28. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature.* 2017;542(7639):115-118. doi:10.1038/nature21056
29. Brinker TJ, Hekler A, Enk AH, et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer.* 2019;119:11-17. doi:10.1016/j.ejca.2019.05.023
30. Haenssle HA, Fink C, Toberer F, et al. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. *Ann Oncol Off J Eur Soc Med Oncol.* 2020;31(1):137-143. doi:10.1016/j.annonc.2019.10.013
31. Chu CS, Lee NP, Adeoye J, Thomson P, Choi S. Machine learning and treatment outcome prediction for oral cancer. *J Oral Pathol Med.* 2020;49(10):977-985. doi:10.1111/jop.13089
32. Ilhan B, Lin K, Guneri P, Wilder-Smith P. Improving Oral Cancer Outcomes with Imaging and Artificial Intelligence. *J Dent Res.* 2020;99(3):241-248. doi:10.1177/0022034520902128
33. Pan X, Zhang T, Yang Q, Yang D, Rwigema JC, Qi XS. Survival prediction for oral tongue cancer patients via probabilistic genetic algorithm optimized neural network models. *Br J Radiol.* 2020;93(1112):20190825. doi:10.1259/bjr.20190825
34. Jeyaraj PR, Samuel Nadar ER. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *J Cancer Res Clin Oncol.* 2019;145(4):829-837. doi:10.1007/s00432-018-02834-7
35. Howard FM, Kochanny S, Koshy M, Spiotto M, Pearson AT. Machine Learning–Guided Adjuvant Treatment of Head and Neck Cancer. *JAMA Netw Open.* 2020;3(11):e2025881. doi:10.1001/jamanetworkopen.2020.25881
36. Zakhem GA, Motosko CC, Ho RS. How Should Artificial Intelligence Screen for Skin Cancer and Deliver Diagnostic Predictions to Patients? *JAMA Dermatol.* 2018;154(12):1383. doi:10.1001/jamadermatol.2018.2714
37. Zakhem GA, Fakhoury JW, Motosko CC, Ho RS. Characterizing the role of dermatologists in developing artificial intelligence for assessment of skin cancer: A systematic review. *J Am Acad Dermatol.* Published online January 20, 2020. doi:10.1016/j.jaad.2020.01.028
38. Bahr OV Markus Eisenhauer, Harald Sundmaeker, Patrick Guillemin, Martin Serrano, Elias Z Tragos, Javier Valiño, Arthur van derWees, Alex Gluhak, Roy. Internet of Things Cognitive Transformation Technology Research Trends and Applications. In: *Cognitive Hyperconnected Digital Transformation.* River Publishers; 2017.
39. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *N Engl J Med.* 2019;380(14):1347-1358. doi:10.1056/NEJMr1814259

40. Benjamens S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *Npj Digit Med.* 2020;3(1):1-8. doi:10.1038/s41746-020-00324-0
41. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. *Transl Vis Sci Technol.* 9(2):14. doi:10.1167/tvst.9.2.14
42. Hastie T, Friedman J, Tibshirani R. Introduction. In: Hastie T, Friedman J, Tibshirani R, eds. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Series in Statistics. Springer; 2001:1-8. doi:10.1007/978-0-387-21606-5_1
43. Du-Harpur X, Watt FM, Luscombe NM, Lynch MD. What is AI? Applications of artificial intelligence to dermatology. *Br J Dermatol.* 2020;183(3):423-430. doi:10.1111/bjd.18880
44. Al Turkestani N, Bianchi J, Deleat-Besson R, et al. Clinical decision support systems in orthodontics: A narrative review of data science approaches. *Orthod Craniofac Res.* 2021;24 Suppl 2(Suppl 2):26-36. doi:10.1111/ocr.12492
45. Schwendicke F, Samek W, Krois J. Artificial Intelligence in Dentistry: Chances and Challenges. *J Dent Res.* 2020;99(7):769-774. doi:10.1177/0022034520915714
46. Vougas K, Sakellaropoulos T, Kotsinas A, et al. Machine learning and data mining frameworks for predicting drug response in cancer: An overview and a novel in silico screening process based on association rule mining. *Pharmacol Ther.* 2019;203:107395. doi:10.1016/j.pharmthera.2019.107395
47. Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. *J Oral Pathol Med Off Publ Int Assoc Oral Pathol Am Acad Oral Pathol.* 2020;49(9):849-856. doi:10.1111/jop.13042
48. Bahng J, Lee CH. Topic Modeling for Analyzing Patients' Perceptions and Concerns of Hearing Loss on Social Q&A Sites: Incorporating Patients' Perspective. *Int J Environ Res Public Health.* 2020;17(17):6209. doi:10.3390/ijerph17176209
49. S P, V L, G A, M M, G T. Topic Modeling and User Network Analysis on Twitter during World Lupus Awareness Day. *Int J Environ Res Public Health.* 2020;17(15). doi:10.3390/ijerph17155440
50. Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. *Quintessence Int Berl Ger 1985.* 2020;51(3):248-257. doi:10.3290/j.qi.a43952
51. Sun W, Cai Z, Li Y, Liu F, Fang S, Wang G. Data Processing and Text Mining Technologies on Electronic Medical Records: A Review. *J Healthc Eng.* 2018;2018:4302425. doi:10.1155/2018/4302425
52. Labrosse J, Lam T, Sebbag C, et al. Text Mining in Electronic Medical Records Enables Quick and Efficient Identification of Pregnancy Cases Occurring After Breast Cancer. *JCO Clin Cancer Inform.* 2019;3:1-12. doi:10.1200/CCI.19.00031

53. Wehner MR, Levandoski KA, Kulldorff M, Asgari MM. Research Techniques Made Simple: An Introduction to Use and Analysis of Big Data in Dermatology. *J Invest Dermatol.* 2017;137(8):e153-e158. doi:10.1016/j.jid.2017.04.019
54. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521(7553):436-444. doi:10.1038/nature14539
55. Cho SI, Sun S, Mun JH, et al. Dermatologist-level classification of malignant lip diseases using a deep convolutional neural network. *Br J Dermatol.* 2020;182(6):1388-1394. doi:10.1111/bjd.18459
56. Haenssle HA, Fink C, Schneiderbauer R, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol Off J Eur Soc Med Oncol.* 2018;29(8):1836-1842. doi:10.1093/annonc/mdy166
57. Maron RC, Weichenthal M, Utikal JS, et al. Systematic outperformance of 112 dermatologists in multiclass skin cancer image classification by convolutional neural networks. *Eur J Cancer Oxf Engl 1990.* 2019;119:57-65. doi:10.1016/j.ejca.2019.06.013
58. Jeong HK, Park C, Henao R, Kheterpal M. Deep Learning in Dermatology: A Systematic Review of Current Approaches, Outcomes, and Limitations. *JID Innov Skin Sci Mol Popul Health.* 2023;3(1):100150. doi:10.1016/j.xjidi.2022.100150
59. Yap J, Yolland W, Tschandl P. Multimodal skin lesion classification using deep learning. *Exp Dermatol.* 2018;27(11):1261-1267. doi:10.1111/exd.13777
60. Guergueb T, Akhloufi MA. Melanoma Skin Cancer Detection Using Recent Deep Learning Models. *Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Int Conf.* 2021;2021:3074-3077. doi:10.1109/EMBC46164.2021.9631047
61. Sagi O, Rokach L. Ensemble learning: A survey. *WIREs Data Min Knowl Discov.* 2018;8(4):e1249. doi:10.1002/widm.1249
62. Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: Automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PLOS ONE.* 2018;13(1):e0191493. doi:10.1371/journal.pone.0191493
63. Tschandl P, Rosendahl C, Kittler H. The HAM10000 dataset, a large collection of multi-source dermoscopic images of common pigmented skin lesions. *Sci Data.* 2018;5:180161. doi:10.1038/sdata.2018.161
64. Marchetti MA, Codella NCF, Dusza SW, et al. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. *J Am Acad Dermatol.* 2018;78(2):270-277.e1. doi:10.1016/j.jaad.2017.08.016
65. Lee D, Yoon SN. Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges. *Int J Environ Res Public Health.* 2021;18(1):271. doi:10.3390/ijerph18010271

66. Mudgal SK, Agarwal R, Chaturvedi J, Gaur R, Ranjan N. Real-world application, challenges and implication of artificial intelligence in healthcare: an essay. *Pan Afr Med J.* 2022;43:3. doi:10.11604/pamj.2022.43.3.33384
67. Abraham A, Sobhanakumari K, Mohan A. Artificial intelligence in dermatology. *J Skin Sex Transm Dis.* 2021;3(1):99-102. doi:10.25259/JSSTD_49_2020
68. De A, Sarda A, Gupta S, Das S. Use of Artificial Intelligence in Dermatology. *Indian J Dermatol.* 2020;65(5):352-357. doi:10.4103/ijd.IJD_418_20
69. Chuchu N, Takwoingi Y, Dinnes J, et al. Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. *Cochrane Database Syst Rev.* 2018;12(12):CD013192. doi:10.1002/14651858.CD013192
70. Hekler A, Utikal JS, Enk AH, et al. Pathologist-level classification of histopathological melanoma images with deep neural networks. *Eur J Cancer.* 2019;115:79-83. doi:10.1016/j.ejca.2019.04.021
71. Mukherjee R, Manohar DD, Das DK, Achar A, Mitra A, Chakraborty C. Automated tissue classification framework for reproducible chronic wound assessment. *BioMed Res Int.* 2014;2014:851582. doi:10.1155/2014/851582
72. Emam S, Du AX, Surmanowicz P, Thomsen SF, Greiner R, Gniadecki R. Predicting the long-term outcomes of biologics in patients with psoriasis using machine learning. *Br J Dermatol.* 2020;182(5):1305-1307. doi:10.1111/bjd.18741
73. Romero G, Garrido JA, García-Arpa M. Telemedicine and Teledermatology (I): Concepts and Applications. *Actas Dermo-Sifiliográficas.* 2008;99(7):506-522. doi:10.1016/S1578-2190(08)70307-X
74. Koka SSA, Burkhart CG. Artificial Intelligence in Dermatology: Current Uses, Shortfalls, and Potential Opportunities for Further Implementation in Diagnostics and Care. *Open Dermatol J.* 2023;17(1). doi:10.2174/18743722-v17-e230505-2022-27
75. Gomes RFT, Schuch LF, Martins MD, et al. Use of Deep Neural Networks in the Detection and Automated Classification of Lesions Using Clinical Images in Ophthalmology, Dermatology, and Oral Medicine-A Systematic Review. *J Digit Imaging.* 2023;36(3):1060-1070. doi:10.1007/s10278-023-00775-3
76. Thomas B, Kumar V, Saini S. Texture analysis based segmentation and classification of oral cancer lesions in color images using ANN. In: *2013 IEEE International Conference on Signal Processing, Computing and Control (ISPC).* ; 2013:1-5. doi:10.1109/ISPC.2013.6663401
77. Fu Q, Chen Y, Li Z, et al. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: A retrospective study. *EClinicalMedicine.* 2020;27:100558. doi:10.1016/j.eclinm.2020.100558
78. Tanriver G, Soluk Tekkesin M, Ergen O. Automated Detection and Classification of Oral Lesions Using Deep Learning to Detect Oral Potentially Malignant Disorders. *Cancers.* 2021;13(11):2766. doi:10.3390/cancers13112766

79. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P. Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *J Oral Pathol Med.* 2021;50(9):911-918. doi:10.1111/jop.13227
80. Shamim MZM, Syed S, Shiblee M, et al. Automated Detection of Oral Pre-Cancerous Tongue Lesions Using Deep Learning for Early Diagnosis of Oral Cavity Cancer. *Comput J.* 2022;65(1):91-104. doi:10.1093/comjnl/bxaa136
81. Anantharaman R, Velazquez M, Lee Y. Utilizing Mask R-CNN for Detection and Segmentation of Oral Diseases. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* ; 2018:2197-2204. doi:10.1109/BIBM.2018.8621112
82. Jurczynszyn K, Kazubowska K, Kubasiewicz-Ross P, Ziółkowski P, Dominiak M. Application of fractal dimension analysis and photodynamic diagnosis in the case of differentiation between lichen planus and leukoplakia: A preliminary study. *Adv Clin Exp Med.* 2018;27(12):1729-1736. doi:10.17219/acem/80831
83. Jurczynszyn K, Kozakiewicz M. Differential diagnosis of leukoplakia versus lichen planus of the oral mucosa based on digital texture analysis in intraoral photography. *Adv Clin Exp Med.* 2019;28(11):1469-1476. doi:10.17219/acem/104524
84. Gorphe P, Jannin C. Regulatory aspects of prospective and retrospective clinical research in France in 2018. *Eur Ann Otorhinolaryngol Head Neck Dis.* 2019;136(2):103-108. doi:10.1016/j.anorl.2019.02.004
85. Zhang X. Melanoma segmentation based on deep learning. *Comput Assist Surg.* 2017;22(sup1):267-277. doi:10.1080/24699322.2017.1389405
86. Dablain D, Jacobson KN, Bellinger C, Roberts M, Chawla NV. Understanding CNN fragility when learning with imbalanced data. *Mach Learn.* Published online April 11, 2023. doi:10.1007/s10994-023-06326-9
87. Welikala RA, Remagnino P, Lim JH, et al. Automated Detection and Classification of Oral Lesions Using Deep Learning for Early Detection of Oral Cancer. *IEEE Access.* 2020;8:132677-132693. doi:10.1109/ACCESS.2020.3010180
88. Dubuc A, Zitouni A, Thomas C, et al. Improvement of Mucosal Lesion Diagnosis with Machine Learning Based on Medical and Semiological Data: An Observational Study. *J Clin Med.* 2022;11(21):6596. doi:10.3390/jcm11216596
89. Cullell-Dalmau M, Otero-Viñas M, Manzo C. Research Techniques Made Simple: Deep Learning for the Classification of Dermatological Images. *J Invest Dermatol.* 2020;140(3):507-514.e1. doi:10.1016/j.jid.2019.12.029
90. Rolnick D, Veit A, Belongie S, Shavit N. Deep Learning is Robust to Massive Label Noise. Published online February 26, 2018. doi:10.48550/arXiv.1705.10694
91. Cassidy B, Kendrick C, Brodzicki A, Jaworek-Korjakowska J, Yap MH. Analysis of the ISIC image datasets: Usage, benchmarks and recommendations. *Med Image Anal.* 2022;75:102305. doi:10.1016/j.media.2021.102305

92. Adamson AS, Smith A. Machine Learning and Health Care Disparities in Dermatology. *JAMA Dermatol.* 2018;154(11):1247-1248. doi:10.1001/jamadermatol.2018.2348
93. Alzubaidi L, Zhang J, Humaidi AJ, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data.* 2021;8(1):53. doi:10.1186/s40537-021-00444-8
94. Li Y, Huang C, Ding L, Li Z, Pan Y, Gao X. Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. *Methods.* 2019;166:4-21. doi:10.1016/j.ymeth.2019.04.008
95. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *J Stat Softw.* 2011;45:1-67. doi:10.18637/jss.v045.i03
96. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. Published online November 24, 2017. doi:10.48550/arXiv.1705.07874
97. Doumard E, Aligon J, Escriva E, Excoffier JB, Monsarrat P, Soulé-Dupuy C. A comparative study of additive local explanation methods based on feature influences. In: Vol 3130. CEUR-WS.org; 2022:31. Accessed July 24, 2023. <https://hal.science/hal-03687554>
98. Abati S, Bramati C, Bondi S, Lissoni A, Trimarchi M. Oral Cancer and Precancer: A Narrative Review on the Relevance of Early Diagnosis. *Int J Environ Res Public Health.* 2020;17(24):9160. doi:10.3390/ijerph17249160
99. Heelan K, McKenna D. A survey of oral medicine education, training and practice among dermatologists in the UK and Ireland. *Br Dent J.* 2016;220(1):17-20. doi:10.1038/sj.bdj.2016.23
100. Lin H, Chen H, Weng L, Shao J, Lin J. Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis. *J Biomed Opt.* 2021;26(8):086007. doi:10.1117/1.JBO.26.8.086007
101. Alkhadar H, Macluskey M, White S, Ellis I, Gardner A. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. *J Oral Pathol Med.* 2021;50(4):378-384. doi:10.1111/jop.13135
102. Khanagar SB, Al-ehaideb A, Maganur PC, et al. Developments, application, and performance of artificial intelligence in dentistry – A systematic review. *J Dent Sci.* 2021;16(1):508-522. doi:10.1016/j.jds.2020.06.019
103. Pacheco AGC, Krohling RA. The impact of patient clinical information on automated skin cancer detection. *Comput Biol Med.* 2020;116:103545. doi:10.1016/j.combiomed.2019.103545
104. Abbott LM, Smith SD. Smartphone apps for skin cancer diagnosis: Implications for patients and practitioners. *Australas J Dermatol.* 2018;59(3):168-170. doi:10.1111/ajd.12758

105. Udrea A, Mitra GD, Costea D, et al. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *J Eur Acad Dermatol Venereol JEADV*. 2020;34(3):648-655. doi:10.1111/jdv.15935
106. El-Khatib H, Popescu D, Ichim L. Deep Learning–Based Methods for Automatic Diagnosis of Skin Lesions. *Sensors*. 2020;20(6):1753. doi:10.3390/s20061753
107. Nasiri S, Helsper J, Jung M, Fathi M. DePicT Melanoma Deep-CLASS: a deep convolutional neural networks approach to classify skin lesion images. *BMC Bioinformatics*. 2020;21(Suppl 2):84. doi:10.1186/s12859-020-3351-y
108. Winkler JK, Sies K, Fink C, et al. Melanoma recognition by a deep learning convolutional neural network-Performance in different melanoma subtypes and localisations. *Eur J Cancer Oxf Engl 1990*. 2020;127:21-29. doi:10.1016/j.ejca.2019.11.020
109. R D S, A S. Deep Learning Based Skin Lesion Segmentation and Classification of Melanoma Using Support Vector Machine (SVM). *Asian Pac J Cancer Prev*. 2019;20(5):1555-1561. doi:10.31557/APJCP.2019.20.5.1555
110. Bajwa MN, Muta K, Malik MI, et al. Computer-Aided Diagnosis of Skin Diseases Using Deep Neural Networks. *Appl Sci*. 2020;10(7):2488. doi:10.3390/app10072488
111. Rotemberg V, Kurtansky N, Betz-Stablein B, et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Sci Data*. 2021;8(1):34. doi:10.1038/s41597-021-00815-z
112. Carvalho TM de, Noels E, Wakkee M, Udrea A, Nijsten T. Development of Smartphone Apps for Skin Cancer Risk Assessment: Progress and Promise. *JMIR Dermatol*. 2019;2(1):e13376. doi:10.2196/13376
113. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. *J Med Internet Res*. 2019;21(4):e12887. doi:10.2196/12887
114. Y R, M M, F F. A facial expression controlled wheelchair for people with disabilities. *Comput Methods Programs Biomed*. 2018;165. doi:10.1016/j.cmpb.2018.08.013
115. Hudec M, Smutny Z. RUDO: A Home Ambient Intelligence System for Blind People. *Sensors*. 2017;17(8):1926. doi:10.3390/s17081926
116. Al-Rawi N, Sultan A, Rajai B, et al. The Effectiveness of Artificial Intelligence in Detection of Oral Cancer. *Int Dent J*. 2022;72(4):436-447. doi:10.1016/j.identj.2022.03.001
117. Hegde S, Ajila V, Zhu W, Zeng C. Artificial intelligence in early diagnosis and prevention of oral cancer. *Asia-Pac J Oncol Nurs*. 2022;9(12):100133. doi:10.1016/j.apjon.2022.100133
118. Yala A, Lehman C, Schuster T, Portnoi T, Barzilay R. A Deep Learning Mammography-based Model for Improved Breast Cancer Risk Prediction. *Radiology*. 2019;292(1):60-66. doi:10.1148/radiol.2019182716

119. Lee S, Kim HS. Prospect of Artificial Intelligence Based on Electronic Medical Record. *J Lipid Atheroscler.* 2021;10(3):282-290. doi:10.12997/jla.2021.10.3.282
120. Top 6 Robotic Applications in Medicine. Accessed August 6, 2023. <https://www.asme.org/topics-resources/content/Top-6-Robotic-Applications-in-Medicine>
121. Alumran A, Alkhalidi O, Aldroorah Z, Alsayegh Z, Alsafwani F, Almaghraby N. Utilization of an Electronic Triage System by Emergency Department Nurses. *J Multidiscip Healthc.* 2020;13:339-344. doi:10.2147/JMDH.S250962
122. Zhang L, Tan J, Han D, Zhu H. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov Today.* 2017;22(11):1680-1685. doi:10.1016/j.drudis.2017.08.010
123. Sangers TE, Wakkee M, Kramer-Noels EC, Nijsten T, Lugtenberg M. Views on mobile health apps for skin cancer screening in the general population: an in-depth qualitative exploration of perceived barriers and facilitators. *Br J Dermatol.* 2021;185(5):961-969. doi:10.1111/bjd.20441
124. Nelson CA, Pérez-Chada LM, Creadore A, et al. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study. *JAMA Dermatol.* 2020;156(5):501-512. doi:10.1001/jamadermatol.2019.5014
125. Byambasuren O, Beller E, Glasziou P. Current Knowledge and Adoption of Mobile Health Apps Among Australian General Practitioners: Survey Study. *JMIR MHealth UHealth.* 2019;7(6):e13199. doi:10.2196/13199
126. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and Facilitators of the Use of Mobile Health Apps From a Security Perspective: Mixed-Methods Study. *JMIR MHealth UHealth.* 2019;7(4):e11223. doi:10.2196/11223
127. Chismar WG, Wiley-Patton S. Does the extended technology acceptance model apply to physicians. In: *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of The.* ; 2003:8 pp.-. doi:10.1109/HICSS.2003.1174354
128. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Q.* 1989;13(3):319-340. doi:10.2307/249008
129. Venkatesh V, Davis F, Morris M. Dead Or Alive? The Development, Trajectory And Future Of Technology Adoption Research. *AIS Educ J.* 2007;8:267-286. doi:10.17705/1jais.00120
130. Gursoy D, Chi OH, Lu L, Nunkoo R. Consumers acceptance of artificially intelligent (AI) device use in service delivery. *Int J Inf Manag.* 2019;49:157-169. doi:10.1016/j.ijinfomgt.2019.03.008
131. Kelly S, Kaye SA, Oviedo-Trespalacios O. What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telemat Inform.* 2023;77:101925. doi:10.1016/j.tele.2022.101925
132. Lin CH, Shih HY, Sher P. Integrating Technology Readiness into Technology Acceptance: The TRAM Model. *Psychol Mark.* 2007;24:641-657. doi:10.1002/mar.20177

