République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mouloud Mammeri de Tizi-Ouzou

Faculté de Génie Électrique et d'Informatique

Département d'Informatique



Projet de Fin d'Etude

En vue de l'obtention du diplôme de Master 2 en Informatique

Spécialité : Système Informatique Intelligent

SII

Un Système d'aide au Diagnostic du Cancer du sein dans la Médecine traditionnelle chinoise

SMAILI Maya et **MORSLI Manel**

Soutenu publiquement le 02/07/2025

Devant le jury composé de

Mme FARHAOUI Chafia	Présidente
Mme BOURKACHE Ghenima	Examinateur
Mme YESLI Yasmine	Encadrante

Promotion 2024/2025

Résumé

Ce mémoire présente le développement d'un système intelligent d'aide au diagnostic précoce du cancer du sein basé sur l'analyse de la langue selon les principes de la Médecine Traditionnelle Chinoise (MTC). Face à l'incidence croissante du cancer du sein, touchant une femme sur huit dans le monde, et aux limitations des méthodes de dépistage conventionnelles, ce travail explore une approche non invasive combinant savoirs ancestraux et technologies modernes.

L'architecture proposée, nommée YOLO_BCTCM, intègre plusieurs composantes innovantes : YOLOv8 pour la détection automatique de la langue (100% de mAP50précision), YOLOv11 pour l'identification de 16 caractéristiques pathologiques spécifiques (71,53% de mAP50précision), et des algorithmes de machine learning (notamment XGBoost 82,69 26% de CV) pour la classification en trois stades cliniques (sain, précoce, avancé). Face à l'absence de datasets publics, nous avons développé une approche novatrice de génération de données synthétiques guidée par l'IA, créant le dataset "Tongue_MTCBC_Hope".

Les résultats expérimentaux démontrent des performances remarquables : diagnostic du stade avancé avec 99,38% de confiance, identification du stade précoce à 95,5%, et classification de l'état sain à 97,4%. L'ensemble est intégré dans une plateforme web interactive "MTC Hybrid Pro", rendant le système accessible aux praticiens. Cette recherche ouvre la voie à une médecine préventive intelligente, particulièrement pertinente pour les régions à accès limité aux technologies du dépistage conventionnelle.

Mots-clés : Cancer du sein, Médecine Traditionnelle Chinoise, Diagnostic par la langue, Intelligence artificielle, YOLO, Vision par ordinateur,Apprentissage automatique, Réseau de neurones profonds, Traitement d'image, Système d'aide au diagnostic précoce, Streamlit, Roboflow, SAM, IA Generative.

Abstract

This thesis presents the development of an intelligent system for early breast cancer diagnosis based on tongue analysis according to Traditional Chinese Medicine (TCM) principles. Facing the increasing incidence of breast cancer, affecting one in eight women worldwide, and the limitations of conventional screening methods, this work explores a non-invasive approach combining ancestral knowledge with modern technologies.

The proposed architecture, named YOLO_BCTCM, integrates several innovative components: YOLOv8 for automatic tongue detection (100% mAPaccuracy), YOLOv11 for identifying 16 specific pathological characteristics (71,53% mAP50), and machine learning algorithms (particularly XGBoost 82,69 % cross validation) for classification into three clinical stages (healthy, early, advanced). Given the absence of public datasets, we developed an innovative approach for AI-guided synthetic data generation, creating the "Tongue_MTCBC_Hope" dataset.

Experimental results demonstrate remarkable performance: advanced stage diagnosis with 99.38% confidence, early stage identification at 95.5%, and healthy state classification at 97.4%. The system is integrated into an interactive web platform "MTC Hybrid Pro", making it accessible to practitioners. This research paves the way for intelligent preventive medicine, particularly relevant for regions with limited access to conventional screening technologies.

Keywords: Breast cancer, Traditional Chinese Medicine, Tongue diagnosis, Artificial intelligence, YOLO, Computer vision, Machine learning, Deep neural networks, Image processing, Early diagnosis support system, Streamlit, Roboflow, SAM, Generative AI.

Avant-propos

Ce mémoire représente l'aboutissement de deux années de formation en Master Systèmes Informatiques Intelligents à l'Université Mouloud Mammeri de Tizi-Ouzou, mais aussi le fruit d'une réflexion plus profonde sur le rôle que peut jouer l'intelligence artificielle dans l'amélioration de la santé publique.

Notre travail a consisté à développer une architecture hybride innovante combinant les technologies de vision par ordinateur les plus récentes (YOLO) avec les principes diagnostiques de la Médecine Traditionnelle Chinoise. Face à l'absence de données cliniques, nous avons relevé le défi en créant un dataset synthétique via l'intelligence artificielle générative, puis en construisant un système complet de détection et de classification des caractéristiques pathologiques de la langue selon la Medecine Traditionnelle chinoise. Ce parcours, jalonné d'obstacles techniques et méthodologiques, nous a enseigné la persévérance et l'importance de l'innovation face aux contraintes.

Nous espérons que ce travail, au-delà de sa contribution académique, ouvrira la voie à de nouvelles approches de diagnostic médical. Si notre système peut contribuer, même modestement, à sauver ne serait-ce qu'une vie par un diagnostic précoce, alors tous nos efforts auront été justifiés.
Ce mémoire est dédié à toutes les femmes qui luttent contre le cancer du sein, à celles qui nous ont quittées trop tôt, et à celles qui, grâce aux avancées technologiques, pourront bénéficier d'un diagnostic plus précoce et d'un meilleur pronostic.

Puisse ce travail inspirer d'autres chercheurs à explorer les ponts entre tradition et modernité, entre sagesse humaine et intelligence artificielle, pour construire ensemble une médecine plus inclusive et plus efficace.

Maya SMAILI et Manel MORSLI
Tizi-Ouzou, Juin 2025.

Remerciement

On tient à exprimer notre sincère gratitude à tous ceux qui ont contribué à la réalisation de ce travail. Votre soutien, vos conseils et vos encouragements ont été essentiels tout au long de ce parcours.

On remercie tout particulièrement notre promotrice, Madame **YESLI Yasmine**, pour son accompagnement, sa bienveillance et ses précieux conseils tout au long de ce travail.

On souhaite également exprimer notre reconnaissance envers l'ensemble des enseignants du **Master Systèmes Informatiques Intelligents (SII)**, pour la qualité de leur enseignement et leur engagement constant à nous transmettre leur savoir.

Notre gratitude va aussi à l'ensemble de la communauté de l'Université Mouloud Mammeri de Tizi-Ouzou.

Enfin, on remercie chaleureusement les membres du **jury** pour l'attention portée à notre travail, ainsi que pour leurs observations et recommandations enrichissantes.

Je dédie ce travail ...

à tous ceux qui ont été ma force en silence.

À mes premiers guides, MAMA et PAPA

À mes précieuses sœurs, SARA et NILDA

À ma chère amie et binôme, MANEL

À mon être cher, MUSTAPHA

`while(distance): heart.keep("connectedfamily")`
© SMAILI Maya -2024/2025-

Je dédie ce travail ...

à tous ceux qui ont été ma force en silence.

À mes premiers guides, MAMA et PAPA

À mes précieuses sœurs, Meriem et Maroua

À mon chèr frère, Mohammed Mehdi

À ma chère amie et binôme, Maya

À mes chères proches amies Houda

Farah, Razika et Lydia

© MORSLI Manel -2024/2025-

"

Table des matières

Liste des figures . . . .
Liste des tableaux . . . .

I Principes de la Médecine Traditionnelle Chinoise (MTC) 5
I.1 Introduction 6
I.2 Histoire de la MTC 6
I.3 Concept fondamental de MTC 7
I.3.1 Le Qi, énergie vitale 7
I.3.1.1 Les types de Qi : 7
I.3.1.2 Nourrir et équilibrer le Qi : 7
I.3.2 Le Yin et le Yang 8
I.3.3 Le système de 5 éléments 8
I.3.3.1 Cycle dengendrement 10
I.3.3.2 Le cycle de contrôle ou de domination 10
I.3.4 Les émotions en MTC 11
I.3.5 Les six excès 12
I.3.6 Les 12 méridiens (Zhang Fu) 12
I.4 L'application de MCT dans le diagnostic médical 14
I.4.1 Les Quatres Méthodes de Diagnostic en MTC 14
I.4.2 Examen de la langue 15
I.4.3 Le Traitement en MTC 17
I.4.3.1 LAcupuncture 17
I.4.3.2 La Moxibustion 18
I.4.3.3 La Pharmacopée Chinoise 18
I.4.3.4 Le Qi Gong et le Tai Chi 18
I.4.3.5 Le Massage Tui Na 18
I.4.3.6 La Diététique Chinoise 18
I.5 Conclusion 19

II Apprentissage automatique 20
II.1 Introduction 21
II.2 Machine Learning (ML) 22
II.2.1 Apprentissage supervisé 23
II.2.1.1 Régression 23
II.2.1.2 classification 24
II.2.2 Apprentissage non supervisé 24
II.2.3 Apprentissage semi-supervisé 24
II.2.4 Apprentissage par renforcement 25
II.3 Réseaux de neurones et Deep Learning 28

II.3.1 Réseaux de neurones Artificiels . . . 28
II.3.1.1 Perceptron simple . . . 29
II.3.1.2 RNA a une seule couche cache . . . 30
II.3.2 Réseaux de neurones profonds et Apprentissage Profond . . . 31
II.3.2.1 Réseaux de neurones profonds . . . 31
II.3.2.2 Apprentissage profond . . . 32
II.3.3 Architectures des RNP . . . 32
II.3.3.1 Réseaux Feedforward . . . 33
II.3.3.2 Réseaux Récurrents (RNN) . . . 34
II.3.3.3 Réseaux Convolutifs . . . 36
II.3.3.4 Réseaux de neurones à auto-encodeurs (Auto-encoders) . . . 39
II.3.3.5 Transformers . . . 39
II.3.3.6 IA Génératif . . . 42
II.4 Processus de création dun RNP . . . 43
II.5 Entraînement dun RNP . . . 43
II.5.1 Mise à jour des paramètres . . . 44
II.5.2 Pour la régression . . . 44
II.5.3 Pour la classification . . . 45
II.6 Évaluation dun modèle . . . 45
II.6.1 La matrice de confusion : . . . 45
II.6.2 Critères dévaluation basés sur la matrice de confusion . . . 46
II.7 Approches avancées de lapprentissage automatique dans les RNP . . . 47
II.7.1 Transfer Learning . . . 47
II.7.2 Few shot Learning . . . 47
II.8 Conclusion . . . 48

III Vision Artificielle et Algorithme Yolo 49
III.1 Introduction . . . 50
III.2 Vision Artificielle . . . 50
III.3 Traitement dimages . . . 51
III.3.1 Limage . . . 51
III.4 Pixel et résolution dimage . . . 52
III.4.1 Représentation dune image en mémoire . . . 52
III.4.2 Résolution dimage . . . 53
III.4.3 Notions fondamentales du traitement dimage . . . 53
III.5 Techniques de traitement dimage . . . 54
III.5.1 Seuillage . . . 55
III.5.2 Histogramme . . . 56
III.5.3 Segmentation dimage . . . 57
III.5.4 Filtrage en traitement dimage . . . 58
III.5.4.1 Filtrage linéaire . . . 58
III.5.4.2 Filtrage non linéaire . . . 58
III.5.5 Techniques de Deep Learning en Vision par Ordinateur . . . 59
III.6 Évaluation des performances des modèles de VA . . . 60
III.6.1 Tâche de segmentation dimages . . . 60
III.6.2 Tâche de détection dobjets . . . 60
III.6.3 Tâche de classification dimages . . . 61

III.7 YOLO: You Only Look Once . . . 62
III.7.1 Fonctionnement de YOLO . . . 62
III.7.1.1 Découpage virtuel de limage en grille . . . 62
III.7.1.2 Réseau neuronal unique et prédictions . . . 63
III.7.1.3 Post-traitement (NMS –Non-Maximum Suppression) . . . 63
III.7.2 Caractéristiques YOLO . . . 63
III.7.3 Versions de Yolo . . . 64
III.7.4 Architecture YOLO . . . 69
III.8 Conclusion . . . 72

IV État de lart sur le diagnostic automatisé du cancer du sein dans la MTC 73
IV.1 Introduction . . . 74
IV.2 Diagnostic du cancer de sein en médecine conventionnelle . . . 74
IV.2.1 Présentation du cancer du sein . . . 74
IV.2.2 Les méthodes actuelles de dépistage . . . 76
IV.2.3 Les méthodes actuelles de dépistage combinée avec l'IA . . . 77
IV.3 Diagnostic de la langue dans la MTC . . . 77
IV.3.1 Travaux de recherche pour le diagnostic automatisé par la langue en MTC . . . 79
IV.3.1.1 Analyse de langue à distance, automatisé et en temps réel avec Multi-Label . . . 79
IV.3.1.2 Une évaluation holistique et personnalisée de l'état de santé du patient . . . 80
IV.3.1.3 Détection des anomalies de santé avec langue via une application mobile iTonguet . . . 81
IV.3.1.4 Diagnostic des marques de dents sur la langue avec lIA . . 81
IV.3.1.5 Reconnaissance automatique des marques dentaires sur la langue par détection faiblement supervisée . . . 82
IV.3.2 Travaux de recherche pour le diagnostic par la langue pour des tumeurs et de maladies spécifiques en MTC . . . 83
IV.3.2.1 Analyse clinique des signes linguaux chez les patients atteints de cancer . . . 83
IV.3.2.2 A Non-Invasive Interpretable NAFLD Diagnostic Method Combining TCM Tongue Features (2023) . . . 85
IV.3.2.3 Analyse objective des syndromes en MTC chez les patients avec le diabète et prédiabète . . . 86
IV.3.3 Synthèse sur les systèmes de diagnostic par la langue en MTC . . . 89
IV.4 Diagnostic du cancer du sein dans la vision MTC . . . 90
IV.4.1 Travaux de recherche sur le diagnostic du cancer du sein avec la langue en MTC . . . 91
IV.4.1.1 Une perspective observation de la langue et cancer du sein avec la MTC . . . 92
IV.4.1.2 Indice de diagnostic du cancer du sein à un stade précoce avec la MTC . . . 93
IV.4.1.3 Une analyse des caractéristiques de la langue chez les patients atteint le cancer du sein : . . . 94

IV.4.1.4 Système de diagnostic et de recommandation de traitement du cancer du sien dans la MTC 95
IV.4.2 Travaux de recherche qui sollicitent lintégration de la MTC dans le traitement des personnes atteintes du Cancer du sein 97
IV.4.2.1 La relation entre la fatigue en MTC et le cancer du sein . 97
IV.4.2.2 Traditional Chinese Medicine for Breast Cancer: A Review 98
IV.4.2.3 Guide de prise en charge des troubles émotionnelle chez les patients atteins le cancer du sein 99
IV.4.3 Synthèse et interprétation en MTC des patients atteints de cancer du sein 100
IV.5 Systèmes daides au diagnostic (SAD) développés pour la MTC 101
IV.5.1 SymMap 101
IV.5.2 Shen-Acupuncture 102
IV.5.3 i-Tongue Diagnosis 103
IV.6 conclusion 104

V Contribution **105**
V.1 Introduction 106
V.2 Jeu de données (Dataset) 106
V.2.1 Problème rencontrée 106
V.2.2 Collecte de données réelles : une solution peu réalisable à court terme 107
V.2.3 Génération des données par lIA 107
V.2.4 Preuve de la faisabilité de lapproche 108
V.2.5 Versions de génération 108
V.2.5.1 Première tentative : un script Python 108
V.2.5.2 Deuxième tentative : Synthèse dimages IA 109
V.2.5.3 Troisième tentative : Réalité augmentée 109
V.2.5.4 Quatrième tentative : GAN –Stable Diffusion et ControlNet 110
V.2.5.5 Cinquième tentative : IA guidée par prompt médical . . . 110
V.3 Metodologie de la recherche 111
V.4 Module de prétraitement du dataset 113
V.4.1 Etape1 : Annotation manuelle de la langue avec RoboFlow 114
V.4.2 Etape2 : Segmentation de la langue avec SAM 116
V.4.3 Etape 3 : Annotation manuelle des caractéristiques de la langue avec RoboFlow 117
V.5 Module de détection de la langue pendant le diagnostique automatisée . . . 119
V.6 Module Détection des caractéristiques pathologiques et diagnostic précoce du cancer du sein 120
V.6.1 Étape 1 : Détection des caractéristiques de la langue avec YOLOv11 121
V.6.1.1 Augmentation des données 121
V.6.2 Étape 2 : Simulation de la cartographie avec les caractéristiques . . 123
V.6.3 Étape 3 : Classification avec les algorithmes de machine learning . . 123
V.7 Modèle étendu 124
V.7.1 Étape 1 : Création du dataset structuré 125
V.7.2 Étape 2 : Détection des zones symptomatiques et interprétation clinique 125
V.7.3 Étape 3 :Dialogue médical personnalisé via modèle de langage . . . 125

V.8 Construction dune plateforme de diagnostic du cancer du sein selon MTC . 126
V.8.1 Besoins Fonctionnels 126
V.8.2 Besoins Non-Fonctionnels 126
V.8.3 Developpement agile et Conception UML 127
V.8.4 Architecture modulaire de la platform 128
V.8.4.1 Phase 1 –Téléversement et détection de la langue 128
V.8.4.2 Phase 2 –Détection des caractéristiques pathologiques . . . 128
V.8.4.3 Phase 3 –Prédiction du stade clinique 129
V.8.4.4 Phase 4 –Restitution du diagnostic 129
V.9 Conclusion 129

VI Implémentation et tests **130**
VI.1 Introduction 131
VI.2 Entraînement de YOLOv8 pour la détection de la langue 131
VI.2.1 Environement de développement 132
VI.2.2 Les hyperparamètres 132
VI.2.3 Évolution des pertes (loss) pendant lentraînement 132
VI.2.4 Évolution des métriques de performance pendant lentraînement . . 133
VI.3 Entraînement de YOLOv11 pour la détection des caractéristiques linguales 135
VI.3.1 Prétraitement automatisé du dataset 135
VI.3.2 Hyperparamètres dentraînement YOLOv11 optimisé 135
VI.3.3 Évolution des pertes (loss) pendant lentraînement 136
VI.3.4 Évolution des métriques de performance pendant lentraînement . . 137
VI.3.5 Performances du modèle préentraîné à détecter les classes dans la validation 138
VI.4 Adaptation du modèle YOLOv11 préentraîné via réentraînement 140
VI.4.1 Performances globales du modèle réentraîné 140
VI.4.2 Performances du modèle réentraîné à détecter les classes dans la validation 142
VI.4.3 Modèle entraîné sur Roboflow 144
VI.5 Système de diagnostic précoce du cancer du sein 145
VI.5.1 Pipeline général du système de diagnostic hybride (YOLO + ML) . 145
VI.5.2 Hybridation du modèle YOLOv11 avec des modèles de classification de ML 146
VI.5.2.1 Méthodologie d'entraînement des classificateurs 146
VI.5.2.2 Paramètres utilisés pour chaque modèle ML 147
VI.5.2.3 Mesure de performance 147
VI.5.3 Test et résultats 148
VI.6 Plateforme interactive pour le diagnostic 151
VI.6.1 Environnement de développement 151
VI.6.2 Architecture de la plateforme 152
VI.6.3 Déploiement et accessibilité 153
VI.6.4 Test de la plateforme 153
VI.7 Conclusion 154

VII Conclusion générale et perspectives **155**

Annexe **163**

A Annexe **163**

I.1 Diagramme de cas d'utilisation 164

I.2 Diagramme de séquences 165

I.3 Diagramme de classe 166

B Annexe **167**

II.1 Annexe:Tests de chargement d'images dans la plateforme 168

Table des figures

1 Les statistiques des cancer chez les femmes en 2022 selon l'IARC 1

I.1 Le Yin et le Yang. 8
I.2 Les 5 Éléments de la MTC. 9
I.3 Le système des 5 éléments Cycle d'engendrement. 10
I.4 Le système des 5 éléments Cycle de contrôle ou de domination. 11
I.5 Les méridiens du corps. 13
I.6 Correspondance entre foyers énergétiques et les différentes parties de la langue. 15
I.7 Les parties de la langue correspondant aux organes. 16
I.8 Illustration des caractéristiques de la langue. 17

II.1 les différents domaines de l'intelligence artificielle. 21
II.2 Types d'apprentissage Automatique 22
II.3 Illustration de l'étiquetage des images pour l'apprentissage automatique . . 23
II.4 Illustration dun neurone biologique. 28
II.5 Le fonctionnement des neurones artificiels. 30
II.6 Un RNA a une seule couche cachée. 31
II.7 La différence entre un RNA avec une seule couche cachée (Single layer), et un RNP avec plusieurs couches cachées. 32
II.8 les différentes architectures des RNPs 33
II.9 Perceptron multicouches avec une couche cachée pour la classification des chiffres manuscrites. 34
II.10 Architecture des réseaux de neurones récurrents 35
II.11 Architecture des CNNs. 37
II.12 Architecture de VGG-16. 38
II.13 L'Architecture de UNet. 39
II.14 Architecture des GANs. 42
II.15 Rétropropagation dans un Réseau de Neurones Artificiels 44

III.1 Image vectorielle Vs Image matricielle. 52
III.2 Représentation dune image couleur en mémoire (RVB)[1] 53
III.3 Différentes techniques de traitement d'image 54
III.4 Processus de conversion d'une image couleur en image binaire par seuillage. 56
III.5 Exemple dhistogrammes. 56
III.6 Exemple de types de segmentation. 57
III.7 Comparaison des effets des filtres linéaires et non linéaires sur une image bruitée. 59
III.8 Comparaison des performances des différents modèles de détection d'objets YOLO en termes de mAP (Mean Average Precision) et de latence. 64

III.9 YOLO11 identifie les anomalies dans les IRM cérébrales [2]. 65
III.10Comparaison des cartes thermiques entre YOLOv10, YOLOv11 et le YOLOv12 proposé. 65
III.11Architecture du modèle YOLO dans sa 1ers version 69
III.12Architecture du modèle YOLO version 11 69
III.13Processus global de fonctionnement de YOLO (pipeline) 70

IV.1 Statistiques du cancer du sein en Algérie. 74
IV.2 Schéma de la structure interne d'un sein 75
IV.3 Illustration des différents stades du cancer du sein. 75
IV.4 Sensibilité et spécificité des principales méthodes de dépistage du cancer du sein 76
IV.5 Les différents aspects observés sur la langue. 78
IV.6 Architecture du système Dr. Tongue. 79
IV.7 Diagramme de l'architecture du modèle Cv-Swin. 80
IV.8 Instrument de diagnostic de la langue TFDA-1 81
IV.9 Architecture de Ammonia-net 82
IV.10Cadre global du modèle. L'approche comprend la première étape de l'extraction automatique de la langue et la deuxième étape de la langue de dents faiblement supervisée détection 83
IV.11Etude clinique sur 250 personnes attiens des tumeurs malignes avant la chimiothérapie 84
IV.12Comparaison des singes de langue avant et après traitement (Chimiothérapie et radiothérapie) 84
IV.13Larchitecture du modèle SelectorNet 86
IV.14Figure (a): les principaux méridiens impliqués dans le développement du cancer du sein , figure (b): une langue typiquement affectée par ces déséquilibres 91
IV.15Les zones de la langue (a) et les composants de ATDS en 2013 (b) 92
IV.16La 2eme version du système ATDS, 2015 94
IV.17Diagramme typique de l'hybridation utilisée dans cet article. 97
IV.18La structure de la BDD SymMapv2 101
IV.19La fig.(a) présente la plateforme de Shen-Acupuncture, et la fig. (b) représente la page d'index de l'application Shen-Acupuncture sous Android. . . 102
IV.20La figure (a) montre le flux complet de l'application iTongue, la figure (b) représente l'application téléchargeable dans l'App Store iOS. 103

V.1 Exemples issus de la première tentative de génération 109
V.2 Exemples de la deuxième tentative de génération 109
V.3 Exemples issus du dataset Tongue_MTCBC_Hope. (1): saine, (2): précoce, (3): avancée 111
V.4 Architecture du système YOLO_BCTCM 112
V.5 Le module prétraitement du dataset 113
V.6 Interface de Roboflow lors de la création du dataset. 114
V.7 Annotation de la figure (a) de langue avec une Bbox dans la figure (b). . . 115
V.8 La différence entre la Bbox et le Polygone, (a):une langue annotée avec une Bbox, (b):une langue annotée avec un polygone. 115

V.9 La figure illustre comment SAM segmente l'image (a) à l'image (b). 116
V.10 Quelques exemples déchecs du modèle SAM à segmenter la langue. 117
V.11 les 16 classes définies dans notre jeu de données, et leur fréquence dapparition 118
V.12 Interface de Roboflow lors de l'annotation des caractéristiques. 118
V.13 Transition directe dune image brute de langue (a) et sa version annotée (b) dans Roboflow. 119
V.14 Exemples issus du dataset Tongue_MTCBC_Hope 120
V.15 Le module détection des caractéristiques et diagnostic 120
V.16 Exemples de transformation avancée utilisée pour le modèle. 122
V.17 Modèle étendue 125
V.18 Schématisation MTC Hybrid Pro 128

VI.1 Prétraitement du dataset de la détection de la langue sur Roboflow. 131
VI.2 Évolution de chaque fonction de perte par epoch YOLO8. 132
VI.3 Evolution des métriques de performance par epoch Yolo8. 134
VI.4 Résultat de validation de YOLOv8. (a) présente la matrice de confusion, figure (b) présente les masque de segmentation sur les données de validation. 134
VI.5 Evolution de chaque fonction de perte par epoch yolo11. 136
VI.6 Evolution des métriques de performance par epoch yolo11. 137
VI.7 Résultats de performance du modèle YOLOv11 par classe. 139
VI.8 Matrice de confusion normalisée sur les classes de caractéristiques linguales 139
VI.9 Les scores des métriques de performances du modèle YOLO11 réentraînée. 140
VI.10 Evolution de la performance du modèle yolo11 réentraînée. 141
VI.11 Matrice de confusion du modèle yolo11 réentraînée. 142
VI.12 les performances du modèle yolo11 réentraîner a détecter chaque classe. . . 143
VI.13 Un exemplaire de validation du modèle yolo11 réentraînée. 143
VI.14 Les différents modèles yolo11 entrainées dans Roboflow. 144
VI.15 Exemples déchec partiel de détection du modèle de Roboflow, confusion entre classes et mauvaise localisation. 144
VI.16 Exemples de détection de notre modèle yolo11 145
VI.17 Performances des modèles ML entraînée pour la classification. 148
VI.18 Diagnostic dun stade avancé avec YOLO_MTCBC. 149
VI.19 Sortie du terminal détaille sur le diagnostic d'un stade avancée avec YOLO_-MTCB 149
VI.20 Diagnostic dun stade précoce avec YOLO_MTCBC. 150
VI.21 Sortie du terminal détaille sur le diagnostic d'un stade précoce avec YOLO_-MTCB. 150
VI.22 Diagnostic d'une personne en bonne santé avec YOLO_MTCBC 151
VI.23 Sortie du terminal détaille sur le diagnostic d'une personne en bonne santé avec YOLO_MTCB. 151
VI.24 L'interface GitHub pour le code accessible. 152
VI.25 Capture de l'Interface d'acceuil de la platform MTC diagnostique Pro . . . 153

A.1 Annexe: Diagramme des cas dutilisation de la plateforme MTC Hybrid Pro 164
A.2 Annexe:Diagramme de sequence de la plateforme MTC Hybrid Pro 165
A.3 Annexe: Diagramme de classes de la plateforme MTC Hybrid Pro 166

B.1 Annexe: Image healthy détectée avec sucés 168

B.2 Annexe:Image healthy et statistiques 169
B.3 Annexe: Image stade avancé détecté avec succès avec les statistiques 170
B.4 Annexe:Image stade avancé détecté avec succès avec les statistiques 171
B.5 Annexe:Image stade avancé avec cartographie 172
B.6 Annexe: Image stade avancé avec détection des caractéristiques sur la langue 173
B.7 Annexe: Image stade avancé avec détails sur les caractéristiques détectées . 173

Liste des tableaux

II.1 Schéma récapitulatif des types dapprentissage automatique. 27
II.2 Une comparaison entre différents modèles de Transformers 41
II.3 Interprétation dune matrice de confusion 46

III.1 Comparaison entre les deux types de traitement d'image. 55
III.2 Les types de segmentation dimage. 57
III.3 Les caractéristiques de YOLO. 63
III.4 Comparaison entre les différentes versions des algorithmes YOLO. 67
III.5 *Explication de larchitecture de YOLO.* 71

IV.1 Maladies et troubles associés à des caractéristiques de la langue en MTC . 78
IV.2 les signes linguaux associée à chaque type de cancer et son interprétation en MTC. 85
IV.3 Les articles proposant un modèle intelligent capable de diagnostiquer de manière non-invasive à partir dimages de langue 88
IV.4 Symptômes lingual des patientes atteintes du cancer du sein selon les articles. 100

V.1 Tableau des Besoins Fonctionnels - Système MTC Diagnostic Pro 126
V.2 Analyse des Besoins Non-Fonctionnels - Système MTC Diagnostic Pro . . . 127

VI.1 Analyse des pertes (loss) du modèle YOLOv8 sur les phases d'entraînement et de validation 133
VI.2 Valeurs finales des métriques de YOLOv8 avec leur interprétation. 133
VI.3 Les hyperparamètres utilisés pour le modèle `YOLOv11s`. 136
VI.4 Les valeurs de chaque métrique à lepoch 49 avec leur interprétation. 138
VI.5 Comparaison entre les performances du modèle **YOLOv11** avant et après le réentraînement. 141
VI.6 Descriptions de chaque algorithme de ML choisi. 146
VI.7 Les hyperparamètres utilisés pour l'entraînement des modèles de ML. . . . 147

Liste des abréviations

Abréviation	Signification
AA	Apprentissage Automatique
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
ATDS	Automatic Tongue Diagnosis System
ATFE	Adaptive Tongue Feature Extraction
BCP	Breast Cancer Patient
BCQ	Body Constitution Questionnaire
BERT	Bidirectional Encoder Representations from Transformers
BFI-T	Brief Fatigue Inventory-Taiwan
CADCM	Class-aware diagnostic correction module
CCMQ	Constitution in Chinese Medicine Questionnaire
CI	Confidence Interval
CNN	Convolutional Neural Networks
COCO	Common Objects in Context
CPU	Central Processing Unit
CRF	Cancer Related Fatigue
CSS	Cascading Style Sheets
CT	Computed Tomography
CVT	Convolutional Vision Transformer
Cv-Swin	Convolutional Vision Swin Transformer
DALL • E	Generative AI model by OpenAI
DL	Deep Learning
DNN	Deep Neural Network
dpi	dots per inch
ELU	Exponential Linear Unit
ESBCP	Early stage Breast Cancer Patient
F2GAN	Few-shot to GAN
FN	False Negative
FP	False Positive
FSL	Few-Shot Learning
GAN	Generative Adversarial Networks
GDDR	Graphics Double Data Rate
GPU	Graphics Processing Unit
GRU	Gated Recurrent Units
GPT	Generative Pre-trained Transformer

Abréviation	**Signification**
HSV	Hue Saturation Value
HT	Healthy Tongue
HTML	HyperText Markup Language
IA	Intelligence Artificielle
IC	Intervalle de Confiance
IDE	Integrated Development Environment
IoT	Internet of Things
IoU	Intersection over Union
IRM	Imagerie par Résonance Magnétique
JSON	JavaScript Object Notation
LLaMA	Large Language Model Meta AI
LSTM	Long Short-Term Memory
LT	Light Tooth-marked
MAE	Mean Absolute Error
mAP	Mean Average Precision
MetaGAN	Meta Generative Adversarial Network
MIL	Multiple Instance Learning
ML	Machine Learning
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MT	Moderate Tooth-marked
MTC	Médecine Traditionnelle Chinoise
MXNet	Apache MXNet framework
NAFLD	Non-Alcoholic Fatty Liver Disease
NBCI	Non Breast Cancer Instance
Neo4j	Neo4j database
NLP	Natural Language Processing
NMS	Non-Maximum Suppression
NP	Normal People
OCR	Optical Character Recognition
ONNX	Open Neural Network Exchange
OpenVINO	Open Visual Inference and Neural network Optimization
OR	Odds Ratio
PDF	Portable Document Format
PIL	Python Imaging Library
PLFE	Pixel-Level Feature Extractor
ppp	points par pouce
QCFLM	Quadruple classification feature learning module
RAM	Random Access Memory
ReLU	Rectified Linear Unit
REPL	Read-Eval-Print Loop
ResNet	Residual Network
RGB	Red Green Blue
RNA	Réseaux de Neurones Artificiels
RNN	Recurrent Neural Networks

Abréviation	**Signification**
RNP	Réseaux de Neurones Profonds
RVB	Rouge Vert Bleu
SAD	Systèmes d'aide au diagnostic
SAM	Segment Anything Model
SMOTE	Synthetic Minority Oversampling Technique
ST	Severe Tooth-marked
SVM	Support Vector Machine
Swin	Shifted Window Transformer
T5	Text-To-Text Transfer Transformer
TanH	Tangente hyperbolique
TEBRS	Tongue Edge and Body Region Separation
TFDA	Tongue and Face Diagnostic Analyzer
TIUO	Tongue Image Upright Orientation Algorithm
TN	True Negative
TP	True Positive
TransH	Translating on Hyperplanes
TRHM	Tongue-image Recommended Herb Model
TTD	Tooth-marked Tongue Diagnosis
UML	Unified Modeling Language
UNet	U-Net
URL	Uniform Resource Locator
VAS	Visual Analogue Scale
VGG	Visual Geometry Group
ViT	Vision Transformer
VS Code	Visual Studio Code
WSL	Windows Subsystem for Linux
WSVM	Weakly Supervised object detection method
YAML	YAML Ain't Markup Language
YOLO	You Only Look Once

Introduction Générale

Le cancer du sein représente aujourd'hui l'une des préoccupations majeures de santé publique mondiale, constituant la forme de cancer la plus fréquente chez les femmes avec une incidence touchant environ une femme sur huit (12,5%)[3]. Cette pathologie grave, qui évolue souvent de façon silencieuse pendant plusieurs années, pose un défi considérable aux systèmes de santé contemporains. L'Algérie n'échappe pas à cette réalité alarmante, où les statistiques de l' Agence Internationale de Recherche sur le Cancer (IARC) en 2022, révèlent que l'incidence du cancer du sein dépasse largement celle des autres types de cancers (voir la figure 1), soulignant l'urgence d'une prise en charge adaptée et efficace. Ces statistiques ne cessent d' augmenter en Algérie, peut s' expliquer, par un manque de sensibilisation au dépistage précoce, en particulier dans les zones éloignées des centres hospitaliers, et par une certaine négligence des femmes face diagnostic de la maladie. Cette situation souligne l' importance cruciale d' une stratégie nationale de prévention et d' éducation à la santé, visant à réduire les ignorances d' accès au diagnostic.

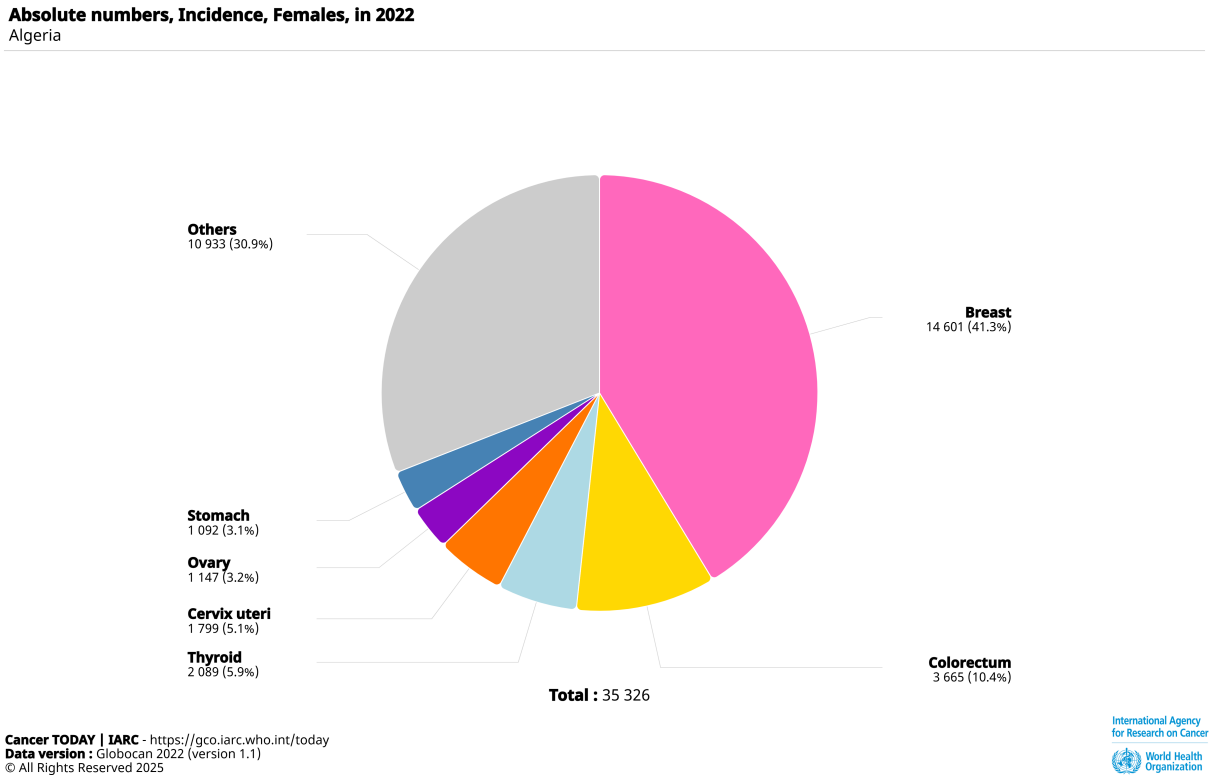


FIGURE 1 – Les statistiques des cancer chez les femmes en 2022 selon l'IARC [4]

Le diagnostic précoce demeure l'élément clé dans la lutte contre cette maladie.

Lorsque le cancer du sein est détecté à un stade avancé, les conséquences sont dramatiques : pertes de vies humaines, souffrance psychologique importante pour les patientes et leurs familles, ainsi que des répercussions sociales et économiques non négligeables. Malgré les avancées significatives de la médecine conventionnelle en matière de dépistage et de traitement, une part importante des patientes est encore diagnostiquée tardivement, limitant considérablement les options thérapeutiques et affectant négativement les taux de survie.

Dans ce contexte, la Médecine Traditionnelle Chinoise (MTC) apporte une approche différente et intéressante. Depuis des milliers d' années, elle utilise l' observation de la langue comme un outil important pour détecter des signes de déséquilibre dans le corps, parfois bien avant l' apparition des vrais symptômes. Cette méthode fait partie des techniques principales de diagnostic en MTC, avec l' observation générale, l' écoute et la palpation. Selon la MTC, le cancer du sein provient d'un déséquilibre dans la circulation de l'énergie vitale (appelée "Qi") dans l'organisme. Cette approche considère que les maladies résultent de blocages énergétiques qui se manifestent bien avant l'apparition de symptômes physiques. L'observation de la langue constitue l'une des méthodes diagnostiques les plus raffinées de cette médecine : la couleur, la texture, la forme et l'aspect de surface de la langue révèlent l'état énergétique des différents organes. Par exemple, des rougeurs sur les côtés de la langue peuvent signaler un déséquilibre au niveau du système hépatique, tandis qu'un aspect sec à la base peut indiquer un affaiblissement du système rénal. Ces signes linguaux offrent une fenêtre unique sur l'état de santé global du patient, permettant potentiellement d'identifier des risques pathologiques du cancer du sein avant leurs manifestations cliniques.

Cette richesse diagnostique de la MTC représente un potentiel considérable pour la médecine préventive moderne. Cependant, l'expertise requise pour interpréter correctement ces signes linguaux demande une formation spécialisée de plusieurs années, limitant ainsi l'accessibilité de cette approche aux professionnels de santé conventionnels. De plus, l'interprétation manuelle reste naturellement subjective, ce qui peut conduire à des variations dans l'évaluation d'un même patient. Cette situation représente une opportunité exceptionnelle pour l'intelligence artificielle : standardiser et objectiver cette expertise millénaire pour la rendre accessible aux professionnels de santé moderne en tant qu'outil complémentaire de pré-diagnostic.

Les technologies d'intelligence artificielle et de vision par ordinateur offrent aujourd'hui des solutions révolutionnaires pour relever ce défi. Les systèmes d'apprentissage automatique particulièrement les réseaux de neurones convolutifs et les algorithmes de détection d'objets comme YOLO (You Only Look Once), ont démontré leur efficacité avec une précision et une constance remarquables dans l'analyse d'images médicales , identifiant des motifs visuels subtils que l'œil humain pourrait manquer ou interpréter différemment. Ces technologies, déjà utilisées avec succès dans l'analyse d'images médicales conventionnelles (radiographies, IRM (Imagerie par Résonance Magnétique), scanners) permettent d'automatiser l'extraction de caractéristiques visuelles complexes, d'objectiver les observations cliniques et de standardiser les processus diagnostiques, elles peuvent être adaptées pour automatiser l'expertise diagnostique de la MTC.

Cette convergence entre sagesse ancestrale et technologies de pointe ouvre la voie à une médecine véritablement intégrative. ce mémoire s'inscrit dans une démarche am-

bitieuse visant à contribuer significativement à l'amélioration du diagnostic précoce du cancer du sein. L'objectif n'est pas de remplacer l'expertise médicale humaine, mais d'offrir aux professionnels de santé un outil d'aide au diagnostic qui combine le meilleur des deux mondes : la profondeur de l'observation traditionnelle chinoise et la précision de l'analyse numérique moderne. En proposant le développement d'un système intelligent révolutionnaire qui transforme l'examen traditionnel de la langue en un outil de pré-diagnostic automatisé et objectif pour améliorer le dépistage précoce du cancer du sein.Créant ainsi un pont entre tradition et modernité au service de la santé des femmes.

Défis et problèmes

- **Enjeu médical, sociétal et limites des méthodes de dépistage conventionnelles**: Malgré les avancées en matière de dépistage du cancer du sein, une part importante des diagnostics reste tardive, en particulier dans les contextes où l'accès aux outils conventionnels est limité ou inadapté. Les méthodes actuelles, comme la mammographie, l'échographie ou l'IRM, présentent plusieurs limites notables. La mammographie, bien qu'efficace, montre une sensibilité réduite chez les femmes ayant une forte densité mammaire, rendant les anomalies plus difficiles à détecter. L'échographie est fortement dépendante de l'opérateur, ce qui peut entraîner une variabilité dans les résultats selon l'expérience du professionnel. Quant à l'IRM, si elle est très sensible, elle génère fréquemment des faux positifs, entraînant des examens invasifs inutiles et une anxiété accrue pour les patientes. Surtout, ces méthodes sont centrées sur la détection de lésions déjà visibles, et non sur les déséquilibres physiologiques précoces, ce qui limite leur efficacité dans une démarche réellement préventive.
- **Potentiel d'innovation de la Médecine Traditionnelle Chinoise** :Parallèlement, la MTC propose une lecture préventive fine du corps à travers l'observation de la langue. Ce potentiel diagnostique remarquable reste toutefois peu exploité, car il repose sur une expertise rare (peu de personnes capables à diagnostiquer correctement), longue à acquérir (langue durée de pratique de MTC pour acquérir ce savoir), et difficilement standardisable (différents diagnostic a plusieurs reprises).

Face à ces multiples enjeux, la question de recherche centrale de ce mémoire peut être formulée ainsi :

Comment développer un système intelligent d'aide au diagnostic précoce du cancer du sein qui combine efficacement l'expertise traditionnelle de l'analyse de la langue en MTC avec les technologies avancées de vision par ordinateur et d'apprentissage automatique, tout en garantissant l'objectivité, la reproductibilité et la fiabilité clinique des résultats ?

Objectif fixés

Ce mémoire vise à répondre à un besoin urgent de dépistage précoce du cancer du sein, en s'appuyant sur une idée innovante : utiliser l'observation de la langue selon la

MTC, rendue accessible grâce à l' intelligence artificielle.

- **Objectif principal :** Développer un outil intelligent, capable d' analyser automatiquement une photo de langue pour détecter des signes précoces de cancer du sein.
- **Objectif technologique :** Rendre l'expertise MTC utilisable par tout professionnel de santé, en la standardisant et en supprimant la subjectivité grâce à la vision par ordinateur.
- **Objectif médical :** Offrir un complément simple, non invasif et économique aux méthodes classiques de dépistage, afin d' identifier plus tôt les patientes à risque.

Structure du mémoire

Ce projet qui ouvre la voie à une médecine intégrative, où la sagesse ancienne rencontre les technologies de pointe pour améliorer la santé des femmes, est structurée en six chapitres, suivant une progression à la fois théorique, technologique et expérimentale :

- **Chapitre 1 –Principes de la Médecine Traditionnelle Chinoise (MTC)**: Présente les fondements théoriques de la MTC, son approche holistique de la santé, et l' importance du diagnostic par la langue.
- **Chapitre 2 –Apprentissage automatique**: Expose les bases de l' intelligence artificielle et du machine learning, en mettant l' accent sur les algorithmes qu'on pourra utiliser dans ce travail.
- **Chapitre 3 –Vision Artificielle et Algorithme YOLO**: Décrit les concepts clés de la vision artificielle et détaille le fonctionnement de l' algorithme YOLO pour la détection d' objets.
- **Chapitre 4 –État de l' art sur le diagnostic automatisé du cancer du sein dans la MTC** Fait le point sur les recherches existantes concernant le diagnostic du cancer du sein, l' usage médical de la MTC et l' apport des technologies d' IA dans le diagnostic par la langue.
- **Chapitre 5 –Contribution :** Présente l' architecture proposée, l' hybridation des algorithmes, la génération du dataset, les choix méthodologiques retenus et la conception d'une plateforme d'aide au diagnostic du cancer du sein.
- **Chapitre 6 –Évaluation et résultats expérimentaux** Expose la phase de test, les performances du système, les limites observées et l'implémentation de la plateforme.

Chapitre I

Principes de la Médecine Traditionnelle Chinoise (MTC)

Sommaire

I.1 Introduction **6**
I.2 Histoire de la MTC **6**
I.3 Concept fondamental de MTC **7**
I.4 L'application de MCT dans le diagnostic médical **14**
I.5 Conclusion **19**

Résumé : *Ce chapitre pose les fondations de la médecine traditionnelle chinoise, une approche holistique fondée sur l' équilibre du Qi, des émotions et des organes. Il explore ses grands principes (Yin-Yang, cinq éléments, méridiens), ses méthodes de diagnostic (notamment l' observation de la langue) et ses traitements naturels.*

I.1 Introduction

La médecine traditionnelle chinoise (MTC) aussi appelée 中医 (Zhng y) en chinois, ne se limite pas à une approche médicale, mais repose sur une vision holistique de l'être humain et de son équilibre. Elle puise autant dans les anciennes sagesses philosophiques que dans les connaissances modernes pour comprendre le lien profond entre le corps, l'esprit et les émotions.
Dans cette approche, le corps et l'esprit sont indissociables. Chaque organe est relié aux autres et fonctionne en harmonie avec l'ensemble du corps. Plutôt que de se focaliser sur les symptômes, la MTC cherche à rétablir l'équilibre énergétique, aussi bien à l'intérieur de l'individu que dans sa relation avec son environnement.
Ce chapitre explore les principes fondamentaux de la médecine traditionnelle chinoise, son approche globale et la manière dont elle envisage la santé comme un état d'équilibre dynamique.

I.2 Histoire de la MTC

La MTC est une pratique ancienne aux origines mythiques (remontent à des temps très anciens) [5, 6, 7, 8], qui s'est développée au fil des siècles pour devenir un pilier de la médecine mondiale.

- **Origines légendaires (IIIe millénaire av. J.-C.) :** la médecine traditionnelle chinoise serait attribuée à trois empereurs mythiques.
- **Époque pré-impériale (2207-722 av. J.-C.) :** peu de traces subsistent, mais cette période témoigne d'une médecine encore primitive.
- **Période des Printemps et Automnes (722-481 av. J.-C.) :** influence du confucianisme et du taoïsme sur la pratique médicale ; distinction entre médecins, prêtres et magiciens.
- **Émergence des concepts médicaux fondamentaux (453-221 av. J.-C.) :** développement des bases théoriques de la médecine savante chinoise.
- **Unification de l'Empire (221 av. J.-C.) :** centralisation des élites médicales, mais destruction de nombreux ouvrages.
- **Dynastie Han (206 av. J.-C. - 220 apr. J.-C.) :** essor important de la médecine et de la pharmacopée, structuration des connaissances.
- **Dynasties suivantes :** avancées en acupuncture et en anatomie, consolidation de la médecine traditionnelle chinoise en tant que système officiel.
- **1929 :** tentative de suppression de la MTC, rapidement abandonnée.
- **Années 1950 :** réformes visant à concilier médecine traditionnelle et médecine occidentale.
- **Introduction en Occident :** connue depuis le XVIe siècle, mais véritable essor à partir des années 1980 grâce à l'ouverture de la Chine et aux échanges universitaires.

I.3 Concept fondamental de MTC

La MTC repose sur des théories qui sont : le yin yang, les cinq éléments, le Qi et les méridiens. Ces quatres éléments sont la base de la philosophie, la culture de cette civilisation chinoise. Ils cherchent à expliquer le fonctionnement de l' univers, de la nature et de l' homme

I.3.1 Le Qi, énergie vitale

Le Qi, ou "énergie vitale" , est un concept central dans la MTC. Il est une force active qui parcourt et stimule le corps humain, distinguant ainsi les êtres vivants de la matière inanimée. Le Qi est l'agent invisible, il parcourt ainsi la surface du corps, puis pénètre en profondeur, dessinant des lignes de force qui relient entre elles chaque partie de notre anatomie tout en faisant le lien entre le corps physique et le psychisme. Ce flux énergétique continu circule dans tout le corps au moyen d'un réseau précis, les voies appelées méridiens. Lorsque le Qi est en quantité suffisante et circule bien, l'organisme est en bonne santé, la pensée claire et les réflexes vifs. Lorsqu'il stagne, on se sent faible, lourd et sans vitalité. Dès que l'harmonie ou l'équilibre sont retrouvés, la santé revient [9]. L'organisme humain est perçu comme une version miniature représentant le vaste système naturel[10].

I.3.1.1 Les types de Qi :

Le Qi existe sous trois formes : le Qi Primaire, le Qi Alimentaire et le Qi des Organes.[6]

- **Le Qi Primaire :** Le Yuan Qi est l' énergie vitale héritée à la conception. Comparable à une réserve limitée, il constitue notre capital santé et notre force de récupération.
- **Le Qi Alimentaire :** Le Gu Qi provient de l' alimentation et alimente notre corps comme un carburant quotidien. Il soutient nos activités et la régénération des tissus.
- **Le Qi des Organes :** Le Zang Fu Qi résulte de la transformation du Qi Primaire et du Qi Alimentaire par les organes. Il circule via les méridiens pour équilibrer le corps.

I.3.1.2 Nourrir et équilibrer le Qi :

Pour nourrir le Qi, il est essentiel d' adopter une alimentation saine et équilibrée. Comme une bougie qui brûle grâce à son huile, notre Qi a besoin d' un carburant de qualité. En chinois, le mot "nourriture" est étroitement lié au Qi, soulignant l'importance de chaque repas dans notre énergie vitale.
Le Qi se renforce aussi par la respiration. En médecine chinoise, le Qi des poumons, acquis par l'air inspiré, est une source essentielle d' énergie. La respiration consciente, l' air frais et la nature contribuent à enrichir notre Qi.
Cependant, nourrir le Qi ne suffit pas, il faut aussi l' équilibrer. En MTC, l' activité et

repos sont essentielle à la circulation du Qi. Des pratiques comme le Qi Gong, le Tai Chi et la méditation aident à le réguler. L' acupuncture et le massage Tui Na favorisent aussi son bon flux en éliminant les blocages et en rétablissant son équilibre [8].

I.3.2 Le Yin et le Yang

Le concept du Yin/Yang, central dans la philosophie chinoise, explique que les forces sont opposées mais complémentaires comme le jour et la nuit, l'ombre et la lumière, le repos et l'activité, sont interdépendantes et en constante interaction. Leur équilibre est crucial pour maintenir l'harmonie dans la vie.
Le Yin se change en Yang et le Yang en Yin. L' été devient hiver, le jour se transforme en nuit ou encore la chaleur en froid. Pour cela, différents facteurs et différentes conditions doivent être réunis pour la transmutation [7].
En effet, selon cette philosophie, un équilibre harmonieux entre le Yin et le Yang est essentiel pour la santé, tandis que les maladies sont le signe d' un déséquilibre. Ce principe souligne l'importance de l'harmonie et de l'équilibre dans tous les aspects de la vie[11]. La figure ci dessous I.1 represente des exemples Yin/Yang.

Le Yin/Yang est caractérisé par quatre aspects : Opposition, Interdépendance, Équilibre, Transmutation [6].



FIGURE I.1 – Le Yin et le Yang.
[12]

I.3.3 Le système de 5 éléments

Les Wu Xing, ou Cinq éléments [6], constituent un concept fondamental de la MTC. Ces cinq éléments –Bois, Feu, Terre, Métal et Eau –(figure I.2) structurent non seulement la physiologie humaine mais également l' ensemble des phénomènes naturels. Chaque

organe du corps est associé à l'un de ces éléments, qui influencent également les saisons :

- **Bois** Printemps
- **Feu** Été
- **Métal** Automne
- **Eau** Hiver
- **Terre** Inter-saisons (périodes de transition entre les saisons)

Les cinq éléments ne sont pas de simples classifications statiques, mais des principes dynamiques en perpétuelle interaction. Tout ce qui existe dans l'univers peut être rattaché à ces cinq éléments, formant un système d'équilibre et de transformation continue. Le chiffre cinq (05) revêt ici une forte symbolique, comme dans de nombreuses cultures : il évoque l'action et la fraternité (les cinq doigts de la main) ou encore les fonctions vitales (les cinq orifices du visage permettant de respirer et se nourrir, donc de vivre).

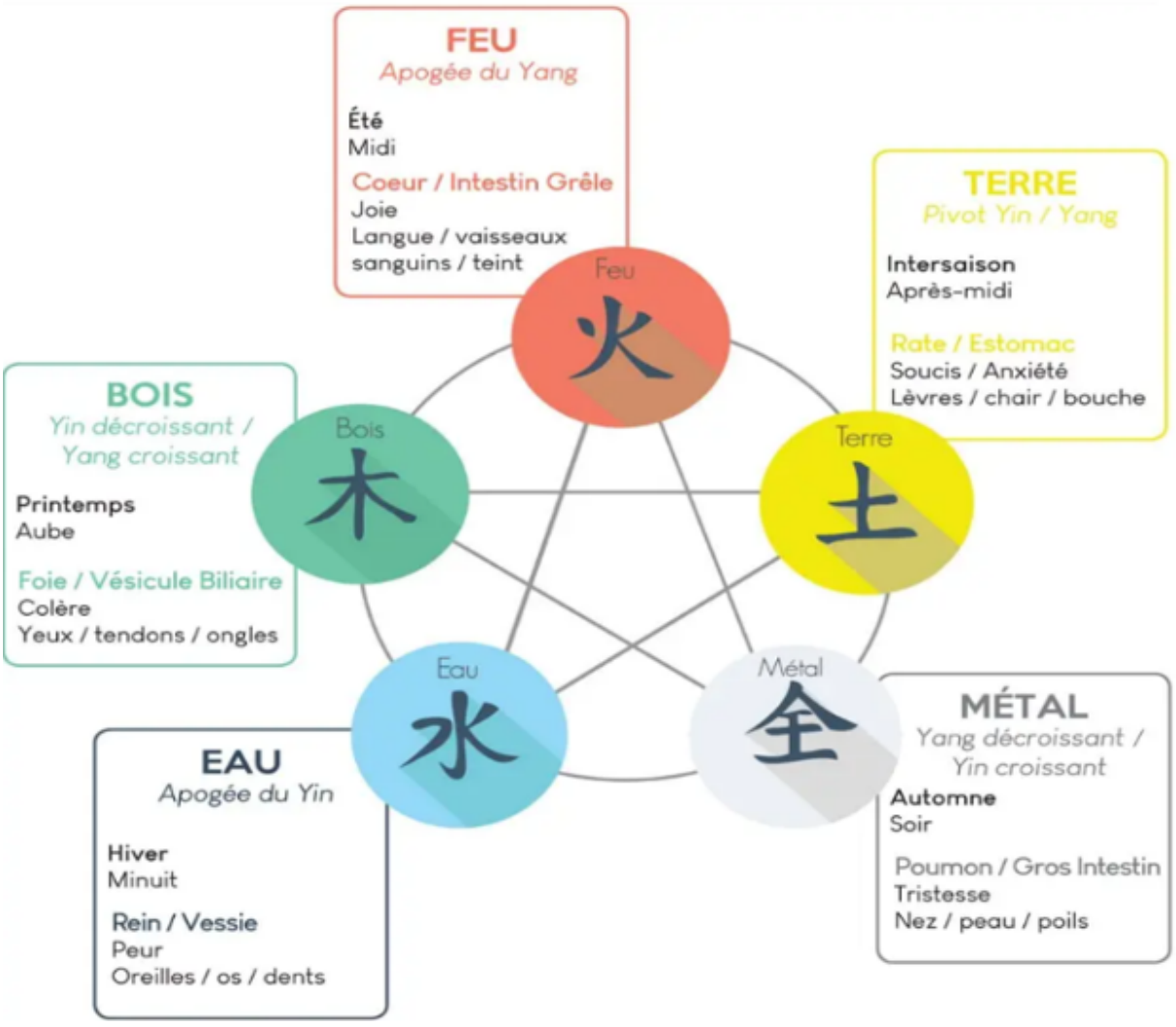


FIGURE I.2 – Les 5 Éléments de la MTC.
[13]

Un patient se plaint de fatigue, de digestion lente et d'inquiétude excessive, indiquant une faiblesse de l'élément **Terre** (Rate/Estomac). Cette faiblesse a un impact sur le Poumon (**Métal**), entraînant une respiration courte et une sécheresse de la peau. Par ailleurs, des signes de chaleur comme des palpitations et une agitation mentale suggèrent que le **Feu** est trop actif, ce qui aggrave encore l'état du Poumon. Le praticien cherchera alors à renforcer la Terre et à apaiser le Feu pour améliorer la respiration et l'état général du patient.

I.3.3.1 Cycle d'engendrement

Ce cycle décrit (figure I.3) comment chaque élément nourrit et soutient l'élément suivant. Le bois alimente le feu, le feu crée la terre (à travers les cendres), la terre produit le métal, le métal collecte l'eau (condensation sur une surface froide métallique), et l'eau nourrit le bois. Ce cycle d'engendrement représente une relation harmonieuse d'entraide et de soutien mutuel entre les éléments [6].

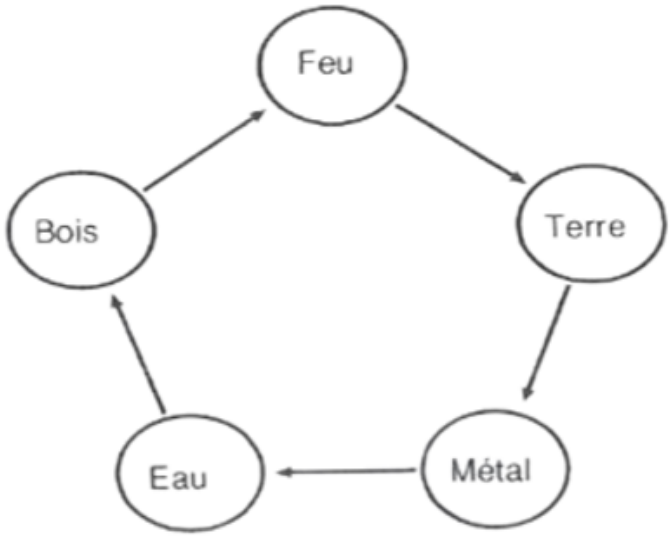


FIGURE I.3 – Le système des 5 éléments Cycle d'engendrement.
[6]

Un patient présente de la raideur musculaire, de l'irritabilité et des maux de tête, signes d'un déséquilibre du Bois (**Foie**). Le praticien observe aussi une grande fatigue, des douleurs lombaires et une frilosité, indiquant une faiblesse de l'**Eau** (Reins). Selon le cycle d'engendrement, l'Eau nourrit le Bois. Ici, le vide de l'Eau affaiblit le Bois. Le traitement visera donc à renforcer l'Eau pour restaurer l'équilibre et soutenir le Bois de façon harmonieuse

I.3.3.2 Le cycle de contrôle ou de domination

Ce cycle illustre comment chaque élément contrôle et modère l'élément qu'il surmonte (figure I.4). Le bois puise dans la terre, la terre absorbe l'eau, l'eau éteint le feu, le

feu fait fondre le métal, et le métal coupe le bois. Cette relation de contrôle assure un équilibre et empêche qu' un élément ne devienne trop puissant [6].

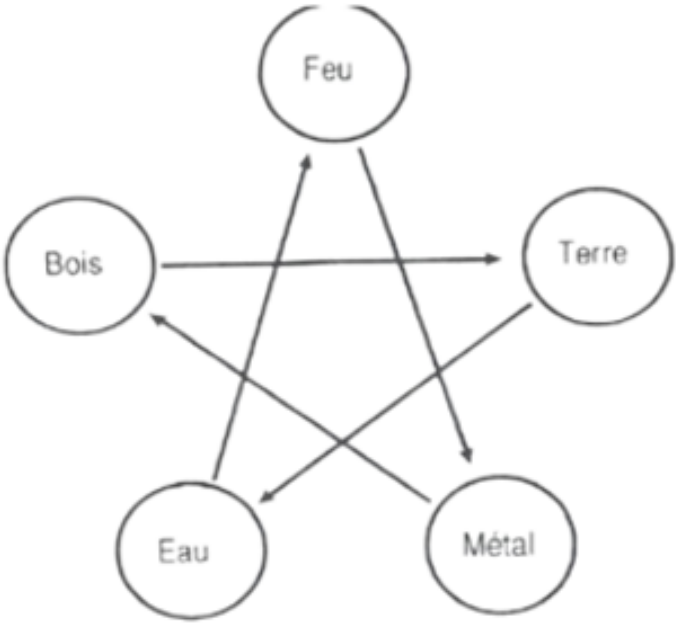


FIGURE I.4 – Le système des 5 éléments Cycle de contrôle ou de domination. [6]

Un patient souffre de toux sèche persistante et de respiration courte, sans signe d' infection. Le praticien observe une faiblesse de l' élément **Métal** (Poumon), qui se manifeste par une voix faible et une pâleur du teint. Il remarque également des signes d' agitation et de rougeur, suggérant un excès de l' élément **Feu** (Cœur). Selon le cycle de domination, le Feu contrôle le Métal. Ici, le Feu est trop fort et affaiblit le Métal. Le traitement visera alors à calmer le Feu afin de permettre au Métal de se rétablir, rétablissant ainsi l' équilibre entre les deux éléments.[6]

I.3.4 Les émotions en MTC

En MTC, les émotions sont directement liées aux organes du corps. Si une émotion devient trop forte ou dure trop longtemps, elle peut perturber la circulation du Qi (l' énergie vitale) et créer des déséquilibres, ce qui peut entraîner des problèmes de santé.[6] Chaque émotion est associée à un organe et à un élément de la nature :

- **Colère**/ Foie (élément Bois): Une colère excessive bloque la circulation du Qi du Foie, provoquant des tensions musculaires, des maux de tête ou des douleurs abdominales.
- **Joie excessive**/ Cœur (élément Feu) :Trop de joie disperse le Qi du Cœur, causant des palpitations, de l' insomnie ou de l' agitation.

- **Trop réfléchir (rumination)**/ Rate (élément Terre) :Trop penser bloque le Qi de la Rate, rendant la digestion difficile et provoquant de la fatigue.
- **Tristesse**/ Poumons (élément Métal): Une grande tristesse affaiblit le Qi des Poumons, entraînant une respiration difficile et une baisse d' énergie.
- **Peur**/ Reins (élément Eau): Une peur intense épuise le Qi des Reins, causant de la faiblesse, des douleurs lombaires et des problèmes urinaires.

Si le Qi d' un organe ne circule pas bien, cela peut créer des blocages ou des maladies dans le corps.

I.3.5 Les six excès

En plus des émotions, la MTC considère que l' environnement peut aussi perturber l' équilibre du Qi. On appelle cela les six excès, qui sont des facteurs climatiques pouvant affecter la santé.[6]

- **Vent**: Déplace le Qi de manière irrégulière, causant des rhumes, des frissons, des douleurs articulaires et des étourdissements.
- **Froid**: Ralentit le Qi, ce qui peut provoquer des douleurs musculaires et une sensation de fatigue.
- **Chaleur**: Agite le Qi, entraînant de la fièvre, une gorge sèche, une transpiration excessive et des inflammations.
- **Humidité**: Ralentit le Qi, rendant le corps lourd, gonflé et provoquant des problèmes respiratoires.
- **Sécheresse**: Consomme les liquides du corps et le Qi, asséchant la peau, la gorge et les poumons, ce qui peut causer de la toux ou des irritations.
- **Feu**: Rend le Qi trop intense, provoquant de l'irritabilité, des rougeurs, des infections et une sensation de brûlure dans le corps.

Si ces éléments sont en excès, ils perturbent le Qi et peuvent entraîner des maladies. Par exemple, le vent et le froid peuvent bloquer le Qi, provoquant un rhume, tandis que la chaleur et l' humidité peuvent stagner le Qi, favorisant l' apparition de boutons ou d' inflammation.

I.3.6 Les 12 méridiens (Zhang Fu)

En MTC, le Qi circule pour assurer le bon fonctionnement des organes [6, 14, 15]. Ces organes sont classés en deux groupes :

- **Les organes Zhang (Yin) :** Ils sont profonds et ont un rôle essentiel dans la production et le stockage de l' énergie. Il s' agit du Cœur, des Poumons, de la Rate, du Foie et des Reins.

- **Les organes Fu (Yang) :** Plus en surface, ils transforment les aliments et éliminent les déchets sans rien stocker. Ils incluent l'Estomac, l'Intestin grêle, le Gros intestin, la Vessie et la Vésicule biliaire.

Chaque organe Zhang est associé à un organe Fu pour maintenir l'équilibre du corps. En plus des organes, le corps est divisé en trois foyers énergétiques :

- **Le foyer supérieur**(Cœur, Poumons) s'occupe de la circulation de l'énergie et du sang.
- **Le foyer médian** (Rate, Estomac) gère la digestion et la production d'énergie.
- **Le foyer inférieur**(Reins, Vessie, Intestins) élimine les déchets.

L'énergie vitale circule à travers des **méridiens** (I.5), des canaux invisibles qui relient les organes entre eux. Il existe **douze méridiens principaux**, chacun correspondant à un organe spécifique. Ces méridiens forment un réseau énergétique comparable à un circuit électrique qui alimente le corps en énergie. En acupuncture [1] , ils sont stimulés pour rétablir l'équilibre et traiter certains troubles.

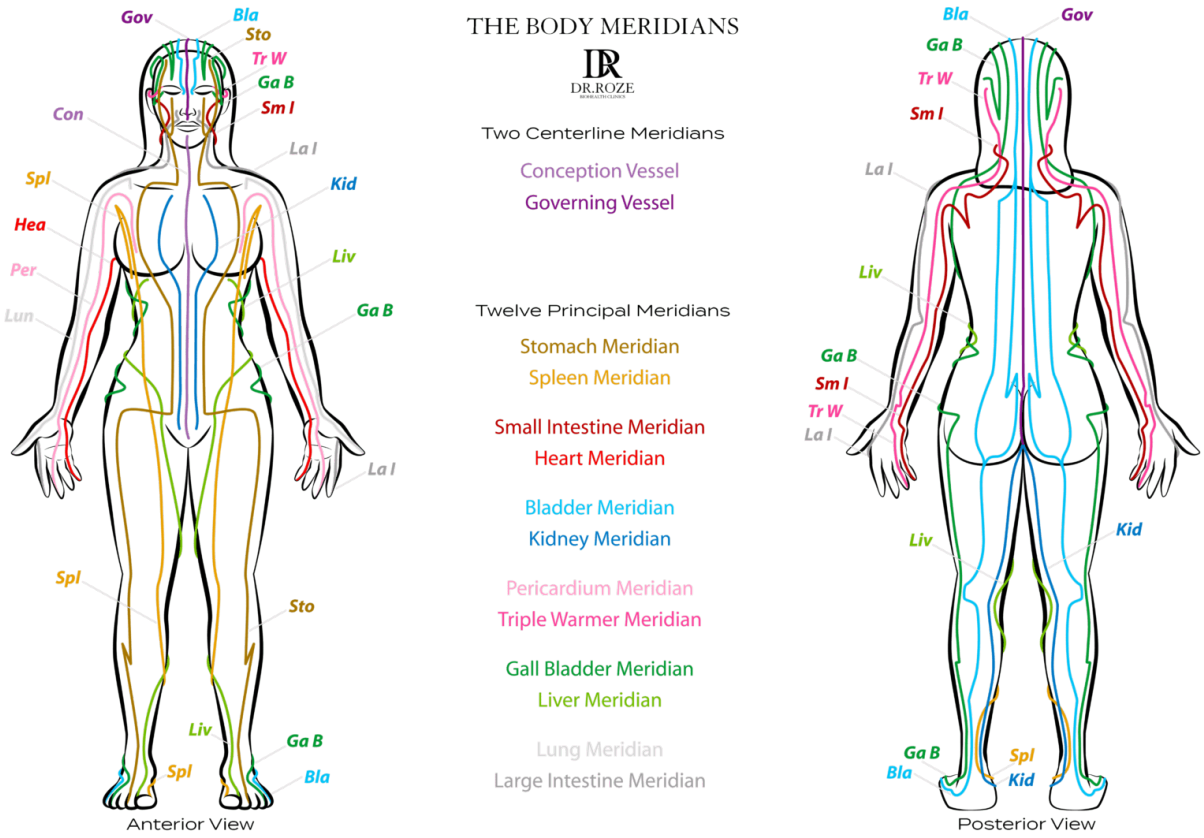


FIGURE I.5 – Les méridiens du corps.
[14]

1. **acupuncture** :Thérapie d'origine chinoise, soignant par l'enfoncement d'aiguilles dans le corps humain.

I.4 L'application de MCT dans le diagnostic médical

Le diagnostic repose sur deux grands principes :

- **Regarde l' extérieur pour voir l' intérieur**: L' état externe du corps (peau, langue, teint) révèle les déséquilibres internes.
- **La partie reflète le tout**:Un simple symptôme peut donner des indices sur l' état énergétique général du patient.

Ainsi, au lieu de se focaliser uniquement sur une maladie spécifique, la MTC cherche à identifier la cause profonde du déséquilibre et à restaurer l' harmonie énergétique du corps.

I.4.1 Les Quatres Méthodes de Diagnostic en MTC

La MTC utilise **quatres méthodes d' examen** pour poser un diagnostic :

- **Observation (Wang Zhen)**:
  - Analyse de l' apparence générale du patient (teint, posture, expressions faciales).
  - Examen de la **langue** : couleur, forme, texture, enduit (par exemple, une langue rouge indique une chaleur interne).
- **Interrogatoire (Wen Zhen)**
  - Questions sur les symptômes, l' état émotionnel, le mode de vie et les antécédents médicaux.
  - Recherche de **déséquilibres énergétiques** à travers les habitudes du patient (sommeil, digestion, état mental).
- **Palpation (Qie Zhen)**
  - Examen des zones douloureuses et des points d' acupuncture.
  - Prise du pouls au poignet (le pouls chinois permet de vérifier l' équilibre du yin et du yang dans le corps).
- **Auscultation et Olfaction (Wen Zhen)**
  - Écoute des bruits corporels (respiration sifflante, battements cardiaques irréguliers).
  - Détection d' odeurs inhabituelles (haleine forte, transpiration).

I.4.2 Examen de la langue

La langue est l'un des éléments les plus importants du diagnostic en MTC, en tant qu'un miroir du corps (Figure I.6), chaque zone de la langue correspond à un foyer énergétique. Par exemple, le bout de la langue est lié au Cœur et aux poumons, tandis que le centre est en lien avec l'estomac et la rate. [5, 7, 16]

Certains Méridiens passent par la langue (comme ceux de la Rate, Vessie, Estomac, Triple Réchauffeur) ou y sont reliés indirectement, permettant à l'état des Organes de se refléter sur des zones spécifiques de la langue.
Deux approches permettent de comprendre la relation langue/Organes :

- ***Répartition en Trois Réchauffeurs :*** La langue est divisée en trois parties :
  - Tiers antérieur:**Réchauffeur Supérieur**(Cœur, Poumon)
  - Tiers moyen : **Réchauffeur Moyen** (Estomac, Rate)
  - Tiers postérieur: **Réchauffeur Inférieur** (Reins, Vessie, Intestins)

  Cette classification reflète des fonctions énergétiques plus que des localisations anatomiques [5]. Par exemple, le Foie peut être rattaché à différents Réchauffeurs selon ses fonctions.

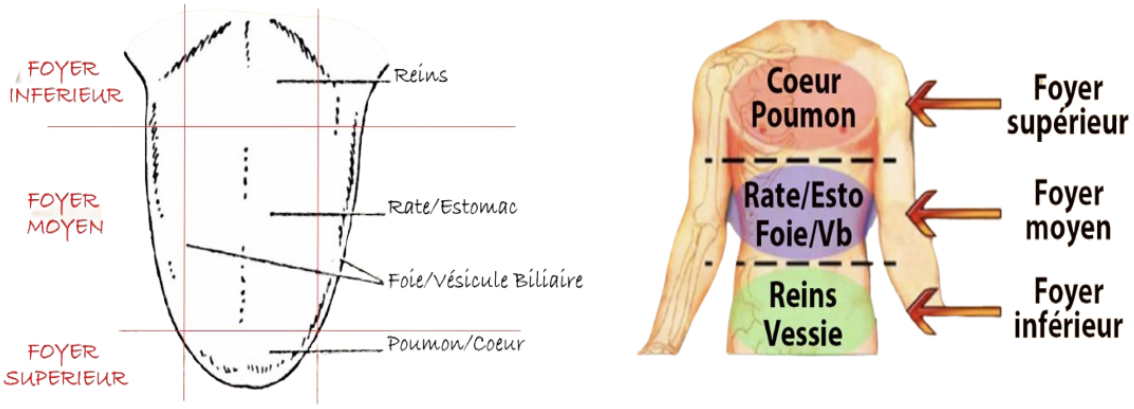


FIGURE I.6 – Correspondance entre foyers énergétiques et les différentes parties de la langue.

[16]

- ***Topographie des Organes sur la langue :***Plusieurs systèmes existent avec de légères variantes. La topographie des organes sur la langue a été montrée dans la figue suivante, ci dessous la figure I.7 qui reflète les details de l'explication par zone:
  - **Pointe** : Cœur (C),
  - **Entre pointe et centre** : Poumon (P),
  - **Centre** : Estomac (E) et Rate (RP),
  - **Base** : Reins (R), Intestins (GI-IG), Vessie (V)et Utérus (chez la femme),

- **Bord gauche** : Foie (F),
- **Bord droit** : Vésicule Biliaire (VB).

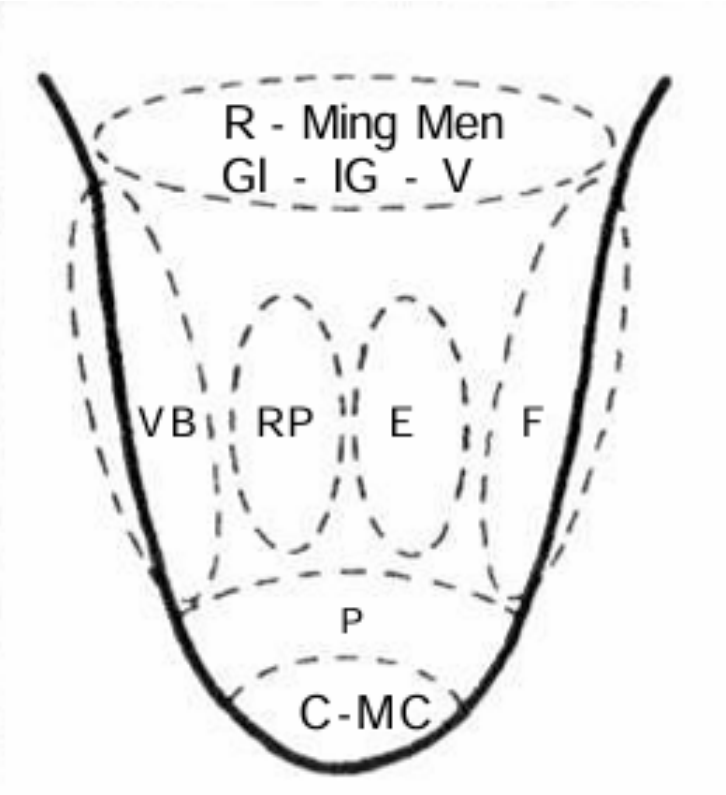


FIGURE I.7 – Les parties de la langue correspondant aux organes.
[5]

La langue reflète directement l'état des organes et de l'énergie vitale sur cela ils font le l'observation Wang Zhen [7]. Un praticien peut observer :

- **La couleur de la langue :** Indique l'état du Qi, du sang et des organes Yin.
  - **Rouge vif :** Excès de chaleur dans le corps.
  - **Pâle :** Déficit de sang ou d'énergie.
  - **Bleuâtre :** Stagnation du sang (mauvaise circulation).
- **La forme de la langue :**Renseigne sur la force du Qi nourricier.
  - **Gonflée :** Rétention d'humidité.
  - **Fine :** Déficit de sang ou de liquides organiques.
- **L'enduit lingual :**Montre l'état des viscères Yang (estomac, intestins, etc.).
  - **Épais et blanc :** Excès d'humidité et de froid.
  - **Jaune :** Excès de chaleur dans le corps.
- **L'humidité de la langue :** Indique l'état des liquides organiques.

– **Sèche :** Déshydratation ou chaleur interne.
– **Trop humide :** Rétention de fluides ou insuffisance digestive.

La figure suivante I.8 illustre les différentes exemples de diagnostic par la langue avec les caractéristiques visuelles.

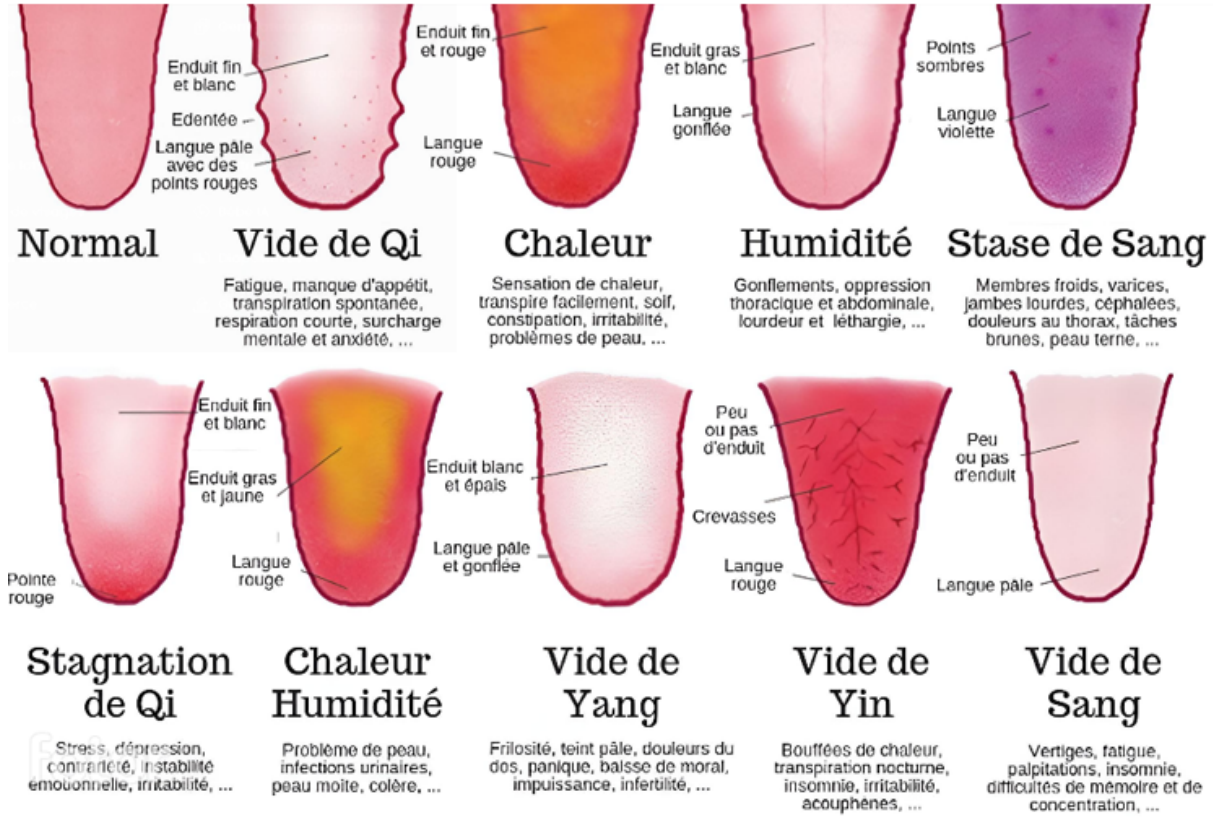


FIGURE I.8 – Illustration des caractéristiques de la langue.

I.4.3 Le Traitement en MTC

Une fois le diagnostic établi, la MTC propose un traitement personnalisé visant à rétablir l'harmonie énergétique [6, 5, 7]. Plusieurs méthodes peuvent être utilisées ensemble.

I.4.3.1 L'Acupuncture

L'acupuncture consiste à insérer de fines aiguilles sur des points précis du corps pour :

- Stimuler le flux du Qi dans les méridiens.
- Soulager la douleur et améliorer la circulation sanguine.
- **Réguler les organes internes** (digestion, respiration, hormones, etc.).

Elle est souvent utilisée pour les douleurs chroniques, les troubles du sommeil, le stress et les déséquilibres hormonaux.

I.4.3.2 La Moxibustion

Cette technique utilise la chaleur produite par la combustion d' armoise (une plante médicinale) pour stimuler les points d' acupuncture. Elle est efficace pour :

- **Les douleurs liées au froid** (arthrose, fatigue chronique).
- Les troubles digestifs.
- La stimulation de l' immunité.

I.4.3.3 La Pharmacopée Chinoise

Les plantes médicinales jouent un rôle clé en MTC. Elles sont classées selon :

- **Leur température** (chaude, froide, neutre).
- **Leur saveur** (amer, sucré, acide, salé, piquant).
- **Leur tropisme** (les organes sur lesquels elles agissent).

Par exemple, le ginseng tonifie le Qi, tandis que la menthe rafraîchit et disperse la chaleur.

I.4.3.4 Le Qi Gong et le Tai Chi

Ce sont des exercices associant mouvements lents, respiration et méditation. Ils aident à améliorer la circulation du Qi, réduire le stress et les tensions, renforcer l' immunité.

I.4.3.5 Le Massage Tui Na

Le Tui Na est un massage thérapeutique qui agit sur les méridiens et les points d' acupuncture. Il est utilisé pour :

- Les douleurs musculaires et articulaires.
- Les troubles digestifs et circulatoires.

I.4.3.6 La Diététique Chinoise

En MTC, l' alimentation est un véritable médicament. Chaque aliment est classé selon son effet énergétique (réchauffant, rafraîchissant, hydratant, etc.). Par exemple : "le gingembre est réchauffant et utile contre le froid" ou "le concombre est rafraîchissant et aide à éliminer la chaleur".

L' objectif est d' adapter son alimentation selon son état énergétique et la saison.

I.5 Conclusion

Ce chapitre a exploré les fondements de la médecine traditionnelle chinoise, une approche holistique où l'équilibre du Qi, des cinq éléments et des interactions organes-émotions est essentiel à la santé. À travers ses méthodes de diagnostic et de traitement, comme l'observation de la langue et l'acupuncture, la MTC illustre une vision globale du bien-être. Le but n'est pas de détrôner la médecine occidentale, mais de la compléter.

Le chapitre suivant présente les principes de l'intelligence artificielle ainsi que deux de ses principaux sous-domaines : l'apprentissage automatique et les réseaux de neurones artificiels. Il aborde également les architectures les plus couramment utilisées dans ce domaine.

Chapitre II

Apprentissage automatique

Sommaire

II.1 Introduction **21**
II.2 Machine Learning (ML) **22**
II.3 Réseaux de neurones et Deep Learning **28**
II.4 Processus de création d' un RNP **43**
II.5 Entraînement d' un RNP **43**
II.6 Évaluation d' un modèle **45**
II.7 Approches avancées de l' apprentissage automatique dans les RNP . **47**
II.8 Conclusion **48**

Résumé : *Ce chapitre explore l'univers fascinant de l'intelligence artificielle, des concepts fondamentaux du Machine Learning aux architectures révolutionnaires des réseaux de neurones profonds. De l'apprentissage supervisé aux Transformers en passant par les CNN et les modèles génératifs, découvrez comment les machines apprennent à partir des données pour résoudre des problèmes complexes. Les quatre types d'apprentissage (supervisé, non supervisé, semi-supervisé et par renforcement) sont détaillés avec des exemples concrets, notamment en diagnostic médical. Cette base théorique solide prépare l'application pratique des techniques d'IA à la vision artificielle et au modèle YOLO dans les chapitres suivants.*

II.1 Introduction

L' intelligence artificielle (IA) ou Artificial Intelligence (AI)[17, 18], vise à permettre aux machines de réaliser des tâches nécessitant normalement l' intelligence humaine. On présente par la figure II.1 la hiérarchie des sous-domaines de l' IA, du Machine Learning (ML) aux réseaux de neurones et au Deep Learning (DL) permettent aux systèmes d' apprendre à partir des données. Elle met en évidence leurs techniques clés et domaines d' application comme la Vision Artificielle, le NLP ou la robotique.

Ce chapitre introduit les concepts clés de l' *apprentissage automatique* en s' appuyant sur les *réseaux de neurones artificiels*, jusqu' aux *architectures profondes*. Il aborde les principes d' optimisation et leur application à la *vision artificielle*.

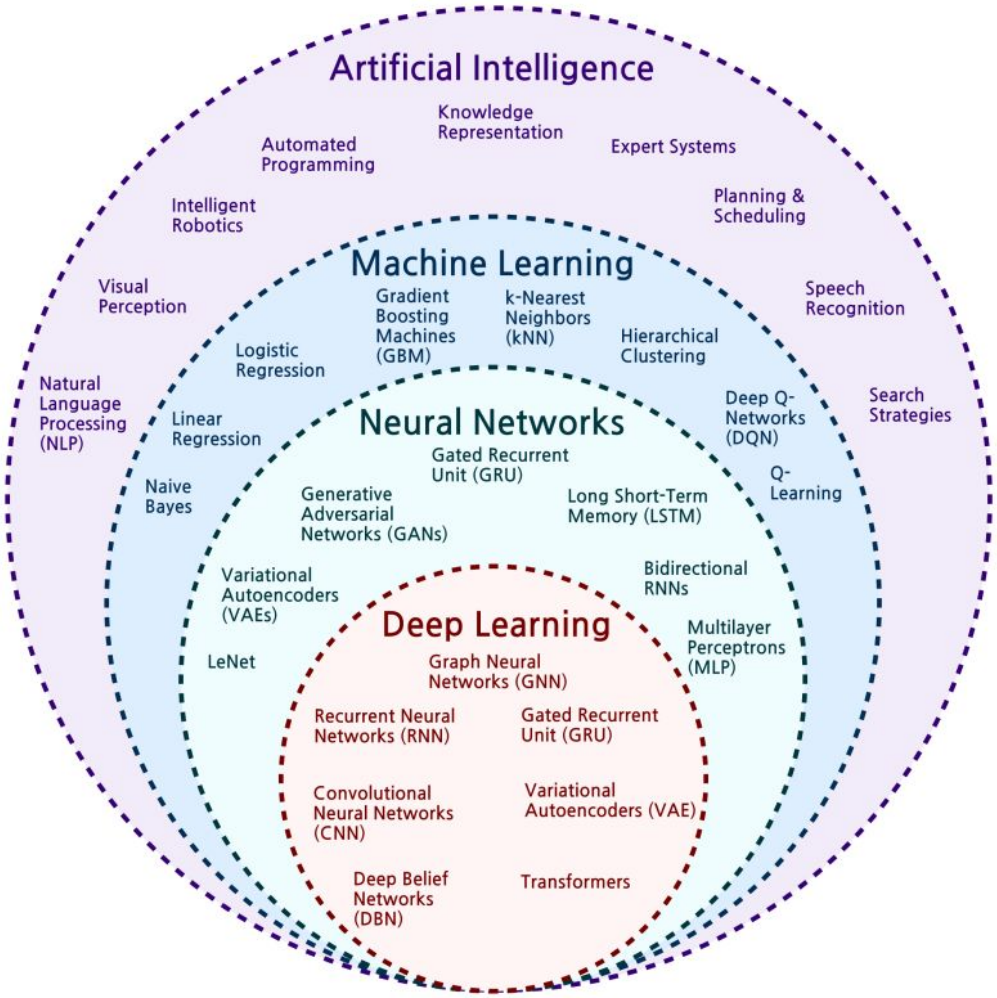


FIGURE II.1 – les différents domaines de l'intelligence artificielle.

II.2 Machine Learning (ML)

Le Machine Learning (ML), ou apprentissage automatique (AA)[17, 18], est une technologie qui permet à un ordinateur d'apprendre à accomplir une tâche en analysant des données, *sans que chaque étape soit programmée à la main*[17]. Concrètement, le système découvre lui-même des règles ou des tendances dans les informations qu'il reçoit. Par exemple, après avoir observé des milliers d'images de chats et de chiens avec leurs étiquettes, un algorithme de ML peut apprendre à reconnaître tout seul si une nouvelle image représente un chat ou un chien.

L'apprentissage automatique se décline en quatre grandes catégories principales II.2, chacune répondant à des types de problèmes bien définis : *l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage par renforcement et l'apprentissage semi-supervisé.* Ces approches diffèrent par la manière dont les données sont utilisées pour entraîner les modèles et par les objectifs visés.

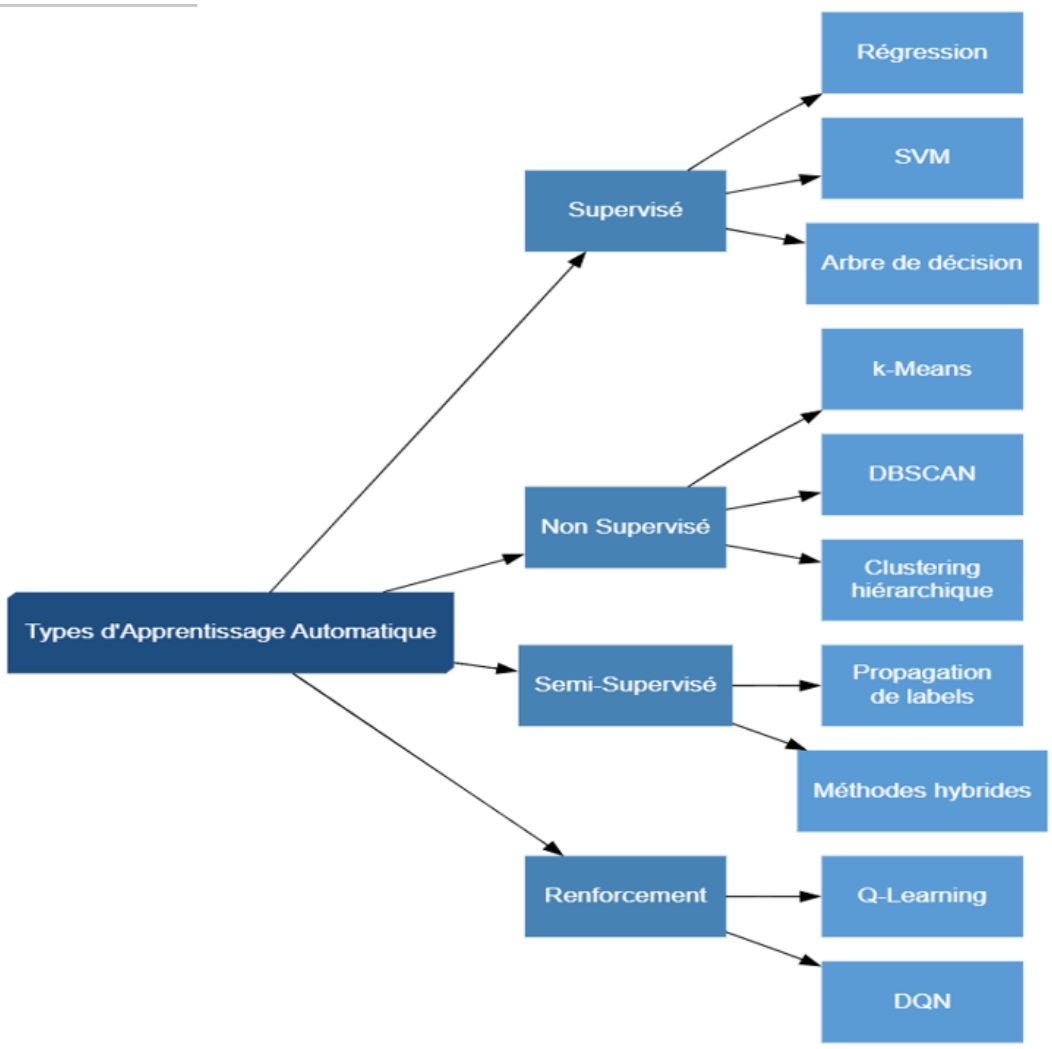


FIGURE II.2 – Types d'apprentissage Automatique

II.2.1 Apprentissage supervisé

L' apprentissage supervisé *(Supervised learning)* est comme un enseignement guidé : on montre à l' ordinateur des exemples de cas pour lesquels la solution est déjà connue [17, 18]. Par exemple, dans le cas d' un diagnostic médical, le programme apprend à faire le lien entre les *données d' entrée* (symptômes, résultats d' analyse, images médicales...) et le diagnostic à poser.
Exemple : *'Le diagnostic du cancer du sein à partir de mammographies* est un exemple d' apprentissage supervisé. Dans ce contexte, un modèle peut être entraîné à reconnaître les signes visuels d' une tumeur en analysant des milliers d' images médicales annotées par des médecins spécialistes. Chaque image utilisée pendant l' apprentissage est associée à une étiquette claire(Figure II.3), comme :

- Image 1 : tumeur bénigne,
- Image 2 : tumeur maligne (cancéreuse),
- Image 3 : aucune anomalie détectée.