133. Choung H, David P, Ross A. Trust in AI and Its Role in the Acceptance of AI Technologies. *Int J Human-Computer Interact.* 2023;39(9):1727-1739. doi:10.1080/10447318.2022.2050543
134. Venkatesh V. Adoption and use of AI tools: a research agenda grounded in UTAUT. *Ann Oper Res.* 2022;308. doi:10.1007/s10479-020-03918-9
135. Poon AIF, Sung JJY. Opening the black box of AI-Medicine. *J Gastroenterol Hepatol.* 2021;36(3):581-584. doi:10.1111/jgh.15384
136. Sariyar M, Holm J. Medical Informatics in a Tension Between Black-Box AI and Trust. *Stud Health Technol Inform.* 2022;289:41-44. doi:10.3233/SHTI210854
137. Rodríguez-Pérez R, Bajorath J. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *J Comput Aided Mol Des.* 2020;34(10):1013-1026. doi:10.1007/s10822-020-00314-0
138. Chaddad A, Peng J, Xu J, Bouridane A. Survey of Explainable AI Techniques in Healthcare. *Sensors.* 2023;23(2):634. doi:10.3390/s23020634
139. Lambert SI, Madi M, Sopka S, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *Npj Digit Med.* 2023;6(1):1-14. doi:10.1038/s41746-023-00852-5
140. Gomolin A, Netchiporouk E, Gniadecki R, Litvinov IV. Artificial Intelligence Applications in Dermatology: Where Do We Stand? *Front Med.* 2020;7. Accessed July 18, 2023. <https://www.frontiersin.org/articles/10.3389/fmed.2020.00100>
141. Du-Harpur X, Arthurs C, Ganier C, et al. Clinically Relevant Vulnerabilities of Deep Machine Learning Systems for Skin Cancer Diagnosis. *J Invest Dermatol.* 2021;141(4):916-920. doi:10.1016/j.jid.2020.07.034
142. Han SS, Kim YJ, Moon IJ, et al. Evaluation of Artificial Intelligence-Assisted Diagnosis of Skin Neoplasms: A Single-Center, Paralleled, Unmasked, Randomized Controlled Trial. *J Invest Dermatol.* 2022;142(9):2353-2362.e2. doi:10.1016/j.jid.2022.02.003
143. Chalutz Ben-Gal H. Artificial intelligence (AI) acceptance in primary care during the coronavirus pandemic: What is the role of patients' gender, age and health awareness? A two-phase pilot study. *Front Public Health.* 2023;10. Accessed July 19, 2023. <https://www.frontiersin.org/articles/10.3389/fpubh.2022.931225>
144. Norton WE, Hosokawa PW, Henderson WG, et al. Acceptability of the decision support for safer surgery tool. *Am J Surg.* 2015;209(6):977-984. doi:10.1016/j.amjsurg.2014.06.037
145. Stifter J, Sousa VEC, Febretti A, et al. Acceptability of Clinical Decision Support Interface Prototypes for a Nursing Electronic Health Record to Facilitate Supportive Care Outcomes. *Int J Nurs Knowl.* 2018;29(4):242-252. doi:10.1111/2047-3095.12178
146. Schulte A, Suarez-Ibarrola R, Wegen D, Pohlmann PF, Petersen E, Miernik A. Automatic speech recognition in the operating room – An essential contemporary tool or a

- redundant gadget? A survey evaluation among physicians in form of a qualitative study. *Ann Med Surg.* 2020;59:81-85. doi:10.1016/j.amsu.2020.09.015
147. Walter S, Gruss S, Frisch S, et al. "What About Automated Pain Recognition for Routine Clinical Use?" A Survey of Physicians and Nursing Staff on Expectations, Requirements, and Acceptance. *Front Med.* 2020;7. Accessed July 31, 2023. <https://www.frontiersin.org/articles/10.3389/fmed.2020.566278>
148. Farhud DD, Zokaei S. Ethical Issues of Artificial Intelligence in Medicine and Healthcare. *Iran J Public Health.* 2021;50(11):i-v. doi:10.18502/ijph.v50i11.7600
149. Floridi L, Taddeo M. What is data ethics? *Philos Trans R Soc Math Phys Eng Sci.* 2016;374(2083):20160360. doi:10.1098/rsta.2016.0360
150. Common ethical challenges in AI - Human Rights and Biomedicine - www.coe.int. Human Rights and Biomedicine. Accessed August 1, 2023. <https://www.coe.int/en/web/bioethics/common-ethical-challenges-in-ai>
151. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. *Big Data Soc.* 2016;3(2):2053951716679679. doi:10.1177/2053951716679679
152. *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS.*; 2021. Accessed August 1, 2023. <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206>
153. Flaten HK, St Claire C, Schlager E, Dunnick CA, Dellavalle RP. Growth of mobile applications in dermatology - 2017 update. *Dermatol Online J.* 2018;24(2):13030/qt3hs7n9z6.
154. Ouellette S, Rao BK. Usefulness of Smartphones in Dermatology: A US-Based Review. *Int J Environ Res Public Health.* 2022;19(6):3553. doi:10.3390/ijerph19063553
155. Jahn AS, Navarini AA, Cerminara SE, et al. Over-Detection of Melanoma-Suspect Lesions by a CE-Certified Smartphone App: Performance in Comparison to Dermatologists, 2D and 3D Convolutional Neural Networks in a Prospective Data Set of 1204 Pigmented Skin Lesions Involving Patients' Perception. *Cancers.* 2022;14(15):3829. doi:10.3390/cancers14153829
156. Uthoff RD, Song B, Sunny S, et al. Point-of-care, smartphone-based, dual-modality, dual-view, oral cancer screening device with neural network classification for low-resource communities. *PLoS One.* 2018;13(12):e0207493. doi:10.1371/journal.pone.0207493
157. Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. *Sci Rep.* 2019;9(1):6994. doi:10.1038/s41598-019-43372-7
158. Dulmage B, Tegtmeyer K, Zhang MZ, Colavincenzo M, Xu S. A Point-of-Care, Real-Time Artificial Intelligence System to Support Clinician Diagnosis of a Wide Range of Skin Diseases. *J Invest Dermatol.* 2021;141(5):1230-1235. doi:10.1016/j.jid.2020.08.027

159. Lakhani NA, Saraiya M, Thompson TD, King SC, Guy GP. Total body skin examination for skin cancer screening among U.S. adults from 2000 to 2010. *Prev Med.* 2014;61:75-80. doi:10.1016/j.ypmed.2014.01.003
160. Rat C, Hild S, Rault Sérandour J, et al. Use of Smartphones for Early Detection of Melanoma: Systematic Review. *J Med Internet Res.* 2018;20(4):e135. doi:10.2196/jmir.9392
161. Cazzaniga S, Castelli E, Di Landro A, et al. Mobile teledermatology for melanoma detection: Assessment of the validity in the framework of a population-based skin cancer awareness campaign in northern Italy. *J Am Acad Dermatol.* 2019;81(1):257-260. doi:10.1016/j.jaad.2019.02.036
162. Choy SP, Kim BJ, Paolino A, et al. Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease. *NPJ Digit Med.* 2023;6:180. doi:10.1038/s41746-023-00914-8
163. Kuo KM, Talley PC, Chang CS. The accuracy of artificial intelligence used for non-melanoma skin cancer diagnoses: a meta-analysis. *BMC Med Inform Decis Mak.* 2023;23(1):138. doi:10.1186/s12911-023-02229-w
164. Albalawi E, Thakur A, Ramakrishna MT, et al. Oral squamous cell carcinoma detection using EfficientNet on histopathological images. *Front Med.* 2024;10. doi:10.3389/fmed.2023.1349336
165. Abdul NS, Shivakumar GC, Sangappa SB, et al. Applications of artificial intelligence in the field of oral and maxillofacial pathology: a systematic review and meta-analysis. *BMC Oral Health.* 2024;24(1):122. doi:10.1186/s12903-023-03533-7
166. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. *Lancet Oncol.* 2002;3(3):159-165. doi:10.1016/s1470-2045(02)00679-4
167. Combalia M, Codella N, Rotemberg V, et al. Validation of artificial intelligence prediction models for skin cancer diagnosis using dermoscopy images: the 2019 International Skin Imaging Collaboration Grand Challenge. *Lancet Digit Health.* 2022;4(5):e330-e339. doi:10.1016/S2589-7500(22)00021-8
168. Narla A, Kuprel B, Sarin K, Novoa R, Ko J. Automated Classification of Skin Lesions: From Pixels to Practice. *J Invest Dermatol.* 2018;138(10):2108-2110. doi:10.1016/j.jid.2018.06.175
169. Tschandl P, Rosendahl C, Akay BN, et al. Expert-Level Diagnosis of Nonpigmented Skin Cancer by Combined Convolutional Neural Networks. *JAMA Dermatol.* 2019;155(1):58-65. doi:10.1001/jamadermatol.2018.4378
170. Tschandl P, Codella N, Akay BN, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol.* 2019;20(7):938-947. doi:10.1016/S1470-2045(19)30333-X

171. Alhazmi A, Alhazmi Y, Makrami A, et al. Application of artificial intelligence and machine learning for prediction of oral cancer risk. *J Oral Pathol Med Off Publ Int Assoc Oral Pathol Am Acad Oral Pathol*. 2021;50(5):444-450. doi:10.1111/jop.13157
172. Alabi RO, Elmusrati M, Sawazaki-Calone I, et al. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. *Int J Med Inf*. 2020;136:104068. doi:10.1016/j.ijmedinf.2019.104068
173. Kuhlman C, Jackson L, Chunara R. No Computation without Representation: Avoiding Data and Algorithm Biases through Diversity. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '20. Association for Computing Machinery; 2020:3593. doi:10.1145/3394486.3411074
174. Arora A, Alderman JE, Palmer J, et al. The value of standards for health datasets in artificial intelligence-based applications. *Nat Med*. 2023;29(11):2929-2938. doi:10.1038/s41591-023-02608-w
175. Saleh S, Boag W, Erdman L, Naumann T. Clinical Collabsheets: 53 Questions to Guide a Clinical Collaboration. In: *Proceedings of the 5th Machine Learning for Healthcare Conference*. PMLR; 2020:783-812. Accessed April 22, 2024. <https://proceedings.mlr.press/v126/saleh20a.html>
176. Ferryman K. Addressing health disparities in the Food and Drug Administration's artificial intelligence and machine learning regulatory framework. *J Am Med Inform Assoc JAMIA*. 2020;27(12):2016-2019. doi:10.1093/jamia/ocaa133
177. Caffery LJ, Rotemberg V, Weber J, Soyer HP, Malvey J, Clunie D. The Role of DICOM in Artificial Intelligence for Skin Disease. *Front Med*. 2020;7:619787. doi:10.3389/fmed.2020.619787
178. Miller J. *Cartesian Genetic Programming*. Vol 43.; 2003. doi:10.1007/978-3-642-17310-3
179. Cortacero K, Mckenzie B, Mueller S, et al. Evolutionary design of explainable algorithms for biomedical image segmentation. *Nat Commun*. 2023;14. doi:10.1038/s41467-023-42664-x
180. Bashi N, Fatehi F, Fallah M, Walters D, Karunanithi M. Self-Management Education Through mHealth: Review of Strategies and Structures. *JMIR MHealth UHealth*. 2018;6(10):e10771. doi:10.2196/10771
181. Masters K. Health professionals as mobile content creators: Teaching medical students to develop mHealth applications. *Med Teach*. 2014;36. doi:10.3109/0142159X.2014.916783
182. Maskara R, Bhootra V, Thakkar D, Nishkalank N. A study on the perception of medical professionals towards artificial intelligence. *Int J Multidiscip Res Dev*. Published online April 1, 2017. Accessed July 21, 2023. <https://www.semanticscholar.org/paper/A-study-on-the-perception-of-medical-professionals-Maskara-Bhootra/dff541ae46d3a41b62c15102cc8a3fca9c41a8bb>