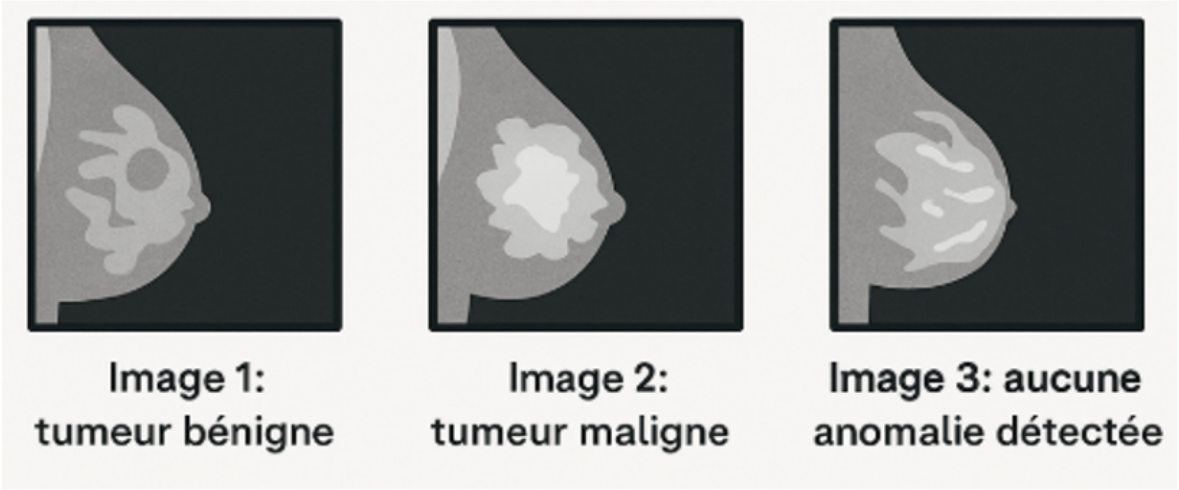


FIGURE II.3 – Illustration de l'étiquetage des images pour l'apprentissage automatique

Le modèle apprend à distinguer les caractéristiques visuelles associées à chaque type de cas : la forme, la texture, les contours de la masse, la densité, etc. Une fois entraîné, il devient capable d' examiner de nouvelles mammographies non annotées et de proposer un diagnostic automatique, en se basant sur les régularités qu' il a apprises.
Ce processus d' apprentissage peut s'appliquer à deux types de tâches, selon le type de sortie attendue :

II.2.1.1 Régression

Quand on souhaite prédire une valeur numérique continue. Par exemple, estimer la probabilité qu' une tumeur évolue en cancer à partir de données cliniques telles que la taille de la lésion, sa densité, ou les antécédents médicaux du patient.[18]

II.2.1.2 classification

Quand l' objectif est de catégoriser un cas dans un ensemble de classes : Identifier si une tumeur du sein est bénigne ou maligne à partir de mammographies, en comparant les nouvelles images à des milliers d'exemples connus.[18]

II.2.2 Apprentissage non supervisé

Contrairement à l' apprentissage supervisé, où les données sont étiquetées (avec une réponse connue), l' apprentissage non supervisé *(Unsupervised learning)* traite des données brutes, sans indication préalable [17, 18, 19]. L' objectif du modèle est d' explorer ces données et de *découvrir par lui-même des structures cachées,* des similitudes, ou des comportements récurrents[20]. On peut comparer cela à une *enquête sans piste initiale*, où l' algorithme cherche à organiser l' information de manière logique et cohérente, sans qu'on lui dise quoi chercher exactement. **Exemple :** "Dans un hôpital, on peut disposer de nombreux dossiers patients contenant des résultats d' examens, des constantes vitales, ou des antécédents médicaux. Si ces données ne sont pas étiquetées (pas de diagnostic confirmé), l' apprentissage non supervisé peut *regrouper les patients présentant des profils médicaux similaires.*" Cela permet, par exemple, d' *identifier de nouveaux sous-groupes de patients* ayant des évolutions cliniques proches, même si ces groupes ne sont pas connus à l' avance. C' est ce qu' on appelle le clustering (regroupement automatique).

II.2.3 Apprentissage semi-supervisé

L' apprentissage semi-supervisé (Semi-Supervised learning) est une approche hybride qui combine les avantages de l' apprentissage supervisé et de l' apprentissage non supervisé [18]. Il repose sur l' utilisation :

- D' un petit volume de données annotées (avec des étiquettes connues),
- Et d' un grand volume de données non annotées, souvent plus faciles à collecter mais plus coûteuses ou complexes à étiqueter.

L' objectif est de tirer parti des données non annotées pour améliorer les performances du modèle, tout en réduisant le besoin en annotation manuelle, souvent longue et exigeante dans les domaines spécialisés comme la médecine.

Exemple : 'Prenons le cas d' un hôpital disposant de *10 000 images médicales* (mammographies, IRM, scanners⋯). Annoter manuellement ces images pour indiquer si elles présentent ou non une pathologie *nécessite l' intervention de spécialistes*, ce qui peut être coûteux et chronophage. Avec l' apprentissage semi-supervisé, seuls 1 000 cas sont annotés par des experts. Le modèle commence son apprentissage sur cette base, puis utilise les *9 000 images restantes* pour affiner sa compréhension, en détectant des similitudes, des regroupements et des motifs partagés avec les images annotées. Cela permet de *réduire considérablement le temps et les ressources humaines nécessaires,* tout en obtenant un modèle performant.'

II.2.4 Apprentissage par renforcement

L'apprentissage par renforcement **(Reinforcement learning)** est une méthode inspirée du comportement humain et animal. Dans cette approche, *un agent (un programme ou un robot)* apprend à *prendre des décisions* en interagissant avec un environnement. À chaque action qu'il effectue, il reçoit une *récompense* ou une *punition* (positive ou négative), selon l'effet de son action [21]. L'objectif est que l'agent apprenne progressivement une stratégie optimale (appelée politique), qui lui permettra de *maximiser la récompense totale* au fil du temps. On peut comparer cela à un *apprentissage par essai-erreur,* où l'agent teste différentes actions, observe les conséquences, et ajuste son comportement pour obtenir les meilleurs résultats possibles[18].

Exemple : Imaginons un système de gestion automatique de traitement pour un patient atteint d'une maladie chronique (comme le diabète). Le modèle agit comme un agent qui doit décider, chaque jour, de la dose optimale d'insuline à administrer, en fonction de l'état du patient (glycémie, alimentation, activité physique$\cdots$). Au fil du temps, le système apprend, grâce aux retours obtenus (taux de glycémie stabilisé ou non), à ajuster ses décisions pour améliorer l'état de santé du patient. L'objectif est ici de minimiser les risques (hypoglycémies, complications) tout en maintenant un bon équilibre thérapeutique.

Le tableau suivant compare les différents types d'apprentissage automatique existant :

Type d'apprentissage	**Données**	**Sortie attendue**	**Exemple d'application**	**Algorithme de base**
Supervisé	Étiquetées	**Oui**	Reconnaissance d'images	• **Régression linéaire** : pour prédire une variable numérique continue. • **Régression logistique** : pour prédire une catégorie binaire (ex. : malade / non malade). • **Arbre de décision** : modèle basé sur une suite de règles conditionnelles. • **Random Forest** : ensemble d' arbres de décision pour plus de robustesse. • **Support Vector Machine (SVM)** : séparation optimale entre classes. • **k-Nearest Neighbors (k-NN)** : classification basée sur la proximité des données.
Non supervisé	Non étiquetées	**Non**	Segmentation de clientèle	• **k-Means** : regroupe les données en k groupes homogènes. • **DBSCAN** : détecte les regroupements denses, en ignorant les points isolés. • **Clustering hiérarchique** : construit une hiérarchie de groupes, visualisable sous forme d' arbre.

Type d'apprentissage	**Données**	**Sortie attendue**	**Exemple d'application**	**Algorithme de base**
Semi-supervisé	Partiellement étiquetées	**Oui**	Classification avec données rares	• **Propagation de labels (Label Propagation / Spreading)** : les étiquettes connues sont "diffusées" dans les données non étiquetées. • **Méthodes hybrides** : utilisent des algorithmes supervisés classiques (ex. : SVM, arbres) avec des mécanismes pour tirer parti des données non étiquetées. • **Réseaux neuronaux semi-supervisés** : combinent apprentissage supervisé + non supervisé dans une même architecture.
Renforcement	Interactions	Récompense	Jeux, robotique	• **Q-Learning** : méthode classique pour apprendre quelles actions donneront la meilleure récompense. • **Policy Gradient** : méthodes qui apprennent directement la meilleure stratégie d'action. • **Actor-Critic** : combinaison entre apprentissage de la stratégie (policy) et de l'évaluation des actions.

TABLE II.1 – Schéma récapitulatif des types d'apprentissage automatique.

II.3 Réseaux de neurones et Deep Learning

Toutes les technologies créées par l'être humain s'inspirent de processus naturels, qu'ils soient manuels ou automatiques. L'IA s'inscrit dans cette logique, en reproduisant certains mécanismes de la pensée humaine. Le cerveau humain fonctionne grâce aux neurones, qui permettent d'apprendre et de raisonner. Les réseaux de neurones artificiels s'inspirent de cette structure : ce sont des modèles simplifiés du fonctionnement des neurones biologiques, utilisant des formules mathématiques pour permettre à la machine d'apprendre et de s'orienter dans ses tâches.

II.3.1 Réseaux de neurones Artificiels

*Les réseaux neuronaux biologique*s forment la structure de base du cerveau humain et d'autres systèmes nerveux. Ces derniers sont constitués de neurones interconnectés qui communiquent à l'aide de signaux électriques et chimiques. Chaque neurone possède trois éléments fondamentaux :

- **Les dendrites**, qui reçoivent les signaux d'entrée ;
- **Le corps cellulaire** (ou soma), qui intègre ces signaux ;
- **L'axone**, qui transmet l'information vers d'autres neurones ou organes.

Les synapses jouent un rôle essentiel dans cette communication, en assurant les connexions entre neurones.[18]

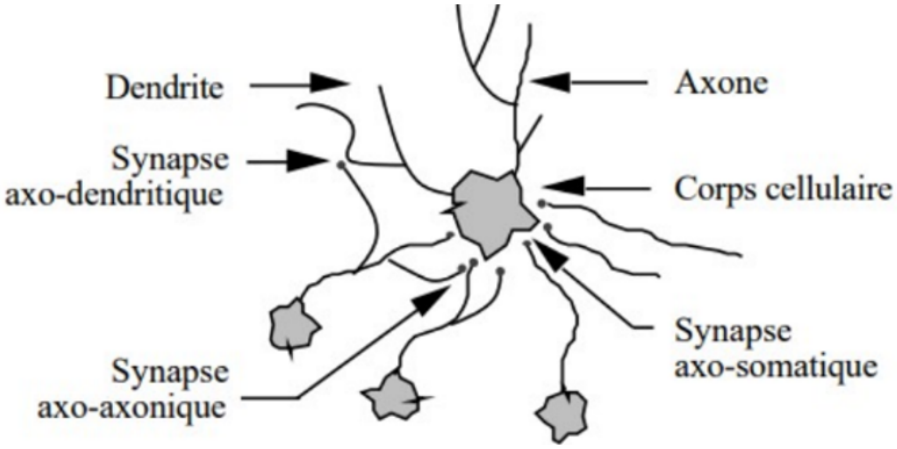


FIGURE II.4 – Illustration d'un neurone biologique.
[20]

Les réseaux neuronaux artificiels (RNA), ou Artificial Neural Network (ANN) une technique d'apprentissage automatique inspiré du fonctionnement biologique du cerveau humain et composées de plusieurs perceptrons ou ń neurones ż qui traitent et transmettent des informations. Ces réseaux acquièrent des connaissances en identifiant des modèles et des relations dans les données, et apprennent grâce à l'expérience et non par la programmation. [18, 20]

Les RNA appelé aussi réseaux de neurones formels [18] sont largement utilisés dans divers domaines tels que la reconnaissance d'images, le traitement du langage naturel, la reconnaissance vocale et la prédiction des prix des actions. Ils sont constitués de neurones interconnectés et peuvent être classés en plusieurs grandes catégories en fonction de leur architecture et de leur mode de fonctionnement.

II.3.1.1 Perceptron simple

Le perceptron simple, mis au point par Frank Rosenblatt en 1957, est une représentation mathématique et informatique d'un neurone biologique, le modèle le plus simple des RNA. Il se compose d' un unique neurone directement relié à une sortie [18, 20]. Chaque neurone formel reçoit plusieurs entrées (Xi), fait le calcul et donne une sortie comme résultats (voir la figure II.5). Son fonctionnement repose sur quatre étapes essentielles :

- **La pondération :** effectue un produit entre chaque entrée Xi et le poids qui lui est associé Wi.
- **La somme :** une somme pondérée des coefficients synaptiques:

$$\sum_{i=1}^{m}(Xi * Wi) \tag{II.1}$$

- **L'ajout du biais :** une valeur numérique est ajoutée à la somme précédente:

$$\sum_{i=1}^{m}(Xi * Wi) + \theta \tag{II.2}$$

- **Activation** le résultat obtenu précédemment est transformé avec une fonction d'activation :

$$\varphi\left(\theta + \sum_{i=1}^{m}(Xi \cdot Wi)\right) \tag{II.3}$$

Une ***fonction d' activation*** est une opération mathématique appliquée à la sortie d' un neurone pour introduire de la *non-linéarité* dans un réseau de neurones[18]. Elle détermine si un neurone doit être activé en fonction de sa sortie, ce qui permet au modèle d' apprendre des relations complexes. Sans elle, le réseau se comporterait comme une simple régression linéaire. **Il existe plusieurs types de fonctions d' activation, chacune adaptée à un contexte particulier :**[18]

- **ReLU** (Rectified Linear Unit) : met à zéro les valeurs négatives ; très efficace dans les réseaux profonds.
- **ELU** (l'unité linéaire exponentielle) : améliore la stabilité de l' apprentissage en gérant mieux les valeurs négatives.
- **Sigmoïde :** comprime la sortie entre 0 et 1, utile pour la classification binaire.

- **TanH** (Tangente hyperbolique) : transforme l' entrée dans l' intervalle [-1, 1], souvent utilisée dans les couches cachées.

- **Softmax :** convertit les scores en probabilités, utilisée en sortie pour les tâches de classification multiclasse.

- **Linéaire :** utilisée uniquement en sortie pour les problèmes de régression.

Le choix de la fonction dépend du type de tâche (classification, régression) et de la structure du réseau [18]. Le perceptron est principalement utilisé dans des tâches de classification binaire, où il permet de séparer des données en deux catégories distinctes (vrai ou faux / 1 ou 0). Exemple : la réalisation des tables de vérités.

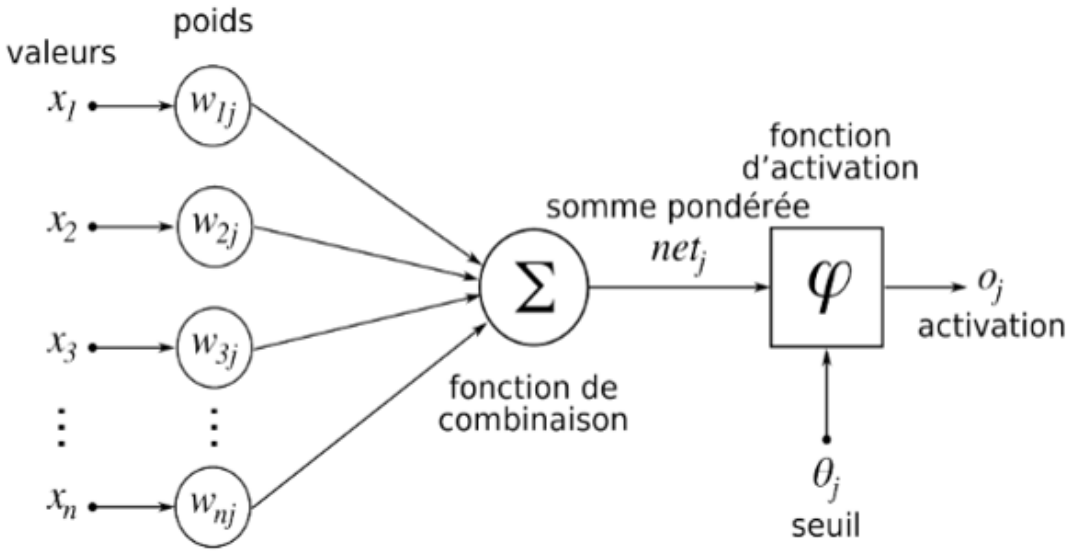


FIGURE II.5 – Le fonctionnement des neurones artificiels.
[20]

II.3.1.2 RNA a une seule couche cache

Un RNA classique, (appelé aussi un perceptron a une seule couche cachée : Single Hidden Layer Neural Network) [21, 18]. Est généralement constitué **d' une seule couche cachée** entre la couche d' entrée et la couche de sortie avec le flux d'information qui circule uniquement **dans un sens**. Cette architecture simple permet de traiter des tâches élémentaires comme la classification linéaire ou la régression, mais montre vite ses limites face à des données complexes ou fortement non linéaires que les méthodes classiques ne pouvaient appréhender. Comme la figure suivante l' illustre un RNA classique est constituée d' une couche d' entrée connectée à une**seule** couche cache, connectée par la suite à une couche de sortie :

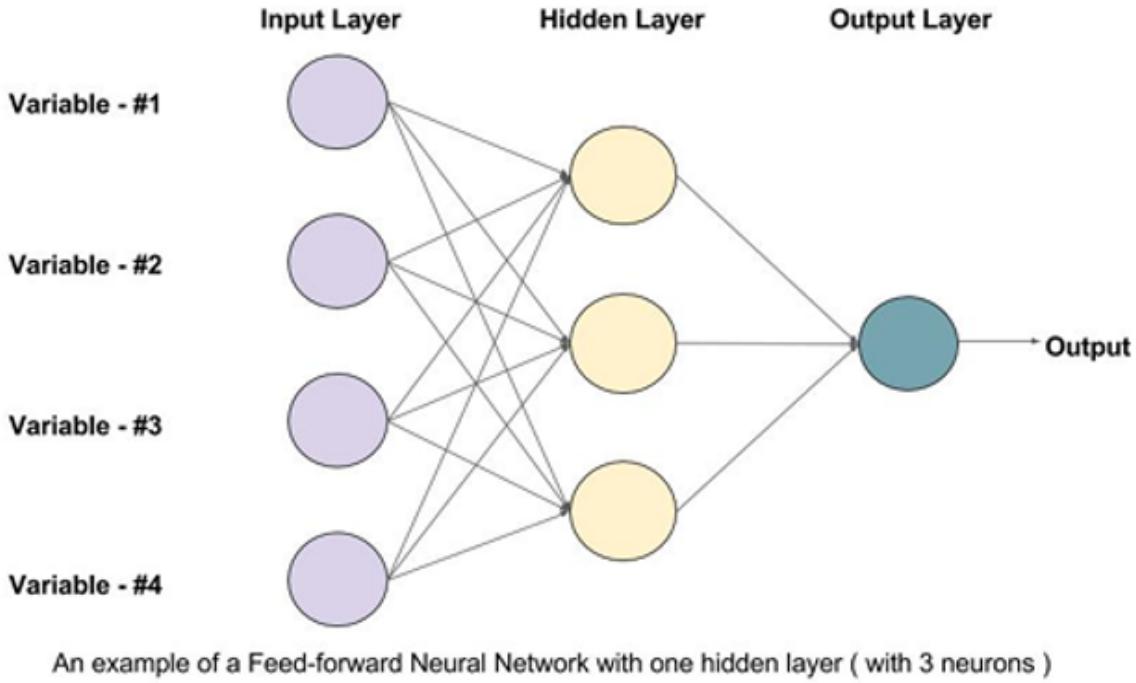


FIGURE II.6 – Un RNA a une seule couche cachée.
[18]

II.3.2 Réseaux de neurones profonds et Apprentissage Profond

II.3.2.1 Réseaux de neurones profonds

Les Réseaux de Neurones Profonds (RNP) ou Deep Neural Network (DNN)[21], une forme avancée des **RNA**. Contrairement aux RNA classique admettant une seule couche cachée, les RNP se constituent de plusieurs couches cachées, 'deux ou plus' couches cachées.[21, 18, 20]

Ces couches permettent au réseau d' apprendre des représentations complexes et abstraites des données en transformant progressivement les données d' entrée à travers différentes étapes de traitement. La couche d' entrée reçoit les données brutes, qui peuvent être des valeurs numériques, des pixels d' image ou du texte. Les données sont transmises à travers une ou plusieurs couches cachées, où chaque couche applique un ensemble de transformations mathématiques à l' aide de poids et de biais pour extraire des caractéristiques et des motifs de l' entrée[18].

La figure II.7 illustre la différence entre un ANN avec une seule couche cachée, et un RNP avec 3 couches cachées.

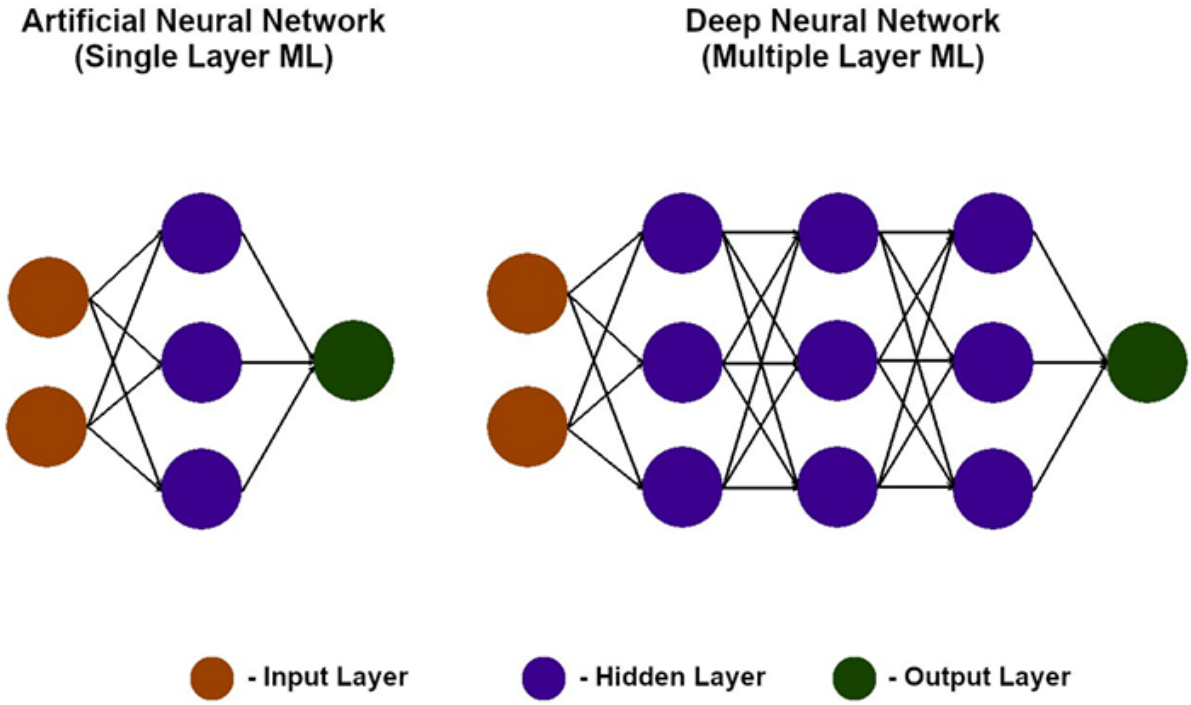


FIGURE II.7 – La différence entre un RNA avec une seule couche cachée (Single layer), et un RNP avec plusieurs couches cachées.
[21]

II.3.2.2 Apprentissage profond

Apprentissage Profond ou Deep Learning (DL) est une branche du Machine Learning qui repose sur des RNPs[21, 22, 18, 20], capables d'apprendre des représentations complexes et extraire des caractéristiques pertinentes à partir de grandes quantités de données brutes. DL a ouvert la voie à des progrès spectaculaires dans de nombreux champs d'application. Parmi :

- La reconnaissance d' images
- Le traitement du langage naturel (NLP)
- La reconnaissance vocale,
- La prédiction de séries temporelles (ex. prix des actions).

La puissance des RNP s'accompagne de besoins importants en données d'entraînement et en ressources computationnelles. Leur succès dans des tâches autrefois considérées comme inaccessibles aux machines a fondamentalement transformé le paysage de l'intelligence artificielle moderne.

II.3.3 Architectures des RNP

Dans ce qui suit, nous allons présenter une classification des architectures des RNP les plus utilisées en intelligence artificielle et DL. La figure suivante présente un résumée

bien structurée des architectures des RNP.

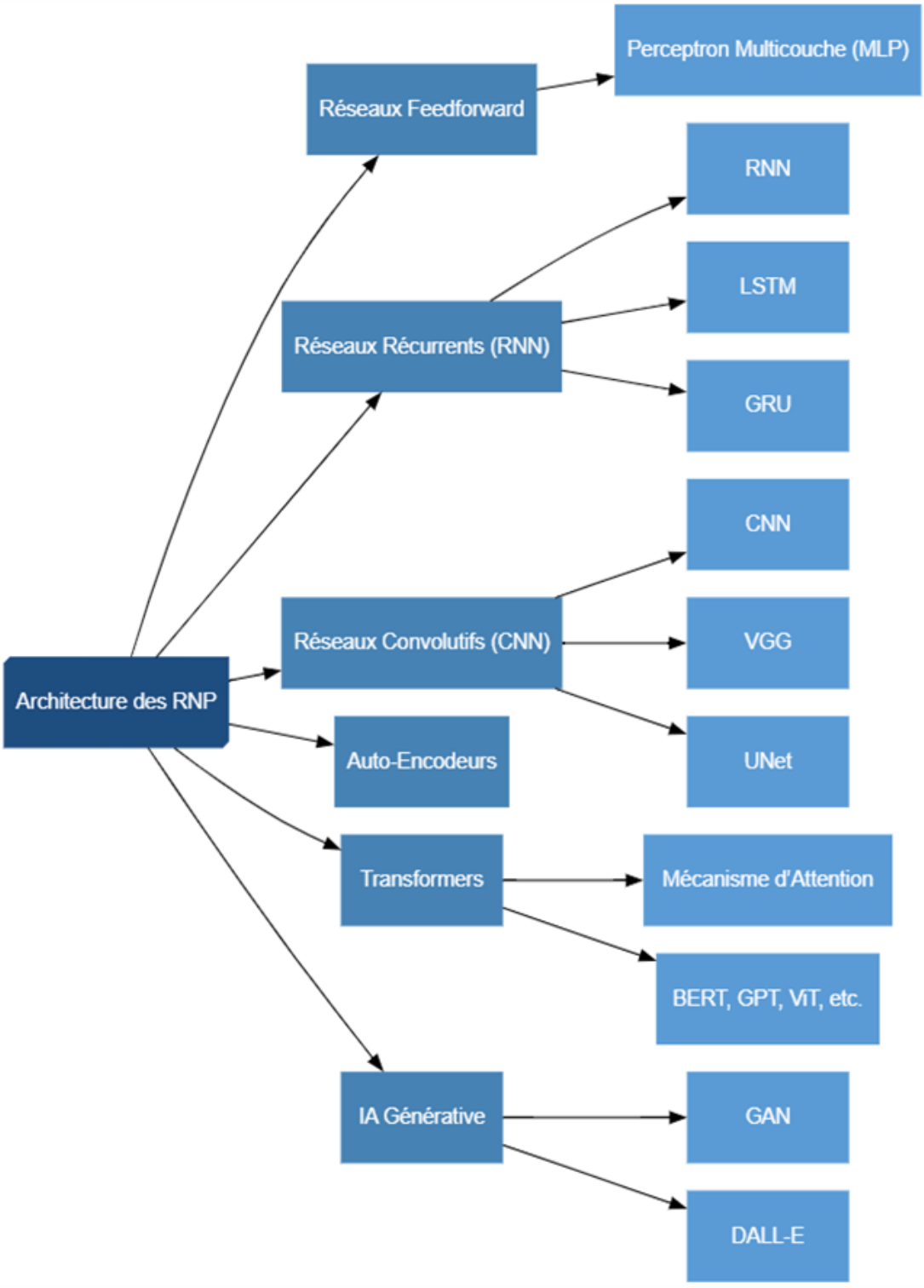


FIGURE II.8 – les différentes architectures des RNPs

II.3.3.1 Réseaux Feedforward

Perceptron Multicouche (ou MLP - Multi-Layer Perceptron), constitue la forme la plus basique de RNP, caractérisée par une connexion directe entre la couche d'entrée et la couche de sortie (appelé Réseaux Feedforward)[21, 18, 20]. Chaque neurone

réalise une combinaison linéaire pondérée des entrées reçues. Cette somme pondérée est ensuite transformée par une fonction d' activation, souvent de type seuil. Le modèle est conçu pour traiter des problèmes linéairement séparables, c'est-à-dire des situations où les données peuvent être divisées par une ligne ou un plan en espaces multidimensionnels[18, 20, 19]. Le RNA a une seule couche cachée, c' est qu' un type de MLP.

Exemple : La figure II.9 illustre ce concept en présentant un MLP doté d'une unique couche cachée. Dans cette illustration, les carrés verts représentent la couche d' entrée, qui correspond aux valeurs des pixels d' images de chiffres manuscrits. La couche intermédiaire est constituée de perceptrons représentés par des ronds gris. Ces perceptrons transforment et transmettent l' information provenant de la couche d' entrée. La couche finale comporte dix neurones, chacun étant associé à une classe de chiffres spécifique. Chaque neurone de sortie fournit la probabilité d'appartenance à sa classe correspondante.[19]

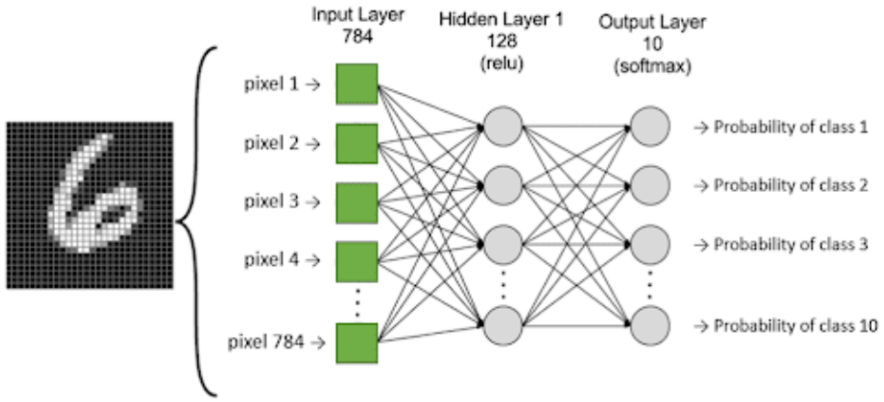


FIGURE II.9 – Perceptron multicouches avec une couche cachée pour la classification des chiffres manuscrites.

[19]

II.3.3.2 Réseaux Récurrents (RNN)

- **Réseaux de neurones récurrents (RNN)**:

  Les réseaux de neurones récurrents (RNN) se distinguent par des boucles de rétroaction qui permettent aux sorties d' un pas de temps d' être réinjectées comme entrée dans les étapes suivantes (voir la figure II.10), leur conférant ainsi une mémoire interne essentielle pour modéliser les dépendances temporelles. Contrairement aux réseaux feedforward qui traitent des données de manière statique, cette architecture dynamique est idéale pour manipuler des séquences ordonnées telles que le texte, la parole ou les séries temporelles. Cette capacité à saisir le contexte rend les RNN particulièrement efficaces pour des tâches comme l' analyse des sentiments, la traduction automatique ou la prédiction de séquences, même si leur apprentissage peut être entravé par le problème du vanishing gradient [1] . Afin de pallier cette difficulté, des variantes avancées comme les architectures LSTM et GRU ont été développées, offrant une meilleure gestion des dépendances à long terme tout en maintenant une

1. Le vanishing gradient dit aussi disparition du gradient : problème fréquent dans l' entraînement des RNNs.

complexité raisonnable pour des applications telles que la reconnaissance vocale et la modélisation du langage [18, 20, 22, 21].

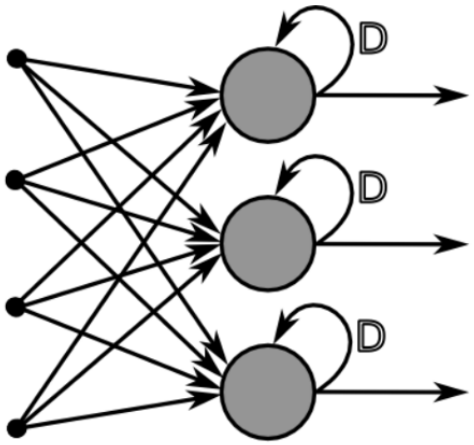


FIGURE II.10 – Architecture des réseaux de neurones récurrents

- **Réseaux à Mémoire Sélective LSTM** :

  Les LSTM (Long Short-Term Memory) sont une amélioration des RNN classiques, conçus pour mieux gérer les dépendances à long terme dans les données séquentielles. Contrairement aux RNN standards qui souffrent de la disparition du gradient, les LSTM intègrent un mécanisme de mémoire à long terme permettant de conserver des informations sur de longues périodes. Leur fonctionnement repose sur trois portes : la porte d' oubli, qui décide quelles informations anciennes éliminer ; la porte d' entrée, qui choisit les nouvelles données à stocker ; et la porte de sortie, qui contrôle l' information transmise à l' état caché pour les prédictions. Grâce à cette structure, les LSTM assurent une gestion efficace et stable des séquences temporelles complexes [22].

- **Unités récurrentes à portes GRU** :

  Les GRU (Gated Recurrent Units), représentent une alternative simplifiée aux LSTM, conçue pour modéliser efficacement les dépendances temporelles dans les séquences tout en réduisant la complexité du modèle. Contrairement aux LSTM qui possèdent une cellule mémoire distincte et trois portes de régulation, les GRU intègrent directement les mécanismes de contrôle dans l' état caché et n' utilisent que deux portes. La porte de réinitialisation permet de contrôler l' influence des états passés sur l'état actuel, tandis que la porte de mise à jour décide quelles informations anciennes doivent être conservées et quelles nouvelles doivent être intégrées. Cette simplification réduit le coût computationnel et accélère l' apprentissage, tout en maintenant des performances comparables à celles des LSTM dans de nombreuses tâches de traitement de séquences. Les GRU sont donc souvent préférés lorsque les ressources sont limitées ou que les dépendances temporelles sont modérées. Ils constituent une solution efficace et légère pour les applications de traitement du langage, des séries temporelles ou de la reconnaissance vocale[22].

II.3.3.3 Réseaux Convolutifs

- **Les Réseaux Neuronaux Convolutifs CNN :**

Les Réseaux Neuronaux Convolutifs **(CNN ou Convolutional neural Networks)** une architecture spécialisée de RNP, optimisée pour le traitement de données structurées spatialement (images, vidéos, cartes thermiques, etc.). Leur conception s' inspire du système visuel biologique, avec une capacité à extraire automatiquement des caractéristiques hiérarchiques (contours, textures, motifs complexes)[21, 18, 20, 22].

Les CNN sont composés de plusieurs couches spécifiques : une couche d' entrée, des couches de convolution, des couches de pooling, des couches entièrement connectées et une couche de sortie. Ils sont conçus pour analyser des données visuelles, telles que des images, en détectant automatiquement des motifs pertinents. Cela inclut les contours, les textures et d'autres caractéristiques essentielles, grâce à l'application de filtres convolutifs[18].

Les CNN se composent principalement de deux phases : *l' extraction de caractéristiques* et *la classification*. Lors de **l' extraction de caractéristiques**, la couche de convolution applique des filtres sur l' image d' entrée pour détecter des motifs locaux (contours, textures) et produire des cartes de caractéristiques, suivie d' une activation ReLU qui introduit de la non-linéarité. Ensuite, la couche de pooling (ou sous-échantillonnage) réduit la taille des cartes tout en conservant l' information essentielle. Deux types de pooling existent : le max pooling, qui retient la valeur maximale d' un champ réceptif, et l' average pooling, qui en calcule la moyenne. Bien que ces couches réduisent certaines informations, elles permettent de diminuer la complexité du modèle, d' améliorer l' efficacité computationnelle et de limiter le surajustement. Ce processus d' extraction est fondamental pour permettre aux CNN de construire des représentations hiérarchiques optimales à partir des données visuelles[18, 20].

En phase de **classification**, les couches entièrement connectées (Fully Connected) interviennent en fin de réseau pour transformer les caractéristiques extraites en prédictions finales. Fonctionnant comme un perceptron multicouche, elles combinent l' ensemble des informations issues des couches précédentes pour effectuer des décisions, telles que l' identification d' une classe dans une tâche de classification. Chaque neurone de ces couches est connecté à tous les neurones de la couche précédente, permettant une intégration globale de l'information extraite. Ce mécanisme finalise le processus du CNN en générant la sortie attendue, comme une probabilité pour chaque classe cible [18, 20].

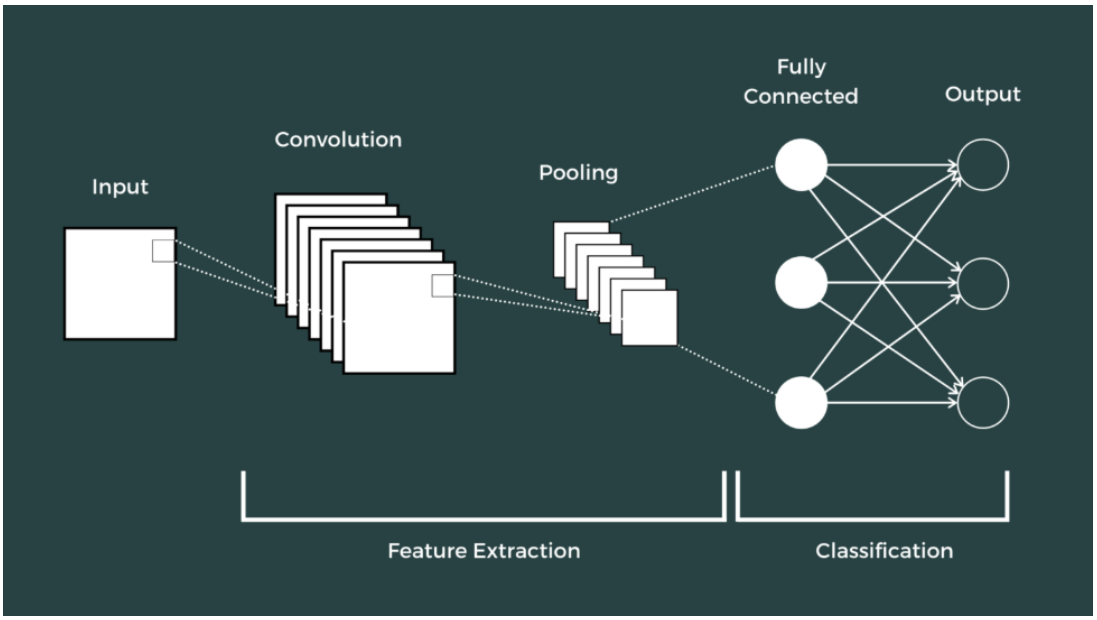


FIGURE II.11 – Architecture des CNNs.
[23]

Les CNN sont spécialement conçus pour traiter des données en deux dimensions et exceller dans des tâches de vision par ordinateur telles que la reconnaissance d' objets, l' analyse d' images médicales et la segmentation d' images. Leur capacité à détecter automatiquement les caractéristiques clés de l' entrée, sans intervention humaine, les rend plus performants que les réseaux traditionnels. Cependant, leur précision dépend de plusieurs facteurs essentiels, tels que le choix des fonctions d' activation, le nombre de couches cachées et la structure globale du réseau, nécessitant une optimisation soigneuse pour maximiser leurs performances.

Il existe plusieurs variantes de CNN, comme VGG, UNET, AlexNet, Xception, Inception et ResNet, chacune offrant des capacités d'apprentissage spécifiques adaptées à divers domaines d'application[18].

- **Visual Geometry Group (VGG) :**

  Le modèle VGG-16 est une architecture CNN, proposé en 2014 par Simonyan et Zisserman, le groupe Visual Geometry Group (VGG) à l'Université d'Oxford. Cette architecture, composée de 16 couches (dont 13 convolutionnelles et 3 entièrement connectées), est reconnue pour sa simplicité, sa profondeur et ses bonnes performances en vision par ordinateur, notamment en classification d' images et en reconnaissance d' objets. L' organisation du modèle repose sur une succession de couches convolutionnelles et de max-pooling, permettant l' apprentissage progressif de représentations visuelles complexes. Bien que plus ancien que d' autres modèles récents, VGG-16 demeure largement utilisé grâce à sa robustesse et sa polyvalence [24].

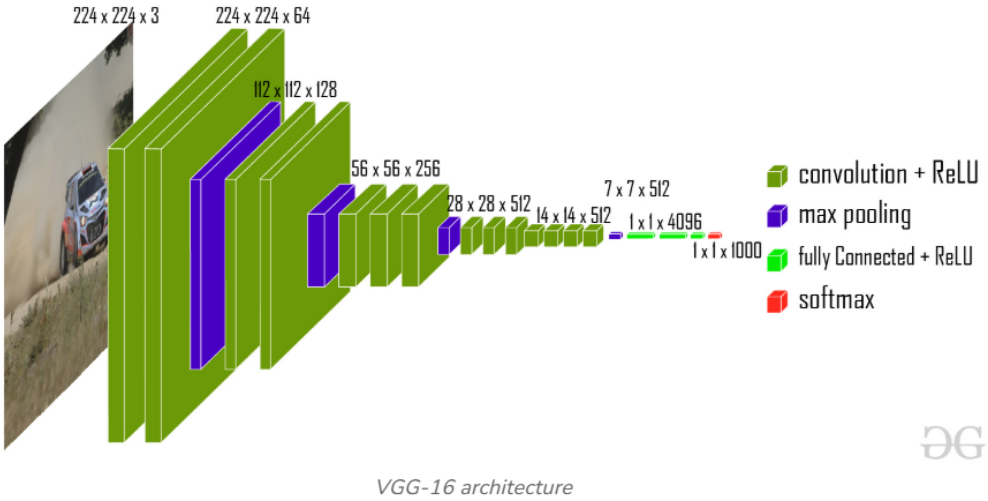


FIGURE II.12 – Architecture de VGG-16.
[24]

- **Les réseaux UNET :**

  Le réseau UNet, dérivé des CNNs, est une architecture en forme de "U" largement utilisée pour la segmentation d'images médicales (IRM, scanner) et de vidéos médicales. Il permet notamment de détecter des anomalies telles que les polypes, les caillots sanguins, ou encore de segmenter des organes comme les poumons [21]. Deux caractéristiques principales font de UNet un modèle particulièrement adapté aux données médicales, notamment vidéos :

  - Efficacité avec peu de données annotées : UNet offre de bonnes performances même avec un ensemble limité d'images d'entraînement.

  - Absence de couches entièrement connectées : cela permet au modèle de traiter des images ou des vidéos de tailles variables, sans contrainte d'entrée fixe. Cette flexibilité est particulièrement utile dans les vidéos médicales riches en contenu, garantissant une segmentation efficace des zones d'intérêt (ROI).

  Grâce à cette architecture, UNet surpasse les approches classiques en segmentant précisément des structures complexes, tout en étant adaptable à différents contextes médicaux[21] [25].

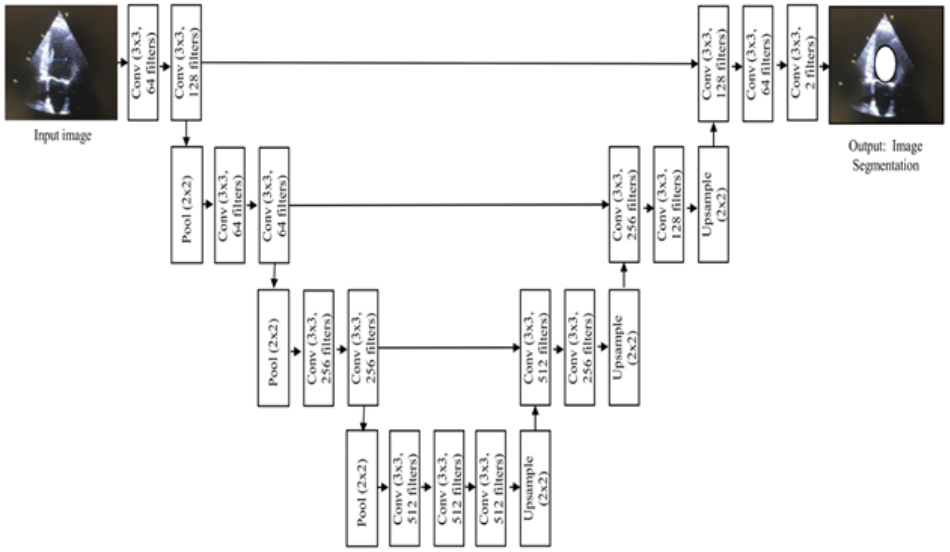


FIGURE II.13 – L'Architecture de UNet.
[25]

II.3.3.4 Réseaux de neurones à auto-encodeurs (Auto-encoders)

Les auto-encodeurs sont des RNP utilisés dans le cadre de l' apprentissage non supervisé. Ils sont particulièrement efficaces pour apprendre des représentations compactes et significatives des données. On peut les voir comme un système intelligent de compression et décompression, capable d' apprendre automatiquement à ne conserver que l' essentiel de l' information. Leur fonction principale consiste à compresser les données tout en préservant leur structure fondamentale, ce qui les rend également utiles pour la détection d' anomalies. Contrairement à des méthodes linéaires comme l'Analyse à Composante Principale (ACP), les auto-encodeurs permettent de capturer des représentations non linéaires, souvent plus adaptées à des données complexes [21] Les auto-encodeurs sont composés de deux éléments clés :

- **L' encodeur :** L' encodeur transforme les données d' entrée en une représentation latente de plus faible dimension. Il extrait les caractéristiques les plus pertinentes tout en éliminant les informations redondantes ou peu utiles. Cette étape permet au réseau de produire une version simplifiée mais informative de l' entrée.
- **Le décodeur :** Le décodeur utilise la représentation latente pour reconstruire les données originales. Il applique une série de transformations inverses dans le but de retrouver la structure initiale. La qualité de la reconstruction témoigne de la capacité du modèle à bien capturer l' essence des données.
  Les auto-encodeurs sont utilisés dans de nombreux domaines : réduction de dimension, détection d' anomalies, génération de données, compression multimédia.

II.3.3.5 Transformers

Avant de parler des Transformers, il est essentiel de comprendre le mécanisme d' attention, qui en constitue le noyau central.