183. Jiang L, Wu Z, Xu X, et al. Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. *J Int Med Res.* 2021;49(3):3000605211000157. doi:10.1177/03000605211000157
184. Baowaly MK, Lin CC, Liu CL, Chen KT. Synthesizing electronic health records using improved generative adversarial networks. *J Am Med Inform Assoc JAMIA.* 2019;26(3):228-241. doi:10.1093/jamia/ocy142
185. Nordling L. A fairer way forward for AI in health care. *Nature.* 2019;573(7775):S103-S105. doi:10.1038/d41586-019-02872-2
186. Crawford K. *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale University Press; 2021. doi:10.2307/j.ctv1ghv45t
187. Hacker P. Teaching fairness to artificial intelligence: Existing and novel strategies against algorithmic discrimination under EU law. *Common Mark Law Rev.* 2018;55(4). Accessed August 1, 2023. <https://kluwerlawonline.com/api/Product/CitationPDFURL?file=Journals\COLA\COLA2018095.pdf>
188. Zajko M. Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. *Sociol Compass.* 2022;16(3):e12962. doi:10.1111/soc4.12962
189. Edgman-Levitan S, Schoenbaum SC. Patient-centered care: achieving higher quality by designing care through the patient's eyes. *Isr J Health Policy Res.* 2021;10:21. doi:10.1186/s13584-021-00459-9
190. Bauchat J, Seropian M, Jeffries P. Communication and Empathy in the Patient-Centered Care Model—Why Simulation-Based Training Is Not Optional. *Clin Simul Nurs.* 2016;12(8). doi:10.1016/j.ecns.2016.04.003
191. Kilbride MK, Joffe S. The New Age of Patient Autonomy: Implications for the Patient-Physician Relationship. *JAMA.* 2018;320(19):1973-1974. doi:10.1001/jama.2018.14382
192. Young AT, Amara D, Bhattacharya A, Wei ML. Patient and general public attitudes towards clinical artificial intelligence: a mixed methods systematic review. *Lancet Digit Health.* 2021;3(9):e599-e611. doi:10.1016/S2589-7500(21)00132-1
193. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial Intelligence and the Future of Primary Care: Exploratory Qualitative Study of UK General Practitioners' Views. *J Med Internet Res.* 2019;21(3):e12802. doi:10.2196/12802
194. Karches KE. Against the iDoctor: why artificial intelligence should not replace physician judgment. *Theor Med Bioeth.* 2018;39(2):91-110. doi:10.1007/s11017-018-9442-3
195. Liu X, Keane PA, Denniston AK. Time to regenerate: the doctor in the age of artificial intelligence. *J R Soc Med.* 2018;111(4):113-116. doi:10.1177/0141076818762648

Annexe

Improvement of Mucosal Lesion Diagnosis with Machine Learning Based on Medical and Semiological Data: An Observational Study

Dubuc A & al. J. Clin.Med. 2022, 11, 6596.<https://doi.org/10.3390/jcm11216596>



Article

Improvement of Mucosal Lesion Diagnosis with Machine Learning Based on Medical and Semiological Data: An Observational Study

Antoine Dubuc, Anissa Zitouni, Charlotte Thomas, Philippe Kémoun, Sarah Cousty, Paul Monsarrat and Sara Laurencin



<https://doi.org/10.3390/jcm11216596>

Article

Improvement of Mucosal Lesion Diagnosis with Machine Learning Based on Medical and Semiological Data: An Observational Study

Antoine Dubuc ^{1,2,†} , Anissa Zitouni ^{3,†}, Charlotte Thomas ^{1,4} , Philippe Kémoun ^{1,5} , Sarah Cousty ^{1,6} , Paul Monsarrat ^{1,5,7,‡}  and Sara Laurencin ^{1,2,*,‡} 

- ¹ School of Dental Medicine and CHU de Toulouse—Toulouse Institute of Oral Medicine and Science, 31062 Toulouse, France
 - ² Center for Epidemiology and Research in POPulation Health (CERPOP), UMR 1295, Paul Sabatier University, 31062 Toulouse, France
 - ³ Oral Surgery and Oral Medicine Department, CHU Limoges, 87000 Limoges, France
 - ⁴ InCOMM, I2MC, UMR 1297, Paul Sabatier University, 31062 Toulouse, France
 - ⁵ RESTORE Research Center, Université de Toulouse, INSERM, CNRS, EFS, ENVT, Université P. Sabatier, CHU de Toulouse, 31300 Toulouse, France
 - ⁶ LAPLACE, UMR 5213 CNRS, Paul Sabatier University, 31062 Toulouse, France
 - ⁷ Artificial and Natural Intelligence Toulouse Institute ANITI, 31013 Toulouse, France
- * Correspondence: laurencin.s@chu-toulouse.fr
† These authors contributed equally to this work.
‡ These authors contributed equally to this work.



Citation: Dubuc, A.; Zitouni, A.; Thomas, C.; Kémoun, P.; Cousty, S.; Monsarrat, P.; Laurencin, S. Improvement of Mucosal Lesion Diagnosis with Machine Learning Based on Medical and Semiological Data: An Observational Study. *J. Clin. Med.* **2022**, *11*, 6596. <https://doi.org/10.3390/jcm11216596>

Academic Editor: Pia Lopez-Jornet

Received: 27 September 2022

Accepted: 2 November 2022

Published: 7 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Despite artificial intelligence used in skin dermatology diagnosis is booming, application in oral pathology remains to be developed. Early diagnosis and therefore early management, remain key points in the successful management of oral mucosa cancers. The objective was to develop and evaluate a machine learning algorithm that allows the prediction of oral mucosa lesions diagnosis. This cohort study included patients followed between January 2015 and December 2020 in the oral mucosal pathology consultation of the Toulouse University Hospital. Photographs and demographic and medical data were collected from each patient to constitute clinical cases. A machine learning model was then developed and optimized and compared to 5 models classically used in the field. A total of 299 patients representing 1242 records of oral mucosa lesions were used to train and evaluate machine learning models. Our model reached a mean accuracy of 0.84 for diagnostic prediction. The specificity and sensitivity range from 0.89 to 1.00 and 0.72 to 0.92, respectively. The other models were proven to be less efficient in performing this task. These results suggest the utility of machine learning-based tools in diagnosing oral mucosal lesions with high accuracy. Moreover, the results of this study confirm that the consideration of clinical data and medical history, in addition to the lesion itself, appears to play an important role.

Keywords: artificial intelligence; oral pathology; oral dermatology

1. Introduction

In the past 30 years, there has been a major increase in interest in Artificial Intelligence (AI) in medicine. Indeed, this craze is explained by the fact that practitioners need help to accelerate treatment, to have a predictive attitude, or to develop telemedicine tools. Recently, this interest has reached the field of dermatopathology [1,2], with over 600 publications referenced in Pubmed covering this subject and its different applications based on classification, segmentation, or feature extraction of lesions [3] and even smartphone based images [4]. Although these applications are currently limited to a few types of lesions, such as skin cancers (particularly melanoma [5–7]), oral squamous cell carcinomas [4,8] (OSCC), ulcers, psoriasis, and atopic dermatitis [9], it has been suggested that

the accuracy of AI-based diagnosis using convolutional neural networks for lesion image analysis may be similar to experienced dermatologists [10], or even better [11]. Machine Learning (ML) can be defined as a branch of AI whose algorithms learn autonomously to make predictions from data and improve their performance over time [9]. ML corresponds to a wide range of algorithm families and complexities, from self-explainable models (e.g., linear regression, polynomial regression) to other models requiring the development of additional explainability methods. (e.g., support vector machine, tree-based models, neural networks). ML models are particularly well suited to capture the complex relationships between variables to extract hidden patterns in the datasets to respond as efficiently as possible to the prediction to be made. Often referred to as “black-boxes” models, the question of how humans can understand the determinants of the prediction is crucial, particularly in the area of healthcare. Explainability is also more than a technological problem, it involves among other ethical, societal, and legal issues [12]. They are therefore of major interest to support the practitioner referring the patient to a specialist (telemedicine), but also in the diagnostic process and in guiding them towards the most appropriate treatment. The practitioner must be able to explain to the patient how the algorithm works and the criteria for the decision process [13]. Although the publications on skin lesions concern most of the work, the contribution of AI in the field of oral mucosa pathology is relatively recent [4,8,14,15]. However, the early diagnosis and management of oral mucosa lesions is a major issue, particularly in the context of oral cancer diagnosis and management. According to the World Health Organization (WHO), lip and oral cavity cancers are the 16th most frequent cancers, corresponding to more than 177,000 deaths worldwide in 2020. Oral cancer and primarily squamous cell carcinoma accounted for half of the overall annual mortality from head and neck cancer in 2018 [16]. Since 2019, some teams have been interested in the contribution of AI algorithms and especially ML in the management of oropharyngeal cancer, mainly in terms of prognosis prediction [17–19] and medical image classification [20], but also in optimizing the treatment of these cancers [21]. However, the lack of practitioners specialized in oral mucosa pathology and the variety of lesions have made differential diagnosis a major challenge and the goal of improving early diagnosis and patient management.

This study aimed to propose a predictive ML-based framework with state-of-the-art explainability techniques based on chair-side patient information collection, such as the semiology of the lesions and patient socio-demographic and medical data for diagnosis of oral mucosa lesions.

2. Materials and Methods

2.1. Selection Criteria

Patients consulting in the specialized oral dermatology consultation of the Toulouse University Hospital (CHU de Toulouse) between January 2015 and December 2020 were considered in this study.

All patients meeting the following criteria were included in the study: (1) Patients consulting in the oral mucosa pathology consultation of the Toulouse University Hospital and presenting a mucosal lesion, (2) Patients already followed for such lesions and requiring a histological re-evaluation, (3) Subjects older than 18 years of age at the time of consultation, (4) Subjects who are able to understand French language, to receive information about the study and understand the information form to participate in the study, (5) Subjects who have given their consent to participate in the study. Patients with the following conditions could not be included in this study: (1) Pregnant and lactating women, (2) Patients under guardianship, curatorship, or safeguard of justice. All patients gave their consent for data collection and the computer processing of personal and medical data was implemented to the results of this research. Patients of legal age were included in the study, patients over 18 years of age.

This study was conducted in accordance with the public health code and bioethics laws governing research in France [22]. This study is retrospective and does not involve

the human person. Therefore, it is not submitted to an ethics committee but complies with the reference methodology (MR-004 of the French National Commission for Informatics and Liberties (CNIL) (number: 2206723 v0)) for which Toulouse University Hospital signed a commitment of compliance. The patient's non-objection was collected.

2.2. Data Collection and Annotation

For each subject, a series of high-resolution, macroscopic photographs of the lesion (either by digital camera or smartphone) were acquired. When for the same patient, several images of the same lesion were taken, only one of them was considered, whereas images of the same patient but representing a different lesion and/or a different representation of the same lesion (e.g., according to topography or color) were treated separately. Each lesion characteristic (type of lesion, color, topography, size, homogeneity, and shape) was annotated by a single investigator (A.D), beforehand calibrated with two experts in oral mucosa pathology (S.C, S.L) before the launch of the study by assessing 50 randomly selected images. The overall inter-rater kappa reached 0.92. Intra-observer reliability was tested by randomly re-examining 30 lesions at a one-month interval of the initial assessment and reached 0.87.

The diagnoses associated with each of the images were previously validated histologically. Anatomopathological analysis of oral lesions as part of the patient's routine care was performed in the anatomopathological laboratory of the Toulouse University Hospital (CHU de Toulouse) and the Toulouse University Cancer Institute (IUCT, Institut Universitaire du Cancer de Toulouse). Demographic (age, gender, alcohol consumption, smoking status), medical data and medications were also recorded. Medical history and medications were, respectively, classified according to the International Classification of Diseases (ICD) and the Anatomical Therapeutic Chemical Classification System (ATC).

2.3. Model Development and Data Analysis

The machine learning pipeline involved (1) an adequate pre-processing step of the database (data encoding, imputing, and normalization), (2) a random train:test dataset splitting of 80:20, (3) the comparison and choice of a suitable predictive model, and (4) the explainability of the prediction results.

Pre-processing step: The binary categorical data (yes/no or male/female) were first label-encoded by integers. The type of elementary lesion was one hot encoded for non-tree-based ML models. Missing socio-demographic and/or medical data were considered as missing at-random, for age (68, 7.1%), gender (61, 6.4%), tobacco (7, 0.7%), alcohol (7, 0.7%) and narcotics (7, 0.7%). Missing data were considered as such for the algorithmic approach being able to deal with (tree-based gradient boosting method LightGBM) or imputed using a multivariate single imputation model [23].

Data modeling : Different Machine Learning strategies were compared: LightGBM (LGB, a tree-based gradient boosting method), Logistic Regression with an elastic net (L1 and L2) penalization (LR), Decision tree (DT), K-Nearest Neighbors (KNN) and Multi-Layer Perceptron (MLP). Models were used from scikit-learn 1.0.1 and lightgbm 3.3.1. To optimize the hyperparameters of each model, a grid search was explored using optuna 2.10.0, with mean ROC AUC one-versus-one after 5-fold cross-validation considered as eval metric.

Model performance: For each model, were computed the true positive rate (the proportion of positive data that are correctly considered as positive), the true negative rate (proportion of negative data that are correctly considered as negative), the positive predictive value (number of correct positive results divided by the number of positive results predicted by the classifier), the negative predictive value (number of correct negative results divided by the number of negative results predicted by the classifier), the overall accuracy (the number of correct predictions divided by the total number of predictions) and the F1-score (the harmonic mean between predictive positive values and true positive rate).

Model explainability: Finally, the influence of each attribute of the dataset on the diagnosis prediction was explored using SHapley Additive exPlanations (SHAP) [24].

Given the high dimensionality of the dataset (high number of variables) and the nature of the final chosen model (tree-based model lightGBM), the model was explained thanks to the TreeSHAP optimization [25].

3. Results

The dataset consisted of a total of 1242 records representing 31 different diagnoses for 299 subjects (39.4% men; 60.6% women; mean aged 57.3 years \pm 19.2). Given the number of cases by diagnosis, 7 categories were considered: oral lichen planus (40.0% of the dataset), bullous diseases (13.9%), aphthous ulcers (8.9%), leukoplakia (7.3%), gingival enlargement (3.7%), oral squamous cell carcinoma (3.3%), and others (22.9%). A detailed description of the final dataset (socio-demographic and medical data) is provided in Table 1. Only 8 (2.7%) subjects were alcohol-dependent and 72 (24.1%) were smokers. The most represented elementary lesions were plaque/papule (27%), with 36% of these lesions being located on the attached gingiva.

Once models were tuned for optimization of hyperparameters and trained on the training dataset, the performance of each model was shown on the test dataset. LightGBM model achieved the best performance with 0.84 for mean accuracy. The other ML models achieved lower performances with 0.54, 0.73, 0.49, and 0.62, respectively, for logistic regression (LR), k-nearest neighbors (K-NN), decision tree classifier (DTC), and multi-layer perceptron models (MLP). The detailed performances of the models are exposed in Table 2. Confusion matrixes are exposed in Figures 1A and S1 for LightGBM and other models, respectively.

Whatever the diagnosis to be predicted and the metrics, LightGBM achieved the best performances. The TreeSHAP method was used to provide insight into the explainability of the predictions. Figure 1B shows the importance of the different features in overall decision-making. The factors influencing the overall decision making are essentially lesion characteristics (type of lesion, color, reticulation...). Although age remains the most important factor in decision-making, other socio-demographic data are considered but with less importance. The model does not take into account the medical history and treatments in the overall explanation of the model behavior. Figures 2 and S2 expose the importance of the different features for the prediction of each diagnosis. If patients' age seems to influence all different lesions, for Oral Lichen planus features such as the reticular form of the lesion, gender, type of elementary lesion, and localization have more influence on the diagnosis than the patient's medical history. On the contrary, for OSSC, the size of the lesion, its localization, and the patient's smoking status are more important to predict the diagnosis than the color of the lesion or the gender. Medical history often has a minor influence on individual prediction, with the exception of a history of dermatologic pathology, which strongly influences the diagnosis of bullous disease.

Finally, Figure 3 illustrates two examples of prediction, respectively, for a carcinoma (Figure 3A) and an oral lichen planus lesion (Figure 3B). The SHAP algorithm provides an understanding of decision making by outlining the factors that positively or negatively influence the different diagnoses.

Table 1. Subject characteristics are used for machine learning model training according to each diagnosis.

Number of Subjects	Lichen Planus (n = 77)	Leukoplakia (n = 41)	Aphthous Ulcer (n = 37)	Bullous Diseases (n = 26)	Gingival Enlargement (n = 12)	Carcinoma (n = 11)	Others (n = 95)	p-Value ^a
Gender (n,%)								
Male	22 (29%)	20 (49%)	19 (51%)	9 (35%)	5 (42%)	5 (45%)	42 (44%)	0.03
Female	51 (66%)	16 (39%)	14 (38%)	15 (58%)	3 (25%)	4 (36%)	53 (56%)	
Mean age	65	65	50	65	43	54	50	<0.001
Alcohol (n,%)	2 (3%)	4 (10%)	1 (3%)	-	-	2 (18%)	2 (2%)	0.007
Smoking (n,%)	21 (27%)	18 (44%)	8 (22%)	3 (12%)	1 (8%)	5 (45%)	16 (17%)	0.002
Number of lesions (n)	497	91	110	173	46	41	252	
Localization (n,%)								
Vermilion	19 (3.8 %)	6 (6.5%)	5 (4.5%)	3 (1.7%)	-	-	16 (6.3%)	
Commissure	1 (0.2%)	-	-	1 (9.6%)	-	-	8 (3%)	
Labial mucosa	8 (1.7%)	3 (3%)	15 (13.6%)	16 (9.2%)	-	1 (2.4%)	21 (8.3%)	
Muco-buccal fold	136 (27.3%)	25 (27.5%)	18 (16.4%)	16 (9.2%)	-	5 (12%)	21 (8.3%)	
Attached gingiva	172 (34.6%)	12 (13.1%)	12 (10.9%)	91 (52.6%)	45 (98%)	16 (39%)	85 (33.7%)	
Buccal mucosa	37 (7.4%)	3 (3.3%)	10 (9.0%)	14 (8.0%)	1 (2.0%)	2 (4.8%)	11 (4.4%)	0.002
Hard palate	22 (4.4%)	13 (14.3%)	5 (4.5%)	10 (5.7%)	-	11 (26.8%)	20 (7.9%)	
Soft palate	1 (0.2%)	4 (4.4%)	5 (4.5%)	19 (10.9%)	-	-	2 (0.7%)	
Tonsillar pillar	3 (0.6%)	1 (1.1%)	1 (0.9%)	1 (0.5%)	-	-	2 (0.7%)	
Dorsal tongue	60 (12.0%)	7 (1.8%)	3 (2.7%)	2 (1.2%)	-	2 (4.8%)	51 (20.2%)	
Ventral tongue	9 (1.8%)	1 (1.0%)	10 (9.0%)	6 (3.5%)	-	1 (2.4%)	2 (0.7%)	
Lateral tongue	24 (4.8%)	14 (15.4%)	13 (11.8%)	-	-	1 (2.4%)	11 (4.4%)	
Floor of mouth	5 (1.0%)	2 (2.2%)	2 (1.8%)	-	-	2 (4.8%)	2 (0.7%)	
Lesion color (n,%)								
White	160 (32.0%)	70 (7.8%)	33 (30.0%)	46 (26.6%)	-	13 (31.7%)	50 (18.8%)	
Red	55 (11.0%)	2 (2.2%)	9 (8.1%)	55 (31.8%)	15 (32,6%)	8 (19.5%)	51 (20.2%)	<0.001
Mixt (red and white)	268 (54%)	19 (20.8%)	66 (60.0%)	69 (39.9%)	5 (11.8%)	16 (39.0%)	79 (31.3%)	
Pigmented	10 (2.0%)	-	1 (0.9%)	-	5 (11.8%)	3 (2.3%)	40 (15.8%)	
No color changes	4 (0.8%)	-	1 (0.9%)	-	21 (46.0%)	1 (2.4%)	32 (12.7%)	
Lesion size (n,%)								
<5 mm	39 (7.8%)	13 (14.3%)	27 (24.5%)	32 (18.5%)	27 (58.6%)	1 (2.4%)	41 (16.2%)	
5 to 10 mm	268 (54.0%)	41 (45.1%)	66 (60.0%)	116 (67.0%)	38 (82.6%)	9 (21.9%)	135 (53.1%)	<0.001
10 to 50 mm	188 (38.0%)	37 (40.7%)	17 (15.5%)	25 (14.5%)	6 (13.0%)	27 (65.8%)	75 (29.7%)	
>50 mm	2 (0.4%)	-	-	-	-	4 (9.7%)	1 (0.4%)	

Table 1. Cont.

Number of Subjects	Lichen Planus (n = 77)	Leukoplakia (n = 41)	Aphthous Ulcer (n = 37)	Bullous Diseases (n = 26)	Gingival Enlargement (n = 12)	Carcinoma (n = 11)	Others (n = 95)	p-Value ^a
Elementary Lesion (n,%)								
Gingival hyperplasia	3 (0.6%)	-	-	1 (0.5%)	40 (86.9%)	-	9 (3.5%)	
Bulla	7 (1.4%)	2 (2.2%)	2 (1.8%)	31 (18.0%)	-	-	16 (6.3%)	
Cellular coating	11 (2.2%)	1 (1.1%)	13 (11.8%)	16 (9.2%)	-	2 (4.8%)	12 (4.7%)	
Erosion	18 (3.6%)	3 (3.3%)	6 (5.5%)	30 (18.3%)	-	1 (2.4%)	8 (3.1%)	<0.001
Macula/patch	184 (37.0%)	12 (13.2%)	11 (10.0%)	79 (45.7%)	6 (13.0%)	4 (9.7%)	78 (30.9%)	
Nodule	13 (2.6%)	2 (2.2%)	2 (1.8%)	-	-	18 (43.9%)	100 (39.6%)	
Papula/plaque	220 (44.3%)	70 (77.0%)	8 (7.3%)	7 (4.0%)	-	13 (31.7%)	23 (9.1%)	
Ulcer	40 (8.0%)	1 (1.1%)	68 (61.8%)	9 (5.2%)	-	3 (7.3%)	4 (1.5%)	
Lesion homogeneity (n,%)	187 (37.6%)	52 (75.1%)	77 (70.0%)	71 (41.0%)	32 (69.5%)	5 (12.1%)	142 (56.4%)	<0.001
Well circumscribed lesion (n,%)	230 (46.3%)	55 (60.4%)	93 (84.5%)	81 (46.8%)	19 (41.3%)	7 (17.0%)	166 (65.8%)	<0.001
Elevated edges of the lesion (n,%)	5 (1.0%)	-	13 (11.8%)	2 (1.2%)	1 (2.0%)	1 (2.4%)	4 (1.6%)	0.29
Reticular lesion (n,%)	175 (35.2%)	3 (3.3%)	1 (0.9%)	5 (2.9%)	-	-	5 (1.9%)	<0.001
Growing lesion (n,%)	13 (2.6%)	2 (2.2%)	-	1 (0.5%)	3 (6.5%)	24 (58.5%)	25 (9.9%)	<0.001
Papillomatus lesion (n,%)	3 (0.6%)	8 (8.8%)	1 (0.9%)	1 (0.5%)	1 (2.0%)	8 (19.5%)	41 (16.2%)	0.007

^a for categorical values, the p-value resulted from a χ^2 test; for numerical values, the p-value resulted from an ANOVA test.