- **Mécanisme d' attention :**

  Le mécanisme d' attention (Attention Mecanism) est un composant des modèles d' intelligence artificielle qui leur permet de focaliser leur attention sur certaines parties importantes d' une entrée jugée pertinente pour une tâche donnée. Il permet au modèle d' attribuer des poids aux différents éléments de l' entrée (images ou texte brute) en dirigeant son attention vers les informations les plus importantes[22, 26].

  Dans l' imagerie médicale, les mécanismes d' attention peuvent être utilisés pour mettre en évidence des régions ou des structures spécifiques dans une image. Par exemple, dans une radiographie pulmonaire, l' attention peut être dirigée vers des zones présentant une pathologie potentielle, aidant ainsi les radiologues dans leur diagnostic et améliorant la précision des algorithmes d' analyse d' images automatisée[25].

  Les mécanismes d' attention peuvent également faciliter la prise de décision clinique en orientant l' attention du modèle vers les informations pertinentes contenues dans les dossiers médicaux des patients. Lors de l' analyse des dossiers de santé électroniques ou de la littérature médicale, le modèle peut se concentrer sur des caractéristiques cliniques essentielles, des symptômes ou des options de traitement, aidant ainsi les professionnels de santé à prendre des décisions éclairées[26].

- **Transformers :**

  Les Transformers, introduits en 2017 par Vaswani et al. dans l'article "Attention Is All You Need"[26], sont des architectures avancées de réseaux neuronaux profonds largement utilisées en traitement automatique du langage naturel (NLP). Contrairement aux RNNs traditionnels comme les LSTM ou GRU, les Transformers s' appuient sur un mécanisme d' attention , en particulier l' auto-attention multi-tête [22], permettant d' évaluer l' importance relative de chaque mot dans une séquence. Cela leur confère une grande efficacité pour modéliser les dépendances longues tout en facilitant la parallélisation du traitement, ce qui est difficilement réalisable avec des architectures séquentielles. Composés de blocs encodeur-décodeur, les Transformers suivent un pré-entraînement sur de larges corpus, puis une phase de fine-tuning adaptée à des tâches spécifiques comme la traduction, la génération de texte ou l' analyse des sentiments. Ils sont également utilisés en reconnaissance vocale, en synthèse vocale et en réponse automatique aux questions. Malgré leurs performances remarquables, ils présentent des limitations : une complexité computationnelle élevée, une grande consommation de mémoire et une forte dépendance à des volumes massifs de données pour un entraînement efficace. Aujourd' hui, la majorité des modèles NLP de pointe (comme BERT, GPT, T5) reposent sur cette architecture puissante [22, 18, 20].
  La plupart des modèles de pointe actuellement utilisés dans le monde sont construits à l'aide de Transformers. Il s'agit de la conversion parole-texte, de la traduction automatique, de la génération de texte, de la paraphrase, de la réponse aux questions, de l'analyse des sentiments, etc. Voici quelques-uns des plus célèbres modèles :

Modèle	Type	Taille	Spécialité	Avantages	Limites
BERT (*Bidirectional Encoder Representations from Transformers*) : Premier modèle à encodeur bidirectionnel pour la compréhension contextuelle profonde.	Encoder	110M–340M	Compréhension contextuelle	Excellente performance en NLP	Unidirectionnel
GPT-4 (*Generative Pre-trained Transformer*) : Famille de modèles génératifs pré-entraînés, pionniers du langage autorégressif.	Decoder	∼1.8T*	Génération de texte	Génération cohérente, multimodale	Fermé, coût élevé
T5 (*Text-To-Text Transfer Transformer*) : Framework unifié de traitement de texte où toutes les tâches sont converties en format texte-à-texte.	Encoder-Decoder	220M–11B	Tâches variées (text-to-text)	Flexible (résumé, traduction, etc.)	Requiert fine-tuning
ViT (*Vision Transformer*) : Transformer pour l'analyse d'images par patches.	Encoder (Vision)	86M–632M	Classification d'images	Performances > CNN sur grands datasets	Gourmand en données
LLaMA-3 (Meta) (*Large Language Model Meta AI*) : *Modèle de Langage Massif de Meta, open-source*	Decoder	8B–70B	Open-weight, efficacité	Performant, libre d'accès	Restrictions d'usage commercial
Mistral : Famille de modèles open-source performants (dont des versions *MoE*), optimisés pour l'efficacité et la qualité.	Decoder	7B–22B	Efficacité computationnelle	Petit mais puissant (*MoE*)	Limité en contexte long
Gemini 1.5 : Famille de modèles multimodaux natifs développés par Google DeepMind.	Multimodal	∼1.56T	Texte, image, vidéo	Contexte ultra-long (1M tokens)	Modèle fermé

TABLE II.2 – Une comparaison entre différents modèles de Transformers

II.3.3.6 IA Génératif

Les **modèles génératifs** [21] occupent une place centrale en intelligence artificielle, car ils permettent de créer des données synthétiques réalistes telles que du texte, des images, des vidéos ou de la musique. Basés sur des architectures avancées utilisant des approches probabilistes ou compétitives, ils sont à l'origine d'innovations majeures dans de nombreux domaines. Parmi les plus utilisés, on trouve **GPT-4** pour la génération de texte multimodal, les **GAN** pour la création d'images visuelles, ou encore **DALL • E 3** [27], capable de transformer des descriptions textuelles en images réalistes. Ces modèles reflètent la progression rapide des technologies génératives et leur fort impact dans la création de contenu, la recherche scientifique et les applications artistiques.

- **Réseaux génératifs adverses (GAN - Generative Adversarial Networks)**

  Les **GAN** (réseaux génératifs adverses) [21] reposent sur une approche compétitive entre deux réseaux : un générateur, chargé de créer des données synthétiques en imitant la distribution des données d'origine, et un discriminateur, qui évalue la qualité des données produites en les comparant aux vraies. Leur utilisation s'étend à la génération d'images, de vidéos et autres contenus numériques[21, 25].

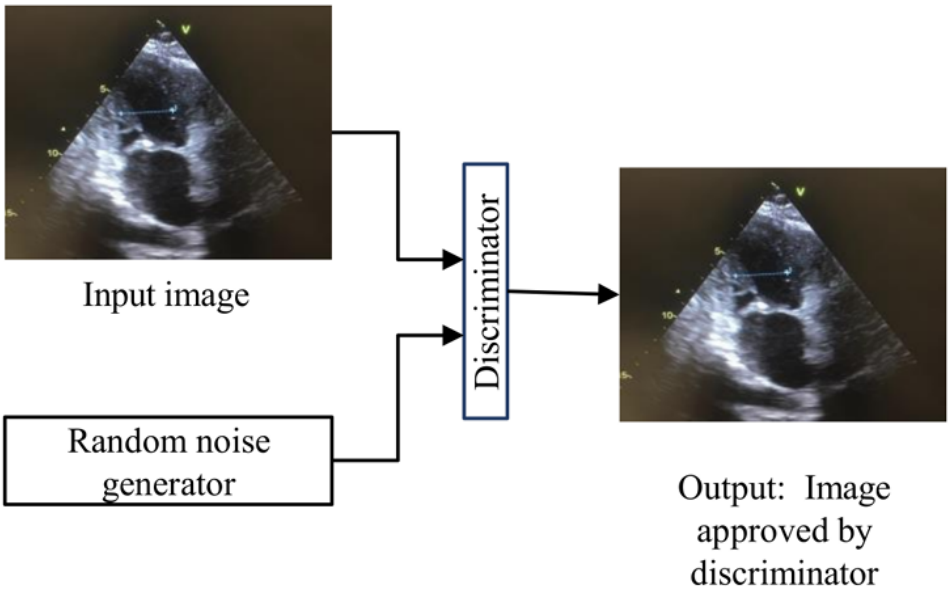


FIGURE II.14 – Architecture des GANs.
[25]

- **DALL • E 3** développé par OpenAI, est un modèle de génération d'images à partir de descriptions textuelles. Il illustre l'intégration avancée de la vision artificielle (VA) et du NLP, reposant sur des architectures de type transformer. Ce modèle est entraîné sur de larges ensembles de données en mode auto-supervisé, en utilisant la rétropropagation pour ajuster les poids des couches neuronales. DALL • E 3 constitue ainsi une application emblématique des réseaux de neurones profonds dans le domaine de la génération multi-modale, mobilisant des concepts tels que le transfert d'apprentissage (Transfer Learning) et l'alignement texte-image[27].

La version la plus récente, DALL • E 3, publiée en octobre 2023, intègre le modèle de langage GPT-4. Cette combinaison permet une compréhension plus fine des consignes textuelles et une génération d' images plus cohérente et détaillée. Le modèle est accessible via l' API OpenAI, mais également aux abonnés "ChatGPT Plus" directement depuis l' interface utilisateur [28].

II.4 Processus de création d' un RNP

Le processus de développement d' un réseau de neurones suit plusieurs étapes clés résumée dans le pipeline présenté suivant. En suivant cette démarche on se permet de développer un bon RNP capable d' apprendre efficacement à partir des données et d' être déployé dans un contexte réel pour répondre à un besoin concret :

1. **Collecte des données :** Il s'agit de rassembler les données nécessaires à l'entraînement du modèle. Ces données doivent être pertinentes, bien étiquetées, et représentatives du problème à résoudre.
2. **Création du réseau de neurones :** On conçoit l' architecture du réseau (nombre de couches, types de neurones, etc.), souvent en fonction de la tâche (classification, régression, génération...).
3. **Configuration du réseau :** On définit les paramètres comme la fonction d' activation, le taux d' apprentissage, l' optimiseur, la fonction de coût, etc.
4. **Initialisation des poids et biais :** Les poids du réseau sont initialisés, souvent aléatoirement. Ensuite, ils seront ajustés grâce à la rétropropagation, un algorithme d' apprentissage fondamental.
5. **Entraînement du réseau :** Le réseau apprend à prédire les bonnes sorties en ajustant ses poids via la rétropropagation, sur la base des données d' entraînement.
6. **Validation du réseau :** On vérifie les performances du réseau sur des données qu'il n' a jamais vues (données de validation), pour détecter un éventuel surapprentissage (overfitting).
7. **Utilisation et application du réseau :** Une fois entraîné et validé, le réseau peut être utilisé dans des applications réelles (reconnaissance d' images, prédiction, etc.).

II.5 Entraînement d' un RNP

L'entraînement d' un RNP repose sur un processus itératif d' apprentissage supervisé[18], visant à ajuster les paramètres internes du modèle pour améliorer ses performances prédictives. Ce processus se décompose en trois étapes principales :

- **Propagation avant (Forward Propagation)** : Les données d' entrée traversent successivement les couches du réseau, produisant une prédiction en sortie via des transformations linéaires suivies de fonctions d' activation non linéaires.

- **Évaluation de l'erreur** : Une fonction de perte calcule l'écart entre la prédiction et la valeur cible, fournissant une mesure quantitative de l'erreur.

- **Propagation inverse (Backpropagation)** : L'erreur est rétropropagée à travers le réseau, permettant de calculer les gradients des poids via la règle de la chaîne.

La figure II.15 présente la Propagation avant en bleu, et la rétropropagation en rouge :

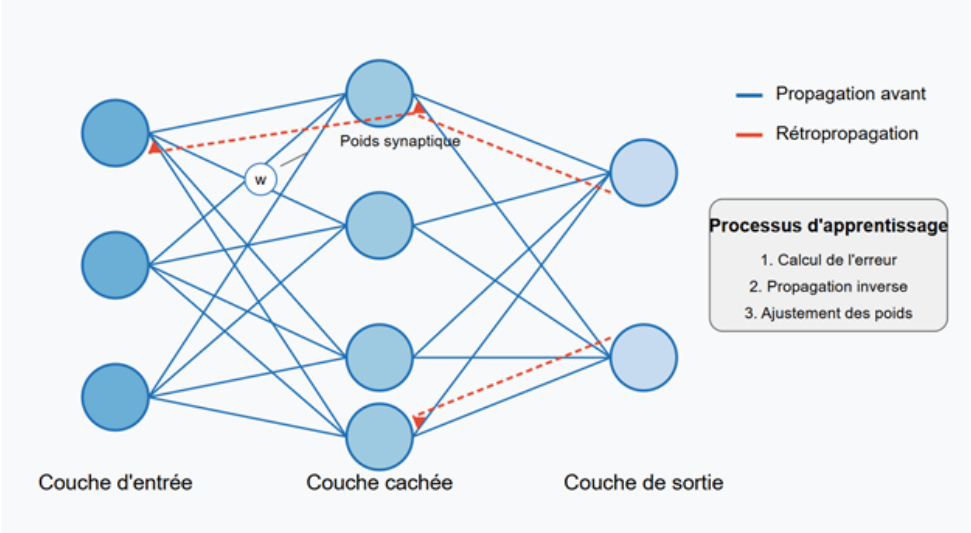


FIGURE II.15 – Rétropropagation dans un Réseau de Neurones Artificiels

II.5.1 Mise à jour des paramètres

Les poids et biais sont ajustés à l'aide d'un algorithme d'optimisation (comme la descente de gradient), afin de minimiser la perte sur l'ensemble des données.

- **Fonctions de coût :** Ces fonctions quantifient une pénalité globale associée à une action ou une séquence d'actions. Elles sont souvent utilisées dans des contextes décisionnels ou dans des modèles d'optimisation plus larges (comme l'apprentissage par renforcement ou les systèmes de planification) [18].

- **Fonctions de perte :** Ces fonctions sont spécifiques à l'entraînement supervisé (régression ou classification) et mesurent l'erreur entre la prédiction du modèle et la valeur réelle. Elles guident l'optimisation des paramètres du modèle [18].

II.5.2 Pour la régression

- **Erreur quadratique moyenne (MSE) :** Pénalise fortement les grandes erreurs.

- **Erreur absolue moyenne (MAE) :** Moins sensible aux valeurs aberrantes, traite toutes les erreurs de manière linéaire.

II.5.3 Pour la classification

- **Entropie croisée binaire :** Utilisée pour classer entre deux catégories (binaire).
- **Entropie croisée catégorielle :** Utilisée pour les tâches multiclasse avec softmax.
- **Algorithmes d' optimisation :** jouent un rôle fondamental dans l' apprentissage automatique, en particulier pour l' ajustement des paramètres des réseaux de neurones. Ils visent à minimiser une fonction coût en se basant sur des itérations successives. Plusieurs variantes ont été développées afin d' améliorer la stabilité, la vitesse de convergence et l' adaptation au contexte de grands ensembles de données : Descente de gradient, Adam, AdamW, etc.

II.6 Évaluation d' un modèle

L' évaluation d' un modèle est une étape cruciale du processus d' apprentissage automatique, car elle permet de mesurer ses performances et sa capacité à généraliser sur de nouvelles données[18]. Cette évaluation repose sur la division du jeu de données en trois ensembles :

- **Ensemble d'entraînement :** utilisé pour apprendre les paramètres du modèle à partir d' exemples.
- **Ensemble de validation :** permet d' évaluer le modèle pendant l' entraînement, de régler les hyperparamètres et de comparer différentes variantes du modèle. Il est utilisé de manière itérative.
- **Ensemble de test :** utilisé une seule fois, à la fin du processus, pour évaluer de manière impartiale la performance finale du modèle sur des données totalement inédites.

Chaque ensemble joue un rôle distinct, et ensemble, ils garantissent une évaluation fiable et représentative des performances réelles du modèle. [2]

II.6.1 La matrice de confusion :

La matrice de confusion est un outil d' évaluation utilisé pour mesurer la performance des modèles de classification. Elle compare les prédictions du modèle aux valeurs réelles sur un ensemble de test. Elle permet d' identifier les bonnes et mauvaises classifications à travers 4 types de résultats : *vrais positifs, vrais négatifs, faux positifs, et faux négatifs* [18].

Valeur réelle	**Valeur prédite = OUI**	**Valeur prédite = NON**
OUI	Vrai Positif (**TP**)	Faux Négatif (**FN**)
NON	Faux Positif (**FP**)	Vrai Négatif (**TN**)

TABLE II.3 – Interprétation d' une matrice de confusion
[18]

II.6.2 Critères d' évaluation basés sur la matrice de confusion

À partir des valeurs de la matrice de confusion (TP, TN, FP, FN), on peut calculer plusieurs métriques d' évaluation essentielles en apprentissage supervisé :

- **Accuracy (Taux de succès) :** C' est la proportion des prédictions correctes par rapport à l' ensemble des prédictions effectuées. Elle est utile pour évaluer la performance globale d' un modèle, mais peut être trompeuse en cas de déséquilibre des classes [18] :

$$\text{Accuracy} = \frac{TP + TN}{FP + FN + TP + TN} \tag{II.4}$$

- **Recall (Rappel ou Sensibilité)**Le rappel mesure la capacité du modèle à identifier tous les cas positifs réels. Utile quand les faux négatifs sont coûteux, comme dans les diagnostics médicaux [18]:

$$\text{Recall} = \frac{TP}{FN + TP} \tag{II.5}$$

- **Precision (Précision)** La précision indique la part des prédictions positives correctes parmi toutes les prédictions positives faites. Elle est cruciale lorsqu' il faut minimiser les faux positifs, comme dans les systèmes de détection de spam[18] :

$$\text{Precision} = \frac{TP}{FP + TP} \tag{II.6}$$

- **F1-Score** Le F1-score est la moyenne harmonique entre la précision et le rappel. Il est particulièrement utile lorsque les classes sont déséquilibrées, car il prend en compte à la fois les faux positifs et les faux négatifs. Un bon compromis quand on veut équilibrer précision et rappel [18] :

$$F1\text{-}score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2.7}$$

II.7 Approches avancées de l'apprentissage automatique dans les RNP

II.7.1 Transfer Learning

Le transfert d'apprentissage (Transfer Learning) est une technique d'apprentissage automatique qui consiste à réutiliser les connaissances acquises par un modèle sur une tâche source pour améliorer la performance sur une tâche cible similaire, souvent avec peu de données annotées. Elle permet de réduire les coûts en temps et en ressources liés à l'entraînement de modèles complexes, notamment en apprentissage profond. Cette approche est largement utilisée en traitement du langage naturel (NLP) et en vision par ordinateur, où des modèles pré-entraînés sur de vastes corpus servent de base pour des tâches plus spécifiques et ciblées[29, 21].

Pour adapter un modèle pré-entraîné à une nouvelle tâche, deux stratégies principales sont employées :

- **Extraction de Caractéristiques (Feature Extraction)**[29] :
  - **Principe** : On garde les premières couches d'un modèle pré-entraîné (ex. CNN), en gelant leurs poids. Les couches finales spécifiques à la tâche source sont retirées.
  - **Utilisation** : Les caractéristiques extraites sont utilisées comme entrée pour un nouveau classifieur (ex. SVM, couche dense).
  - **Avantages** : Économie de calcul, réutilisation facile sur d'autres tâches.
- **Affinage (Fine-tuning) :** [29, 21]
  - **Principe :** On dégèle certaines ou toutes les couches du modèle pré-entraîné pour les réentraîner sur les nouvelles données.
  - **Utilisation :** Idéal pour adapter le modèle à une tâche spécifique, à partir d'un bon point de départ.
  - **Avantages :** Meilleure adaptation à la tâche cible, convergence plus rapide.

  En extraction de caractéristiques, seuls les poids du nouveau classificateur final sont entraînés. En affinage, les poids du modèle pré-entraîné (en partie ou en totalité) sont également modifiés lors de l'entraînement sur la nouvelle tâche. Le choix entre ces stratégies, ou une combinaison des deux, dépend de la similarité entre les tâches source et cible et de la quantité de données disponibles pour la tâche cible.

II.7.2 Few shot Learning

Le Few-Shot Learning (FSL) est une méthode d'apprentissage automatique permettant à un modèle de traiter la tâche d'apprendre de nouveaux concepts ou catégories à partir de quelques exemples étiquetés limitées [21], souvent entre 1 et 10 par classe. Cette approche s'oppose aux techniques traditionnelles qui nécessitent de grandes quantités de données annotées pour obtenir de bonnes performances. FSL est particulièrement utile

dans des contextes où les données sont rares, coûteuses ou difficiles à collecter, comme la médecine, la reconnaissance faciale, ou certaines applications industrielles spécialisées [30]. Il en existe deux formes de FSL :

- **Zero-Shot Learning :** apprentissage à partir de zéro exemple (aucun exemple n' est présent pour l' apprentissage)
- **One-Shot Learning :** apprentissage à partir d' un seul exemple.
- **Few-Shot Learning :** apprentissage à partir de quelques exemples (souvent 5 ou 10).

Few-Shot Learning basé sur l' augmentation de données. Pour limiter le surapprentissage, on augmente les données via deux stratégies :

- Sur-échantillonnage (Oversampling) :
  - **SMOTE et H-SMOTE :** génèrent des exemples synthétiques pour équilibrer les classes minoritaires.
  - **Active PETs :** sélectionne les échantillons les plus utiles à annoter, en combinant sur-échantillonnage et comités de modèles.
- Méthodes génératives :
  - F2GAN et MetaGAN : génèrent des images réalistes à partir de peu d'exemples grâce aux GANs.
  - Méthodes bayésiennes : utilisent l' inférence probabiliste pour améliorer la généralisation à de nouvelles tâches [30]

Le Transfer Learning, avec ses deux approches principales, l' apprentissage inductif et l' apprentissage transductif, est aujourd' hui l' un des piliers de l' intelligence artificielle. Ces techniques font face à des défis majeurs, tels que l' adaptation à des domaines fortement divergents, la réduction du sur–apprentissage, ou encore la nécessité d' optimiser les performances avec peu de données annotées dans la tâche cible.

II.8 Conclusion

Ce chapitre a posé les bases théoriques essentielles de l' intelligence artificielle, en explorant les différentes formes d' apprentissage automatique et les principales architectures de réseaux de neurones les plus utilisées et incontournables. Il a permis de comprendre comment ces modèles apprennent à partir des données et comment ils sont évalués dans des contextes réels.

Dans le chapitre suivant, nous nous focaliserons sur un domaine d' application spécifique et central : la vision artificielle, avec une attention particulière portée au modèle YOLO, reconnu pour ses performances en détection d' objets.

Chapitre III

Vision Artificielle et Algorithme Yolo

Sommaire

III.1 Introduction **50**
III.2 Vision Artificielle **50**
III.3 Traitement d' images **51**
III.4 Pixel et résolution d' image **52**
III.5 Techniques de traitement d' image **54**
III.6 Évaluation des performances des modèles de VA **60**
III.7 YOLO : You Only Look Once **62**
III.8 Conclusion **72**

Résumé : *Ce chapitre explore la vision artificielle moderne, des techniques classiques de traitement d'images aux réseaux de neurones convolutifs avancés. L'algorithme YOLO révolutionne la détection d'objets en analysant une image entière en une seule passe, alliant vitesse exceptionnelle et précision remarquable. Cette base technique solide prépare l'application de ces technologies de pointe au diagnostic médical, transformant l'analyse visuelle automatisée en outil clinique innovant pour la détection précoce de pathologies.*

III.1 Introduction

La vision artificielle constitue aujourd'hui l'un des piliers majeurs de l'intelligence artificielle appliquée. Elle vise à doter les machines d'une capacité d'analyse visuelle similaire à celle de l'être humain, leur permettant de détecter, comprendre et interpréter automatiquement le contenu d'images ou de vidéos. Ce domaine, à la fusion du traitement d'images, de l'apprentissage automatique et des réseaux de neurones profonds, joue un rôle central dans des secteurs aussi variés que la santé, la sécurité, l'automobile ou encore la robotique.

Ce chapitre présente les fondements du traitement d'image, les types de représentations numériques, les opérations classiques comme le filtrage ou le seuillage, et les étapes essentielles de l'analyse visuelle. Les techniques avancées de Deep Learning sont ensuite abordées, avec un focus sur les modèles de détection d'objets en temps réel, en particulier YOLO (You Only Look Once), qui révolutionne l'interprétation visuelle par sa rapidité et sa précision. Cette exploration pose les bases nécessaires à la compréhension des choix technologiques du système de diagnostic automatisé.

III.2 Vision Artificielle

La vision artificielle (VA), ou vision par ordinateur (Computer vision en anglais) [1], est un sous-domaine de l'IA qui vise à simuler la perception visuelle humaine à l'aide de la puissance de calcul des ordinateurs. Elle repose sur des algorithmes avancés permettant d'accomplir des tâches telles que la *reconnaissance de formes*, la *détection d'objets* ou encore l'*analyse de mouvements*. En combinant les techniques de traitement d'images numériques avec l'apprentissage automatique, la vision artificielle ouvre la voie à des applications innovantes dans de nombreux secteurs, redéfinissant ainsi notre interaction avec le monde numérique.[18, 19, 1]

Son objectif va bien au-delà de la simple reconnaissance visuelle : il s'agit aussi d'interpréter les scènes, de comprendre les contextes et d'agir en conséquence. Cette technologie s'intègre aujourd'hui dans de nombreux secteurs, de la médecine à la sécurité, en passant par la robotique ou la réalité augmentée, transformant profondément notre relation aux systèmes numériques[1].

Les applications de la vision par ordinateur sont diverses, parmi lesquelles :

- **Détection d'objets** : Localiser et identifier des objets spécifiques dans une image (voiture, piéton, animal, etc.). Utilisée notamment en sécurité, vidéosurveillance ou véhicules autonomes.
- **Vision embarquée** : Intégrée dans les drones, robots, ou caméras intelligentes, pour analyser l'environnement en temps réel.
- **Reconnaissance faciale** : Identifier une personne à partir de son visage, à des fins d'authentification, de contrôle d'accès, ou de sécurité publique.
- **Reconnaissance d'activité** : Interpréter les gestes et actions humaines dans une vidéo (ex. : marcher, courir, tomber), utile en réalité virtuelle ou vidéosurveillance

intelligente.

- **Analyse d' images médicales** : Détecter des anomalies dans des radiographies, IRM ou scanners, pour assister les professionnels de santé dans le diagnostic.

- **OCR –Reconnaissance optique de caractères** : Lire automatiquement du texte à partir d' images (documents scannés, plaques d' immatriculation⋯), pour des usages comme la numérisation de documents ou la lecture automatique.

III.3 Traitement d' images

L' Imagerie ou traitement d'image s'appuie sur les progrès technologiques en informatique et en graphisme. Il regroupe l' ensemble des techniques permettant de créer, traiter et analyser des images numériques, dans le but d' en extraire des informations pertinentes selon le contexte d' application (médical, scientifique, satellitaire, etc.). [31]

III.3.1 L' image

Une image est une représentation visuelle en deux dimensions (2D) d' une scène ou d' un objet, souvent issue d' une réalité tridimensionnelle (3D). D' un point de vue mathématique, elle est modélisée comme une matrice de pixels, chaque pixel représentant une petite portion de l' image à une position donnée (X, Y) [18, 1] Il existe deux grandes catégories d' images numériques :

- **Les images vectorielles** : elles utilisent des formes géométriques simples (lignes, courbes, polygones) pour représenter des objets. Elles sont idéales pour les dessins techniques ou les schémas. *Exemple* : dessins animés.

- **Les images matricielles (ou bitmap)** : elles sont constituées d' un tableau de points ou pixels et sont les plus courantes pour les photographies ou les images réalistes. *Exemple* : une photo de paysage naturel.

Une image vectorielle peut être redimensionnée sans perte de qualité, contrairement à une image matricielle. Plus la densité de pixels est élevée, plus l' image contient d' informations. En conséquence, la taille en mémoire et le temps de traitement augmentent également.[1]

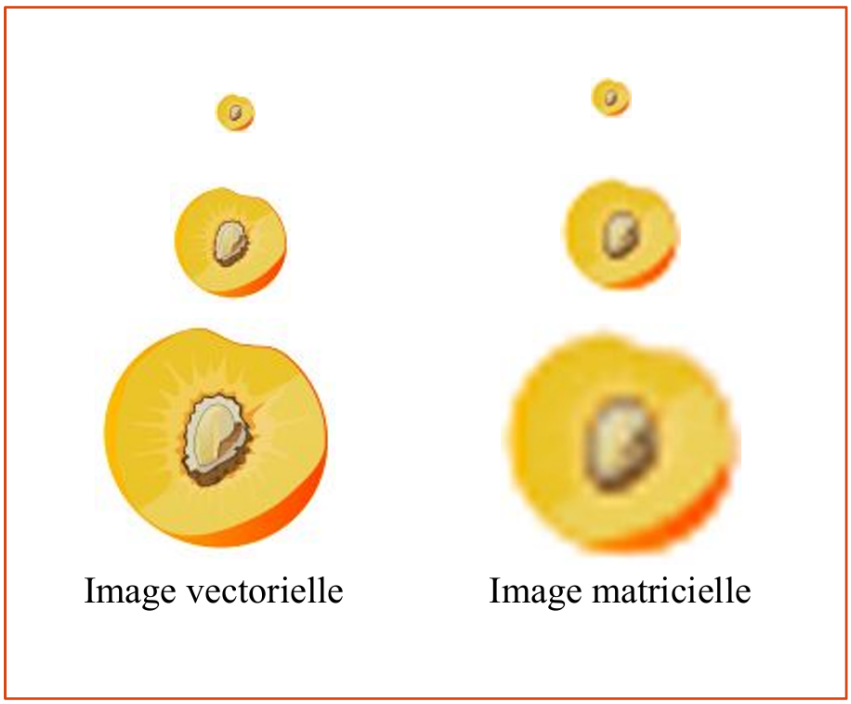


FIGURE III.1 – Image vectorielle Vs Image matricielle.
[1]

III.4 Pixel et résolution d'image

Un **pixel** (abréviation de *picture element*)[1] désigne l'unité élémentaire constituant une image numérique. Il s'agit du plus petit point lumineux visible à l'écran. Organisés en grille bidimensionnelle, les pixels forment une matrice régulière recouvrant toute la surface de l'image. Chaque pixel est directement connecté à ses voisins par les bords, ce qui permet une représentation continue du contenu visuel [1, 18].

III.4.1 Représentation d'une image en mémoire

En mémoire, une image est stockée sous forme de matrice 2D (deux dimensions). La structure de codage varie selon le type d'image [1] :

- **Image binaire (monochrome)** : chaque pixel est codé sur 1 bit, avec 0 pour noir et 1 pour blanc.
- **Image en niveaux de gris** : chaque pixel est codé sur 1 octet (8 bits), ce qui permet de représenter 256 nuances de gris, allant de 0 (noir) à 255 (blanc).
- **Image couleur (RVB)** : chaque pixel est codé sur 3 octets, représentant les intensités des composantes rouge, verte et bleue (R, G, B). Chaque composante est codée sur 1 octet (256 niveaux), et leur combinaison détermine la couleur finale. Cela constitue une représentation tridimensionnelle implicite, chaque pixel étant un vecteur de trois intensités.

La figure III.2 illustre une représentation d'une image en couleur avec des matrices.

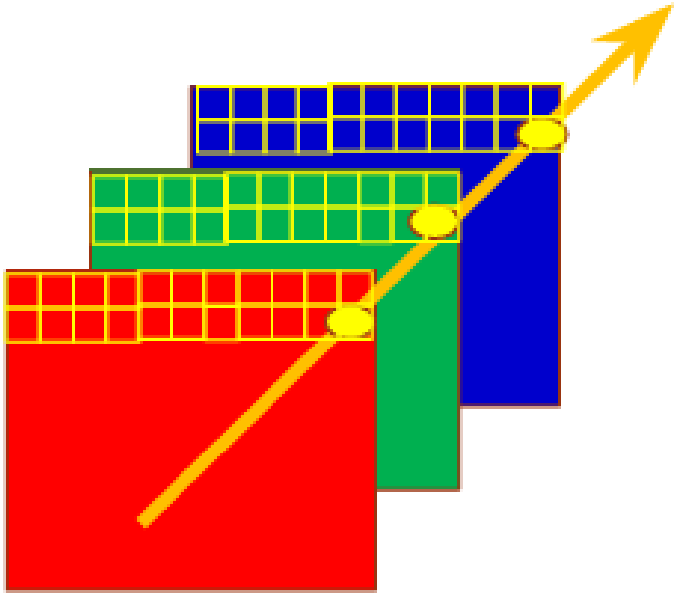


FIGURE III.2 – Représentation d' une image couleur en mémoire (RVB)[1]

III.4.2 Résolution d' image

La **résolution** d' une image correspond à la densité de pixels par unité de longueur, généralement exprimée en *dpi* (*dots per inch*) ou *ppp* (*points par pouce*), sachant qu' un pouce équivaut à 2,54 cm. Une résolution élevée traduit un niveau de détail plus important et une meilleure qualité visuelle. [1]

Trois caractéristiques fondamentales permettent de décrire une image numérique [1] :

- **Définition** : nombre total de pixels (exemple : 1280 Œ 960).
- **Dimensions physiques** : taille réelle de l' image, exprimée en centimètres ou en pouces.
- **Résolution** : densité de pixels par unité de surface.

La relation mathématique entre ces éléments est la suivante :

$$\text{Résolution (en pixels/pouce)} = \frac{\text{Définition}}{\text{Taille physique (en pouces)}}$$

III.4.3 Notions fondamentales du traitement d' image

- **Luminance** : intensité lumineuse perçue, représentant la brillance de l' image [31].
- **Contraste** : différence d' intensité lumineuse entre les zones claires et les zones sombres [31, 18].

- **Région homogène** : zone de l'image dans laquelle les pixels partagent des caractéristiques similaires, telles que l'intensité ou la couleur.[1, 31]
- **Bruit** : perturbation parasite introduisant des pixels incohérents dans l'image. Il peut être atténué par des techniques de filtrage adaptées.[1, 31]
- **Contours** : frontières entre deux régions distinctes de l'image, généralement détectées à l'aide de gradients ou de dérivées de l'intensité.[1]
- **Effet de bord** : distorsion observée au niveau des contours, souvent causée par le bruit ou par les traitements de filtrage[1]. Plusieurs méthodes de *padding* (remplissage) permettent de limiter cet effet :
  - **Zéros** : ajout de pixels noirs autour des bords.
  - **Reflet** : duplication symétrique des bords.
  - **Répétition** : extension des valeurs des pixels en bordure.

III.5 Techniques de traitement d'image

Le traitement d'image regroupe l'ensemble des opérations appliquées à une image numérique dans le but de l'améliorer, d'en extraire de l'information ou de faciliter son interprétation. Il existe deux grandes catégories de traitements [1]: **le traitement de bas niveau**, qui agit directement sur les pixels, et **le traitement de haut niveau**, qui cherche à comprendre le contenu de l'image de manière sémantique. La figure suivante présente une arborescence des principales techniques de traitement d'image, en distinguant clairement les approches de bas niveau (prétraitement visuel) et de haut niveau (interprétation sémantique).[1]

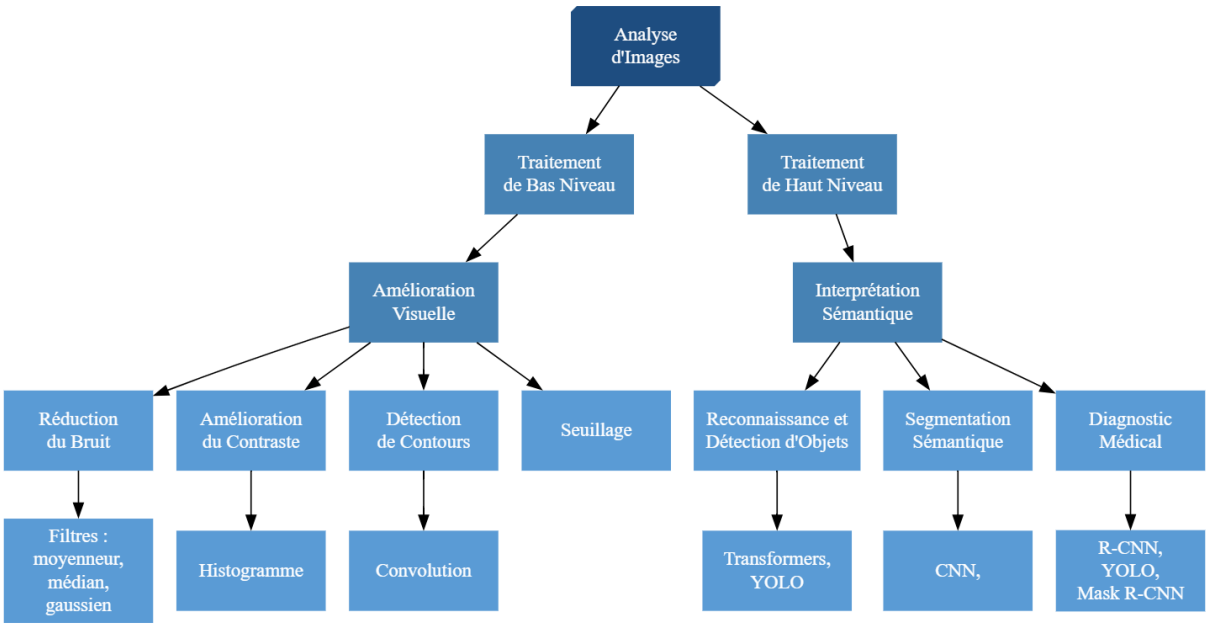


FIGURE III.3 – Différentes techniques de traitement d'image

Dans le tableau suivant présente la différence majeur entre le traitement bas niveau et le haut niveau dans le traitement d'image.

Aspect	Traitement de bas niveau	Traitement de haut niveau
Objectif	Améliorer la qualité visuelle, prétraiter l'image	Extraire une signification, comprendre le contenu
Niveau d'analyse	Pixel par pixel	Interprétation sémantique
Opérations typiques	Réduction du bruit des filtres (Filtre moyenneur, Filtre médian, Filtre gaussien) Amélioration du contraste (Histogramme) Détection de contours (Convolution) Seuillage (binarisation)	Reconnaissance et détection d'objets (YOLO, Transformers), Segmentation sémantique (CNN), Diagnostic médical (R-CNN, YOLO, Mask R-CNN).
Outils / techniques	Convolution, histogramme, transformation d'intensité	CNN, R-CNN, YOLO, Mask R-CNN, Transformers, réseaux profonds
Applications	Réduction du bruit, amélioration du contraste	Vision par ordinateur, reconnaissance faciale, diagnostic médical
Dépendance à l'apprentissage	Non (basé sur des règles simples ou filtres)	Oui (nécessite un entraînement sur de grands jeux de données)

TABLE III.1 – Comparaison entre les deux types de traitement d'image.

III.5.1 Seuillage

Le **seuillage** est une méthode simple de segmentation d'images qui consiste à classer les pixels en deux catégories selon une valeur seuil [31]:

- Les pixels inférieurs au seuil sont mis à zéro (noir),
- Les pixels supérieurs ou égaux sont mis à la valeur maximale (blanc).

Le résultat est une **image binaire**, ce qui permet de faire ressortir des objets ou des formes dans l'image (on parle aussi de binarisation).[1]

La figure III.4 illustre la conversion d'une image couleur en une image binaire via un processus de seuillage. Les pixels sont classés en noir ou blanc en fonction d'une valeur seuil définie.

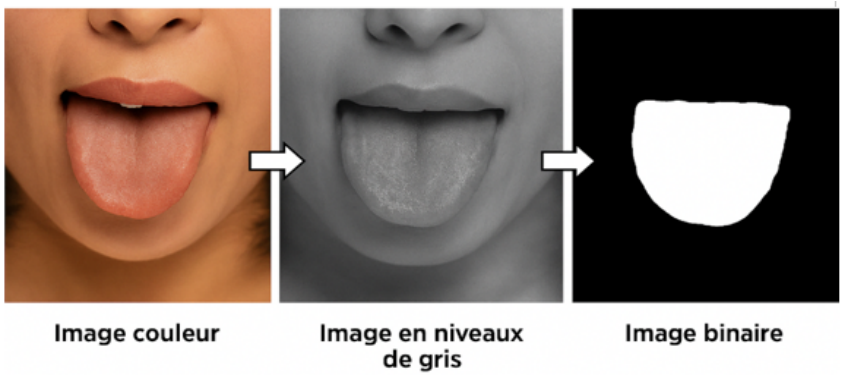


FIGURE III.4 – Processus de conversion d'une image couleur en image binaire par seuillage.

III.5.2 Histogramme

Un **histogramme d' image** représente la répartition des niveaux de gris dans une image. Pour chaque intensité lumineuse (de 0 à 255 pour une image codée sur 8 bits), il indique combien de pixels possèdent cette valeur.[31, 1]

Cela permet de **visualiser la dynamique** de l' image (contraste, luminosité) et d' appliquer différents traitements comme : l'amélioration du contraste, l' étalement de l' histogramme ou encore son égalisation.[1]

La figure suivante III.5 illustre deux histogrammes d' images, respectivement avant et après un étalement de l' histogramme.

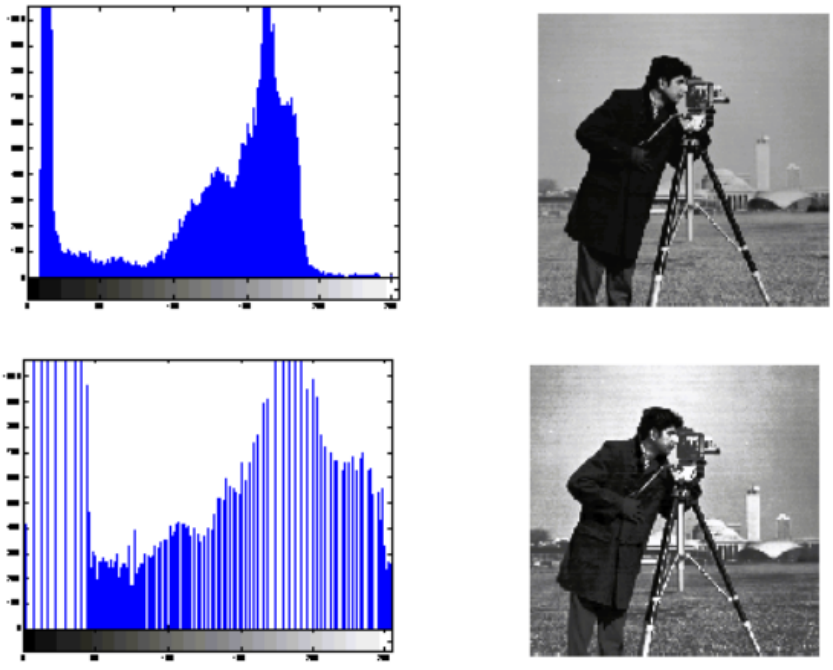


FIGURE III.5 – Exemple d' histogrammes.
[1]

III.5.3 Segmentation d' image

La segmentation d' image consiste à diviser une image en régions significatives pour en faciliter l' analyse[1]. Elle représente une étape essentielle dans les systèmes de vision par ordinateur, notamment en imagerie médicale ou en reconnaissance d' objets.[1, 31, 18]

Le tableau suivant présente les types de segmentations, leur principe ainsi les méthodes courantes utilisée pour chaque type:

Type de segmentation	**Principe**	**Méthodes courantes**
Par contours	Recherche de discontinuités dans l'image (bords, gradients)	Détection de contours (**Sobel**, **Canny**, **Laplacien**)
Par région	Regroupe les pixels ayant des propriétés similaires	Croissance de région, regroupement, seuillage
Hybride (par classification)	Combine les approches précédentes avec des algorithmes d' apprentissage	SVM, CNN, k-means, U-Net

TABLE III.2 – Les types de segmentation d' image.

Cette Figure III.6 illustre trois types de segmentation d'image : par contours, par région, et hybride.

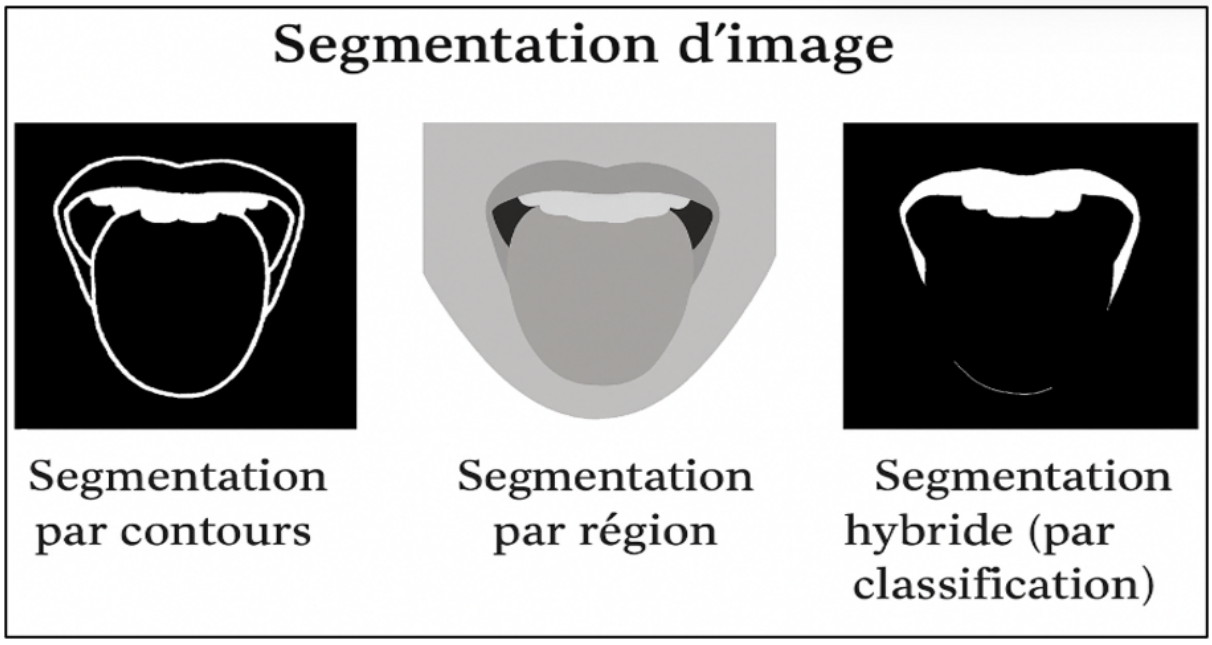


FIGURE III.6 – Exemple de types de segmentation.

III.5.4 Filtrage en traitement d' image

Une opération qui consiste à modifier les valeurs des pixels d' une image en tenant compte de leur voisinage, dans le but de réduire le bruit, améliorer la qualité visuelle, restaurer une image dégradée ou mettre en évidence certains détails (comme les contours). Le filtrage intervient principalement en phase de prétraitement, pour faciliter l' analyse ou l' interprétation d' une image. [31, 1]

III.5.4.1 Filtrage linéaire

Le **filtrage linéaire** repose sur une *convolution* entre l' image et un masque de valeurs fixes (appelé *noyau*) [1]. Il est particulièrement adapté pour *adoucir une image* ou *atténuer les perturbations globales*, mais il a pour inconvénient de flouter les contours.

Exemples les plus utilisés :

- **Filtre passe-bas** : Ce filtre sert à "flouter" l' image. Il enlève les petits détails et le bruit (les petits points gênants) pour rendre l' image plus lisse. On l' utilise souvent avant d' analyser une image, pour éviter que des éléments inutiles ne perturbent les résultats. L' image devient un peu floue, mais plus propre.[1]

- **Filtre passe-haut** : Ce filtre fait le contraire : il rend l' image plus "nette" en mettant en évidence les contours. Il supprime les zones uniformes et fait ressortir les bords, ce qui aide à mieux voir les formes. C' est très utile pour détecter les objets ou les séparations dans une image.[1]

III.5.4.2 Filtrage non linéaire

Le **filtrage non linéaire** remplace chaque pixel par une valeur calculée selon une règle non linéaire [1]. Il est plus robuste face aux valeurs aberrantes, telles que le bruit impulsionnel (de type sel et poivre).

Exemples les plus utilisés :

- **Filtre médian** : remplace un pixel par la médiane de ses voisins —idéal contre le bruit impulsionnel.[1]

- **Filtre min/max** : remplace un pixel par la valeur minimale ou maximale de son voisinage —utile pour atténuer ou accentuer les extrêmes.[1]

- **Filtre moyenneur** : calcule la moyenne locale —adoucit les variations douces sans trop déformer l' image.[1]

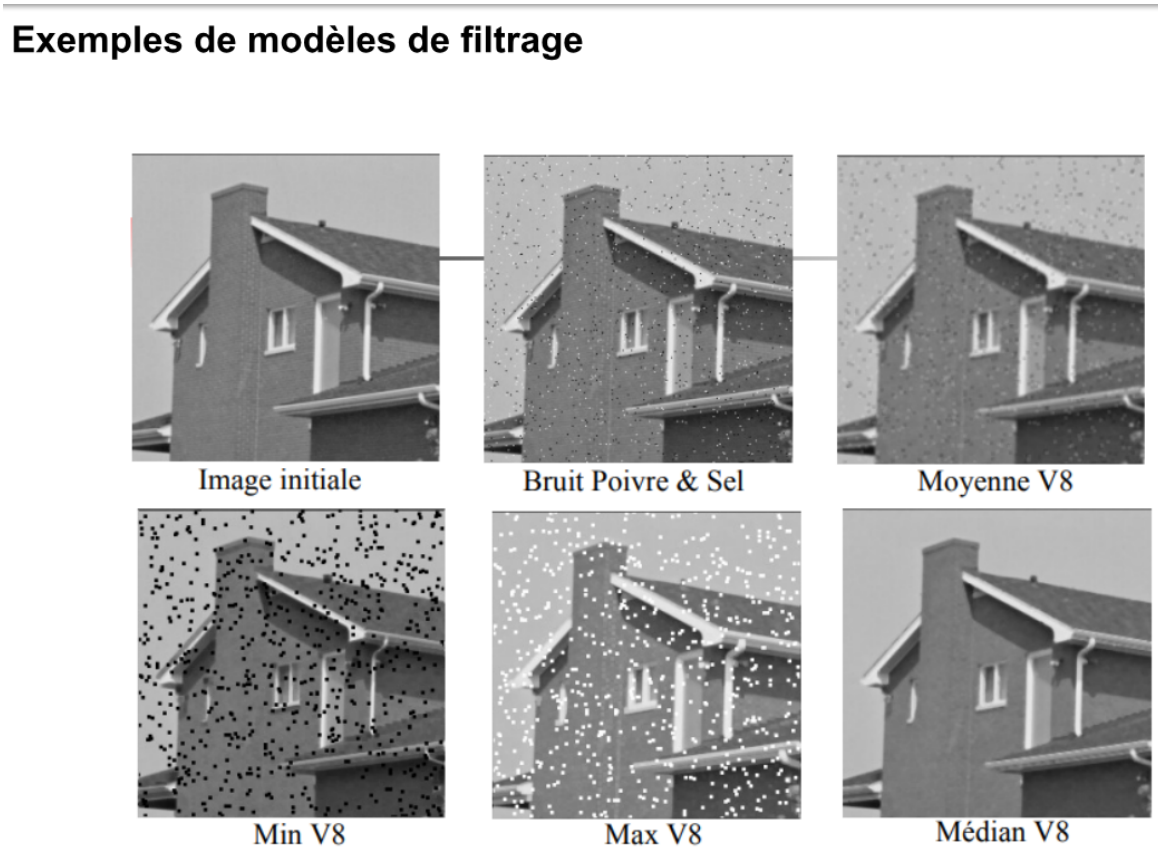


FIGURE III.7 – Comparaison des effets des filtres linéaires et non linéaires sur une image bruitée.

[1]

III.5.5 Techniques de Deep Learning en Vision par Ordinateur

Grâce au Deep Learning, et plus précisément aux CNN, la vision par ordinateur a connu des avancées majeures dans plusieurs domaines [20, 19]. Ces modèles permettent aujourd'hui de réaliser des tâches complexes telles que le classification d'images, reconnaissance faciale, détection d'objets ou encore la segmentation sémantique.

Plusieurs architectures de CNN ont marqué les évolutions récentes[18, 20, 1], les plus plus connue et utilisés sont :

- **VGG (Visual Geometry Group)** : Architecture simple et profonde basée sur des couches convolutives empilées. Utilisée principalement pour la classification d'images standards.
- **ResNet (Residual Network)** : Introduit les connexions résiduelles pour faciliter l'apprentissage dans des réseaux très profonds et éviter la perte d'information.
- **Inception** : Permet une extraction multi-échelle des caractéristiques grâce à des blocs combinant plusieurs tailles de filtres en parallèle.
- **EfficientNet** : Optimise l'efficacité du réseau en équilibrant sa profondeur, sa largeur et sa *résolution*. Offre un excellent rapport performance/coût.

- **YOLO (You Only Look Once)** : Une approche innovante pour la détection d'objets en temps réel. L'image est analysée en une seule passe par le réseau, ce qui permet à YOLO d'être extrêmement rapide et efficace. Il sera détaillé dans la section suivante.

III.6 Évaluation des performances des modèles de VA

Pour mesurer la qualité des modèles en VA, on utilise différentes métriques selon la nature de la tâche : segmentation, détection ou classification [20, 18, 1]. Ces indicateurs permettent de comparer les résultats du modèle avec les données réelles (vérités terrain), et d'ajuster les performances en fonction des besoins (précision, couverture, équilibre$\cdots$).

III.6.1 Tâche de segmentation d'images

La segmentation vise à attribuer un label à chaque pixel de l'image. Les principales métriques utilisées [32] sont :

- **mIoU (Mean Intersection over Union)** : mesure la qualité de la prédiction pixel par pixel en comparant l'intersection et l'union entre la zone prédite P et la zone réelle G :
$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \frac{|P_i \cap G_i|}{|P_i \cup G_i|}$$

- **Dice Coefficient (ou F1-score en segmentation)** : proche du mIoU mais donne plus de poids aux pixels bien segmentés :
$$\text{Dice} = \frac{2 \cdot |P \cap G|}{|P| + |G|}$$

- **Pixel Accuracy (PA)** : pourcentage de pixels correctement classés par rapport au total :
$$\text{PA} = \frac{\text{Nombre de pixels corrects}}{\text{Nombre total de pixels}}$$

- **Boundary F1-Score (BF-score)** : évalue la qualité des contours prédits, particulièrement utile pour des tâches où la précision des bords est cruciale.

III.6.2 Tâche de détection d'objets

Dans la détection, l'objectif est de localiser les objets dans une image avec des boîtes englobantes [32]. On utilise principalement :

- **IoU (Intersection over Union)** : compare une boîte prédite B_p à une boîte réelle B_r :

$$\text{IoU} = \frac{\text{Aire}(B_p \cap B_r)}{\text{Aire}(B_p \cup B_r)}$$

Un objet est considéré comme bien détecté si l'IoU dépasse un seuil (souvent 0.5).

- **AP (Average Precision)** : moyenne des précisions calculées à différents rappels. C'est l'aire sous la courbe précision-rappel pour une classe donnée :

$$\text{AP} = \int_0^1 p(r)\, dr \quad \text{(approximée par sommation discrète dans la pratique)}$$

- **mAP (Mean Average Precision)** : moyenne des AP pour toutes les classes :

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i$$

On parle de $mAP@50$ pour un seuil d'IoU = 0.5, et $mAP@[.5 : .95]$ pour la moyenne des AP avec des seuils d'IoU variant de 0.5 à 0.95 (par pas de 0.05).

- **Précision (Precision)** : part des prédictions correctes parmi toutes les prédictions positives.

- **Rappel (Recall)** : part des objets réels bien détectés.

III.6.3 Tâche de classification d'images

Dans la classification, on cherche à prédire la bonne étiquette pour une image entière.[33]

- **Accuracy (précision globale)** : pourcentage de bonnes prédictions.

- **Top-k Accuracy** : Une prédiction est considérée correcte si la bonne classe se trouve parmi les k classes les plus probables (ayant les plus hauts scores de prédiction). La formule est :

$$\text{Top-}k \text{ Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left(y_i \in \hat{Y}_i^{(k)}\right)$$

où :

  - N est le nombre total d'exemples,
  - y_i est la classe réelle de l'exemple i,
  - $\hat{Y}_i^{(k)}$ est l'ensemble des k classes les plus probables prédites pour l'exemple i,
  - $\mathbb{1}(\cdot)$ est la fonction indicatrice qui vaut 1 si la condition est vraie, sinon 0.

- **ROC-AUC (Receiver Operating Characteristic –Area Under Curve)** : Cette métrique évalue la capacité du modèle à distinguer les classes positives et négatives. Elle est basée sur la courbe ROC qui trace :
  - l'axe des abscisses : **FPR (False Positive Rate)** $= \frac{FP}{FP+TN}$
  - l'axe des ordonnées : **TPR (True Positive Rate)** = **Recall** $= \frac{TP}{TP+FN}$

  L'aire sous cette courbe (AUC) est calculée comme :

$$\text{AUC} = \int_0^1 \text{TPR}(x)\, dx$$

- **F1-Score** : moyenne harmonique entre précision et rappel, utile surtout en cas de classes déséquilibrées.

III.7 YOLO: You Only Look Once

YOLO signifie **"You Only Look Once"**, c'est-à-dire "vous ne regardez qu'une seule fois". Ce nom reflète le concept fondamental du modèle : analyser toute l'image en une seule passe à travers un réseau de neurones, pour détecter et localiser plusieurs objets à la fois. Contrairement aux méthodes classiques qui découpent l'image en régions ou passent par plusieurs étapes (proposition, classification, ajustement), YOLO effectue toutes les opérations simultanément, ce qui le rend extrêmement rapide. [20, 2]

III.7.1 Fonctionnement de YOLO

Le fonctionnement de Yolo peut être résumé en trois étapes principales : *le découpage en grille, le traitement par réseau neuronal, puis le post-traitement.*

III.7.1.1 Découpage virtuel de l'image en grille

Dès le début du processus, YOLO divise virtuellement l'image en une grille carrée (par exemple 13Œ13 ou 20Œ20 cellules). Chaque cellule de cette grille est responsable de prédire :

- S'il y a un objet dans sa zone (oui ou non),
- Les coordonnées de la boîte englobante : (x, y) pour la position du centre de l'objet, (w, h) pour sa largeur et sa hauteur,
- Et la classe de l'objet (par exemple : personne, chat, voiture...).

Ce découpage ***n'est pas visuel***, l'image n'est pas "coupée" en morceaux, c'est une structure mathématique sur laquelle les prédictions sont organisées. Plusieurs cellules peuvent repérer le même objet et proposer ***des boîtes englobantes similaires***.

III.7.1.2 Réseau neuronal unique et prédictions

Une fois la grille définie, YOLO fait passer l'image à travers un CNN. Ce réseau extrait des caractéristiques visuelles (contours, formes, textures...) puis utilise des couches entièrement connectées pour générer les prédictions.

Pour chaque cellule de la grille, le réseau renvoie un vecteur de sortie comprenant :

- Les coordonnées de l'objet détecté (bounding box),
- La probabilité qu'un objet soit réellement présent (score de confiance),
- Et les probabilités associées à chaque classe possible.

Ainsi, YOLO donne en sortie un ensemble de boîtes localisant les objets dans l'image, accompagnées de scores de confiance et d'étiquettes de classe.

III.7.1.3 Post-traitement (NMS –Non-Maximum Suppression)

Comme plusieurs cellules peuvent détecter le même objet, le modèle génère souvent plusieurs boîtes très proches et se recouvrant fortement. Pour éviter les doublons, YOLO applique une méthode appelée ***NMS (Non-Maximum Suppression).***

Cette méthode : Garde uniquement la boîte ayant le score de confiance le plus élevé, Et supprime les autres boîtes similaires, si elles se chevauchent au-delà d'un certain seuil. Grâce à cette étape, le résultat final est plus propre, clair et précis, avec une seule boîte bien positionnée par objet détecté.

III.7.2 Caractéristiques YOLO

Nous résumons les caractéristiques de YOLO [2] dans le tableau suivant:

Caractéristique	Description
Vitesse	YOLO peut traiter **30 à 60 images par seconde**, ce qui permet une **détection en temps réel**, même sur des vidéos.
Précision	Il détecte plusieurs objets dans une même image avec un bon **compromis entre précision et rapidité**.
Simplicité	Le réseau est **entièrement intégré**, donc plus facile à entraîner, déployer et adapter.
Polyvalence	YOLO fonctionne sur de nombreux types d'images : surveillance, route, médecine, industrie, etc.

TABLE III.3 – Les caractéristiques de YOLO.

III.7.3 Versions de Yolo

[H] La famille YOLO compte à ce jour douze versions majeures, depuis la première introduite en 2016 par Joseph Redmon avec YOLOv1, jusqu' à la dernière version YOLOv12 publiée en 2025. Les versions sont indiquées par les suffixes n (nano), s (small), m (medium), l (large), et x (extra-large)[2]. Ces tailles correspondent à différents niveaux de complexité, de vitesse et de précision. Par exemple, les modèles ń n ż et ń s ż sont conçus pour des applications nécessitant une très faible latence et une faible consommation de ressources matérielles, idéaux pour les appareils portables ou embarqués, mais au prix d'une précision moindre. Les modèles ń m ż offrent un compromis équilibré, adaptés à un large éventail d'applications médicales générales. Enfin, les modèles ń l ż et ń x ż sont les plus précis mais nécessitent plus de ressources matérielles et sont destinés aux environnements spécialisés, tels que les centres hospitaliers, les laboratoires médicaux nécessitant une très haute précision en diagnostic automatisé.[20, 2]

Ce graphe donne un aperçu des variantes du modèle YOLO11, présentant leur applicabilité dans des tâches spécifiques et la compatibilité avec des modes opérationnels tels que l'inférence, la validation, la formation et l'exportation. Cette flexibilité rend YOLO11 adapté à un large éventail d'applications dans la vision par ordinateur, de la détection en temps réel aux tâches de segmentation complexes.[2]

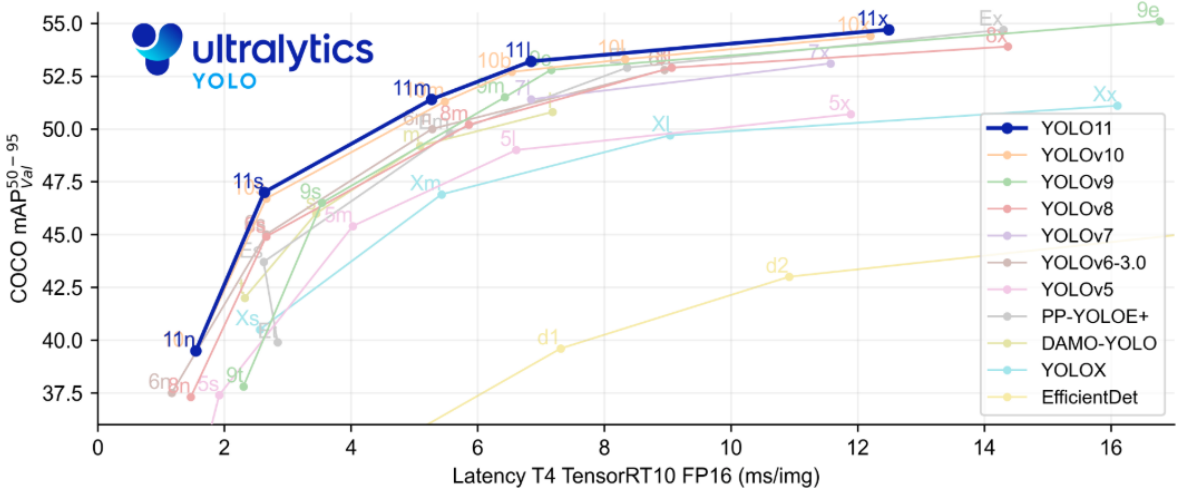


FIGURE III.8 – Comparaison des performances des différents modèles de détection d'objets YOLO en termes de mAP (Mean Average Precision) et de latence.
[2]

Par exemple, YOLO11 peut être entraîné à détecter des anomalies potentielles dans les IRM, telles que des tumeurs, des anomalies vasculaires ou une croissance irrégulière des tissus.
En mettant en évidence les zones préoccupantes, il permet aux radiologues de donner la priorité aux cas nécessitant une attention immédiate [2].

YOLO12, grâce à son architecture centrée sur l'attention, améliore la détection d'objets en temps réel, comme illustré ici. Les résultats montrent une précision accrue dans l'identification des anomalies potentielles, tout en maintenant une vitesse d'inférence es-

sentielle pour des applications critiques telles que l'analyse d'IRM ou la surveillance médicale[2].

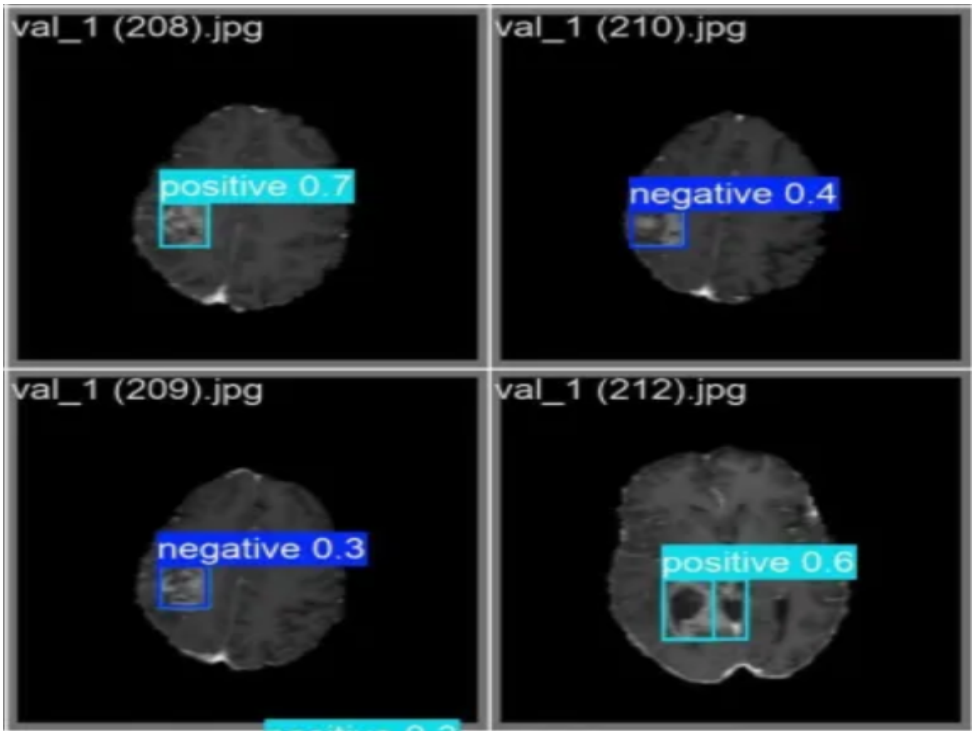


FIGURE III.9 – YOLO11 identifie les anomalies dans les IRM cérébrales [2].

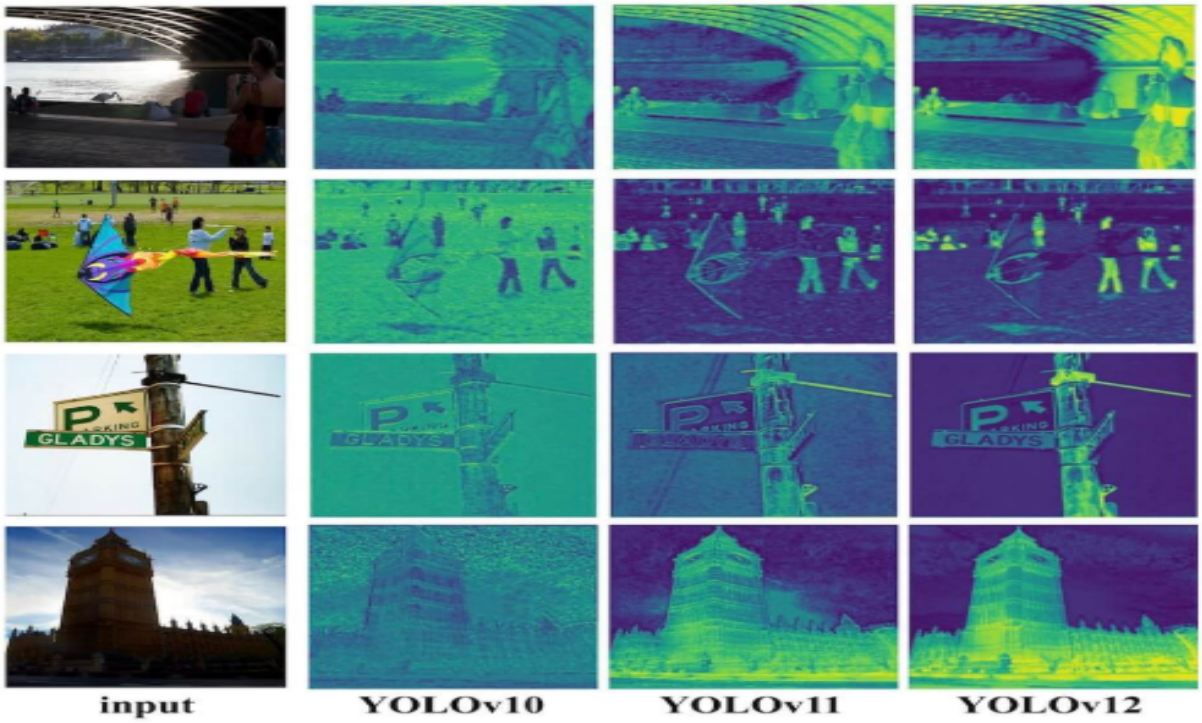


FIGURE III.10 – Comparaison des cartes thermiques entre YOLOv10, YOLOv11 et le YOLOv12 proposé.

[2]

Le tableau récapitulatif suivant représente toutes les versions de YOLO :

TABLE III.4 – Comparaison entre les différentes versions des algorithmes YOLO. [2]

Version	Année	Description	Valeur ajoutée	Limites
YOLOv1	2016	Première version de YOLO, détection en une seule passe (single-shot).	Vitesse de traitement élevée	Faible précision, problèmes avec petits objets
YOLOv2	2017	Amélioration avec batch normalisation, ancrage des boîtes, et détection multi-échelle.	Meilleure précision et vitesse	Plus complexe
YOLOv3	2018	Ajout de plusieurs améliorations : prédiction multi-échelles, détection fine.	Bonne précision sur différents types d'objets	Lourd en calculs, plus lent
YOLOv4	2020	Optimisation du réseau avec CSPDarknet [1], PANet [2], etc.	Très bonne précision / vitesse selon la configuration	Configuration très technique
YOLOv5	2020	Développé par Ultralytics [3], pas officiellement "YOLO", mais adopté largement.	Facile à utiliser, documentation claire, nombreux pré-entraînements	Moins personnalisable
YOLOv6	2021	Développé par Meituan, optimisé pour déploiement industriel.	Bonne performance sur GPU NVIDIA, pipeline simplifiée	Moins populaire, communauté moindre
YOLOv7	2022	Propose l'expansion dynamique du modèle (model scaling) et réutilisation des données.	Excellente précision sans sacrifier la vitesse	Complexité accrue, difficile à configurer

Suite à la page suivante

1. **CSPDarknet :** Architecture de réseau qui améliore l' efficacité du calcul et la qualité des caractéristiques extraites.

2. **PANet :** Méthode de fusion des caractéristiques à différentes échelles pour améliorer la détection, surtout des petits objets.

3. **Ultralytics** est une entreprise spécialisée en intelligence artificielle et vision par ordinateur, connue principalement pour le développement et la maintenance des versions récentes du modèle YOLO.

Table III.4 – suite de la page précédente

Version	Année	Description	Valeur ajoutée	Limites
YOLOv8	2023	Version native Ultralytics, intégration segmentation et suivi objets.	API[4] intuitive. Bon équilibre vitesse/précision	Personnalisation limitée. Précision modérée sur objets très petits
YOLOv9	2024	Améliorations de précision, gestion des données difficiles (petits objets, faible contraste).	Stabilité améliorée. Détection améliorée sur petits objets	Modèle plus lourd Entraînement plus exigeant
YOLOv10	2024	Multitâche unifié : détection, segmentation, suivi en une seule sortie. Simplification du pipeline[5] multitâche	Bonne performance globale Documentation limitée	Nécessite GPU puissant pour l'entraînement
YOLOv11	2024	Optimisation extrême pour dispositifs embarqués et mobiles avec intégration ONNX[6]/OpenVINO[7].	Latence très faible Idéal pour applications temps réel embarquées	Configuration initiale plus complexe
YOLOv12	2025	Spécialisé pour l' imagerie médicale avancée (3D IRM/CT) et utilisation clinique.	Précision maximale (particulièrement en imagerie médicale complexe) Adapté aux grands volumes de données	Très récent, documentation encore limitée

4. **API (Application Programming Interface) :** interface logicielle qui permet à différents systèmes ou programmes informatiques de communiquer entre eux et d'échanger des données ou des fonctionnalités.

5. **Pipeline :** en science des données ou en apprentissage automatique, un pipeline désigne une suite organisée d' étapes de traitement (comme la préparation des données, l'entraînement d' un modèle, puis la prédiction), automatisées et enchaînées pour exécuter un flux de travail complet.

6. **ONNX (Open Neural Network Exchange)** est un format standard permettant l'échange de modèles entre différents frameworks de Deep Learning.

7. **OpenVINO (Open Visual Inference et Neural Network Optimization)** est un outil logiciel développé par ntel permettant d' optimiser et accélérer l' exécution des modèles d' IA sur divers appareils à ressources limitées.

III.7.4 Architecture YOLO

L'algorithme YOLO repose sur une architecture unifiée qui permet de détecter plusieurs objets dans une image en une seule passe à travers un réseau de neurones convolutif [34]. Contrairement aux méthodes classiques à deux étapes (comme R-CNN), YOLO effectue la détection en temps réel, ce qui en fait une solution idéale pour les systèmes embarqués et les applications médicales sensibles au temps.[35]

Les figure III.11 illustrent d'abord l'architecture simplifiée du YOLOv11. [34]

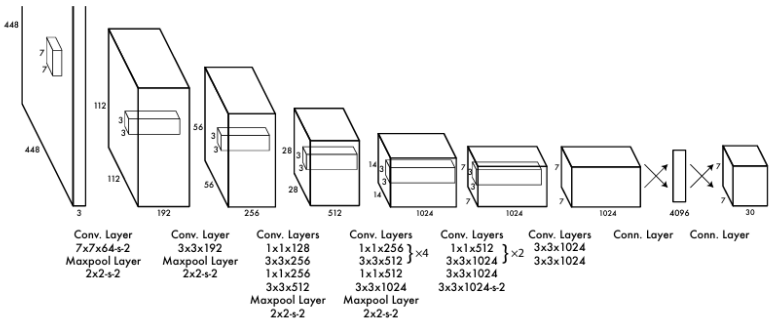


FIGURE III.11 – Architecture du modèle YOLO dans sa 1ers version [35]

Les figure III.12 illustrent d'abord l'architecture simplifiée du YOLOv11. [35]

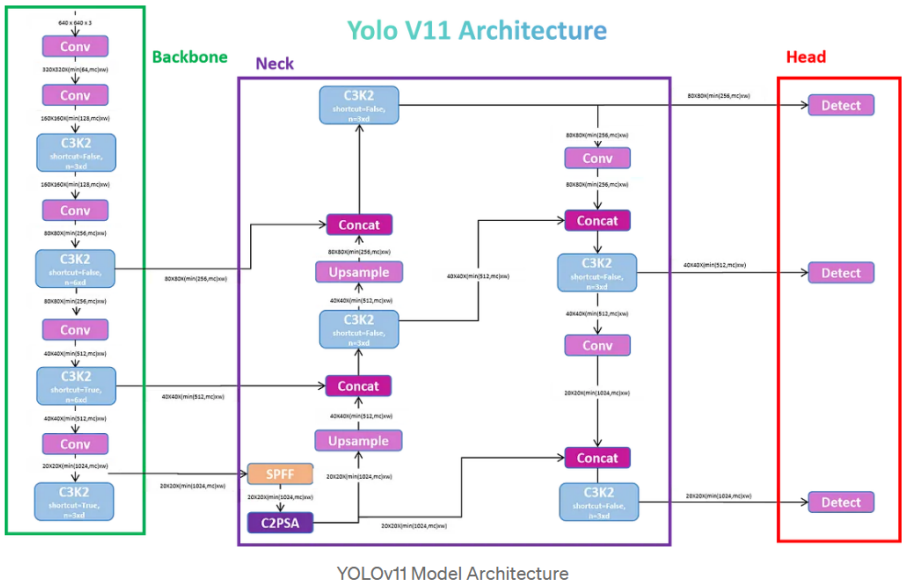


FIGURE III.12 – Architecture du modèle YOLO version 11 [35]

Son fonctionnement repose sur une succession de modules spécialisés :

- un **Backbone** chargé d'extraire les caractéristiques visuelles de l'image,

- un **Neck** permettant de fusionner ces caractéristiques à différentes échelles,
- un **Head** produisant les prédictions finales (classes, coordonnées, scores).

La Figure III.13 illustre le processus global de fonctionnement de YOLO, depuis l'entrée de l'image jusqu'à la sortie finale avec les objets détectés.

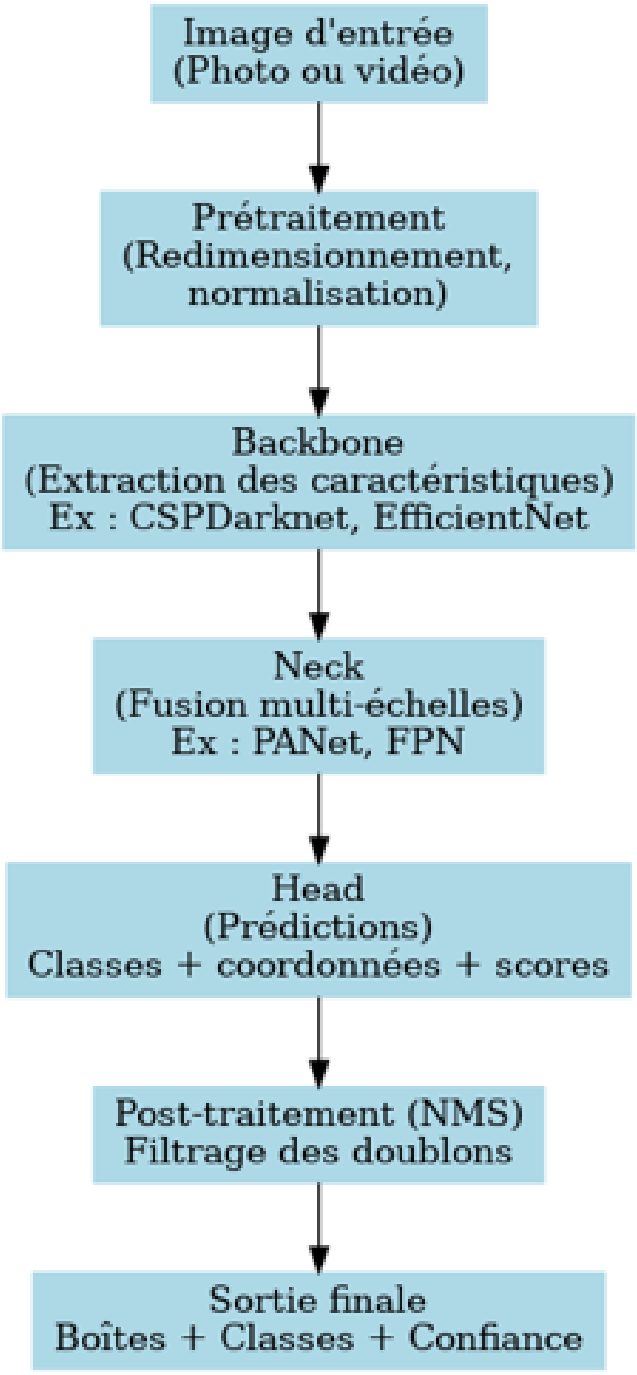


FIGURE III.13 – Processus global de fonctionnement de YOLO (pipeline)

Le tableau suivant détaille chaque étape de ce processus :

Étape	Rôle	Exemple ou image mentale
Image d'entrée	C'est la photo ou la vidéo que l'on donne au modèle.	Une image de radiographie, de langue ou de caméra de sécurité.
Prétraitement	On prépare l'image pour qu'elle soit adaptée au réseau : redimensionnement (ex : 640×640), normalisation des couleurs.	Comme recadrer et nettoyer une photo avant de l'imprimer.
Extraction (Backbone)	C'est la première grande étape du réseau : elle **repère les formes, les couleurs, les textures**. C'est ici que YOLO "regarde" ce qui se passe dans l'image.	Comme un il artificiel qui observe l'image et en extrait ce qui semble important.
Fusion (Neck)	Cette étape **combine les informations** détectées à plusieurs niveaux (petits et grands objets).	Imagine que tu regardes une image de loin, puis de près, et que tu combines les deux points de vue.
Prédiction (Head)	YOLO décide **où sont les objets**, ce **qu'ils sont**, et **à quel point il en est sûr**.	Il dit : « à cet endroit, il y a 92 % de chances que ce soit un chat. » Cela veut dire, **dans cette cellule de la grille**, le modèle a détecté un objet avec une grande confiance.
Résultat final (Sortie)	YOLO affiche les objets détectés sous forme de **boîtes + étiquettes + pourcentage de confiance**.	Tu vois un rectangle autour d'un objet, avec écrit "chat (92 %)" ou "vélo (85 %)".

TABLE III.5 – *Explication de l'architecture de YOLO.*

Grâce à sa conception modulaire reposant sur un *Backbone* (extraction de caractéristiques), un *Neck* (fusion multi-échelles) et un *Head* (prédictions finales), YOLO s'impose comme une architecture de détection rapide, précise et efficace. Elle est largement adoptée dans divers domaines, notamment la médecine, où elle facilite la détection automatique d'anomalies, de tumeurs ou de signes cliniques discrets, difficiles à repérer par des méthodes conventionnelles.

III.8 Conclusion

Ce chapitre présente les fondements théoriques indispensables pour comprendre l'évolution de la vision artificielle, depuis les opérations de bas niveau jusqu'aux modèles avancés de détection en temps réel. En combinant des techniques classiques de traitement d'images à des réseaux neuronaux profonds comme YOLO, l'intelligence artificielle ouvre de nouvelles perspectives dans l'analyse automatique d'images complexes, notamment dans le domaine médical.

Les différentes versions de YOLO, et leurs performances, montrent à quel point cette famille de modèles est devenue incontournable dans les applications nécessitant un traitement rapide et fiable, telles que le diagnostic assisté par image.

Cette base théorique constitue une étape essentielle avant d'aborder, dans le chapitre suivant, l'application concrète de ces technologies à une problématique médicale spécifique : le diagnostic du cancer du sein à partir de l'analyse de la langue selon les principes de la Médecine Traditionnelle Chinoise.

Chapitre IV

État de l'art sur le diagnostic automatisé du cancer du sein dans la MTC

Sommaire

IV.1 Introduction **74**
IV.2 Diagnostic du cancer de sein en médecine conventionnelle **74**
IV.3 Diagnostic de la langue dans la MTC **77**
IV.4 Diagnostic du cancer du sein dans la vision MTC **90**
IV.5 Systèmes d'aides au diagnostic (SAD) développés pour la MTC . . **101**
IV.6 conclusion **104**

Résumé : *Ce chapitre dévoile le potentiel révolutionnaire du diagnostic précoce du cancer du sein par l'analyse de la langue en MTC, une pathologie touchant une femme sur huit dans le monde. Malgré des performances IA impressionnantes atteignant 87% de précision sur d'autres pathologies, seulement quatre études mondiales ciblent spécifiquement le cancer du sein, révélant un vide scientifique critique. L'analyse de ces rares travaux identifie des caractéristiques linguales distinctives entre stades précoces et avancés, établissant les bases théoriques d'un système automatisé capable de détecter les déséquilibres énergétiques précurseurs de cette maladie silencieuse.*

IV.1 Introduction

Le cancer du sein représente la forme de cancer la plus fréquente chez les femmes à travers le monde, et l'Algérie n'échappe pas à cette réalité. Comme le montre la figure IV.1, l'incidence du cancer du sein dépasse largement celle des autres types de cancers féminins en Algérie, soulignant l'urgence d'une prise en charge adaptée et efficace.[4]

Malgré les avancées médicales en matière de dépistage et de traitement, une part significative des patientes est encore diagnostiquée à un stade avancé, limitant ainsi les options thérapeutiques et affectant les taux de survie. Cette situation met en évidence la nécessité de renforcer les stratégies de prévention et de sensibilisation, tout en améliorant l'accès à des méthodes de diagnostic précoce et à des traitements innovants.

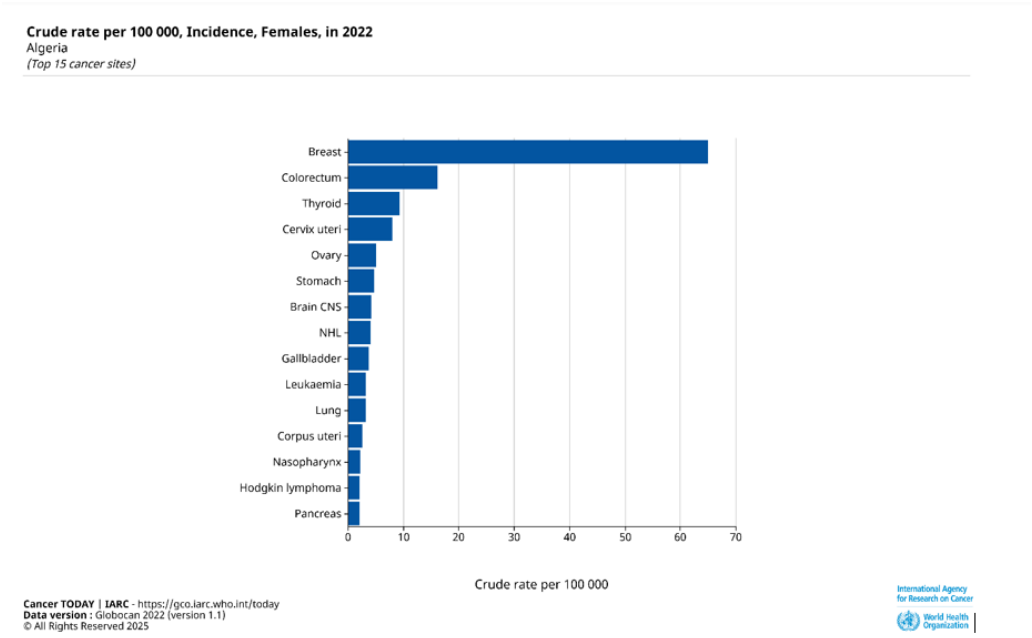


FIGURE IV.1 – Statistiques du cancer du sein en Algérie.
[4]

La MTC, quant à elle, propose une approche préventive et holistique du diagnostic, notamment à travers l'examen de la langue. Ce chapitre présente une synthèse des travaux existants sur le diagnostic du cancer du sein, en croisant les perspectives de la médecine conventionnelle, de la MTC et des nouvelles technologies d'intelligence artificielle.

IV.2 Diagnostic du cancer de sein en médecine conventionnelle

IV.2.1 Présentation du cancer du sein

Le cancer du sein est une maladie due à une multiplication anormale des cellules mammaires, formant une tumeur qui peut être ***bénigne*** **non cancéreuse,** peut être gênante mais reste localisée, ou une tumeur ***maligne*** qui signifie une tumeur **cancéreuse,**

agressive et dangereuse car elle peut envahir l'organisme (on parle des *métastases*). Il s'agit du cancer le plus fréquent chez la femme, touchant environ une femme sur huit (12,5% de femmes) [36].

Il se développe généralement dans les canaux galactophores (carcinome canalaire) ou dans les lobules (carcinome lobulaire). Le cancer peut rester localisé au sein ou se propager à d'autres parties du corps sous forme de métastases [37].
La figure IV.2 ci-dessous montre la structure interne d' un sein en mettant en valeur la partie où le cancer peut se développer.

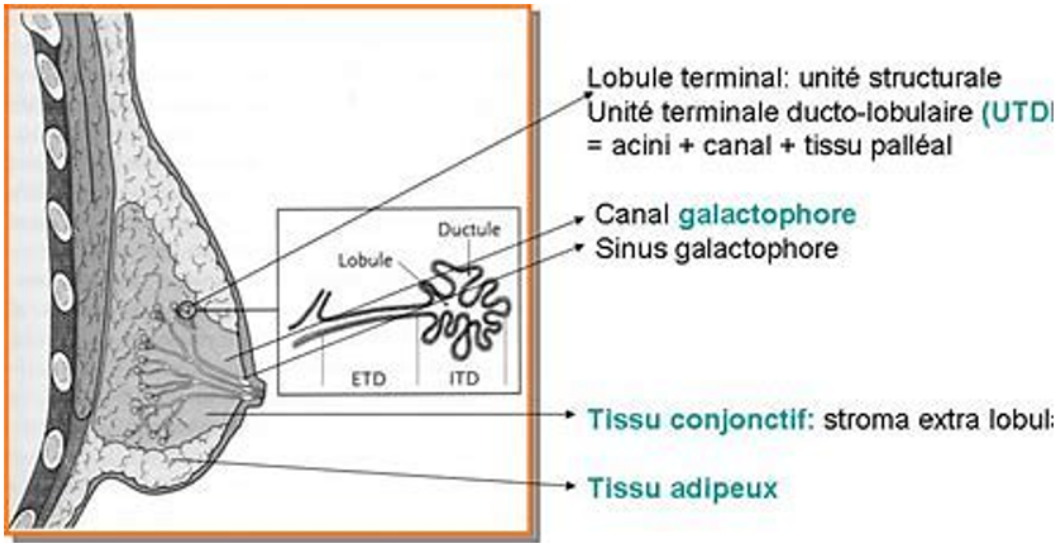


FIGURE IV.2 – Schéma de la structure interne d'un sein
[37]

Les causes sont variées : facteurs génétiques, déséquilibres hormonaux, et facteurs liés au mode de vie (obésité, tabac, alimentation, etc).

Le cancer du sein est classé en plusieurs stades, de 0 à 4, selon la taille de la tumeur. La figure IV.3 suivante illustre les différents stades du cancer du sein. Au stade métastatique, la maladie atteint d'autres organes, rendant la guérison beaucoup plus difficile.

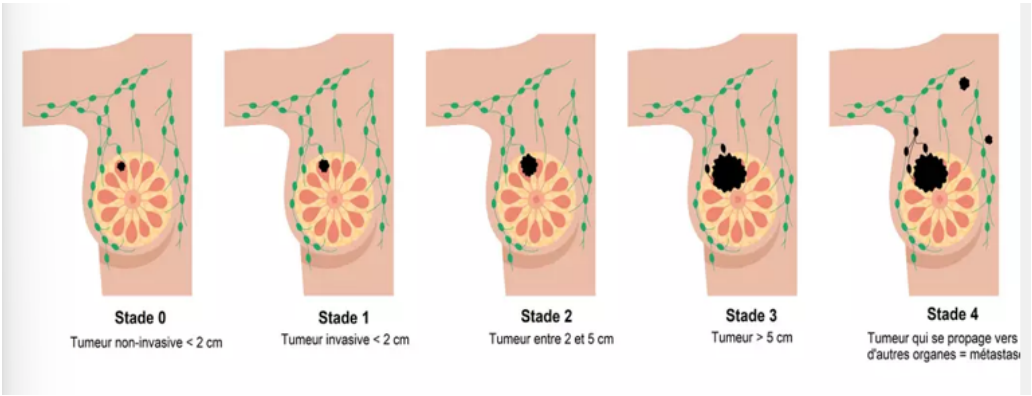


FIGURE IV.3 – Illustration des différents stades du cancer du sein.
[38]

IV.2.2 Les méthodes actuelles de dépistage

Le dépistage du cancer du sein repose sur plusieurs techniques complémentaires, visant à identifier des anomalies à un stade précoce.

Mammographie : Elle constitue la méthode de référence pour le dépistage organisé. Il s'agit d'une radiographie des seins, recommandée tous les deux ans chez les femmes de 50 à 74 ans. Elle permet de détecter des microcalcifications ou des masses suspectes. Cependant, sa sensibilité est limitée chez les femmes ayant une forte densité mammaire, ce qui peut entraîner des faux négatifs [39].

Échographie : Utilisée en complément de la mammographie, elle permet une meilleure visualisation des tissus chez les femmes jeunes ou à seins denses. Elle est également utile pour caractériser les masses détectées. Toutefois, elle est opérateur-dépendante et moins efficace en dépistage de masse [39].

Imagerie par Résonance Magnétique (IRM) : Elle est réservée aux patientes à haut risque (antécédents familiaux, mutation BRCA1/2). L'IRM est très sensible, mais elle génère un nombre important de faux positifs, ce qui peut conduire à des examens supplémentaires inutiles [39].

Biopsie : Elle n'est pas une technique de dépistage mais de confirmation. Réalisée en cas de suspicion, elle permet une analyse histologique du tissu prélevé. C'est l'examen de référence pour confirmer un diagnostic de cancer[39].

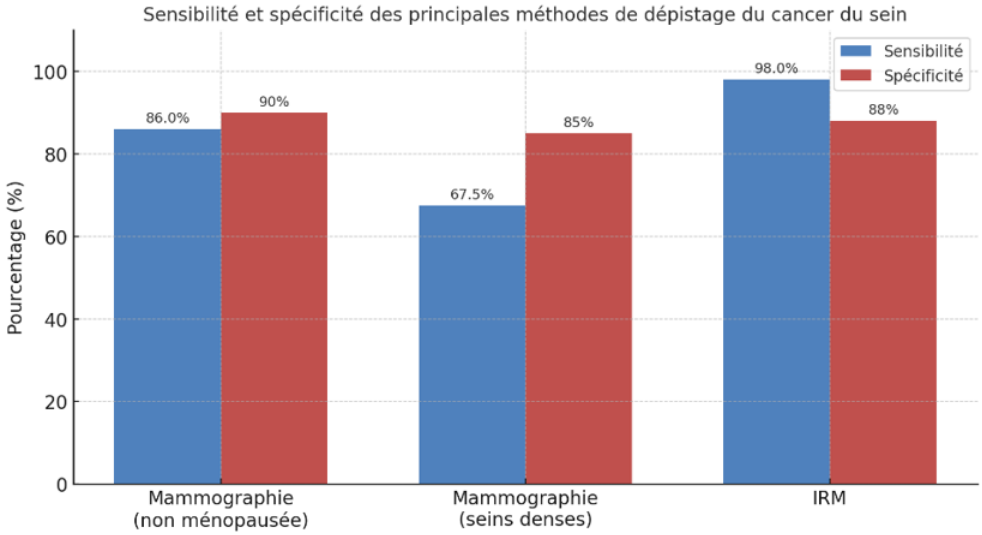


FIGURE IV.4 – Sensibilité et spécificité des principales méthodes de dépistage du cancer du sein

Ce graphique montre que l' IRM est la méthode la plus performante pour détecter un cancer, avec un taux de sensibilité de 98 %, ce qui signifie qu' elle identifie presque tous les cas existants. En revanche, sa spécificité est légèrement inférieure, à 88 %, ce qui indique qu' elle peut parfois signaler un cancer alors qu' il n' y en a pas. La mammographie, de son côté, reste efficace chez les femmes non ménopausées, c' est-à-dire celles qui ont encore leurs règles et sont généralement plus jeunes, avec une sensibilité de 86 %.