Table 2. Models’ performances for each diagnosis.

Diagnostics	LightGBM					Elastic Net Regression					K-Nearest Neighbors					Decision Tree					Multilayer Perceptron				
	TPR	TNR	PPV	NPV	F1	TPR	TNR	PPV	NPV	F1	TPR	TNR	PPV	NPV	F1	TPR	TNR	PPV	NPV	F1	TPR	TNR	PPV	NPV	F1
Gingival enlargement	0.92	1.00	1.00	1.00	0.96	0.92	0.99	0.85	1.00	0.88	0.83	0.99	0.71	0.99	0.77	0.67	0.97	0.47	0.99	0.55	0.83	1.00	1.00	0.99	0.91
Carcinoma	0.90	1.00	0.90	1.00	0.90	0.60	0.92	0.21	0.99	0.32	0.70	0.99	0.64	0.99	0.67	0.80	0.96	0.40	0.99	0.53	0.50	0.98	0.50	0.98	0.50
Leukoplakia	0.78	0.98	0.75	0.98	0.77	0.74	0.91	0.40	0.98	0.52	0.57	0.94	0.42	0.96	0.48	0.74	0.89	0.36	0.98	0.49	0.35	0.98	0.62	0.95	0.44
Oral lichen planus	0.89	0.89	0.85	0.92	0.87	0.38	0.92	0.77	0.68	0.51	0.85	0.88	0.83	0.89	0.84	0.27	0.98	0.92	0.66	0.42	0.84	0.70	0.66	0.86	0.74
Bullous diseases	0.72	0.96	0.76	0.95	0.74	0.70	0.81	0.38	0.94	0.49	0.70	0.96	0.75	0.95	0.72	0.74	0.76	0.34	0.95	0.46	0.35	0.95	0.56	0.90	0.43
Apthous ulcers	0.75	0.99	0.84	0.97	0.79	0.54	0.96	0.56	0.95	0.55	0.46	0.97	0.59	0.95	0.52	0.75	0.92	0.50	0.97	0.60	0.50	0.98	0.74	0.95	0.60
Other lesions	0.89	0.97	0.89	0.97	0.89	0.60	0.94	0.72	0.90	0.66	0.67	0.93	0.72	0.91	0.69	0.46	0.93	0.64	0.87	0.54	0.73	0.92	0.70	0.93	0.71
Mean accuracy	0.84					0.54					0.73					0.49					0.67				

Abbreviations : TPR, true positive rate; TNR, true negative rate; PPV, positive predictive value; NPV, negative predictive value. True positive rate: proportion of positive data that are correctly considered as positive. True negative rate: proportion of negative data that are correctly considered as negative. Positive predictive value: number of correct positive results divided by the number of positive results predicted by the classifier. Negative predictive value: number of correct negative results divided by the number of negative results predicted by the classifier. F1-score: the harmonic mean between predictive positive values and true positive rate.

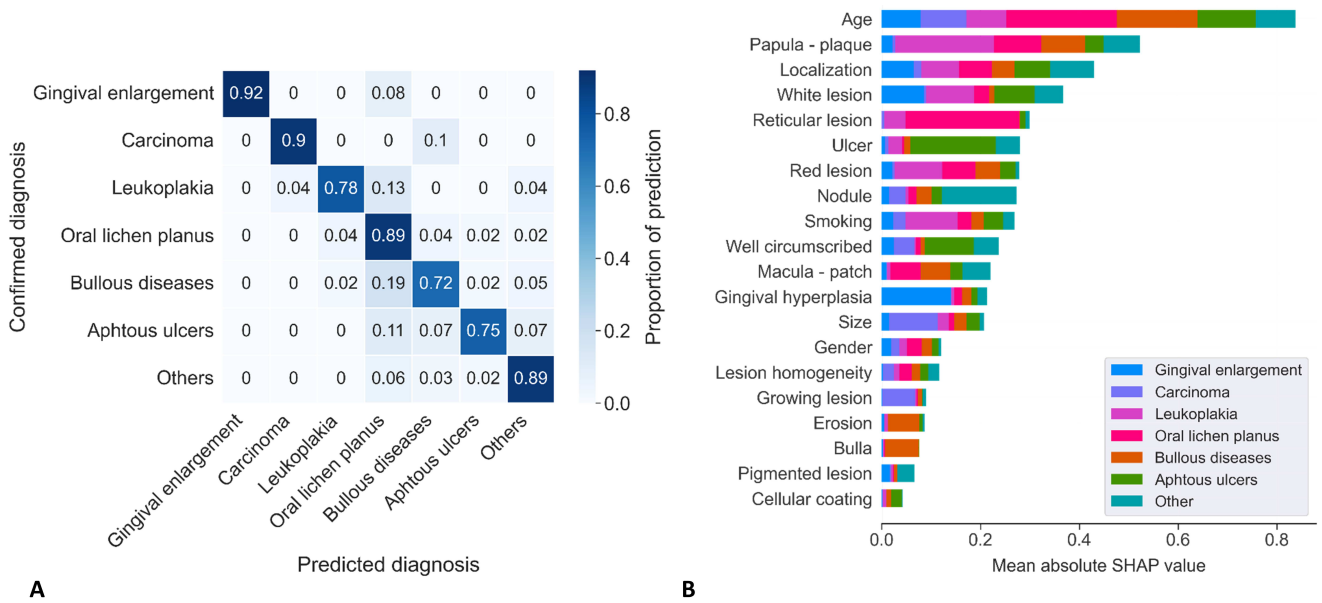


Figure 1. Confusion matrix for LightGBM model and Feature importance for overall predictions. (A) The confusion matrix describes the performance of the LightGBM classifier on the test dataset. The number of correctly classified lesions ranged from 92% to 72% depending on the different diagnoses evaluated. (B) SHAP feature importance for the LightGBM trained model for predicting oral mucosa lesions measured as the mean absolute SHAP values. Abbreviations: ATC A, Anatomical Therapeutic Chemical classification: Alimentary tract and metabolism ; ATC C, Anatomical Therapeutic Chemical classification: Cardiovascular system ; ICD E00-E89, in-ternational classification of diseases: Endocrine, nutritional and metabolic diseases; ICD I00-I99, international classification of diseases: Diseases of the circulatory system; ICD K00-K95, international classification of diseases: Diseases of the digestive system; ICD L00-L99, international classification of diseases: Diseases of the skin and subcutaneous tissue; ICD M00-M99, international classification of diseases: Diseases of the musculoskeletal system and connective tissue.

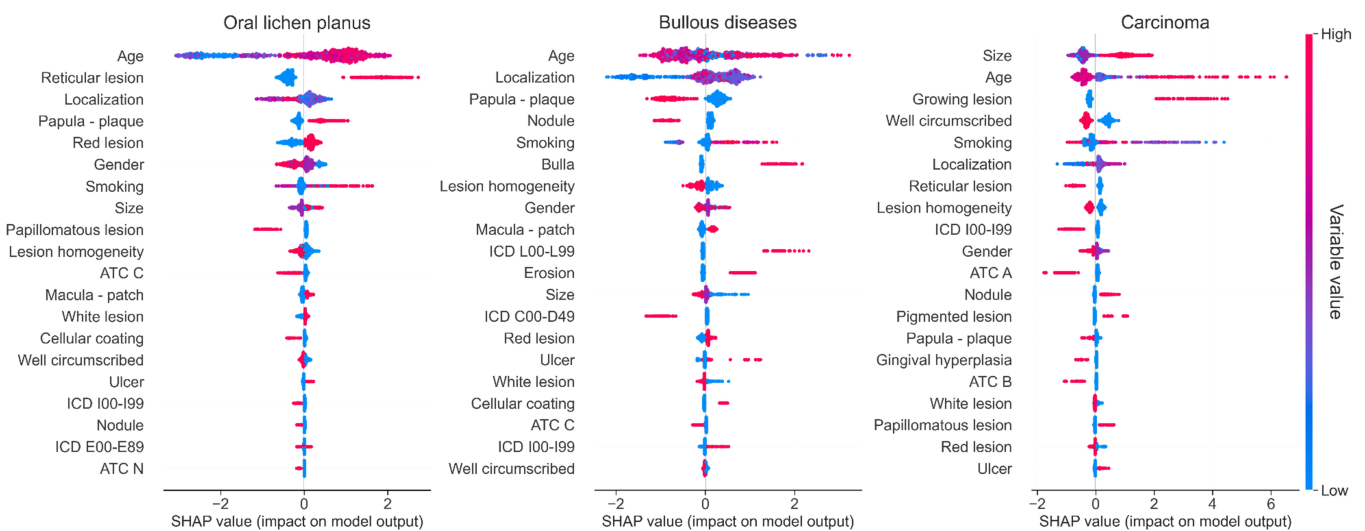


Figure 2. Summary plot of feature importance for each diagnostic prediction for Oral lichen planus, Bullous diseases and Carcinoma. Each point on the summary plot is a SHAP value for a given feature and instance. The position on the y-axis is determined by the feature and on the x-axis by the SHAP value. The color represents the value of the feature from low to high. Overlapping points are jittered in y-axis direction, so we get a sense of the distribution of the SHAP values per feature. The features are ordered according to their importance. Abbreviations: ATC A, Anatomical Therapeutic Chemical

classification: Alimentary tract and metabolism; ATC C, Anatomical Therapeutic Chemical classification: Cardiovascular system; ATC H, Anatomical Therapeutic Chemical classification: Systemic hormonal preparations; ATC M, Anatomical Therapeutic Chemical classification: Musculo-skeletal system; ATC N, Anatomical Therapeutic Chemical classification: Nervous system; ATC R, Anatomical Therapeutic Chemical classification: Respiratory system; ICD C00-D49, international classification of diseases: Neoplasms; ICD E00-E89, international classification of diseases: Endocrine, nutritional and metabolic diseases; ICD F00-F99, international classification of diseases: Mental and behavioral disorders; ICD G00-G99, international classification of diseases: Diseases of the nervous system; ICD I00-I99, international classification of diseases: Diseases of the circulatory system; ICD K00-K95, international classification of diseases: Diseases of the digestive system; ICD L00-L99, international classification of diseases: Diseases of the skin and subcutaneous tissue; ICD M00-M99, international classification of diseases: Diseases of the musculoskeletal system and connective tissue; ICD N00-N99, international classification of diseases: Diseases of the genitourinary system.