Toutefois, chez les femmes ayant des seins denses, caractérisés par une forte concentra-

tion de tissu mammaire compact visible à l' imagerie, cette sensibilité diminue nettement pour atteindre 67,5 %. Face à ces limites, il devient pertinent de recourir à d' autres méthodes, comme l' intelligence artificielle, capable de détecter des anomalies très discrètes que l' il humain ne perçoit pas, et ainsi de réduire les risques d' erreurs dans le diagnostic.

IV.2.3 Les méthodes actuelles de dépistage combinée avec l'IA

L' IA transforme progressivement le dépistage du cancer du sein en renforçant la précision et la rapidité des diagnostics. Des modèles hybrides comme CNN, sont largement utilisés pour analyser les mammographies et classer les images en fonction de leur caractère bénin ou malin. Ces réseaux permettent d' extraire automatiquement des caractéristiques visuelles pertinentes, souvent indétectables à l' il nu.

Leur performance s' est montrée remarquable sur plusieurs bases cliniques, avec des taux de précision dépassant parfois 90 %, notamment grâce à des architectures comme ResNet ou EfficientNet. Des modèles plus récents, comme YOLO (You Only Look Once), sont utilisés pour localiser en temps réel les anomalies suspectes dans les images médicales. Leur rapidité d' exécution permet une détection efficace des tumeurs avec un temps de traitement minimal, ce qui en fait un outil prometteur dans les environnements cliniques à forte demande.

Malgré ces avancées, l' intégration clinique de ces systèmes nécessite encore des validations rigoureuses, des bases de données plus représentatives, ainsi qu' une meilleure transparence des algorithmes pour garantir la confiance des professionnels de santé.

IV.3 Diagnostic de la langue dans la MTC

Les modifications observées sur l' aspect de la langue peuvent refléter de manière objective l' état de santé global du corps, signalant aussi bien sa vitalité que son affaiblissement. Elles peuvent également révéler la gravité d' une maladie ou la présence d' un déséquilibre pathologique.

Des recherches ont démontré que les patients atteints de cancers présentent davantage d' altérations anormales de la langue que les personnes en bonne santé ou atteintes d' autres types de pathologies [40].

Le *diagnostic de la langue* est une méthode fondamentale en MTC permettant d'évaluer l'état de santé global d'un patient. La figure IV.5 montre les différentes observations des praticiens maîtres en MTC sur la langue pour établir un diagnostic par la langue.

En MTC, le diagnostic de la langue ne détecte pas les *maladies spécifiques au sens occidental,* mais identifie des *déséquilibres énergétiques* pouvant être associés à des troubles.

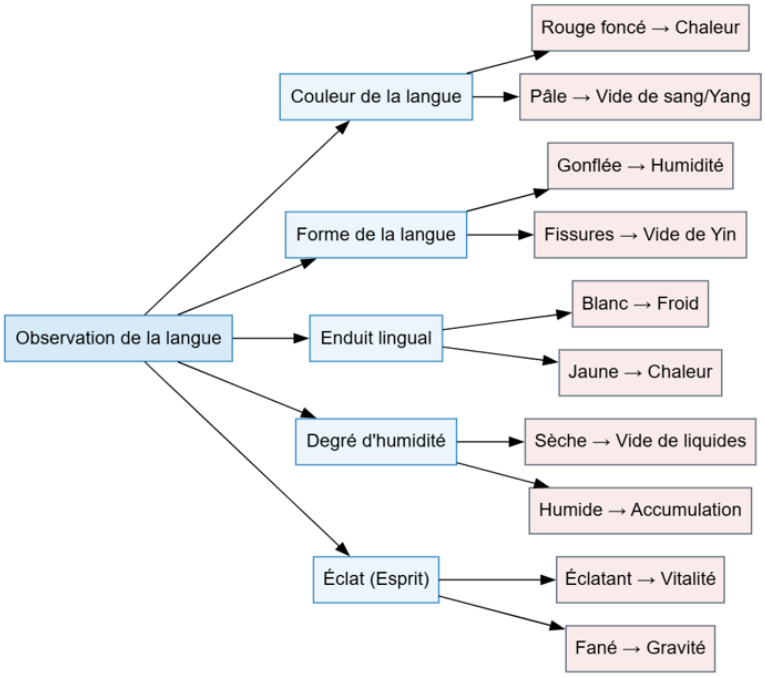


FIGURE IV.5 – Les différents aspects observés sur la langue. [5]

Le tableau suivant présente l' interprétation de quelque caractéristiques visuelles :

Caractéristique de la langue	**Interprétation en MTC**	**Maladies / troubles associés**
Langue pâle	Vide de sang ou de Qi	Anémie, fatigue chronique, hypotension
Langue rouge foncé	Chaleur excessive ou stase de sang	Hypertension, troubles inflammatoires, stress intense
Langue violette	Stagnation de sang / mauvaise circulation	Troubles cardiovasculaires, douleurs chroniques
Langue gonflée avec marques de dents	Vide de Qi ou humidité de la Rate	Hypothyroïdie, rétention d'eau, digestion lente
Langue fine et sèche	Vide de Yin ou déshydratation	Diabète, syndrome sec, anxiété
Langue avec fissures	Vide Yin, chaleur interne ou constitution faible	Bouffées de chaleur, sécheresse, cancer, troubles hormonaux
Langue tremblante	Vide de Qi ou atteinte du système nerveux	Fatigue extrême, anxiété, maladie de Parkinson
Enduit épais et jaune	Chaleur interne ou humidité-chaleur	Infections, troubles digestifs, inflammation
Enduit fin blanc	Normal (si modéré) ou froid interne (si excessif)	Refroidissement, troubles gastro-intestinaux bénins

TABLE IV.1 – Maladies et troubles associés à des caractéristiques de la langue en MTC [5].

IV.3.1 Travaux de recherche pour le diagnostic automatisé par la langue en MTC

Dans la section suivante on va présenter les études les plus récente de qui automatise le diagnostic automatisé par la langue en MTC avec les modèles les plus puissants de l'IA. Dans ce qui suit 3 études qui traite la langue comme indicateur de Santé, et 2 autres études qui analyse la langue pour détecter les marques de dents dans la langue.

IV.3.1.1 Analyse de langue à distance, automatisé et en temps réel avec Multi-Label

Les auteurs **Yuwei Wang et al.** publient en 2025 l'article intitulé *"Dr. Tongue: Sign-Oriented Multi-Label Tongue Image Analysis for Remote TCM Diagnosis"* [41]. Ils proposent un système automatisé **'Dr. Tongue'**, conçu pour analyser des images de langue à distance dans le cadre de la MTC. L'objectif est de détecter plusieurs signes cliniques à partir d'une seule image, en s'appuyant sur une approche **multi-étiquettes orientée signes** (une instance peut avoir plusieurs étiquettes simultanément, pas seulement une seule).

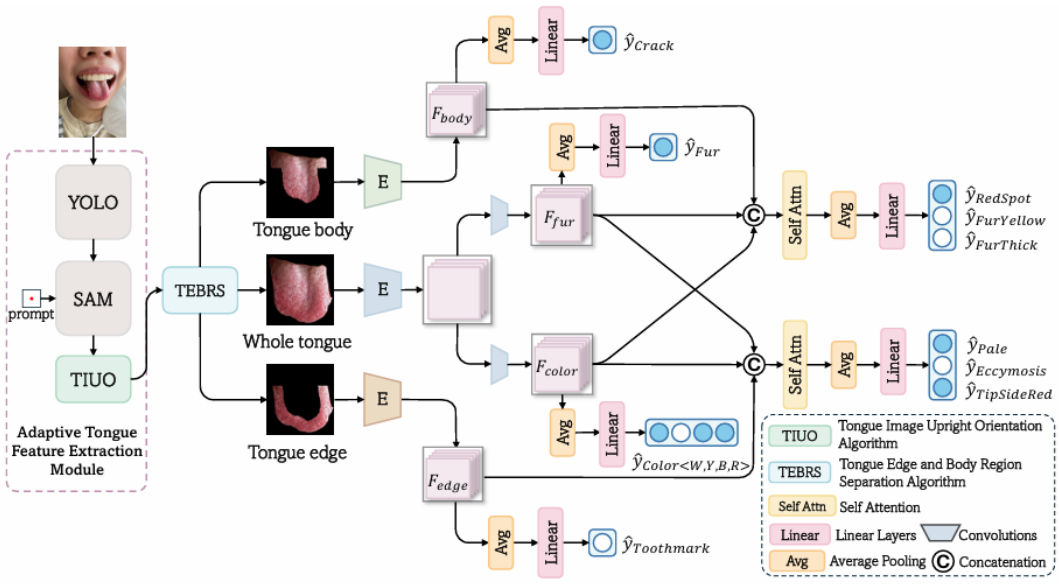


FIGURE IV.6 – Architecture du système Dr. Tongue.
[41]

Le système repose sur un module d'extraction adaptative des caractéristiques, nommé **ATFE**, qui combine trois algorithmes puissants : **YOLOv5**, pour la détection de la langue, SAM, pour la segmentation automatique, et **TIUO** (Tongue Image Upright Orientation Algorithm) pour l'unification des images en position verticale. Une fois l'image segmentée, le module **TEBRS** (Tongue Edge and Body Region Separation) divise la langue en deux zones distinctes : le corps et les **bords**. Chacune est ensuite analysée par

un réseau profond appelé **SignNet** (sign-oriented network) vise à reconnaître simultanément plusieurs attributs linguaux visuels, comme la couleur de la langue, les fissures, la nature de l' enduit ou d' autres signes subtils.

Le modèle atteint une précision **86.68±0.30%** et une F1-score de **72.87±0.63%**, avec des scores supérieurs à 70% sur la majorité des caractéristiques. Cette approche montre un fort potentiel pour le dépistage automatisé et en temps réel des troubles de santé.

IV.3.1.2 Une évaluation holistique et personnalisée de l'état de santé du patient

Les auteurs Lanyu Jia et al. publient en 2024, l' article intitulé *"Modernizing Tongue Diagnosis : AI Integration with Traditional Chinese Medicine for Precise Health Evaluation"* [42]. Ils y présentent un système intelligent de diagnostic de la langue fondé sur les principes de la MTC, marquant une avancée majeure vers l' intégration pratique de la MTC avec l' intelligence artificielle. On y trouve une méthode rigoureuse, objective et modernisée pour le diagnostic clinique pour une évaluation holistique et personnalisée de l' état de santé du patient. Leur architecture, baptisée **Cv Swin Transformer**, combine deux approches puissantes: d' un côté, le Convolutional Vision Transformer (CVT), performant pour détecter des motifs locaux; de l' autre, le Shifted Window Transformer (Swin), qui segmente les images en fenêtres décalées afin de capter les corrélations globales. On bénéficie ainsi d' un traitement à la fois local et global des images, organisé selon une structure hiérarchique.

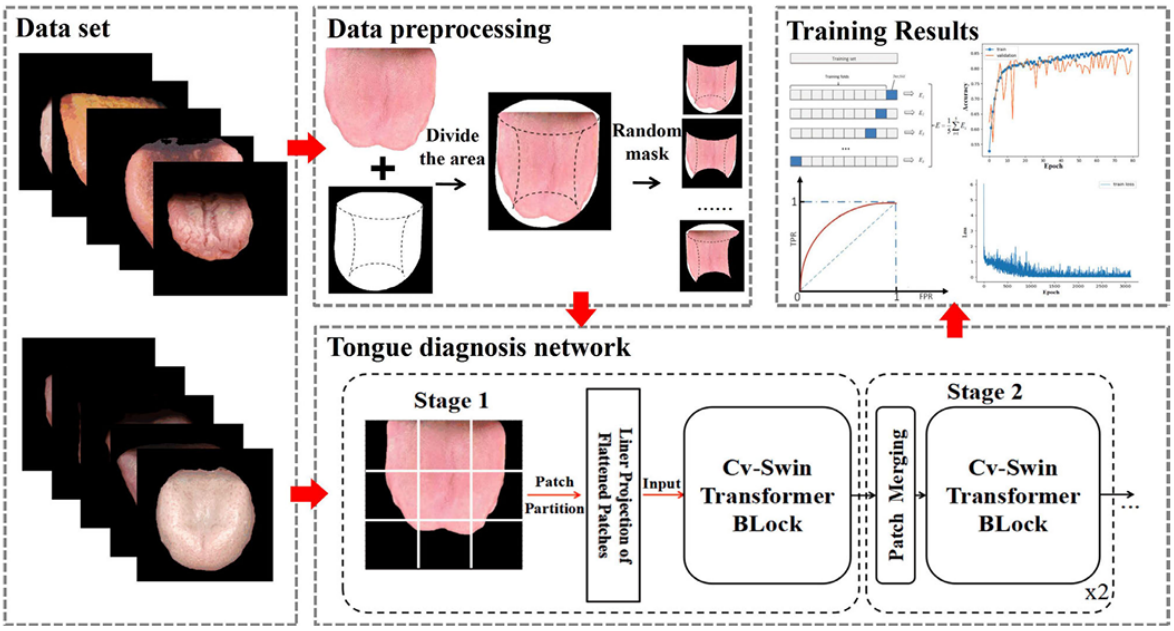


FIGURE IV.7 – Diagramme de l'architecture du modèle Cv-Swin. [42]

Le modèle a été entraîné sur 5365 images classées en dix catégories cliniques standardisées. Il atteint une Accuracy de 87,37%, ce qui surpasse plusieurs modèles de référence. Pour affiner le diagnostic, on utilise une segmentation de la langue en racine, pointe, côtés

et centre, associée à l'extraction de caractéristiques visuelles et à l'analyse de données démographiques telles que l'âge, le sexe ou les habitudes de vie.

IV.3.1.3 Détection des anomalies de santé avec langue via une application mobile iTonguet

Les auteurs **J. Xie et al.** publient en 2022 l'article intitulé *"Digital Tongue Image Analyses for Health Assessment"* [43]. Ils explorent l'analyse automatisée d'images de langue pour évaluer l'état de santé général et détecter précocement certaines maladies. L'approche repose sur une application mobile baptisée **iTongue**, qui permet de collecter des images et de les traiter à l'aide de modèles d'apprentissage profond. Les chercheurs ont fine-tuné deux modèles de référence, **ResNet50** et **VGG19**, sur un corpus de 4 787 images. Parmi celles-ci, 2 215 collectée par un professionnel avec l'appareil **TDFA-1** pour la capture et l'analyse de l'image de la langue présenté dans la figure 3-8, tandis que 2 572 ont été directement collectées via l'application. Lors de l'évaluation, ResNet50 atteint un Accuracy de **64,52%**, légèrement supérieure à celle de **VGG19**, qui s'élève à **62,86%**. Aucune architecture détaillée était présentée dans cet article, mais on trouve le pipeline de fonctionnement de l'application mobile iTong IV.19



FIGURE IV.8 – Instrument de diagnostic de la langue TFDA-1 [6]

Des liens significatifs sont observés entre certaines caractéristiques de la langue et des pathologies comme le diabète, le cancer ou les troubles digestifs. Cependant, même si l'idée est prometteuse, l'étude présente plusieurs limites. On ne dispose pas d'informations précises sur l'architecture complète du système, et l'ensemble souffre d'un manque de standardisation méthodologique. De plus, comme la majorité des participants sont issus de la population chinoise, cela limite la généralisation des résultats à d'autres contextes culturels ou médicaux.

IV.3.1.4 Diagnostic des marques de dents sur la langue avec l'IA

Les auteurs S.Sun et al. publient en 2023 une étude intitulée *"Ammonia-Net: A Multi-task Joint Learning Model for Multi-class Segmentation and Classification in Tooth-marked Tongue Diagnosis"* [44]. Dans ce travail, on cherche à diagnostiquer les marques de dents sur la langue selon quatre niveaux de sévérité : normal (HT), léger (LT), modéré

(MT) et sévère (ST). Pour cela, les chercheurs proposent un modèle multi-tâches baptisé **Ammonia-Net**, qui associe une architecture **U-Net** pour la segmentation des images à un classificateur **ShuffleNetV2** pour la reconnaissance des niveaux. Ce dernier, basé sur un réseau de neurones convolutif léger, permet un bon compromis entre vitesse de traitement et précision, grâce à une structure optimisée combinant mélange de canaux et goulot d' étranglement.
Le modèle est testé sur le jeu de données **TTD (Tooth-marked Tongue Diagnosis)**, publié en 2020, qui contient 856 images réparties en quatre catégories : 291 en HT, 133 en LT, 233 en MT et 199 en ST. Lors de l' évaluation, on atteint une précision remarquable de **99,06 %** en classification binaire et **80,02 %** en classification à quatre classes. Mais malgré ces performances, certaines limites apparaissent. On observe notamment des confusions fréquentes entre des niveaux visuellement proches, comme entre LT et MT, ou entre LT et ST. On remarque aussi une forte dépendance du modèle aux annotations fournies par des experts, ce qui freine sa capacité à être généralisé à d' autres contextes cliniques.
On présente l' architecture simplifiée de ce modèle dans la figure suivante : Tels que le module PLFE (Pixel-Level Feature Extractor) est un réseaux U-net pour extraire les marques de dents, le module BCFLM (Binary Classification Feature Learning Module) est un ShufflenetV2 pour la classification binaire, et le module QCFLM (Quadruple classification feature learning module) pour la classification des 4 classes. Ainsi que le module CADCM (Class-aware diagnostic correction module), qui combine les résultats de BCFLM et QCFLM pour approuver les résultats de la classification.

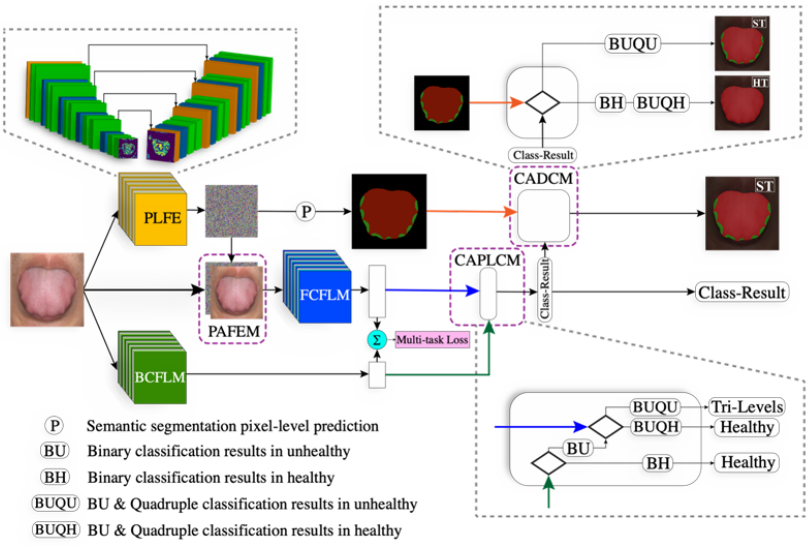


FIGURE IV.9 – Architecture de Ammonia-net
[44]

IV.3.1.5 Reconnaissance automatique des marques dentaires sur la langue par détection faiblement supervisée

Les auteurs **Y. Zhang et al**. publient en 2024 l' article intitulé *"Weakly Supervised Object Detection for Automatic Tooth-marked Tongue Recognition"* [45]. Ce travail cherche à automatiser la détection des marques par les dents sur la langue, en s' appuyant sur une méthode de **détection faiblement supervisée** (données étiquetées de manière peu

fiable). L' approche repose sur un modèle appelé **WSVM (Weakly Supervised** object detection method based on Vision transformer (ViT) and **Multiple** instance learning (MIL), qui combine plusieurs techniques récentes : **YOLOv8n** pour détecter la langue sur l' image, **SAM** pour la segmentation, puis un double mécanisme pour identifier les régions présentant des marques dentaires, basé sur le **ViT** (Transformer des images, inspiré des Transformers de NLP) et un apprentissage par instances multiples **MIL** (une approche d' apprentissage supervisé qui utilise des ensembles d' instances pour la classification ou la détection, au lieu d'analyser chaque donnée séparément).

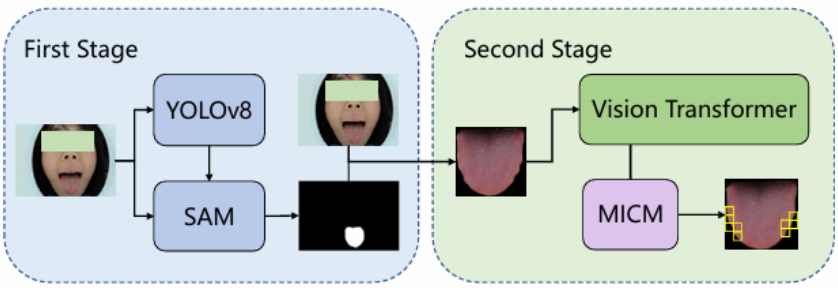


FIGURE IV.10 – Cadre global du modèle. L'approche comprend la première étape de l'extraction automatique de la langue et la deuxième étape de la langue de dents faiblement supervisée détection

[45]

Le modèle a été entraîné sur un total de 1 745 images, dont 255 montrent des langues cadrées, 556 présentent des marques de dents, et 1 188 sont considérées comme normales. Ces images proviennent à la fois d' un jeu de données privé et d' un jeu public disponible sur Kaggle. En évaluation, la performance du détecteur **YOLOv8n atteint une mAP de 100 %**, ce qui signifie qu' il détecte très efficacement la langue dans les images. Pour la reconnaissance des marques dentaires, le modèle atteint **84,2% de précision** sur le jeu privé et 89 % sur celui de Kaggle. La performance moyenne globale reste à **80,2 %**, ce qui montre que le modèle est encore perfectible. On présente l' architecture simplifiée de ce modèle dans la figure IV.10.

IV.3.2 Travaux de recherche pour le diagnostic par la langue pour des tumeurs et de maladies spécifiques en MTC

Ces dernières années, les tumeurs malignes sont devenues une maladie qui met gravement en danger la vie et la santé des personnes dans le monde entier.

IV.3.2.1 Analyse clinique des signes linguaux chez les patients atteints de cancer

L' auteur **Wu Junde** publie en 2011, à l' Université de Médecine Chinoise de Pékin, une étude intitulée *"Étude clinique des caractéristiques de l' image de la langue chez les patients atteints de cancer"* [40]. Dans ce travail, on s' intéresse à l' observation des images de langue chez des patients atteints de tumeurs malignes, dans le but d' identifier des

signes cliniques spécifiques selon la MTC. Cette étude a été répartie en deux études
La première étude clinique :

Étudier la distribution pondérée des signes de langue en fonction des types de cancer, nous l' avons transformée les donnes trouvées dans le résumée de l' article en un graphique qui permet de visualiser, pour chaque type de cancer, avec la fréquence d' apparition des différents signes de langue observés chez les patients.

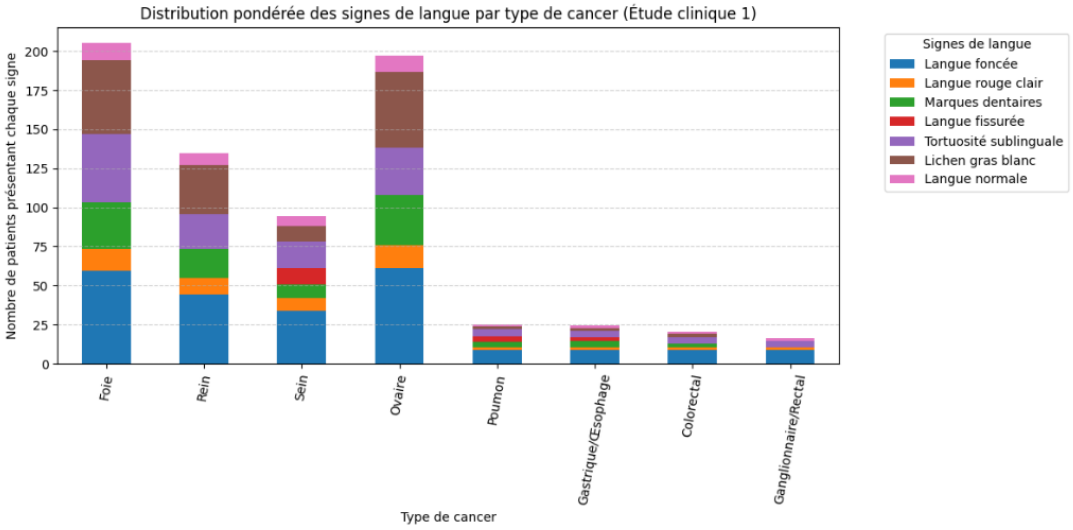


FIGURE IV.11 – Etude clinique sur 250 personnes attiens des tumeurs malignes avant la chimiothérapie

La deuxième étude clinique Le graphique ci-dessous illustre les résultats de la deuxième étude clinique portant sur l' évolution des signes de langue **avant** et **après** *traitement par chimiothérapie et radiothérapie* :Il présente la comparaison des signes les plus fréquemment observés chez les patients à quatre étapes distinctes : avant chimiothérapie, après chimiothérapie, avant radiothérapie et après radiothérapie.

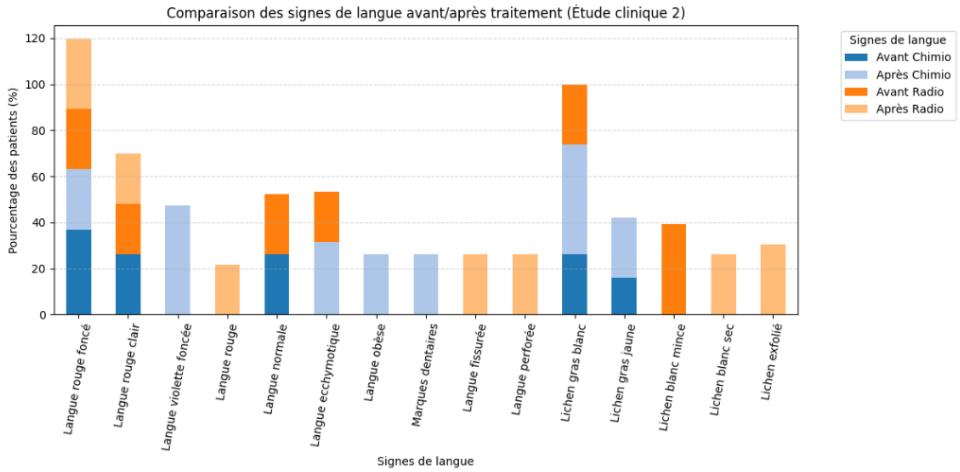


FIGURE IV.12 – Comparaison des singes de langue avant et après traitement (Chimiothérapie et radiothérapie)

Cette représentation met en évidence les modifications linguistiques associées aux

effets secondaires des traitements, soulignant l'intérêt d'un suivi visuel de la langue dans l'évaluation de l'état de santé des patients[40]

Nous synthétisons l'analyse de la langue de cette étude dans ce tableau

Type de cancer	**Signes de langue dominants**	**Interprétation en MTC**
Foie	Langue violette et foncée, veines sublinguales tortueuses, ecchymoses, lichen gras blanc, ligne biliaire hépatique	Stase sanguine, accumulation de mucosités, trouble de la circulation hépatique
Rein	Langue violette, mousse grasse blanche, lichen gras jaune	Carence en yin des reins, chaleur humide, mauvaise transformation de l'eau
Sein	Langue piquante (pétéchies), lichen gras jaune	Feu du foie, mucosités et humidité, troubles émotionnels
Ovaire	Langue piquante, lichen gras blanc	Humidité excessive, relation avec l'ascite, chaleur interne
Poumon	Langue fissurée, langue avec marques dentaires, desquamation, lichen gras	Sécheresse pulmonaire, carence en yin, déficience de la rate, humidité accumulée
Estomac / sophage	Marques dentaires, langue fissurée, peu de lichen ou lichen absent	Déficience de la rate et de l'humidité, carence en yin de l'estomac
Ganglionnaire / Rectal	Lichen gras jaune, langue ecchymotique	Chaleur humide et stase sanguine
Après chimiothérapie	Langue violette, ecchymoses, marques dentaires, mousse grasse blanche, langue enflée	Stase sanguine aggravée, atteinte de la rate et de l'estomac, troubles métaboliques de l'eau et du sang
Après radiothérapie	Fissures, lichen exfolié	Consommation du yin des poumons et de l'estomac, syndrome de chaleur-sécheresse selon la MTC

TABLE IV.2 – les signes linguaux associée à chaque type de cancer et son interprétation en MTC.

Ce travail s'inscrit dans une démarche clinique qui cherche à objectiver les observations traditionnelles en les rattachant à des profils pathologiques précis. Il ouvre ainsi la voie à une meilleure reconnaissance des modèles linguaux typiques des affections cancéreuses, en vue d'un diagnostic plus précoce et mieux ciblé

IV.3.2.2 A Non-Invasive Interpretable NAFLD Diagnostic Method Combining TCM Tongue Features (2023)

Les auteurs **Shan Cao et al.** publient en 2023 l'article intitulé "*A Non Invasive Interpretable NAFLD Diagnostic Method Combining TCM Tongue Features*" [46]. Ils pré-

sentent une méthode non invasive, pour diagnostiquer la stéatose hépatique non alcoolique (NAFLD : Non-Alcoholic Fatty Liver Disease). On s' appuie uniquement sur des signes linguaux issus de la MTC (forme, couleur, revêtement de la langue) et sur des indicateurs physiologiques faciles à mesurer (âge, sexe, taille, poids, circonférences abdominales et des hanches). Le modèle a été entrainé sur 1778 images de langue (831 cas de NAFLD, et 947 non-NAFLD) collectée avec l' appareil TFDA-1 présenté dans l' article précédent (3.3.1.3). Les images de langues (segmentées avec TongueSAM [46] un modèle conçu pour la segmentation de la langue) et les mesures physiologiques, sont fusionnées dans un réseau neuronal, SelectorNet (modèle de fusion attentionnel), qui utilise à la fois des mécanismes d' attention et de sélection de caractéristiques pour améliorer la précision et rendre le modèle interprétable. Le modèle atteint ainsi une Accuracy de 77,22 %, tout en générant des matrices d' interprétation visuelles des caractéristiques les plus influentes, ce qui permet de comprendre pourquoi un diagnostic est posé.

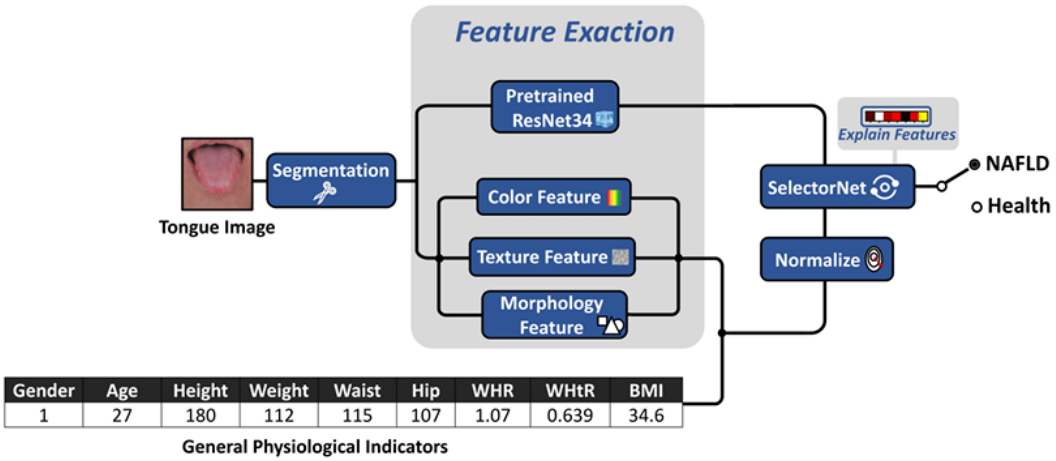


FIGURE IV.13 – L' architecture du modèle SelectorNet [46]

L'étude met l'accent sur l'intégration de l'analyse automatique des images de la langue avec l' IA, et propose un dispositif peu coûteux pour un dépistage précoce et accessible de la NAFLD, sans recourir à l' échographie ou la biopsie. Toutefois, la généralisation du modèle reste limitée par la dépendance au dispositif d'imagerie TFAD-1, et à des conditions de capture contrôlées. En outre, sa spécificité encore modérée (70,28 % ś 3,55 %) peut entraîner des faux positifs, réduisant sa fiabilité clinique dans un contexte réel.

IV.3.2.3 Analyse objective des syndromes en MTC chez les patients avec le diabète et prédiabète

Les auteurs **Ng et al**. publient en 2024 une étude intitulée " *Objective Analysis of Traditional Chinese Medicine Syndrome Differentiation of Patients With Diabetes and Prediabetes* " [47], dans laquelle ils proposent un protocole exploratoire pour standardiser le diagnostic syndromique en MTC, appliqué au diabète et au prédiabète. L' approche repose sur l' utilisation conjointe de questionnaires, d' un sphygmographe (l' analyse du pouls) et d' un dispositif d' imagerie linguale, dans une logique de croisement entre tradition clinique et intelligence technologique.

L' étude s' organise autour de cinq groupes équilibrés (n = 250 au total), allant des sujets sains aux patients atteints de diabète avec comorbidités. Les outils numériques utilisés permettent d' objectiver les caractéristiques du pouls, et de la langue (couleur, humidité, texture, collatéraux), en les comparant avec les constitutions corporelles évaluées par questionnaire (CCMQ : Constitution in Chinese Medicine Questionnaire). L' analyse repose sur des données collectées dans des conditions strictement standardisées, et une reconnaissance automatique fondée sur l' intelligence artificielle intégrée aux dispositifs.

Bien que les résultats définitifs soient attendus en 2025, cette étude marque un pas important vers l' objectivation du diagnostic en MTC. Les auteurs anticipent une amélioration notable de la fiabilité du diagnostic, en particulier à Singapour, avec une précision renforcée dans la différenciation des syndromes comme la chaleur humide, le vide de yin ou de qi. L' étude vise aussi à examiner si les profils observés diffèrent significativement de ceux établis sur la base des chinoises. Néanmoins, la nécessité d' une validation clinique à grande échelle et la dépendance à des configurations matérielles spécifiques peuvent constituer des freins à la généralisation immédiate.

Le tableau ci-dessous présente une comparaison claire et synthétique des travaux de recherche récents sur les méthodes modernes d' analyse de la langue à l' aide de l' intelligence artificielle. Il regroupe exclusivement des études proposant un modèle intelligemment automatisé pour le diagnostic non invasif basé sur les images linguales, avec une présentation classée du plus récent au plus ancien.

Titre de l' article	Objectif principal	Modèle	Données utilisées	Résultats	Limitations
Dr. Tongue [2025] [41]	Diagnostic à distance de la langue en MTC	YOLO + MobileSAM + TIUO + SignNet	Données non spécifiées	Précision : 86.68±0.30 %, F1-score : 72.87±0.63%	Dépendance à la qualité des images, variabilité des conditions de prise de vue
WSOD for Tooth-marked Tongue [2024] [45]	Détection et reconnaissance des marques dentaires	WSVM basé sur YOLOv8n + SAM + ViT + MIL	1 745 images + dataset Kaggle	YOLO : 100 %, MIL : privée 84,2%, Kaggle 89 %	Dépendance aux experts
Modernizing TD [2023] [42]	Diagnostic intelligent basé sur la MTC	Cv-Swin Transformer	5 365 images classées	Accuracy : 87,37 %	Ressources très coûteuses pour obtenir cette précision
NAFLD & TCM Tongue Features [2023] [46]	Diagnostic non invasif et interprétable de la stéatose hépatique	SelectorNet (fusion attentionnelle) + TongueSAM (segmentation)	1 778 images : 831 NAFLD, 947 non-NAFLD	Accuracy : 77,22 %, matrices interprétables des caractéristiques influentes	Dépendance à l' instrument TFAD-1, restreint au peuple chinois, spécificité modérée entraînant des faux positifs
Ammonia-Net [2023] [46]	Diagnostic des marques de dents	U-Net + ShuffleNetV2	TTD (856 images)	Classification : 2 classes : 99,06 %, 4 classes : 80,02 %	Confusion entre classes, dépendance aux experts
DTIA for Health Assessment [2022] [43]	Analyse et détection précoce de maladies (iTongue)	Fine-tuning de VGG et ResNet	4 787 images de langues diverses	ResNet50 : 64,52 %, VGG19 : 62,86 %	Aucune architecture fournie, absence de standardisation

TABLE IV.3 – Les articles proposant un modèle intelligent capable de diagnostiquer de manière non-invasive à partir d' images de langue

IV.3.3 Synthèse sur les systèmes de diagnostic par la langue en MTC

L'ensemble des études analysées souligne une tendance claire vers la modernisation du diagnostic de la langue en MTC à travers l'intégration de technologies avancées d' IA. La conception d' un système intelligent de diagnostic basé sur l' image de la langue s' articule autour de **11 étapes clés**, allant de la définition du besoin clinique à l' évaluation continue du modèle. Chaque étape repose sur une interaction entre les savoirs traditionnels de la MTC et les technologies modernes d' analyse d' image et d' IA. **On présente le pipeline de conception de ce système comme suit :**

- **1. Définir le cas d' usage** : Il s' agit d' identifier la pathologie cible ou les syndromes MTC à évaluer, comme le diabète, la NAFLD, ou le cancer du sein, etc.
- **2. Collaborer avec des praticiens MTC** : Un partenariat clinique est indispensable pour valider les critères de diagnostic utilisés en MTC (forme, enduit, couleur, fissures, etc.) et assurer leur transposition en variables observables sur image, ainsi qu' un médecin compétent de la médecine classique est indispensable pour assurer les pathologies de la MTC. On parle là de la combinaison entre la MTC et la médecine classique dans un cadre de recherche clinique hospitalier.
- **3. Capturer des images standardisées :** Les images de langue doivent être prises dans des conditions contrôlées (éclairage, distance, orientation), à l' aide d' appareils spécialisés ou une simple caméra mobile, comme le propose iTongue.
- **4. Annoter et segmenter les images :** Les zones de la langue sont annotées manuellement ou automatiquement (pointe, racine, côtés, centre), puis segmentées avec des outils comme SAM ou TongueSAM, afin d' isoler les zones cliniquement pertinentes.
- **5. Labelliser les images cliniquement :** Chaque image est associée à un diagnostic ou à plusieurs étiquettes simultanées (approche multi-label), en lien avec les syndromes MTC décrits par les praticiens.
- **6. Enrichir avec des données physiologiques :** L' ajout d' informations comme l' âge, le poids, le sexe, la constitution corporelle (CCMQ) permet d' améliorer la précision du modèle et de refléter la logique holistique propre à la MTC.
- **7. Prétraiter automatiquement les données :** Les images sont normalisées (taille, couleur), redressées si nécessaire à l' aide d' algorithmes comme TIUO, pour garantir une cohérence dans l' entrée du modèle. On construit alors un jeu de données dédié à notre étude.
- **8. Choisir une architecture d' IA adaptée :** On sélectionne le ou les modèles les plus performants selon le contexte: réseaux convolutifs (ResNet, VGG), Transformers (ViT, Swin), U-Net pour la segmentation, ou MIL pour les cas faiblement supervisés.
- **9. Entraîner et valider le modèle :** fine-tuning du modèle avec le jeu de données, un suivi d'évaluation comme l'accuracy, le F1-score, ou la sensibilité sur des nouvelles données.

- **10. Déployer une version de test clinique :** Le système est intégré dans une application mobile ou web utilisée en conditions réelles, et suivi d' indicateurs d' évaluation comme l' accuracy, le F1-score ou la sensibilité sur de nouvelles données.

- **11. Assurer une évaluation continue :** Enfin, les retours cliniques sont exploités pour enrichir la base de données, affiner les règles de prédiction et améliorer la robustesse du système dans des contextes variés.

Pour construire un bon système de diagnostic par la langue, on doit respecter toutes les étapes. Certaines peuvent être contournées ou externalisées grâce aux travaux scientifiques existants. En tant qu' ingénieurs en informatique, on commence souvent notre travail à partir de l' étape 7, car les étapes de 1 à 6 nécessitent l' expertise de praticiens médicaux et sont coûteuses en temps et en ressources.

Même lorsque des recherches ont été menées dans ce sens, les équipes ne partagent que rarement leur protocole complet, et leurs jeux de données restent le plus souvent privés. Cela freine considérablement les avancées dans ce domaine. C' est notamment le cas pour le **cancer du sein**, pour lequel **aucun dataset public n' existe à ce jour.** Cette rareté peut s' expliquer par la sensibilité des données cliniques. Par ailleurs, la généralisation de ces approches est également limitée par la population étudiée, qui est souvent restreinte géographiquement ou culturellement.

IV.4 Diagnostic du cancer du sein dans la vision MTC

En MTC, le cancer du sein résulte d'un déséquilibre énergétique complexe impliquant principalement une *stagnation du Qi et du sang*, une *accumulation de mucosités-phlegmeS* et de *chaleur toxique*. Selon le Dr Nan Lu, spécialiste de la MTC, le cancer du sein nécessite une compréhension des *causes profonde*s plutôt qu'un traitement purement symptomatique. Il souligne que les maladies reflètent un *déséquilibre global de la vie*, où les *facteurs émotionnels* et le stress jouent un rôle critique dans le développement et la guérison du cancer du sein. Sa vision met l'accent sur l'importance de la *circulation du Qi* et établit un lien direct entre l'équilibre émotionnel et la santé physique[48]. Les manifestations du cancer du sein varient d'une personne à l'autre, notamment en raison des fluctuations énergétiques entre le Yin et le Yan, compliquant ainsi le diagnostic. Les syndromes couramment associés au cancer du sein comprennent :

- **Stagnation du Qi du Foie** : Liée au stress émotionnel, provoque des douleurs mammaires et des masses.

- **Stase de sang** : formation des masses dures et douloureuses.

- **Accumulation de mucosités-phlegmes et de chaleur toxique** : formation de nodules et inflammation locale.

- **Vide de Yin des Reins** : Se manifeste par des douleurs sourdes, bouffées de chaleur, langue sèche à la base, et fatigue chronique.

Diagnostic par observation de la langue de ces déséquilibres

- **Rougeurs ou fissures sur les côtés**Stagnation du Foie
- **Langue gonflée avec enduit gras au centre** Accumulation Humidité/Mucosités (Estomac)
- **Langue sèche à la base** Vide de Yin des Reins
- **Pointe rouge** Troubles émotionnels (Cœur/ Poumon)

Les méridiens du Péricarde (PC), Estomac (ST), Rein (KD) et Foie (LV) traversent la région mammaire, créant des voies pathologiques privilégiées selon les zones anatomiques affectées. La figure suivante représente des méridiens qui favorisent l'apparition du cancer du sein avec un pourcentage de chaque zone.[48]

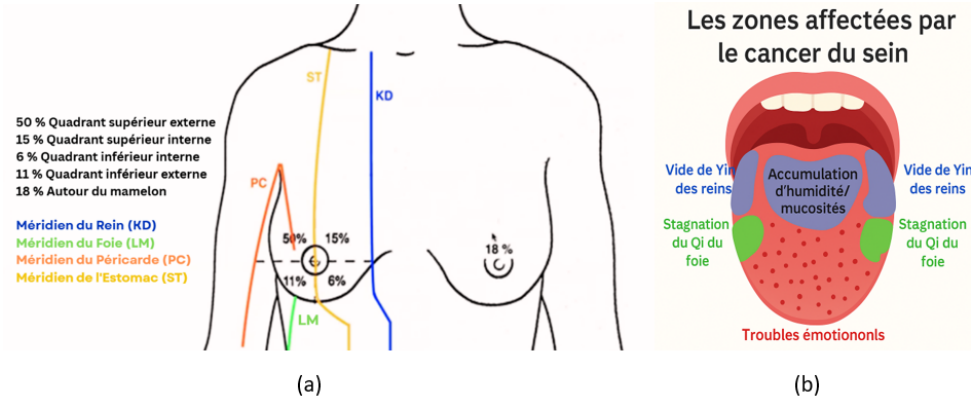


FIGURE IV.14 – Figure (a): les principaux méridiens impliqués dans le développement du cancer du sein , figure (b): une langue typiquement affectée par ces déséquilibres [48]

Le Dr Lu privilégie une approche préventive traitant les déséquilibres émotionnels avant leur manifestation physique, encourageant une connexion consciente avec les émotions corporelles. La MTC propose des traitements complémentaires (acupuncture, Qi-gong, pharmacopée) personnalisés selon la *localisation anatomique* du cancer, chaque zone reflétant des problématiques énergétiques spécifiques.[48]

Différents praticiens peuvent émettre des jugements variés en observant une même langue, et un même praticien peut aboutir à des diagnostics inconsistants s'il examine la langue à des moments différents [49] . Ce qui souligne la nécessité de concevoir des outils de diagnostic qui soient à la fois objectifs et uniformisés. Par conséquent, on observe des divergences dans les diagnostics présentés dans différentes publications.

IV.4.1 Travaux de recherche sur le diagnostic du cancer du sein avec la langue en MTC

Dans ce qui suit on va présenter trois études tentent de mettre au point des méthodes de diagnostic du cancer du sein basées sur l'observation de la langue en MTC, mais en se

limitant à des approches statistiques fixes, sans recourir à des modèles d' IA. Ainsi qu' un travail non accessible seulement sont résumée, disponible en chinois. Cela souligne à quel point ce sujet est peu exploré, et restreint au asiatiques.

IV.4.1.1 Une perspective observation de la langue et cancer du sein avec la MTC

Les auteurs **Lun-Chien Lo et al.** publient en 2013 l'article intitulé *"Breast Cancer Index: A Perspective on Tongue Diagnosis in Traditional Chinese Medicine"* [49], qui explore la possibilité de distinguer les patientes atteintes de cancer du sein (BCP : Breast Cancer Patent) des personnes saines (NP : Normal People) à l'aide de l'observation de la langue dans la MTC.

Échantillon : 60 BCP et 70 NP.

Outil utilisé *:un système de diagnostic automatique de la langue* (**ATDS**), créer par l' Université de Taipei à Taïwan, un outil permettant une analyse standardisée et non subjective, développé pour *capturer des images de la langue et extraire automatiquement des caractéristiques* de maladies, notamment le cancer du sein de manière statistique fiable pour aider au diagnostic des praticiens de la MTC. Il extrait *9 caractéristiques* de la langue (la couleur de la langue, la qualité de la langue, les *fissures* de la langue, le pelage de la langue, les points rouges, les ecchymoses, *les marques de dents*, la *salive* et la *forme* de la langue). Les caractéristiques identifiées ont été subdivisées en fonction des zones de la langue localisées (*rate-estomac, foie-vésicule biliaire gauche, foie-vésicule biliaire droite, rein, et Cœur-poumons*) [49] comme la figure suivante le montre.

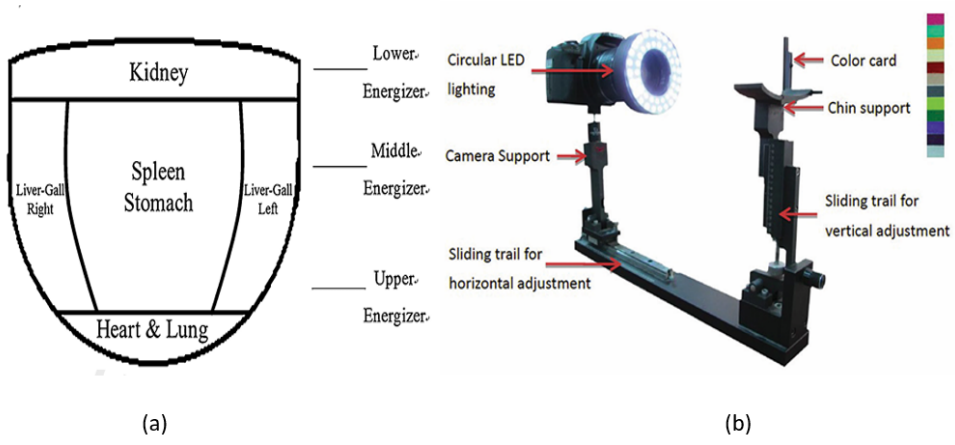


FIGURE IV.15 – Les zones de la langue (a) et les composants de ATDS en 2013 (b) [49]

Analyse statistique :

- Le test de Mann–Whitney (test statistique non paramétrique) a été utilisé pour détecter les différences significatives entre les deux groupes BCP et NP, en utilisant la métrique P (p-value variable de probabilité que le résultat soit prédit au hasard, la différence significative dans cet article est $P<0.05$).

- Trois modèles de régression logistique ont été construits pour prédire la probabilité d' un cancer selon les caractéristiques de la langue en utilisant trois métriques principales : la P, l' Odds Ratio (OR), qui estime le niveau de risque associé à une caractéristique (un OR > 1 indique une augmentation du risque, tandis qu' un OR < 1 reflète une réduction), et l' Intervalle de Confiance (IC) à 95%, qui encadre l' estimation de l' OR afin d' évaluer la précision du modèle.

Les patientes présentent des différences significatives par rapport aux personnes saines sur les points suivants :

- Quantité d'enduit lingual (souvent plus élevée)
- Enduit localisé dans la zone rate-estomac
- Surface maximale de l'enduit (plus étendue)
- Enduit mince (plus fréquent chez les patientes)
- Nombre de marques dentaires (tooth marks)
- Points rouges localisés dans les zones : Rate-estomac, Foie-vésicule biliaire gauche et droite, Cœur-poumon

Ces signes sont interprétés en MTC comme des manifestations de **stase de sang**, **chaleur interne**, **vide de yin**, ou **déséquilibres énergétiques** dans les organes concernés. Cette étude est la première à utiliser cette méthode non invasive statistique pour différencier des patientes atteintes d' un cancer du sein précoce des personnes saines.

IV.4.1.2 Indice de diagnostic du cancer du sein à un stade précoce avec la MTC

Dans la continuité des travaux antérieurs de 2013, les auteurs Lun-Chien Lo et al. publient en 2015 un article intitulé " TCM Tongue Diagnosis Index of Early-Stage Breast Cancer " [50],dans le a exploré la possibilité de distinguer les patientes atteintes de cancer du sein stade précoce 0 et 1 (ESBCP : Early stage BCP) des personnes saines (NBCI : Non BC Instance) à l' aide de l' observation de la langue dans la médecine chinoise traditionnelle (MTC). Sa méthodologie est la suivante : **Échantillon** : 67 ESBCP (stades 0 et 1) et 70 NBCI. **Outil utilisé** : le système ATDS. La 2eme version de ce système intègre un écran affichant les résultats de classification en temps réel. L'intégration d' un écran dans le système ATDS a amélioré la fiabilité et la précision du diagnostic linguistique en MTC, grâce à une analyse plus rapide, standardisée et facile à interpréter comme la figure suivante le montre :

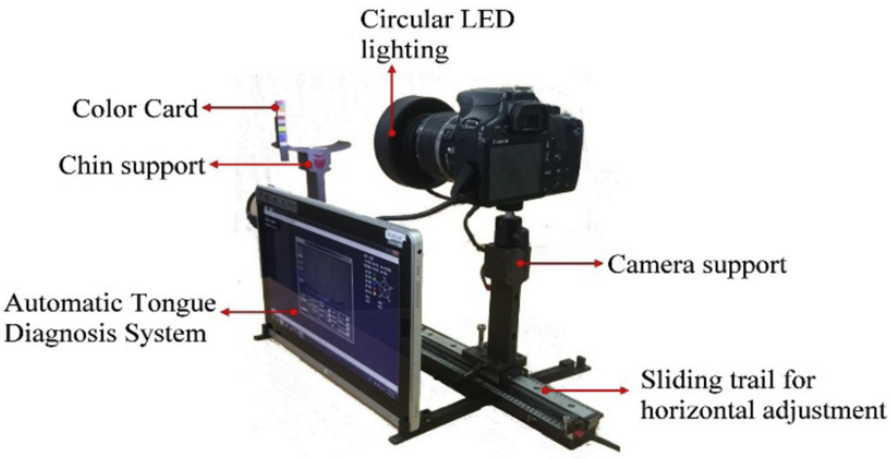


FIGURE IV.16 – La 2eme version du système ATDS, 2015 [50]

Analyse statistique :

- Le test de Mann–Whitney a été utilisé pour détecter les différences significatives entre les deux groupes ESBCP et NBCI. Une différence significative est définie par $P<0.05$.
- Trois modèles de régression logistique ont été construits pour prédire la probabilité d'un cancer du sein en fonction des caractéristiques de la langue. Les métriques utilisées incluent P, OR (Odds Ratio) et CI (Intervalle de confiance).

Les patientes atteintes de cancer du sein stade précoce présentaient :

- Une **quantité d'enduit lingual plus importante,**
- Une surface maximale de l'enduit plus étendue, et Une présence plus fréquente d'un enduit mince. De plus,
- La surface moyenne et maximale des points rouges sur la langue était significativement plus élevée, en particulier dans les zones correspondant à la rate–estomac et au Cœur–poumon.

En Médecine Traditionnelle Chinoise, ces signes sont interprétés comme le reflet de chaleur interne, de vide de yin, de stase sanguine ou de déséquilibres énergétiques affectant les organes associés. Cette étude est la première à utiliser cette méthode non invasive statistique pour différencier des patientes atteintes d'un cancer du sein précoce des personnes saines.

IV.4.1.3 Une analyse des caractéristiques de la langue chez les patients atteint le cancer du sein :

Dans la continuité des travaux antérieurs de 2013 et 2015, les auteurs **Lun-Chien Lo et al.** publient en **2021** un article intitulé " *A Perspective on Tongue Diagnosis*

in Patients with Breast Cancer " [50], approfondit l'analyse des caractéristiques de la langue chez les patientes atteintes de cancer du sein, en utilisant un système automatisé de diagnostic par l'image appelé **ATDS** et évaluer leur lien avec le stade de la maladie. L'objectif et la méthodologie de cette étude est la suivante :

- Étude rétrospective et contrôlée, menée de 2012 à 2016 à Taïwan.
- Analyse de 122 BCP et 155 femmes en bonne santé.
- Extraction automatique de caractéristiques linguales avec ATDS.
- Utilisation de tests statistiques (T-test pour les variables continues (ex : nombre de fissures), Chi-Square test pour les variables catégorielles (ex : couleur de la langue) et de régression logistique multiple (une méthode statistique qui modélise la probabilité d'un événement binaire en fonction de plusieurs variables explicatives). Utilisant les métriques P, OR et CI.

L'article ne donne pas explicitement une valeur globale de précision, mais il montre que plusieurs caractéristiques de la langue sont significativement différentes entre les deux groupes. Les patientes atteintes de cancer du sein présentent de façon significative les caractéristiques linguales suivantes : Petite langue, Couleur pâle, Enduit jaune, Salive humide, Fissures, Plus d'ecchymoses dans la zone Cœur-poumon. Ainsi que des signes cliniques classiques. Voici les manifestations les plus courantes :

- Fatigue chronique, douleurs mammaires ou thoraciques
- Troubles du sommeil, altérations de l'humeur (anxiété, dépression)
- Effets secondaires post-chimiothérapie : nausées, perte d'appétit, chute de cheveux

Ces signes seraient associés, selon la théorie de la MTC, à une déficience en yin, un manque de sang, une stagnation de chaleur, ainsi qu'à une accumulation de mucosités ou de stase sanguine. Certaines différences apparaissent entre *stades précoces* et *stades avancés* :

- *Moins de salive humide* dans les stades avancés
- *Plus d'ecchymoses dans la zone foie-vésicule biliaire droite* dans les stades avancés

L'étude montre que l'observation automatisée de la langue peut offrir un soutien au diagnostic préliminaire du cancer du sein, en complément des méthodes biomédicales classiques. Elle renforce également l'intérêt d'intégrer la MTC dans une prise en charge globale, notamment pour les patientes en traitement ou en phase de récupération post-chirurgicale ou post-chimiothérapie.

IV.4.1.4 Système de diagnostic et de recommandation de traitement du cancer du sien dans la MTC

L'auteur **Fan Baochao publie en 2021**, l'article intitulé " *Recherche sur le diagnostic de la langue MTC et la technologie de recommandation de la MTC pour le*

cancer du sein basée sur un graphe de connaissances et un réseau neuronal " [51] propose un système de diagnostic et de recommandation de traitement du cancer du sien dans la MTC en combinat 3 modules principales :

Graphe de connaissances en MTC : Un graphe de connaissances hiérarchisé est construit à partir du principe de la MTC. Ce graphe repose sur une ontologie médicale et intègre des données structurées (manuels médicaux, dossiers cliniques, etc.). Il modélise les relations *Théorie* de quatre niveaux (*symptômes*) loi (*syndrome*) *Prescription (organe touchées) Médecine (traitement)*, modélisant le raisonnement clinique (un chemin logique de choix de medicaments). Ce graphe intègre des données non structurées issues de manuels professionnels, normes nationales et dossiers cliniques. Les entités (attributs linguaux, maladies, organes, syndromes, herbes, etc.) et les relations sémantiques (causalité, correspondance, traitement) sont vectorisées à l' aide de la méthode **TransH** (Translating on Hyperplanes : transforme les graphes en vecteurs), puis stockées dans une base de données de type **Neo4j** (Une base de données orientée graphe), avec 2 686 nuds et 19 215 relations.

Analyse et reconnaissance de l' image de langue : Un CNN basé sur des variantes de ResNet (ResNet-18, ResNet-34, ResNet-50) est implémenté dans le cadre MXNet (framework open-source de DL, pour construire et entraîner un modèle de RNP) pour traiter les images de langue. Il permet :

- L' identification automatique de **22 attributs visuels** (couleur, forme, mousse, texture, etc.) spécifiques à la langue tumorale.
- L' extraction de **vecteurs de caractéristiques profond**s, en représentation numérique utile au diagnostic.
- Une évaluation comparative des performances des modèles (ResNet-50 atteignant **87,19** % de précision globale), sélectionné comme noyau du système de reconnaissance

Système de recommandation intelligent TRHM (Tongue-image Recommended Herb Model) Ce module combine les deux sorties des deux précédentes étapes :

- Les chemins cliniques et logiques extraits du graphe de connaissances MTC,
- Le vecteur de caractéristiques visuelles de la langue issue de ResNet-50.

Ensuite :

- Les données sont exploitées en fonction de la rétroaction implicite, c' est-à-dire les réponses indirectes des utilisateurs (comme l' acceptation ou l' efficacité d' un traitement), afin d' affiner les recommandations futures.
- Un filtrage collaboratif est utilisé : il recommande un traitement à un patient en se basant sur les profils de patients similaires (ayant présenté des caractéristiques linguistiques et cliniques proches), et sur les traitements qui se sont révélés efficaces pour eux.

Ainsi, le système peut recommander automatiquement des médicaments, de manière personnalisée, en tenant compte à la fois des données cliniques de la MTC et des images médicales. La figure suivante présente le diagramme typique de cette hybridation utilisée :

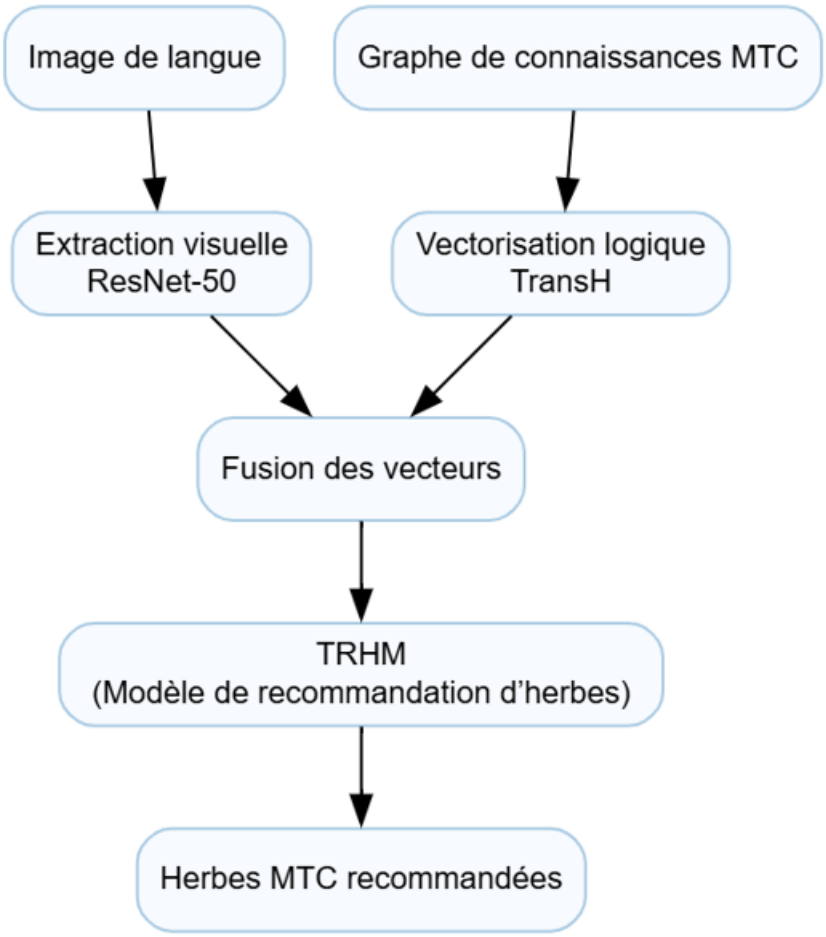


FIGURE IV.17 – Diagramme typique de l'hybridation utilisée dans cet article.

IV.4.2 Travaux de recherche qui sollicitent l' intégration de la MTC dans le traitement des personnes atteintes du Cancer du sein

On va présenter dans ce qui suit trois études, qui sollicite l'ajout de la MTC aux soins des patientes touchées par un cancer du sein. Ces travaux visent à améliorer la qualité de vie, à réduire la fatigue liée au cancer ou à proposer des approches complémentaires holistiques, là où les traitements classiques montrent des limites.

IV.4.2.1 La relation entre la fatigue en MTC et le cancer du sein

Les auteurs **Hsiao-Ye Liu et al**. publient en 2020 une étude intitulée "*Association between cancer-related fatigue and traditional Chinese medicine body constitution in female patients with breast cancer*" [52]. Elle explore le lien entre la fatigue liée au cancer (CRF

: Cancer Related Cancer) et les constitutions corporelles MTC chez 110 femmes atteintes de cancer du sein sous chimiothérapie, suivies entre octobre 2017 et mars 2018. Trois questionnaires ont été utilisés :

- Le BFI-T (Brief Fatigue Inventory-Taiwan) pour évaluer la sévérité de la fatigue entre 0 et10, le seuil retenu pour définir une fatigue cliniquement significative était un score 4.
- Le BCQ (Body Constitution Questionnaire) pour identifier les types de constitutions :
  - Yang-Qi deficienc (Insuffisance de l' énergie chaude, fatigue, frilosité, digestion faible).
  - Phlegm-Stasis syndrome (douleurs, masse, surpoids, métastases possibles).
  - Yin-Xue deficiency (Chaleur excessif sécheresse, l' insomnie et la pâleur).
- Le VAS (Visual Analogue Scale) pour mesurer l'intensité de la douleur et les troubles du sommeil, avec les métriques OR, CI et P.

Les résultats montrent que la fatigue traduit un déséquilibre énergétique, surtout une déficience de Yang-Qi, fortement associée à la fatigue persistante. Le syndrome Phlegm-Stasis, quant à lui, était fréquent chez les patientes atteintes depuis plus de 18 mois ou présentant des métastases. L' étude recommande d' intégrer les outils comme le BFI-T, le VAS et le BCQ à un suivi thérapeutique MTC personnalisé, afin d' adapter les traitements et améliorer la qualité de vie des patientes.

IV.4.2.2 Traditional Chinese Medicine for Breast Cancer: A Review

Les auteurs Yuan Jiang et al. publient en 2023 l' article de revue intitulé *"Traditional Chinese Medicine for Breast Cancer : A Review"* [53]. Ce travail passe en revue les effets bénéfiques potentiels de la MTC lorsqu' elle est utilisée en complément des traitements conventionnels du cancer du sein. L' article identifie quatre grands syndromes MTC associés au cancer du sein :

- **Stagnation du Qi du foie :** douleurs mammaires, distension thoracique, pouls tendu, langue rouge avec un enduit blanc mince ou jaune mince.
- **Accumulation de chaleur toxique :** rougeur, ulcération, fièvre, langue rouge foncé, avec des ecchymoses, avec enduit jaune mince ou épais.
- **Déficience de Qi et de sang :** fatigue, langue pâle avec un fur blanc épais, insomnie, pouls faible.
- **Déséquilibre de Chong et Ren :** bouffées de chaleur, sueurs nocturnes, langue rouge avec peu d' enduit blanc ou jaune, faible pouls. Le Chong et Ren des méridiens énergétiques responsable de la régulation hormonale, la reproduction du cycle de vie (grossesse, ménopause), et l' équilibre énergétique global, précisément chez les femmes.

Selon l' article, certaines formules traditionnelles en MTC à base de plantes peuvent aider à soulager la fatigue, les douleurs, les troubles du sommeil, les nausées et renforcer les défenses immunitaires. Des extraits de plantes ont montré qu' ils pouvaient empêcher les cellules cancéreuses de se multiplier et aider le corps à s' en débarrasser. L' acupuncture est également citée comme méthode efficace pour améliorer le bien-être global. Toutefois, les auteurs insistent sur le besoin de **preuves cliniques supplémentaire**s afin de valider scientifiquement ces résultats ainsi qu' ils soulignent un manque de standardisation des formules et des critères d'évaluation. Cette étude ouvre ainsi la voie à une complémentarité prometteuse entre MTC et médecine moderne.

IV.4.2.3 Guide de prise en charge des troubles émotionnelle chez les patients atteins le cancer du sein

Les auteurs **Junde Wu et al**. publient en 2024 un article de guide, intitulée "*Guideline of Traditional Chinese Medicine Diagnosis and Treatment for Breast Cancer Patients with Emotional Disorders*" [54]. Ils proposent une prise en charge spécifique aux *troubles émotionnels (dépression, anxiété, stress post-traumatique)* chez les patientes atteintes de cancer du sein, souvent négligés dans les traitements classiques. Le traitement est individualisé selon le *stade clinique du cancer* (péri-opératoire, chimio/radiothérapie, consolidation, traitement de rattrapage) et la *sévérité du trouble émotionnel* (léger, modéré ou sévère). L' approche de la MTC repose sur la **différenciation syndromique rigoureuse**, associant observation clinique (langue, pouls, symptômes) à des traitements ciblés : formules phytothérapie avec des plantes médicinales, acupuncture, thérapie musicale des cinq tons, relaxation, soutien familial et social. Chaque syndrome est défini par des symptômes principaux

- **Stagnation du Qi du Foie :** Dépression, oppression thoracique, douleurs costales.
- **Stagnation du Qi du Foie avec déficience de la Rate :** Troubles digestifs + humeur dépressive.
- **Stagnation du Qi du Foie transformée en feu :** Irritabilité, bouche amère, insomnie.
- **Accumulation de mucosités et d' humidité :** Sensation de boule dans la gorge, lourdeur mentale.
- **Déficience du Cœur et de la Rate :** Fatigue mentale, palpitations, troubles du sommeil.
- **Déficience du Yin du Foie et des Reins :** Agitation, sueurs nocturnes, sécheresse oculaire.

Chacun de ces syndromes donne lieu à une stratégie thérapeutique spécifique. Le guide appelle enfin à des recherches cliniques de plus grande ampleur pour valider scientifiquement ces approches intégratives. Cette directive a pour objectif de normaliser les pratiques et de proposer une stratégie holistique alliant tradition et médecine contemporaine pour soutenir la santé mentale et physique des patientes, tout en appelant à plus d' études cliniques pour en confirmer les effets.

IV.4.3 Synthèse et interprétation en MTC des patients atteints de cancer du sein

Les études analysées suggèrent que l' observation de la langue en MTC offre un complément intéressant pour évaluer l' état énergétique des patientes atteintes de cancer du sein, en lien avec le stade de la maladie et les déséquilibres internes. Sur la base de ces études on synthétise plusieurs caractéristiques linguistiques distinctes :

Symptôme lingual	Stade précoce (stade 0–1)	Stade avancé (stade 2)
Forme	Petite ou normale	Petite ou normale
Couleur	Pâle ou légèrement rouge	Rose, parfois rouge foncé
Enduit (fur)	Mince, parfois épais, jaunâtre ou blanc	Épais, jaune (plus marqué), surface plus étendue
Salive	Humide (fréquent)	Normale ou humide
Fissures	Peu présentes	Plus nombreuses
Ecchymoses	Rarement présentes	Fréquentes, surtout zones Cœur–Poumon ou Foie-VB
Points rouges	Zone Cœur–Poumon ou Foie-VB, Parfois fréquent.	Très fréquent Zone Cœur–Poumon ou Foie-VB.

TABLE IV.4 – Symptômes lingual des patientes atteintes du cancer du sein selon les articles.

Ces signes suggèrent une déficience de Qi et de Sang, ainsi qu' une accumulation d' humidité débutante, chez les patientes au stade précoce. Au stade avancé, une augmentation des ecchymoses est notée, surtout dans les zones Cœur–Poumons et Foie–Vésicule biliaire, indiquant un blocage énergétique plus marqué. L' augmentation des points rouges dans les zones Foie–Rate ou la zone Cœur–Poumon est remarquée, traduisant un stress très élevé.
Bien que plusieurs signes aient été rapportés dans les études, certaines caractéristiques comme les marques dentaires ou les points rouges localisés dans la zone de l' estomac n' ont pas été intégrées au tableau, car elles apparaissaient moins fréquemment. Seuls les signes les plus récurrents et significatifs ont été retenus, afin de refléter une synthèse fiable et éviter toute interprétation erronée.

Afin d' offrir un point de référence aux observations pathologiques précédentes, il est essentiel de définir les attributs d' une langue considérée comme saine.Les attributs d' une langue saine ont été définis :

- Forme : normale à légèrement gonflée,
- Couleur : rose uniforme,
- Enduit : blanc, mince et homogène,
- Fissures : parfois légères,
- Points rouges : discrets chez les patients stressés.

Les *caractéristiques de la langue* sont des *indicateurs complémentaires* dans l'évaluation globale, mais ne constituent pas un *diagnostic* à elles seules. En complément, ces observations nécessitent d'autre observations comme l'état physique et psychologique des patients, une interprétation doit être *individualisée* et intégrée à l'ensemble des *signes linguaux et symptômes cliniques*. Des recherches cliniques complémentaires sont encore nécessaires pour consolider leur validité scientifique

IV.5 Systèmes d'aides au diagnostic (SAD) développés pour la MTC

IV.5.1 SymMap

SymMap, est une base de données intégrative qui relie la médecine traditionnelle chinoise (MTC) à la médecine moderne en cartographiant les symptômes. Les auteurs ont manuellement curé 1 717 termes de symptômes issus de la MTC et les ont associés à 499 herbes ainsi qu'à 961 termes de symptômes utilisés en médecine moderne. La base de données comprend également 5 235 maladies, 19 595 constituants d'herbes et 4 302 gènes cibles, formant un vaste réseau hétérogène pour ça 1ère version. SymMap permet ainsi une intégration des connaissances sur les apparences des maladies et moléculaires entre la MTC et la médecine moderne, facilitant la découverte de médicaments basée sur les symptômes. [55] La figure suivante illustre un aperçu de la structure de la base de données de la 2ème version.

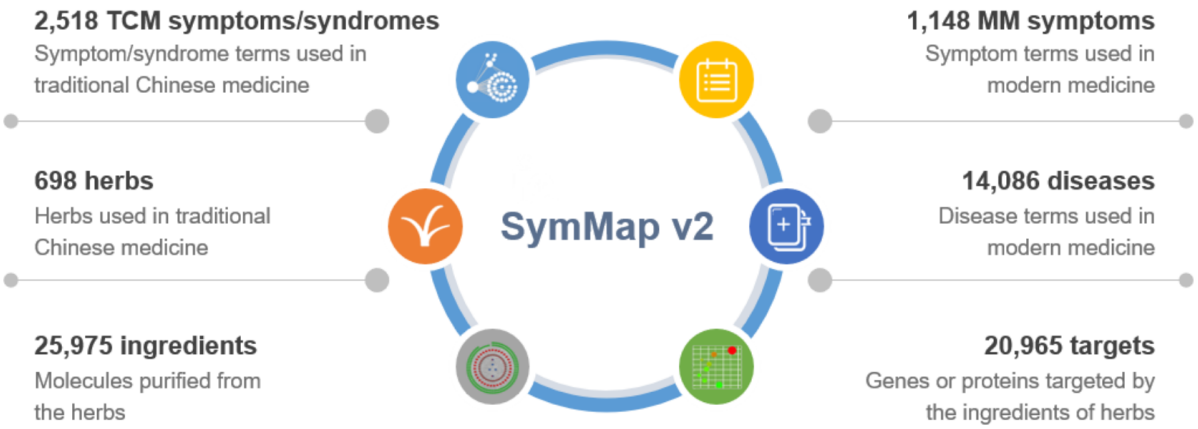


FIGURE IV.18 – La structure de la BDD SymMapv2 [56]

Cette base de données peut également servir de ressource précieuse pour le développement de systèmes d'IA, notamment dans l'analyse des relations entre symptômes, herbes médicinales et cibles moléculaires, en vue d'identifier des traitements personnalisés, comme cela a été illustré dans l'étude menée en 2021 par Fan Baochao sur le cancer du sein.

IV.5.2 Shen-Acupuncture

Shen-Acupuncture est une application mobile développée en 2020 par Juergen Bschaden [57], conçue comme *un outil de référence complet* pour les praticiens en MTC. Elle propose un *dictionnaire détaillé de plus de 4 000 symptômes*, incluant leur interprétation selon les principes MTC, ainsi que des indications sur les *zones de la langue associées à différents organes*. Bien que l'analyse de la langue soit réalisée *manuellement* par l'utilisateur, l'application offre un support visuel précieux pour l'observation clinique, notamment dans le suivi des patients atteints de **cancer**, comme le cancer du sein. Shen-Acupuncture est particulièrement utile pour guider le diagnostic énergétique et la sélection des traitements par acupuncture ou pharmacopée. Elle est disponible sur toutes les plateformes mobiles. On présente les deux figures suivantes : (a) présente la plateforme de Shen-Acupuncture, retirer du site officiel de Shen-Acupuncture [57] , tandis que la figure (b) représente la page d'index de l'application Shen-Acupuncture sous Android, téléchargé et capturée par nous-même.[58]

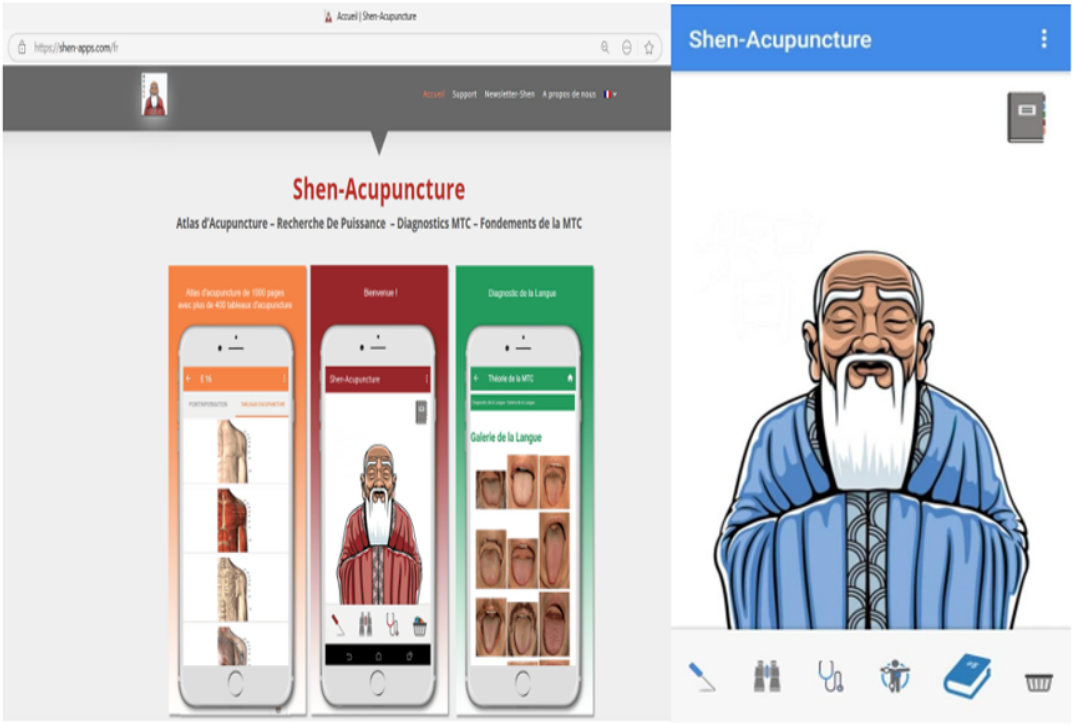


FIGURE IV.19 – La fig.(a) présente la plateforme de Shen-Acupuncture, et la fig. (b) représente la page d'index de l'application Shen-Acupuncture sous Android.
[58, 57]

Cette application donne aux spécialistes de la médecine traditionnelle chinoise la possibilité de renforcer leur savoir-faire et leurs méthodes, tout en proposant au public un moyen simple de découvrir un monde à la fois passionnant et complexe. Même si elle n'est pas totalement offerte, elle met à disposition des propositions gratuites, puis devient payante quand on désire obtenir des bilans poussés, dans une logique de valorisation du travail fourni.

IV.5.3 i-Tongue Diagnosis

L' application mobile *iTongue Diagnosis*, développée en Chine et mentionnée dans l' article [43], conçue pour assister le diagnostic en MTC à partir de l'analyse de la langue et d'un questionnaire physique. Elle fonctionne en deux volets : l' utilisateur peut soit répondre à un questionnaire d' évaluation traditionnelle, soit envoyer une photo de sa langue via l' application. Un mécanisme automatique de détection garantit que l' image transmise est bien une langue, dans le cas contraire, l' application invite l' utilisateur à soumettre une nouvelle image valide.
Une fois l' image validée, elle est transmise à un serveur où un modèle d' apprentissage automatique préalablement entraîné classe l' utilisateur selon son **profil énergétique en MTC**. Le système fournit ensuite des informations personnalisées sur la constitution, les déséquilibres potentiels, les tendances pathologiques ainsi que des conseils de prévention ou de rééquilibrage. Un module d' historique permet également de visualiser les diagnostics précédemment réalisés et de suivre l' évolution de son état.
L' application est actuellement disponible via des plateformes chinoises comme WeChat et sur l' App Store iOS, et elle est proposée en chinois et en anglais. Deux figures illustratives accompagnent cette description : la figure (a) présente le flux complet du diagnostic via l' application (tirée de l' article), et la figure (b) montre l' interface réelle de l' application iOS, telle que téléchargée et testée par nous.

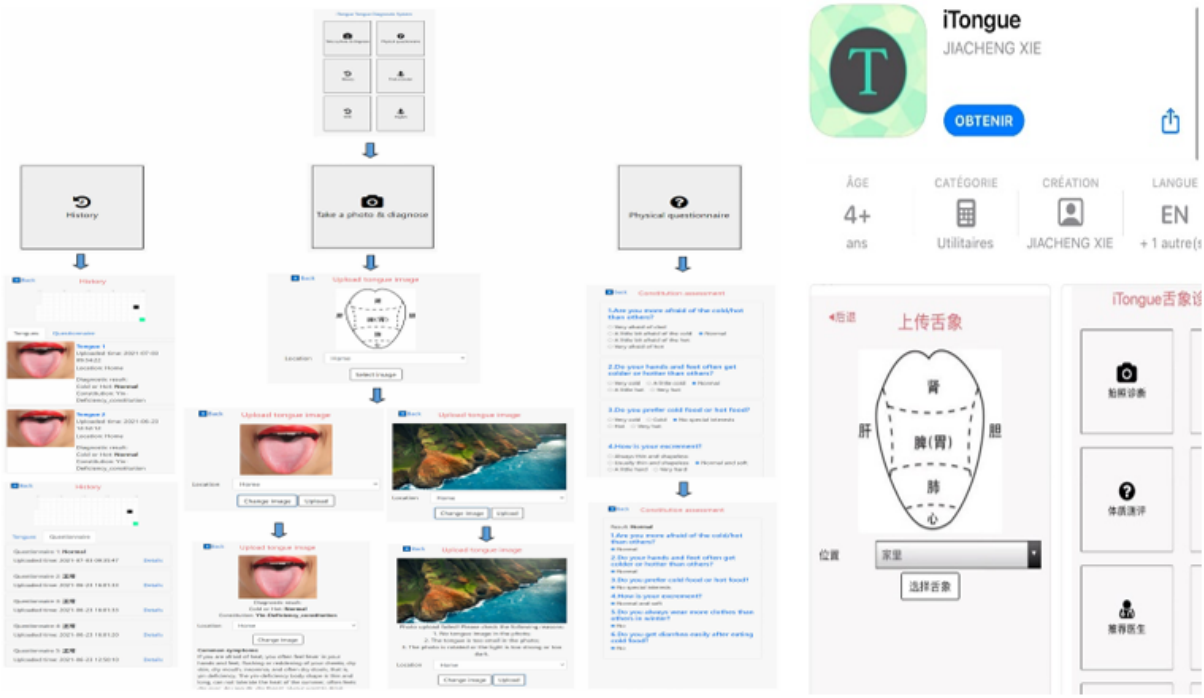


FIGURE IV.20 – La figure (a) montre le flux complet de l'application iTongue, la figure (b) représente l'application téléchargeable dans l'App Store iOS.
[43]

Cette application permet un diagnostic complet, clair et efficace de l' état de santé globale. Toutefois, son usage reste limité par sa disponibilité restreinte (uniquement en

chinois et anglais), avec des mises à jour peu fréquentes, ce qui pourrait freiner son intégration dans une pratique clinique plus large ou dans des recherches internationales.

IV.6 conclusion

L' examen de la langue en Médecine Traditionnelle Chinoise révèle des déséquilibres énergétiques subtils, souvent invisibles aux approches biomédicales classiques. Cette singularité ouvre des perspectives concrètes pour développer des systèmes de diagnostic automatisés, capables de repérer des signaux précoces de pathologies majeures comme le cancer du sein. On l' a vu tout au long de ce chapitre : la langue devient un miroir du corps, et l' intelligence artificielle, un outil pour en décrypter les reflets. On a résumé les travaux les plus récents consacrés à l' analyse automatisée de la langue, en identifiant 11 étapes IV.3.3 clés pour concevoir un système d' aide au diagnostic fiable. Ces étapes, allant de la définition du besoin clinique jusqu' au déploiement en conditions réelles, posent un cadre structurant et rigoureux. Néanmoins, nous avons aussi souligné une réalité frappante : malgré l' abondance des recherches en MTC, seules quatre études ciblent explicitement le cancer du sein par la langue, trois accessibles en anglais, une seule en chinois et dont le résumé seulement est fourni. Ce déséquilibre illustre la rareté des efforts engagés sur un sujet aussi sensible, qui touche une femme sur huit dans le monde[36], et reste particulièrement inquiétant en Algérie.

En parallèle, nous avons présenté deux applications mobiles " iTongue " et " Shen-Acupuncture " qui matérialisent cette volonté d' appliquer les savoirs traditionnels à des solutions technologiques actuelles. Ces exemples restent limités, mais ils montrent que l' innovation est en marche.Ce chapitre aura permis de poser les fondations théoriques de notre réflexion.

Dans le prochain, on détaillera notre propre contribution: un système d' aide au diagnostic précoce du cancer du sein, fondé exclusivement sur l' image de la langue, en se basant sur les étapes 1 a 8 du pilpeline IV.3.3 de création d' un système de diagnostic par la langue. On y exposera une approche innovante, pensée pour dépasser les limites actuelles, notamment l' indisponibilité des données cliniques à l' étape 7 du pipeline IV.3.3. On montrera qu' en l' absence de bases spécifiques pour le cancer du sein, il est encore possible d' avancer, à condition de penser autrement.

Chapitre V

Contribution

Sommaire

V.1 Introduction **106**

V.2 Jeu de données (Dataset) **106**

V.3 Metodologie de la recherche **111**

V.4 Module de prétraitement du dataset **113**

V.5 Module de détection de la langue pendant le diagnostique automatisée **119**

V.6 Module Détection des caractéristiques pathologiques et diagnostic précoce du cancer du sein **120**

V.7 Modèle étendu **124**

V.8 Construction d' une plateforme de diagnostic du cancer du sein selon MTC **126**

V.9 Conclusion **129**

Résumé : *Face à l'absence de données cliniques ouvertes, ce chapitre révèle comment l'intelligence artificielle a permis de créer un dataset synthétique pour le diagnostic du cancer du sein par analyse de la langue. En combinant les principes de la Médecine Traditionnelle Chinoise avec les technologies YOLO, nous proposons un système de détection précoce intégré dans une plateforme interactive "MTC Hybrid Pro". Cette innovation transforme une approche diagnostique ancestrale en outil numérique moderne, ouvrant la voie à une médecine préventive intelligente et non-invasive. Le système allie tradition millénaire et intelligence artificielle pour révolutionner le dépistage précoce.*

V.1 Introduction

Ce chapitre constitue le Cœur de notre contribution. Il s' inscrit dans une volonté de répondre à un enjeu médical majeur : le cancer du sein, une pathologie grave qui évolue souvent de façon silencieuse pendant plusieurs années. Lorsqu' il est diagnostiqué tardivement, il peut entraîner des pertes de vies humaines, une souffrance psychologique importante, et des conséquences d' ordre social non négligeables. Dans ce contexte, ce mémoire vise à contribuer à l' amélioration du diagnostic précoce de cette maladie en proposant un système intelligent fondé sur l' analyse de la langue selon les principes de la MTC.

Ce système s' inscrit dans un pipeline complet de conception de solutions d' aide au diagnostic basé image, articulé autour de 11 étapes (titre , chap , page): de la définition clinique du besoin à l' évaluation continue du modèle, en passant par la capture, l' annotation, le prétraitement, l' entraînement de modèles IA adaptés et le déploiement d' une plateforme testable. Toutefois, en l' absence de données cliniques ouvertes et d' expertise médicale directe, certaines de ces étapes (notamment les phases cliniques initiales) ont été contournées au profit d' une démarche expérimentale centrée sur l' ingénierie des données et l' intelligence artificielle.

Dans ce chapitre, nous allons d' abord présenter la méthode de génération du dataset, puis détailler l' architecture globale du système proposé, en décrivant chacun de ses modules fonctionnels, de la segmentation à la classification, jusqu' à la mise en place d' une plateforme de test.

V.2 Jeu de données (Dataset)

V.2.1 Problème rencontrée

Dans tout projet d' apprentissage automatique, la qualité et la pertinence du jeu de données constituent un pilier fondamental. Avant même de concevoir une architecture ou de choisir un algorithme, il est essentiel de disposer de données fiables, représentatives et correctement annotées. Or, dans notre cas, la première difficulté rencontrée a été l' absence de dataset public adéquat répondant à notre problématique spécifique : l' analyse de la langue pour le diagnostic du cancer du sein selon les critères de la Médecine Traditionnelle Chinoise (MTC). Les principales plateformes spécialisées comme Kaggle [59], Roboflow [60] ou OpenML [61] ne proposent aucun jeu de données correspondant à ce double niveau d' exigence. Pour élargir notre recherche, nous avons également exploré d'autres bases scientifiques ouvertes reconnues, telles que : Dataverse Harvard [62], ScienceDirect [63], IEEE DataPort [64], AI Studio Baidu [65], Europe PMC [66]. Malgré cette recherche élargie, nous n' avons identifié aucun dataset exploitable pour notre cas d' usage. De nombreuses études analysées dans notre état de l' art mentionnent l' utilisation de bases de données pertinentes, mais celles-ci sont privées, souvent détenues par des institutions hospitalières ou des consortiums académiques. Malgré plusieurs tentatives de prise de contact par e-mail avec les auteurs de ces publications, nous n' avons obtenu aucune réponse favorable de leur part pour accéder aux données. Ce manque d'ouverture constitue un obstacle majeur à la reproductibilité des recherches dans ce domaine émergent.

V.2.2 Collecte de données réelles : une solution peu réalisable à court terme

Face à cette indisponibilité, une option envisagée fut la constitution d' un jeu de données réel. Cependant, cette solution s' est vite heurtée à une série d' obstacles pratiques et réglementaires. En effet, créer un jeu de données cliniques original implique une démarche longue, coûteuse, et encadrée juridiquement. Il aurait fallu :

- **Obtenir une autorisation de recherche clinique** auprès d'un centre hospitalier, avec le soutien de l' université et du ministère de la santé publique ;
- **Convaincre des patientes atteintes** de cancer du sein de participer volontairement à l' étude en acceptant d' être photographiées dans un cad**re strictement médical ;**
- **Acquérir du matériel spécialisé** pour assurer des conditions de prise de vue normalisées (lumière, humidité, angle de langue, distance) ;
- **Mobiliser des praticiens expérimentés en MTC** (au moins deux ou trois) pour valider les observations visuelles selon les critères traditionnels (couleur, forme, enduit, fissures⋯);
- **Faire intervenir des médecins conventionnels,** capables de croiser les diagnostics MTC avec des données médicales modernes (dossiers patients, analyses biologiques⋯) ;
- **Disposer de techniciens et d' ingénieurs compétent**s pour la prise d' images et l' annotation rigoureuse et reproductible des observations.

Ce protocole, bien qu' idéal sur le plan méthodologique, ne pouvait être mis en uvre dans les délais impartis par notre projet, faute de ressources humaines, matérielles et administratives suffisantes. C' est pourquoi nous avons opté pour une alternative plus rapide, tout en veillant à respecter la rigueur scientifique.

V.2.3 Génération des données par l' IA

Face aux difficultés de collecte réelle et à l' absence de données ouvertes, nous avons fait le choix innovant de générer un dataset synthétique via l' intelligence artificielle, en nous appuyant sur le Chapitre IV
Notre démarche s'appuie sur les caractéristiques cliniques les plus fréquemmentes associées aux patientes atteintes de cancer du sein selon la MTC (voir le Tableau IV.4 du Chapitre IV).

En croisant ces informations avec les éléments visuels récurrents observés dans la littérature scientifique [49, 50, 53], nous avons conçu un protocole de génération contrôlée d' images de langues synthétiques.
Chaque image générée repose sur une combinaison de variables :

- Zones linguales affectées (racine, centre, bords, pointe),

- Caractéristiques visuelles (coloration, enduit, humidité, fissures),
- Et leur correspondance avec les syndromes MTC liés au cancer du sein.

Nous avons utilisé des outils de génération d'image assistée par IA, complétés par une validation manuelle des images synthétiques, afin de nous assurer qu'elles respectaient à la fois la cohérence visuelle et les critères diagnostiques décrits. Cette méthode nous a permis de constituer un jeu de données original, parfaitement adapté à notre cas d' utilisation, tout en surmontant les limitations éthiques, pratiques et temporelles de la collecte réelle. Bien que synthétique, ce dataset représente une base solide pour le prototypage de notre système, et pourra être enrichi ou complété ultérieurement dans un cadre clinique plus encadré.

V.2.4 Preuve de la faisabilité de l' approche

Comme le souligne l' article de Chen et al. (2021), intitulé « Synthetic data in machine learning for medicine and healthcare », publié dans la revue *Nature Biomedical Engineering* [67], l' utilisation de données synthétiques générées par intelligence artificielle est aujourd' hui scientifiquement reconnue et validée dans le domaine médical. Cette publication de haut niveau, largement citée par la communauté scientifique, démontre que les données artificielles produites par des modèles de type GAN (Generative Adversarial Networks) ou d' autres architectures génératives peuvent remplacer les données réelles tout en maintenant un niveau de fiabilité élevé pour le diagnostic, la classification ou la segmentation.

L' étude conduite par Hu et al. (2024), intitulée « A Survey of Synthetic Medical Image Generation for Improving Disease Classification », menée à Stanford University [68], explore la génération d' images de fond d' il pour le diagnostic de rétinopathie diabétique à l' aide de GAN. Les auteurs ont ajusté un StyleGAN préentraîné sur des visages humains à un jeu d' images médicales. Malgré un volume initial limité, les images synthétiques ont enrichi le dataset. Les tests montrent une meilleure robustesse des classifieurs, sans sacrifier la qualité visuelle. Bien que certains détails fins soient mal restitués, les résultats sont prometteurs.

Dans une étude de Van Booven et al. (2024), intitulée « Synthetic Genitourinary Image Synthesis via Generative Adversarial Networks », publiée dans le *Journal of Personalized Medicine* [69],nous avons démontré qu' on pouvait générer des images histologiques synthétiques de haute qualité pour la pathologie génito-urinaire. En utilisant des bases publiques, les chercheurs ont entraîné leurs modèles GAN (notamment StyleGAN) sans fine-tuning préalable. 80% des images synthétiques ont satisfait les critères cliniques d' acceptabilité selon des pathologistes certifiés.

V.2.5 Versions de génération

V.2.5.1 Première tentative : un script Python

Un script Python a été développé par nous même. Il définit les formes, couleurs, et effets associés à chaque type de syndrome (pâleur, rougeur, enduit, fissures, etc.), à l' aide

des bibliothèques `os` [1], `PIL` [2], `random` [3], et `yaml` [4]. Malgré la génération de 2000 images, le réalisme clinique était absent.