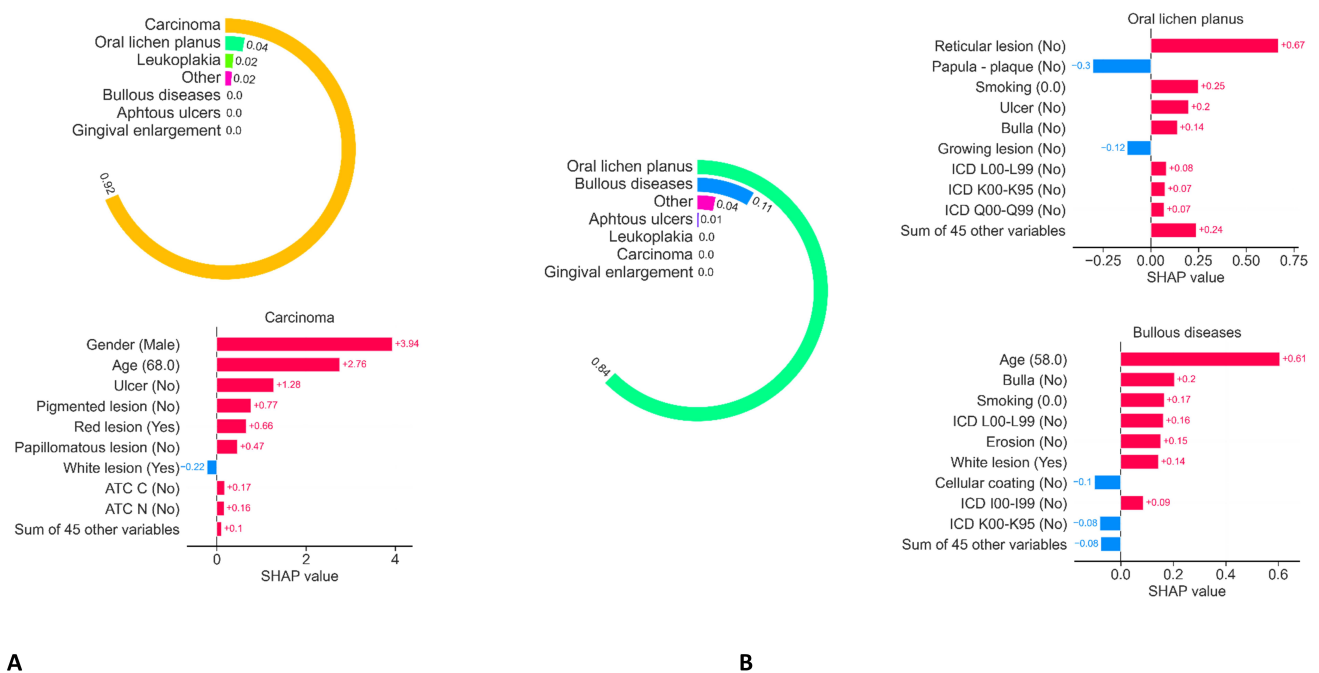


Figure 3. Illustration of two examples of prediction, respectively for a carcinoma (A) and an oral lichen planus lesion (B). The SHAP algorithm provides an understanding of decision making by outlining the factors that positively or negatively influence the different diagnoses. Abbreviations: ATC A, Anatomical Therapeutic Chemical classification: Alimentary tract and metabolism; ATC C, Anatomical Therapeutic Chemical classification: Cardiovascular system; ATC H, Anatomical Therapeutic Chemical classification: Systemic hormonal preparations; ATC M, Anatomical Therapeutic Chemical classification: Musculo-skeletal system; ATC N, Anatomical Therapeutic Chemical classification: Nervous system; ATC R, Anatomical Therapeutic Chemical classification: Respiratory system; ICD C00-D49, international classification of diseases: Neoplasms; ICD E00-E89, international classification of diseases: Endocrine, nutritional and metabolic diseases; ICD F00-F99, international classification of diseases: Mental and behavioral disorders; ICD G00-G99, international classification of diseases: Diseases of the nervous system; ICD I00-I99, international classification of diseases: Diseases of the circulatory system; ICD K00-K95, international classification of diseases: Diseases of the digestive system; ICD L00-L99, international classification of diseases: Diseases of the skin and subcutaneous tissue; ICD M00-M99, international classification of diseases: Diseases of the musculoskeletal system and connective tissue; ICD N00-N99, international classification of diseases: Diseases of the genitourinary system.

4. Discussion

Based on the lesion topography and medical and socio-demographic patients features, LightGBM model allowed a prediction of the 7 oral mucosa lesions selected in this study with an overall accuracy of approximately 0.84. The specificity (true negative rate) and sensitivity (true positive rate) remain very high, ranging from 0.89 to 1.00 and 0.72 to 0.92, respectively; thus, illustrating the discriminatory capacity of our model. Therefore, this tool seems suitable for non-specialist practitioners to guide them in their diagnostic and therapeutic process.

The early diagnosis of oral lesions presents a challenge in treating them promptly and reducing therapeutic penalties, particularly in the case of cancer. Especially because therapeutic progress has increased the life expectancy of these cancer patients. Although oral dermatology is at the crossroads of several medical specialties (Dentists, Ear-Nose-Throat specialists, Oral and maxillofacial surgeons, dermatologists), oral lesions remain underdiagnosed, especially for precancerous and cancerous lesions [26,27]. Indeed, general practitioners often lack specific training and patients face a lack of referral to specialists in the field [28]. On the other hand, while the contribution of AI is booming in medicine and especially in dermatology, only a few models have been developed in oral dermatology [4,8,29]. To answer this problem, we have demonstrated that we can accurately predict the diagnosis of an oral mucosal lesion using an algorithmic ML strategy based on routine oral medicine consultation. Similarly, other models created for the automatic detection of skin cancers have shown the importance of including patient clinical information to optimize these tools [30].

Machine learning algorithms are considered as “black boxes”. Explainability aims at highlighting the importance of each data in the decision-making process to improve transparency, questioning, and understanding.

This explainability increases the confidence of the models by ensuring that the decision making is coherent with the semiological data classically accepted in the literature. Furthermore, the identification of new correlations must be supported by evidence to eliminate misleading decisions. Finally, understanding the models is not only a research objective; it allows the practitioner to be guided in his clinical thinking, to identify variables of high importance and, therefore, optimize the diagnosis of lesions. Moreover, while the explainability of a model can highlight important interactions between certain data, it does not allow us to conclude that there are causal links.

The development of applications accessible to the general public for preventive and diagnostic purposes is an important issue [31]. In 2018, the systematic review by Chuchu et al. [32] evaluated the sensitivity and specificity of smartphone apps used for the diagnosis of melanoma. Although promising, the results showed a lack of precision that led to diagnostic errors compared to the diagnosis made by expert dermatologists. Two years later, Udréa et al. in 2020 [33] improved the system by developing a smartphone application based on a machine learning algorithm that allows the detection of cancerous or precancerous skin lesions with a sensitivity of 95.1% and a specificity of 77.3%. In the diagnosis of oral lesions, by allowing the detection of “high risk cancer patients” using patient’s socio-demographic and behavior data associated with elementary lesion description, the present algorithm can thus constitute a simple tool for specialists and non-specialists to use, with a high precision accuracy.

Today, the most powerful AI models used for diagnosis in dermatology are now based on deep convolutional neural networks. They allow the image of the lesion to be used considering the pixels that constitute it as features in their own right [4,34–36], even associating information about the patient [37]. The use of the photograph itself to determine the diagnosis is the logical continuation of our work, especially to refine the prediction.

Based on the results of such study, our algorithm could be used in routine practice to assist in diagnosis and perform a first triage. In fact, it could make it easier to detect patients with lesions at risk of malignant development. Of course, it does not allow one to date to replace the specialist due to some limitations.

Some issues are currently being raised to train machine learning models. Unlike skin lesions, there is currently no public database available for oral lesions [38,39]. There is also a very important clinical polymorphism of certain oral mucosa lesions, which can complicate the prediction. Finally, as endo-buccal photos are difficult to standardize, it will be necessary to normalize the images as much as possible to compare them [40]. This is an essential step in the development and generalization of this device to predict not only cancerous or precancerous lesions, but also other often misdiagnosed oral mucosal lesions that also have an impact on patient quality of life and well-being.

A limitation of the present study is the limited size of the database and the imbalance between the different diagnoses chosen to predict. Even if this has been taken into account and compensated for in the algorithm, it would be preferable to train it on a larger scale with more cases and a more balanced dataset. However, to date there is no international database on oral mucosal lesions, unlike databases on skin lesions.

5. Conclusions

The use of ML models in medicine is promising, especially due to their ability to compute a very large amount of data to make predictions. Although there are some limitations in our database (size of the database, picture standardization), the results of our study are promising in that they show a high accuracy for predicting cancerous lesions (0.90 accuracy) and an overall good accuracy of 0.84 for five of the other most frequent oral mucosal lesions (Oral Lichen planus, Leukoplakia, bullous diseases, aphthous ulcers, gingival enlargements). More studies with much larger databases should be carried out to validate the accuracy of our model. Moreover, the use of deep learning models could be particularly interesting and should be investigated to optimize diagnostic devices to predict directly from the image of the lesion. These tools create new opportunities for democratization and improvement of patient care. However, the implementation of AI-based applications raises other questions about issues such as the acceptability of AI in health or data protection.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/jcm11216596/s1>, Figure S1: Confusion matrices for the different machine learning models tested on the test dataset. The confusion matrix describes the performance of the difference classifier models on the test dataset. Figure S2: Feature importance for each diagnostic prediction. Each point in the summary plot is an SHAP value for a feature and an instance. The position on the y-axis is determined by the feature and on the x-axis by the SHAP value. The color represents the value of the feature from low to high. The overlapping points are jittered in the y-axis direction, so we get a sense of the distribution of the SHAP values per feature. Features are ordered according to their importance.

Author Contributions: A.D., A.Z., C.T., S.L. and S.C. contributed to acquisition data. A.D., P.M., S.L. and S.C. contributed to analyses and interpretation of data. A.D., A.Z., P.M. and S.L. contributed to the conception and design of the study and the manuscript draft. P.K., P.M., S.L. and S.C. critically revised manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: We gratefully acknowledge the Programme d'Investissements d'Avenir, and the Agence Nationale pour la Recherche for the support of the national infrastructure "ECELLFrance: Development of mesenchymal stem cell-based therapies" (PIA-ANR-11-INBS-005) and for the grant ANR-16-CE18-0019-01.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request from the authors.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
ML	Machine learning
CNN	Convolutional neural network
DL	Deep learning
ICD	International classification of diseases
ATC	Anatomical therapeutic chemical classification system
CNIL	Commission for informatics and liberties
LR	Logistic regression
KNN	K-nearest neighbors
DTR	Decision tree regressor
MLP	Multilayer perceptron
TNR	True negative rate
TPR	True positive rate
PPV	Predictive positive value
PNV	Predictive negative value
ROC	Receiver Operator Characteristic
AUC	Area under curve

References

- Hogarty, D.T.; Su, J.C.; Phan, K.; Attia, M.; Hossny, M.; Nahavandi, S.; Lenane, P.; Moloney, F.J.; Yazdabadi, A. Artificial Intelligence in Dermatology—Where We Are and the Way to the Future: A Review. *Am. J. Clin. Dermatol.* **2019**, *21*, 41–47. [[CrossRef](#)]
- Olsen, T.G.; Jackson, B.H.; Feeser, T.A.; Kent, M.N.; Moad, J.C.; Krishnamurthy, S.; Lunsford, D.D.; Soans, R.E. Diagnostic Performance of Deep Learning Algorithms Applied to Three Common Diagnoses in Dermatopathology. *J. Pathol. Inform.* **2018**, *9*, 32. [[CrossRef](#)] [[PubMed](#)]
- Zahia, S.; Zafirain, M.B.G.; Sevillano, X.; González, A.; Kim, P.J.; Elmaghaby, A. Pressure injury image analysis with machine learning techniques: A systematic review on previous and possible future methods. *Artif. Intell. Med.* **2020**, *102*, 101742. [[CrossRef](#)] [[PubMed](#)]
- Lin, H.; Chen, H.; Weng, L.; Shao, J.; Lin, J. Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis. *J. Biomed. Opt.* **2021**, *26*, 086007. [[CrossRef](#)]
- Day, G.R.; Barbour, R.H. Automated skin lesion screening—A new approach. *Melanoma Res.* **2001**, *11*, 31–35. [[CrossRef](#)] [[PubMed](#)]
- Han, S.S.; Lim, W.; Kim, M.S.; Park, I.; Park, G.H.; Chang, S.E. Interpretation of the Outputs of a Deep Learning Model Trained with a Skin Cancer Dataset. *J. Investig. Dermatol.* **2018**, *138*, 2275–2277. [[CrossRef](#)] [[PubMed](#)]
- Burki, T.K. Machine learning model in melanoma. *Lancet. Oncol.* **2018**, *19*, e340. [[CrossRef](#)]
- Alkhadar, H.; Macluskey, M.; White, S.; Ellis, I.; Gardner, A. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. *J. Oral Pathol. Med.* **2021**, *50*, 378–384. [[CrossRef](#)]
- Thomsen, K.; Iversen, L.; Titlestad, T.L.; Winther, O. Systematic review of machine learning for diagnosis and prognosis in dermatology. *J. Dermatol. Treat.* **2020**, *31*, 496–510. [[CrossRef](#)]
- Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]
- Brinker, T.J.; Hekler, A.; Enk, A.H.; Berking, C.; Haferkamp, S.; Hauschild, A.; Weichenthal, M.; Klode, J.; Schadendorf, D.; Holland-Letz, T.; et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur. J. Cancer* **2019**, *119*, 11–17. [[CrossRef](#)] [[PubMed](#)]
- Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V.I.; Precise4Q Consortium. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med. Inform. Decis. Mak.* **2020**, *20*, 310. [[CrossRef](#)]
- Diprose, W.K.; Buist, N.; Hua, N.; Thurier, Q.; Shand, G.; Robinson, R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. *J. Am. Med. Assoc. JAMIA* **2020**, *27*, 592–600. [[CrossRef](#)] [[PubMed](#)]
- Adeoye, J.; Tan, J.Y.; Choi, S.W.; Thomson, P. Prediction models applying machine learning to oral cavity cancer outcomes: A systematic review. *Int. J. Med. Inform.* **2021**, *154*, 104557. [[CrossRef](#)]
- Idrees, M.; Farah, C.S.; Shearston, K.; Kujan, O. A machine-learning algorithm for the reliable identification of oral lichen planus. *J. Oral Pathol. Med.* **2021**, *50*, 946–953. [[CrossRef](#)] [[PubMed](#)]
- Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)]
- Chu, C.S.; Lee, N.P.; Adeoye, J.; Thomson, P.; Choi, S.W. Machine learning and treatment outcome prediction for oral cancer. *J. Oral Pathol. Med.* **2020**, *49*, 977–985. [[CrossRef](#)]

18. Ilhan, B.; Lin, K.; Guneri, P.; Wilder-Smith, P. Improving Oral Cancer Outcomes with Imaging and Artificial Intelligence. *J. Dent. Res.* **2020**, *99*, 241–248. [[CrossRef](#)]
19. Pan, X.; Zhang, T.; Yang, Q.; Yang, D.; Rwigema, J.C.; Qi, X.S. Survival prediction for oral tongue cancer patients via probabilistic genetic algorithm optimized neural network models. *Br. J. Radiol.* **2020**, *93*, 20190825. [[CrossRef](#)]
20. Jeyaraj, P.R.; Samuel Nadar, E.R. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *J. Cancer Res. Clin. Oncol.* **2019**, *145*, 829–837. [[CrossRef](#)]
21. Howard, F.M.; Kochanny, S.; Koshy, M.; Spiotto, M.; Pearson, A.T. Machine Learning-Guided Adjuvant Treatment of Head and Neck Cancer. *JAMA Netw. Open* **2020**, *3*, e2025881. [[CrossRef](#)] [[PubMed](#)]
22. Gorphe, P.; Jannin, C. Regulatory aspects of prospective and retrospective clinical research in France in 2018. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* **2019**, *136*, 103–108. [[CrossRef](#)] [[PubMed](#)]
23. Buuren, S.V.; Groothuis-Oudshoorn, K. Mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **2011**, *45*, 1–67. [[CrossRef](#)]
24. Lundberg, S.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874.
25. Doumard, E.; Aligon, J.; Escriva, E.; Excoffier, J.; Monsarrat, P.; Soulé-Dupuy, C. A Comparative Study of Additive Local Explanation Methods based on Feature Influences. In Proceedings of the 24th International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data (DOLAP) Co-Located with the 25th International Conference on Extending Database Technology and the 25th International Conference on Database Theory (EDBT/ICDT 2022), Edinburgh, UK, 29 March 2022; Stefanidis, K., Golab, L., Eds.; Volume 3130, pp. 31–40.
26. Abati, S.; Bramati, C.; Bondi, S.; Lissoni, A.; Trimarchi, M. Oral Cancer and Precancer: A Narrative Review on the Relevance of Early Diagnosis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 9160. [[CrossRef](#)] [[PubMed](#)]
27. Seoane, J.; Alvarez-Novoa, P.; Gomez, I.; Takkouche, B.; Diz, P.; Warnakulasiruya, S.; Seoane-Romero, J.M.; Varela-Centelles, P. Early oral cancer diagnosis: The Aarhus statement perspective. A systematic review and meta-analysis. *Head Neck* **2016**, *38* (Suppl. 1), 2182–2189. [[CrossRef](#)]
28. Heelan, K.; McKenna, D. A survey of oral medicine education, training and practice among dermatologists in the UK and Ireland. *Br. Dent. J.* **2016**, *220*, 17–20. [[CrossRef](#)]
29. Khanagar, S.B.; Al-Ehaideb, A.; Maganur, P.C.; Vishwanathaiah, S.; Patil, S.; Baeshen, H.A.; Sarode, S.C.; Bhandi, S. Developments, application, and performance of artificial intelligence in dentistry—A systematic review. *J. Dent. Sci.* **2021**, *16*, 508–522. [[CrossRef](#)]
30. Pacheco, A.G.C.; Krohling, R.A. The impact of patient clinical information on automated skin cancer detection. *Comput. Biol. Med.* **2020**, *116*, 103545. [[CrossRef](#)]
31. Abbott, L.M.; Smith, S.D. Smartphone apps for skin cancer diagnosis: Implications for patients and practitioners. *Australas. J. Dermatol.* **2018**, *59*, 168–170. [[CrossRef](#)]
32. Chuchu, N.; Takwoingi, Y.; Dinnes, J.; Matin, R.N.; Bassett, O.; Moreau, J.F.; Bayliss, S.E.; Davenport, C.; Godfrey, K.; O’Connell, S.; et al. Smartphone applications for triaging adults with skin lesions that are suspicious for melanoma. *Cochrane Database Syst. Rev.* **2018**, *12*, CD013192. [[CrossRef](#)] [[PubMed](#)]
33. Udrea, A.; Mitra, G.D.; Costea, D.; Noels, E.C.; Wakkee, M.; Siegel, D.M.; de Carvalho, T.M.; Nijsten, T.E.C. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *J. Eur. Acad. Dermatol. Venereol. JEADV* **2020**, *34*, 648–655. [[CrossRef](#)] [[PubMed](#)]
34. El-Khatib, H.; Popescu, D.; Ichim, L. Deep Learning-Based Methods for Automatic Diagnosis of Skin Lesions. *Sensors* **2020**, *20*, 1753. [[CrossRef](#)]
35. Nasiri, S.; Helsper, J.; Jung, M.; Fathi, M. DePicT Melanoma Deep-CLASS: A deep convolutional neural networks approach to classify skin lesion images. *BMC Bioinform.* **2020**, *21*, 84. [[CrossRef](#)] [[PubMed](#)]
36. Winkler, J.K.; Sies, K.; Fink, C.; Toberer, F.; Enk, A.; Deinlein, T.; Hofmann-Wellenhof, R.; Thomas, L.; Lallas, A.; Blum, A.; et al. Melanoma recognition by a deep learning convolutional neural network-Performance in different melanoma subtypes and localisations. *Eur. J. Cancer* **2020**, *127*, 21–29. [[CrossRef](#)] [[PubMed](#)]
37. D, S.R.; A, S. Deep Learning Based Skin Lesion Segmentation and Classification of Melanoma Using Support Vector Machine (SVM). *Asian Pac. J. Cancer Prev. APJCP* **2019**, *20*, 1555–1561. [[CrossRef](#)]
38. Bajwa, M.N.; Muta, K.; Malik, M.I.; Siddiqui, S.A.; Braun, S.A.; Homey, B.; Dengel, A.; Ahmed, S. Computer-Aided Diagnosis of Skin Diseases Using Deep Neural Networks. *Appl. Sci.* **2020**, *10*, 2488. [[CrossRef](#)]
39. Rotemberg, V.; Kurtansky, N.; Betz-Stablein, B.; Caffery, L.; Chousakos, E.; Codella, N.; Combalia, M.; Dusza, S.; Guitera, P.; Gutman, D.; et al. A patient-centric dataset of images and metadata for identifying melanomas using clinical context. *Sci. Data* **2021**, *8*, 34. [[CrossRef](#)]
40. Cullell-Dalmau, M.; Otero-Viñas, M.; Manzo, C. Research Techniques Made Simple: Deep Learning for the Classification of Dermatological Images. *J. Investig. Dermatol.* **2020**, *140*, 507–514.e1. [[CrossRef](#)]

Titre : Utilisation de nouvelles approches en intelligence artificielle pour le diagnostic de lésions de la muqueuse orale

Mots clés : Intelligence artificielle, Muqueuse buccale, Cancers oraux

Résumé : Les pathologies de la muqueuse buccales présentent un polymorphisme clinique important. Cela concerne les dermatoses bénignes mais également les lésions cancéreuses et pré-cancéreuses. Un diagnostic précoce reste un élément clé de la prise en charge, en particulier des cancers de la cavité buccale. Cependant ce diagnostic demeure aujourd'hui trop tardif, que cela soit lié à la difficulté inhérente à la présentation clinique ou aux délais de rendez-vous en consultations de recours auprès de spécialistes du domaine, en nombre insuffisant. Il paraît donc utile de proposer des outils d'aide au diagnostic, fiables et destinés aux professionnels de la santé orale. Bien que l'intelligence artificielle utilisée dans le diagnostic dermatologique soit en plein essor, son application à la pathologie de la cavité buccale reste plus anecdotique. Nous avons adapté et évalué différentes approches d'intelligence artificielle puis développé un outil à destiné des professionnels de santé.

Title: Use of innovative artificial intelligence approaches for the diagnosis of oral mucosal lesions

Key words: Artificial Intelligence, Oral mucosa, Oral cancer

Abstract: Clinical manifestations of oral mucosal pathologies exhibit significant polymorphism. This includes benign dermatoses, cancerous or even pre-cancerous lesions. Early diagnosis remains a key element of management, especially for oral cavity cancers. However, this diagnosis is currently often delayed, whether due to the inherent difficulty in clinical presentation or insufficient availability of specialist consultations, leading to extended waiting times. Therefore, it appears useful to propose reliable diagnostic support tools for oral healthcare professionals. While artificial intelligence in dermatological diagnosis is advancing rapidly, its application to oral cavity pathology is still relatively limited. We adapted and evaluated various artificial intelligence approaches and developed a tool intended for healthcare professionals.