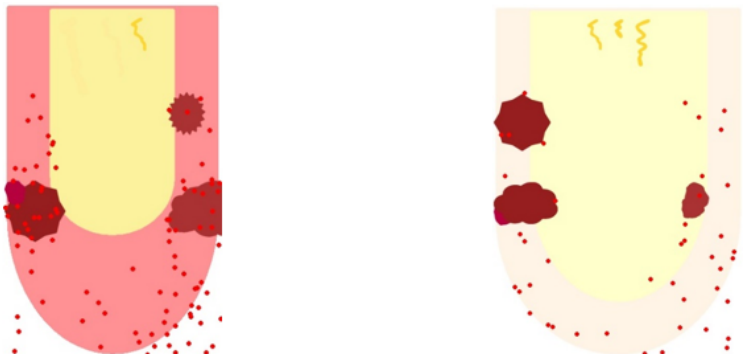


FIGURE V.1 – Exemples issus de la première tentative de génération

V.2.5.2 Deuxième tentative : Synthèse d' images IA

La deuxième tentative, c'était de faire une synthèse d'images par l'IA.nous avonstesté cette fois –ci un code python généré automatiquement par l' IA dont le but est de produire des langues artificielles avec certains syndromes, selon nos critères. Mais, le résultat n'était pas convaincant. Les images ne ressemblaient pas à des langues humaines : plutôt à des formes floues, texturées comme des ufs, avec des couleurs incohérentes. Le réalisme était complètement absent. La figure suivante illustre quelques rendus obtenus à cette étape.

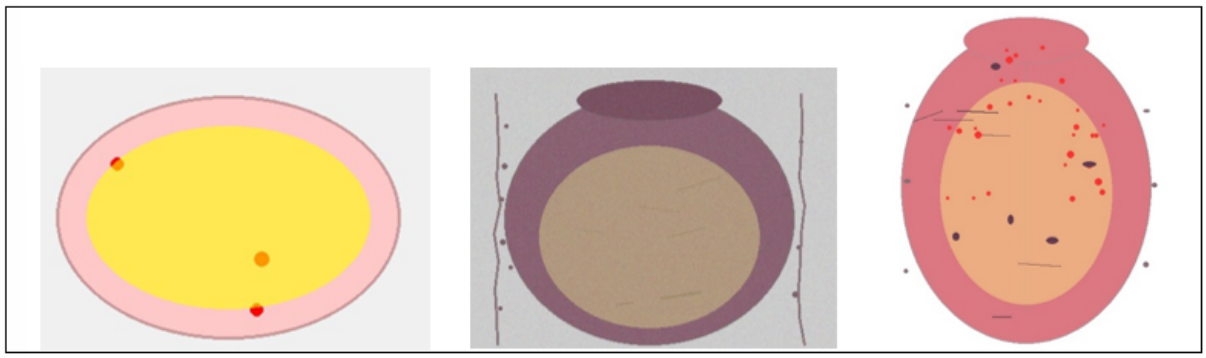


FIGURE V.2 – Exemples de la deuxième tentative de génération

V.2.5.3 Troisième tentative : Réalité augmentée

Dans cette tentative, l' idée était d' appliquer une forme de réalité augmentée, en superposant artificiellement des marqueurs pathologiques sur des images réelles. L'objectif est d'enrichir visuellement les images pour simuler des cas de langues pathologiques sans créer une image entièrement artificielle. Cependant, le résultat visuel s' est avéré peu

1. **Os :** pour gérer les chemins de fichiers et les dossiers,
2. **PIL:** Python Imaging Library, dédiée à la manipulation d' images en Python
3. **Random :** pour générer des valeurs aléatoires, notamment dans les processus de variation ou d' augmentation
4. **yaml**: utilisée pour lire ou écrire des fichiers de configuration au format YAML

satisfaisant. Les ajouts paraissaient dessinés, déconnectés de la texture de la langue, ce qui rendait l' image finale ni crédible ni exploitable pour un système d' analyse. Le rendu faisait penser à un montage maladroit, et non à une langue réellement atteinte.

V.2.5.4 Quatrième tentative : GAN –Stable Diffusion et ControlNet

Dans cette tentative,nous avonsfait recours à la génération d' images par l' IA. Deux modèles génératifs (GAN) ont été utilisés. Le premier, Stable Diffusion, est un modèle préentraîné de type diffusion permettant de générer des images artificielles à partir de descriptions textuelles (approche text-to-image). Il repose sur un processus de génération progressive, partant d' un bruit aléatoire et guidé par un *prompt médical*[5] décrivant les caractéristiques d' une langue associée au cancer du sein selon la MTC. Parfois, il génère rien car il n' arrive pas à trouver le lien correct entre le prompt et les données sur lequel il a été entrainée.

Le second modèle, ControlNet, a été utilisé pour l' approche image-to-image. Il s' appuie sur la même base que Stable Diffusion, mais ajoute une capacité de contrôle guidé par une image source.nous avonsfourni une langue saine comme image d' entrée, et le modèle applique les transformations nécessaires (forme, texture, couleurs) pour simuler visuellement des symptômes pathologiques en se basant sur un prompt défini, malgré que c' est une approche avancée comparé à la précédente texte-to-texte, mais elle ne donne toujours pas des bons résultats.

V.2.5.5 Cinquième tentative : IA guidée par prompt médical

Après plusieurs échecs pour générer des images réalistes de langues (proches de la réalitée), on a décidé de changer totalement d' approche. Au lieu de générer les langues à partir de zéro ou d'altérer des images de base,nous avonsconfié cette tâche directement à l'IA (ChatGpt [70] et Copilot [71]). Cette fois, plus de lignes de code, juste une idée précise traduite en un prompt médicale, détaillé et documenté avec une image de base d' une langue humaine, suivie d' une demande pour générer une version réaliste correspondant à un état clinique bien précis. Nous avons pris soin de rassembler toutes les connaissances acquises au fil des lectures, notamment les syndromes récurrents, les localisations des signes cliniques (zones Foie–Rate, Cœur–Poumons, etc.), les types de langues(petite , normale , gonflée) , les couleurs d' enduit (jaune et blanc), les fissures, et bien d' autres encore. Chaque symptôme a été soigneusement intégré au prompt.nous avonsveillé à ce que l' IA tienne compte des trois classes visées : *langue saine, stade précoce et stade avancé du cancer du sein*, principalement basée sur les caractéristiques synthétisée et présentée dans la synthèse IV.4.3.

Nous avons répété ce processus manuellement, image par image, afin de conserver un contrôle total sur la cohérence clinique de chaque sortie générée, ainsi que de choisir des images réelle de langues des personnes saines. Ce travail méticuleux a permis de constituer un jeu de données synthétique structuré comme suit :

5. Un prompt médical désigne la requête adressée à l' IA dans un contexte médical. Dans ce projet, il s' agit d' instructions permettant de générer des images de langue selon les caractéristiques définies par la MTC.

- 21 images **ABC (stade avancé),**
- 22 images **EBC (stade précoce),**
- 24 images **HEALTHY (langue saine),**
- 21 images **réelles (sources ouvertes)**.

Chaque image générée respecte les critères de segmentation de la langue et les caractéristiques visuelles attendues (coloration, enduit, fissures⋯), garantissant ainsi la validité clinique minimale nécessaire pour l' entraînement et la validation de notre système. Ce processus de génération de données n' a pas été sans difficulté. L' IA a parfois refusé de générer certaines images, invoquant des contenus sensibles ou des limites éthiques. D'autres fois, elle produisait des résultats incohérents, voire inutilisables.nous avonsdû s' adapter, affiner les prompts, patienter, faire preuve de résilience. Cette étape, bien que chronophage, a été essentielle pour donner naissance à un dataset crédible et exploitable. C' est pourquoi le dataset final a été nommé : *Tongue_MTCBC_Hope*.

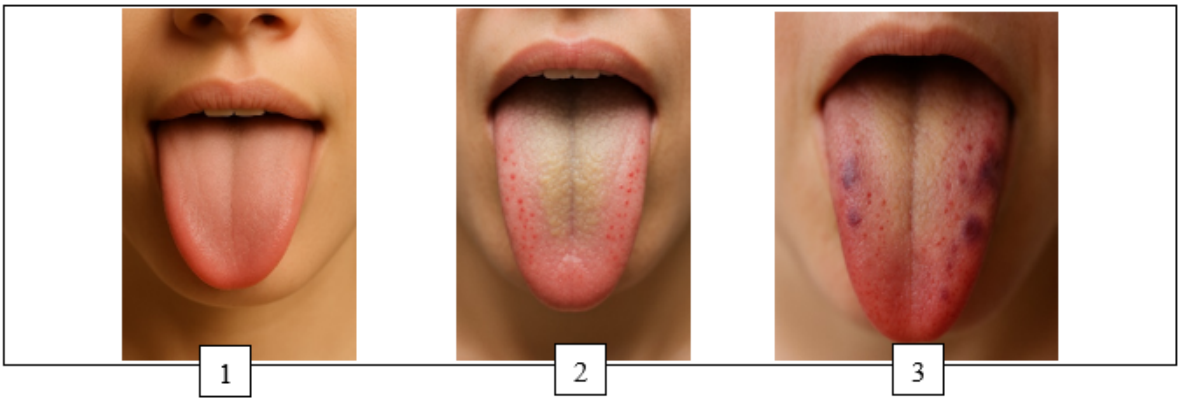


FIGURE V.3 – Exemples issus du dataset Tongue_MTCBC_Hope. (1): saine, (2): précoce, (3): avancée

Malgré les difficultés, ce dataset constitue une contribution majeure à la recherche sur le cancer du sein en MTC.

V.3 Metodologie de la recherche

À la suite d' un état de l' art approfondi sur les approches de diagnostic basées sur la langue en MTC, nous proposons une architecture complète et originale de système intelligent de diagnostic du cancer du sein, que nous avons nommée YOLO_BCTCM (YOLO for Breast Cancer in Traditional Chinese Medecine). Cette architecture repose sur une combinaison de techniques de vision par ordinateur, d' annotation experte, de détection d' objets, de classification supervisée et d' interprétation médicale du cancer du sein selon la MTC.

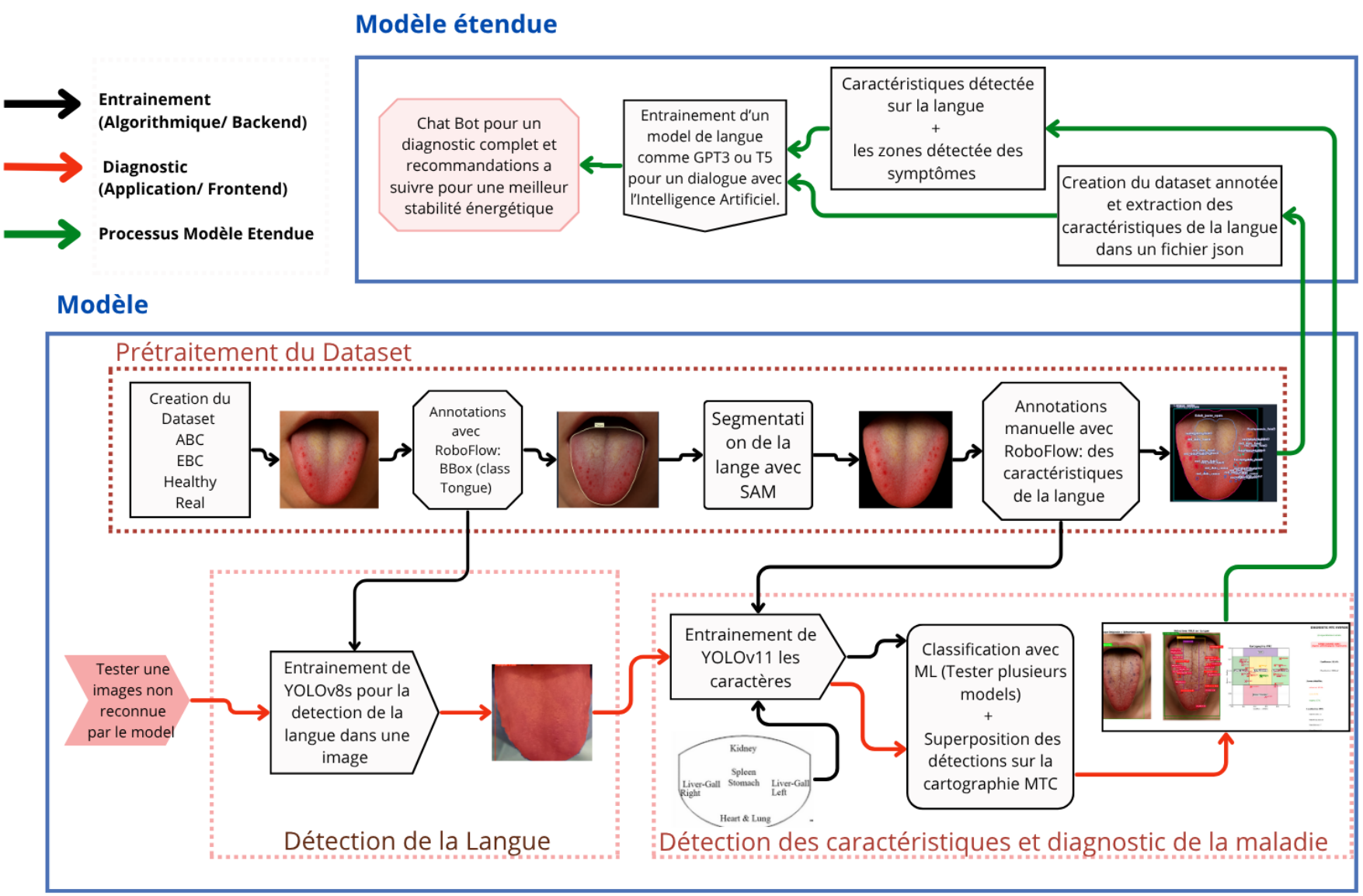


FIGURE V.4 – Architecture du système YOLO_BCTCM

Cette architecture repose principalement sur 4 modules principale qui sont : Cette architecture repose principalement sur 4 modules principaux qui sont :

- **Prétraitement du dataset** : annotation manuelle avec Roboflow, et segmentation de la langue avec le modèle SAM.
- **Détection de la langue** : entraînement de YOLOv8s pour localiser automatiquement la langue dans une image.
- **Détection des caractéristiques pathologiques** : entraînement de YOLOv11 pour identifier les symptômes visibles sur la langue, suivi d' une classification par modèles de Machine Learning (SVM, XGBoost, etc.) avec superposition sur la cartographie MTC.
- **Modèle étendu** : génération automatique d'un dataset JSON enrichi, détection des zones cliniques affectées, et intégration d' un chatbot médical basé sur GPT-3/T5 pour un diagnostic interactif et des recommandations de stabilisation énergétique.

Nous allons détailler par la suite chaque module en profondeur. La Figure V.4 présente notre architecture

V.4 Module de prétraitement du dataset

La première étape consiste à faire une segmentation sémantique grâce à YOLO, pour localiser automatiquement la langue dans chaque image. À cet effet, les images sont d' abord annotées avec une boîte englobante autour de la langue.

Pour obtenir un contour plus précis, nous avons appliquées l' algorithme SAM (Segment Anything Model), qui affine la zone détectée en extrayant uniquement la langue, enlever le fond.

Chaque image est annotée manuellement dans Roboflow pour indiquer les caractéristiques pathologiques observées (zones rouges, fissures, enduits⋯). Ces informations serviront à entraîner le modèle de détection des symptômes.

Ce processus est illustré dans la figure suivante.

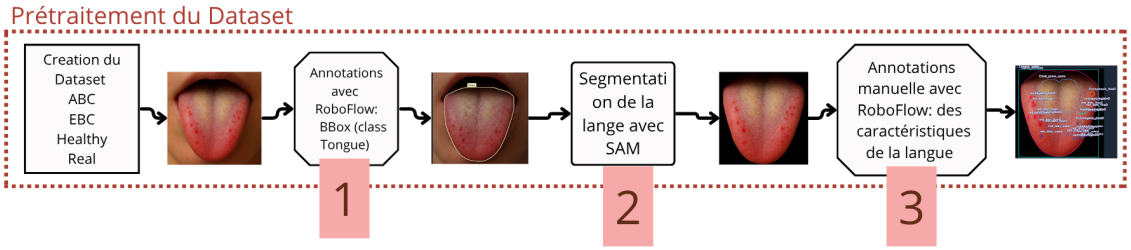


FIGURE V.5 – Le module prétraitement du dataset

V.4.1 Etape1 : Annotation manuelle de la langue avec Roboflow

La première étape clé, c' était la labellisation (L'annotation). En d' autres termes, il fallait identifier sur chaque image ce qu' on voyait, puis attribuer une étiquette (ou "label") claire à chaque langue repéré.
L'annotation a été effectué sur ń Roboflow ż : une plateforme cloud qui gère les données d' image, l'annotation, la conversion de formats, la préparation des datasets et l'entraînement des modèles proposés aux utilisateurs. Cet outil en ligne propose des modèles intelligents pour assister les humains dans ce travail fastidieux. Il supporte plus de 30 formats d' annotation: YOLO, JSON, COCO, etc. Il permet aussi de diviser les données en ensembles train/val/test, attribuer des tags, et exportées dans le format adapté au modèle souhaité. L' interface est simple. D' abord, nous avons créé un projet avec type d' annotations, ensuite nous avons téléversé nos images, dans un projet. Une fois les fichiers chargés,nous avonspu accéder à l' onglet Annotate, où l' on retrouve toutes les images prêtes à être labellisées. Là, on peut soit utiliser un modèle automatique proposé par Roboflow (Auto-Labeling), soit annoter manuellement les zones d' intérêt. La figure V.6 illustre l' interface de Roboflow dans l' onglet Dataset ainsi que la possibilité de création des versions du dataset avec diverses variations.

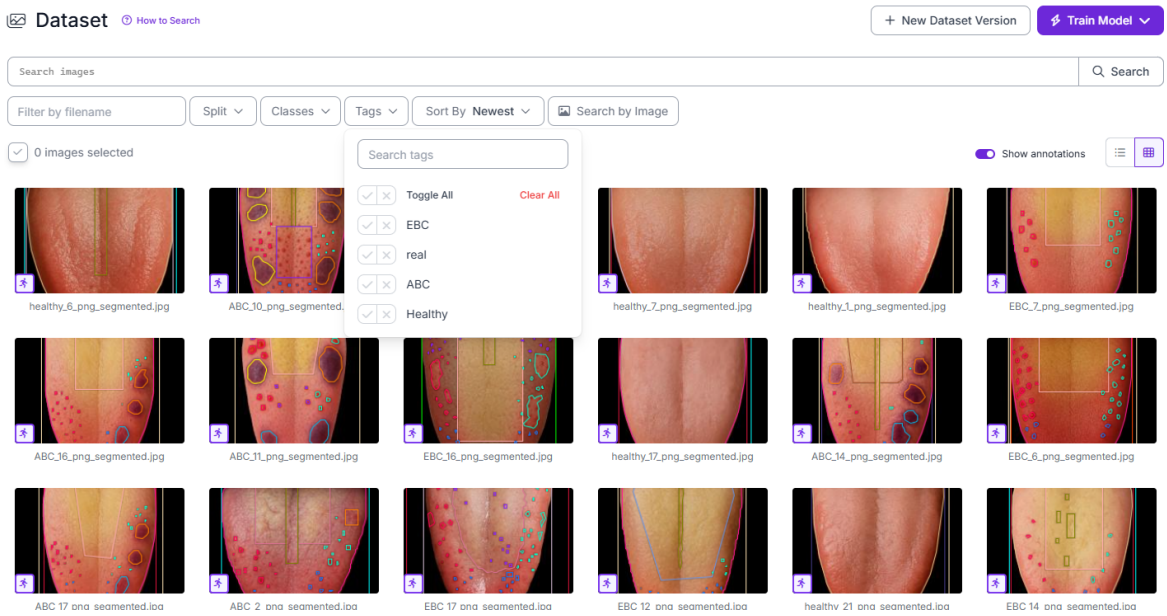


FIGURE V.6 – Interface de Roboflow lors de la création du dataset.

Dans notre projet, bien qu' il existe plusieurs types d' annotations possibles, on s' est concentré principalement sur deux approches : la boîte englobante (bounding box ou BBox) et le polygone.

La bounding box, c' est un rectangle qu' on dessine autour d' une zone importante de l' image, par exemple dans la figure [?]nous avonsannoté la langue avec une BBox. En la nommant, on indique à l' algorithme ce qu' il doit apprendre à reconnaître. C' est une méthode simple mais efficace pour entraîner des modèles à détecter la présence d' un élément.

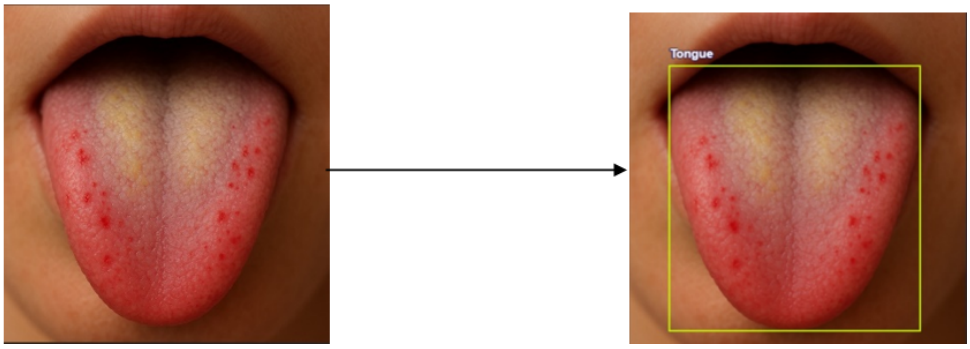


FIGURE V.7 – Annotation de la figure (a) de langue avec une Bbox dans la figure (b).

À côté de cela,nous avonsaussi utilisé l' annotation par polygone une approche qui intervient à la limite de Bbox dans la détection d' objets, qui permet de dessiner précisément le contour d' une zone, point par point. C' est ce qu' on appelle la détection d' instance : chaque objet (ici, chaque langue ou caractéristique) est isolé avec précision, comme si on découpait son contour. Voici une comparaison claire entre les deux :

- **Détection d' objet (BBox)** : permet de repérer les objets dans l' image à l' aide de rectangles englobants. Cette méthode est rapide et efficace, mais elle ne fournit pas la forme exacte des objets.
- **Détection d' instance (polygone)** : aussi appelée *segmentation d' instance*, elle va plus loin en traçant la forme réelle de chaque objet, en isolant chaque langue individuellement à l' aide de masques précis, définis pixel par pixel.

Prenons un exemple : Dans notre expérimentation ,nous avons rencontré un problème de segmentation, malgré annoté les langues avec Bbox, le modèle de segmentation ne segmente pas correctement la langue, il se trompe entre la langue et le visage car la couleur est similaire. Il fallait avoir les contours exacts pour segmenter la langue correctement. La figure V.8 présente clairement la différence entre la Bbox et le Polygone.

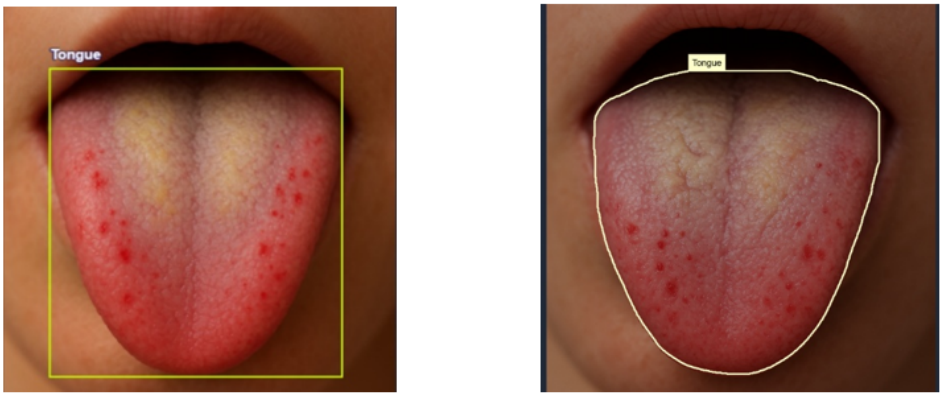


FIGURE V.8 – La différence entre la Bbox et le Polygone, (a):une langue annotée avec une Bbox, (b):une langue annotée avec un polygone.

Cette hybridation d'approche nous permet donc de combiner rapidité et précision, en fonction des besoins de chaque phase du projet.

En exportant le dataset a de roboflow, un choix de formats est proposés.nous avons-donc choisit les annotations 'COCO segmentation' pour la segmentation avec SAM, qui est format est très utilisé avec les modèles de segmentations. Et pour la détection de la langue dans une image,nous avonschoisis le format 'YOLOv8' .

V.4.2 Etape2 : Segmentation de la langue avec SAM

SAM : Segment Anything Model [2], le modèle le plus utilisée par la communauté scientifique pour la segmentation des objets, développé par MetaAI et entraîné sur un très grand jeu de données, est un modèle de segmentation dit "promptable" : on lui fournit un point, une boîte ou un masque, et il génère un masque précis sans apprentissage supplémentaire (avec un zero-shot). La version SAM2 étend ces capacités à la vidéo, grâce à un mécanisme de mémoire spatio-temporelle qui suit les objets dans le temps, gère les occlusions et permet une segmentation fluide sur plusieurs frames.[2]

Nous avons projeté SAM pour éliminer l'arrière-plan et ne laisser que la langue visible sur un fond noir: à l' aide des annotations COCO extraite de Roboflow auparavant, SAM segmente automatiquement son contour et supprime tout le reste (on l' appelle aussi Background removal). La figure suivante illustre comment SAM arrive à segmenter l' image correctement.

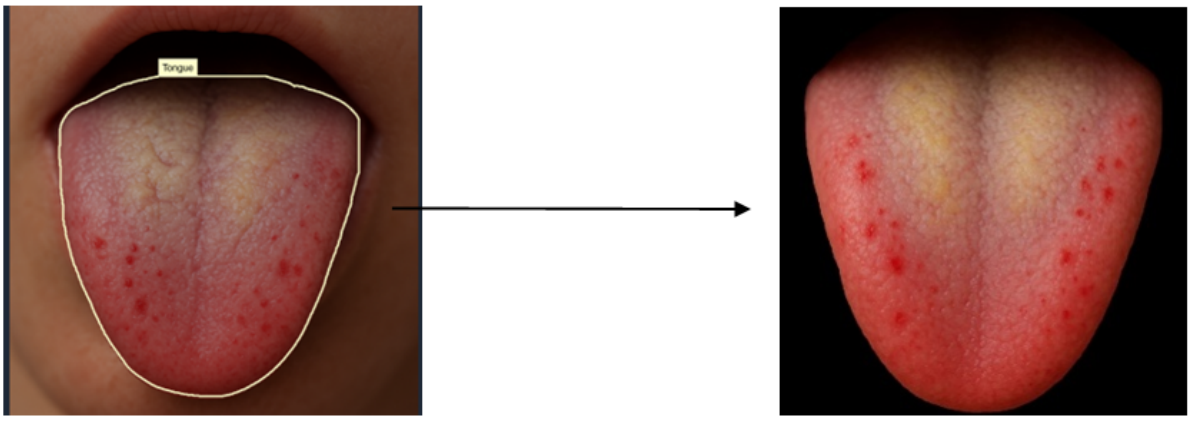


FIGURE V.9 – La figure illustre comment SAM segmente l'image (a) à l'image (b).

Nous avons également constaté que SAM n' était pas capable de segmenter correctement la langue de manière autonome. Sans indication préalable comme une Bbox ou un polygone, le modèle échoue à identifier la zone pertinente. C' est pourquoi nous avons dû compléter l' étape '1' avec des annotations manuelles via Roboflow, en traçant les BBox ou polygones pour guider la segmentation.

La figure V.10 illustre quelques exemples d' échecs du modèle SAM lorsqu' il tente de segmenter la langue sans aucune aide visuelle ou annotation préalable.



FIGURE V.10 – Quelques exemples d' échecs du modèle SAM à segmenter la langue.

La sortie de ce modèle est une image de la langue seule sans arrière-plan, avec dé-mentions réelles, et format d'image 640.

V.4.3 Etape 3 : Annotation manuelle des caractéristiques de la langue avec RoboFlow

Il n' existe à ce jour aucun modèle préentraîné spécifiquement dédié à l' annotation fine des caractéristiques pathologiques de la langue, notamment dans un contexte de diagnostic basé sur la MTC. Cette absence de modèle adapté, nous a poussées à réaliser manuellement l' ensemble des annotations à l' aide de la plateforme Roboflow. On s' est principalement basée sur les caractéristiques synthétisée et présentée dans le Tableau de la synthèse IV.4.3.

Ce travail minutieux mais essentiel consiste à examiner une à une les images du dataset, afin de délimiter manuellement les différentes caractéristiques visuelles de la langue à l' aide des outils de Roboflow. Nous avons utilisé principalement :

Ce travail minutieux mais essentiel consiste à examiner une à une les images du dataset, afin de délimiter manuellement les différentes caractéristiques visuelles de la langue à l' aide des outils de Roboflow. Nous avons utilisé principalement :

- **Les polygones**, pour annoter les éléments localisables avec précision, comme :

  Les enduits jaunes (épais ou minces), les enduits blancs (épais ou minces), les ecchymoses, les red_dot (points rouges), ainsi que l' état de la salive (normale ou humide).

- **Les Bboxes**, pour des formes plus générales ou diffusées telles que :

  Les fissures, les formes de la langue (normale, petite, ganflée), les différentes teintes de langue (pâle, rose, rouge, rouge foncée)

Ces annotations sont illustrées dans la Figure V.12 et la Figure V.11, qui présentent les 16 classes définies dans notre jeu de données, ainsi que leur fréquence d' apparition. Ce travail de repérage précis permet de fournir au modèle une base d' apprentissage fiable et structurée, essentielle pour la suite du processus de détection automatique.

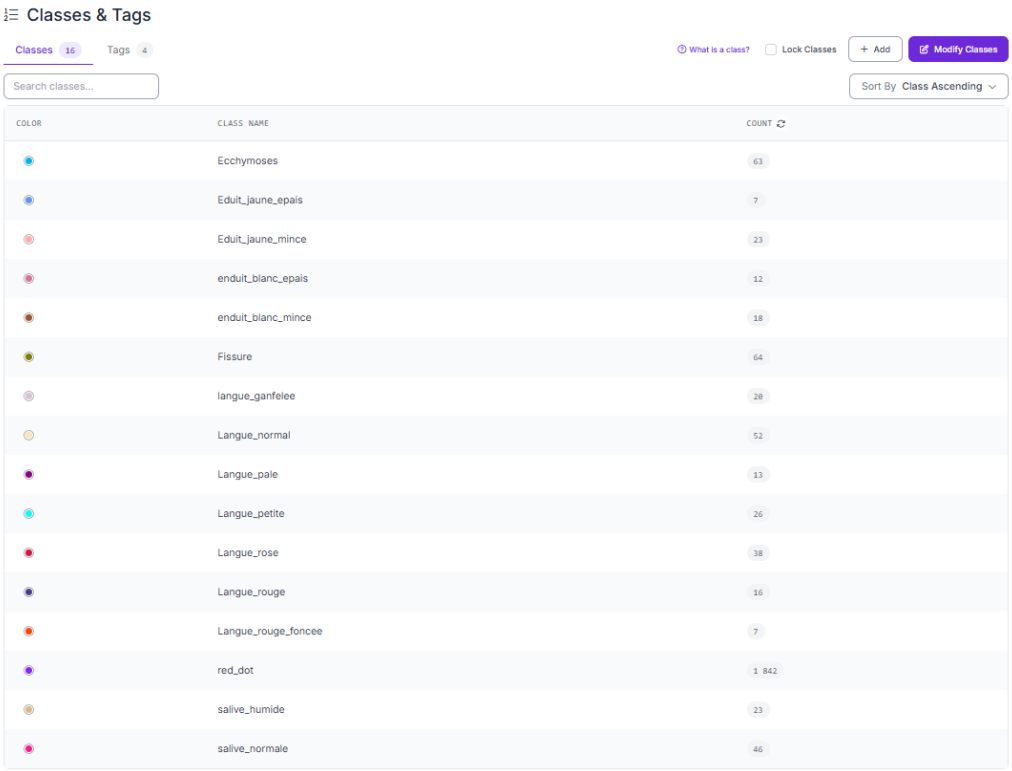


FIGURE V.11 – les 16 classes définies dans notre jeu de données, et leur fréquence d'apparition

La FigureV.12 illustre l'interface de Roboflow en cours de réalisation d'une annotation détaillée sur une langue d' une patiente atteinte du cancer du sein : chaque polygone entoure une zone marquée, et le nom de la classe associée est inscrit à proximité. On peut observer par exemple des zones de "red_dot", des "ecchymoses", ou encore un enduit jaune mince, clairement identifiés et localisés.

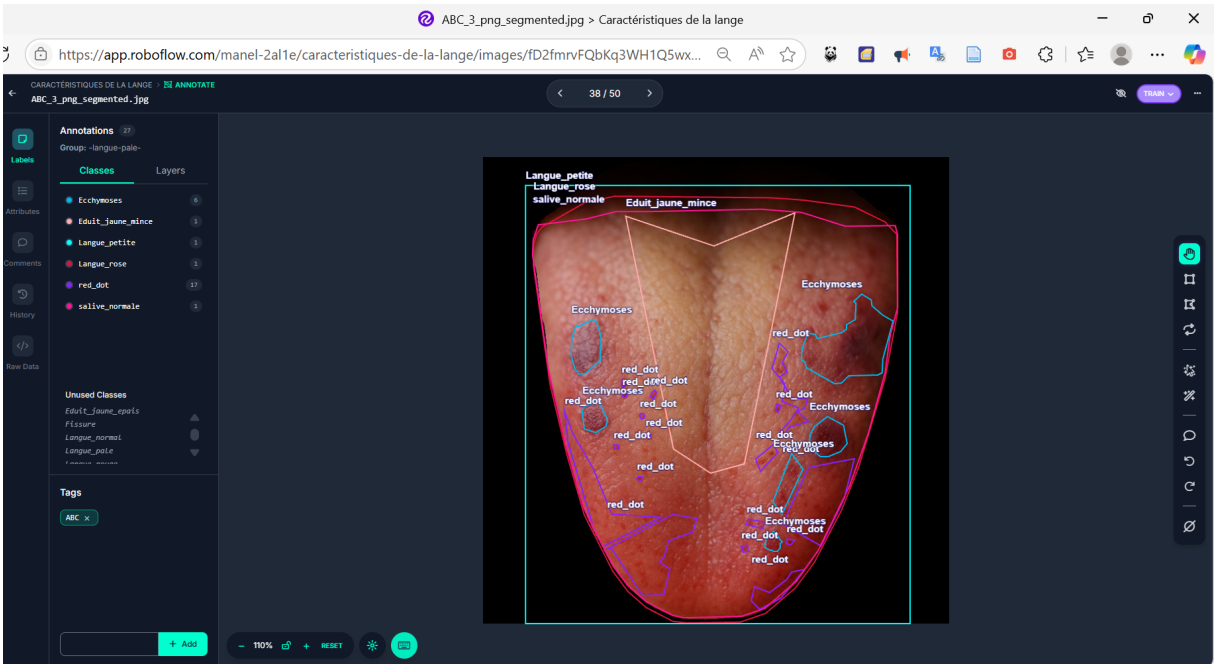
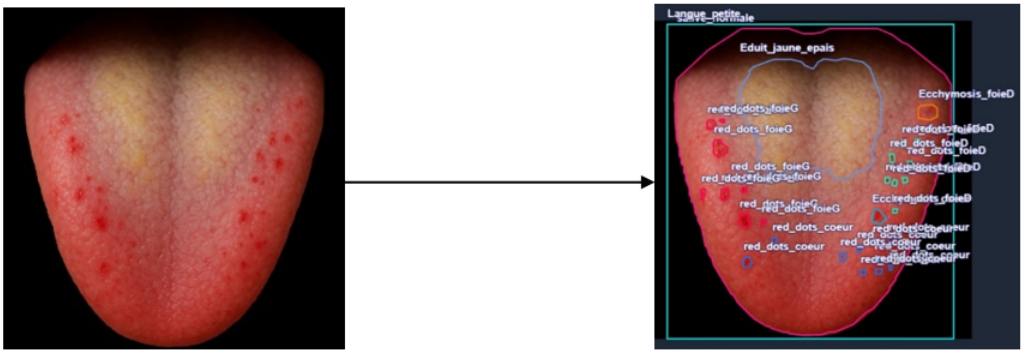


FIGURE V.12 – Interface de Roboflow lors de l'annotation des caractéristiques.

Cette étape d'annotation est indispensable pour constituer un dataset d'apprentissage supervisé. Une fois annoté, ce jeu de données a été utilisé pour entraîner un modèle de détection sur Roboflow. Toutefois, cette méthode n' est pas totalement fiable, car la précision de l' apprentissage dépend fortement de la qualité des annotations initiales, du nombre d' images, et de la diversité des cas représentés. La figure V.13 présente un exemple d' une transition directe d' une image brute de langue (a) et sa version annotée (b) dans Roboflow.



(a) (b)

FIGURE V.13 – Transition directe d' une image brute de langue (a) et sa version annotée (b) dans Roboflow.

Cette transition montre à quel point l' annotation manuelle permet d' enrichir les images en informations médicalement exploitables. Chaque zone colorée ou altérée visuellement devient une source de connaissance pour entraîner un modèle.
Malgré les limites rencontrées lors de ce travail, cette approche nous a permis de poser les bases solides d' un système intelligent de détection des signes cliniques linguaux du cancer de sein dans la MTC, en attendant l' enrichissement futur du dataset et l' implémentation de modèles plus robustes

V.5 Module de détection de la langue pendant le diagnostique automatisée

Ce module a pour objectif de détecter la langue dans une image pendant le diagnostic, dans le but de prétraiter la langue avant l'extraction des caractéristiques et le diagnostic (Figure V.14)

Pour détecter automatiquement la langue dans les images, nous avons entraîné un modèle YOLOv8 Small en version segmentation d' instance (yolov8s-seg.pt). Ce choix permet d' obtenir une délimitation précise du contour de la langue, grâce à des masques.

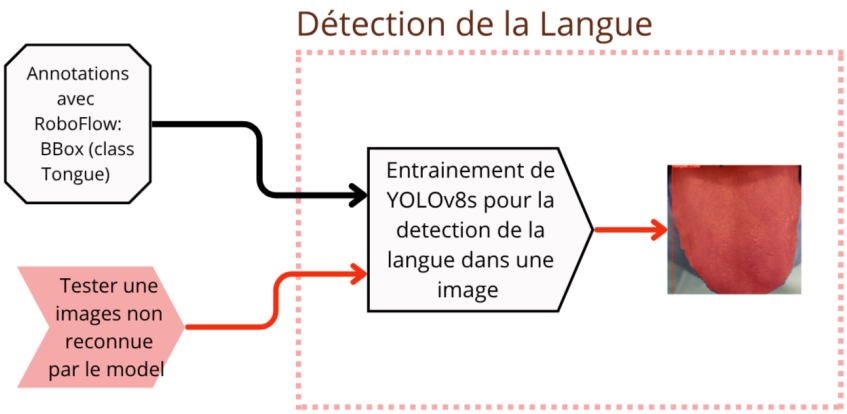


FIGURE V.14 – Exemples issus du dataset Tongue_MTCBC_Hope

Dans l'architecture YOLOv8, deux types de modèles sont utilisées.

- **YOLOv8 (sans suffixe)** : utilisé pour la détection classique d'objets à l'aide de bounding boxes uniquement (exemple de modèle : `yolov8s.pt`).
- **YOLOv8-seg** : destiné à la segmentation d'instance, combinant bounding boxes et génération de masques précis pour chaque objet (exemple : `yolov8s-seg.pt`).

V.6 Module Détection des caractéristiques pathologiques et diagnostic précoce du cancer du sein

Notre système de diagnostic repose sur une approche hybride combinant deux composantes clés, la détection automatique des signes pathologiques visibles sur la langue à l'aide du modèle YOLOv11, et la classification du stade clinique (sain, précoce, avancé) à l'aide d'algorithmes de machine learning supervisés. L'ensemble est enrichi par une cartographie MTC permettant de contextualiser chaque anomalie en fonction de sa localisation sur la langue.

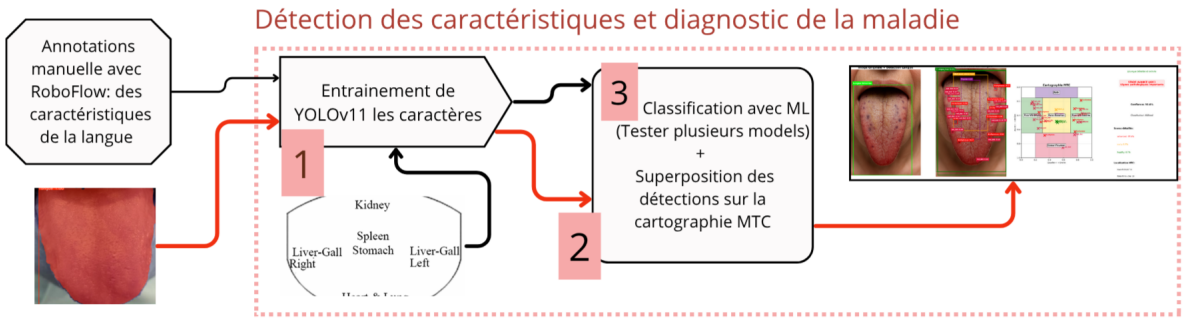


FIGURE V.15 – Le module détection des caractéristiques et diagnostic

V.6.1 Étape 1 : Détection des caractéristiques de la langue avec YOLOv11

La première étape consiste à détecter automatiquement les signes cliniques à partir d'une image de langue. Pour cela, nous avons entraîné un modèle YOLOv11s sur notre dataset annoté. Le modèle a été calibré pour détecter 16 classes caractéristiques spécifiques, définies selon les critères cliniques issus de la MTC.

Notre choix s'est appuyé sur une comparaison approfondie des performances des différentes versions de YOLO. Comme l'illustre la Figure III.8, qui compare les modèles en termes de mAP (Mean Average Precision) et de latence, YOLOv11s atteint une mAP50-95 de 47.0%, avec une très faible latence.

Cela le rend idéal pour les applications embarquées ou en temps réel, même sur des machines avec ressources limitées (comme un ordinateur portable ou un Raspberry Pi avec accélération matérielle). L'un de ses atouts majeurs est sa capacité à être entraîné sur CPU, ce qui n'est pas le cas des modèles plus volumineux.

Une augmentation par un facteur de $\times 4$ ou $\times 6$ s'est révélée insuffisante pour garantir une robustesse satisfaisante du modèle. C'est précisément ce constat qui nous a poussés à explorer davantage de variations d'augmentation, afin d'enrichir significativement le dataset et améliorer la capacité de généralisation du modèle

V.6.1.1 Augmentation des données

Afin de pallier le nombre limité d'images annotées dans notre dataset initial et d'améliorer la généralisation du modèle YOLOv11s, nous avons mis en place une stratégie d'augmentation avancée des données. Le facteur d'augmentation choisi est de $\times 8$, ce qui signifie que chaque image originale est utilisée pour générer 8 variantes augmentées, portant ainsi le volume global du dataset à plus de 8 fois sa taille initiale. Voici les transformations appliquées :

- **Rotation + flip horizontal** : rotation de 90 degrés dans le sens horaire, suivie d'un retournement horizontal de l'image.
- **Flip horizontal simple** : symétrie gauche/droite sans modification géométrique, utile pour augmenter la diversité tout en conservant la structure.
- **Ajustement de la luminosité et du contraste** : variation contrôlée des niveaux de lumière pour simuler des conditions d'éclairage différentes.
- **Rotation légère + saturation** : rotation aléatoire dans l'intervalle $[-15°, +15°]$, combinée à un ajustement des couleurs dans l'espace HSV (teinte, saturation, valeur).
- **Flou gaussien + bruit aléatoire** : ajout de flou et de bruit pour simuler des imperfections de prise de vue ou de capteurs.
- **Déformation de perspective** : modification angulaire simulant une image prise avec un angle incliné, altérant les proportions pour plus de réalisme.

- **Correction gamma** : ajustement de la courbe de luminance afin d'imiter les variations de capteurs ou de dispositifs d'acquisition.
- **Combinaison complexe** : enchaînement de plusieurs transformations (rotation, variation HSV, changements de teinte et de couleur) pour enrichir fortement le dataset.

En plus des transformations classiques (rotation, contraste, etc.), nous avons utilisé trois techniques spécifiques à YOLO lors de l'entraînement pour améliorer la qualité du modèle :

- **Mosaic augmentation** : assemble quatre images différentes en une seule. Chaque quadrant de l'image résultante contient une portion d'image originale, ce qui permet au modèle d'observer plusieurs objets dans des contextes variés. Cette technique améliore la capacité de généralisation, notamment pour les objets de tailles et de positions différentes.
- **MixUp** : fusionne deux images et leurs annotations en combinant leurs pixels. Cela génère des exemples flous ou hybrides, réduisant le risque de surapprentissage en forçant le modèle à apprendre des représentations plus robustes et généralisables.
- **Copy-Paste** : extrait un objet (ex. une zone contenant des taches rouges ou des fissures) d'une image et l'intègre dans une autre. Cette méthode permet de créer des combinaisons réalistes sans nécessiter une réannotation manuelle complète.

La figure V.16 illustre une explication comment ces techniques fonctionnent pour mieux comprendre :

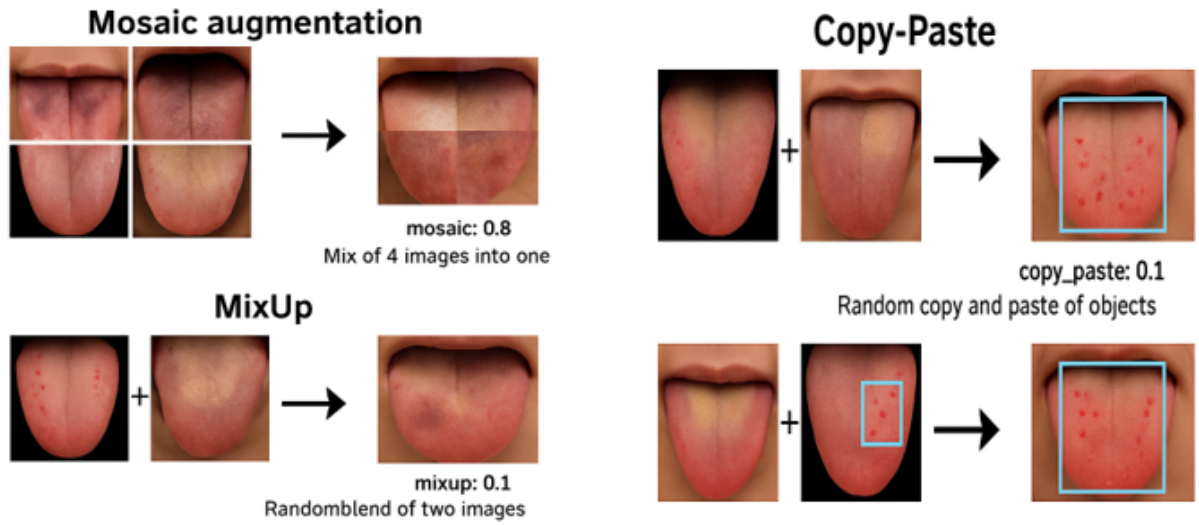


FIGURE V.16 – Exemples de transformation avancée utilisée pour le modèle.

Chaque transformation est soigneusement appliquée de manière contrôlée, avec recalcul des coordonnées pour que les fichiers .txt de labels YOLO restent cohérents. Ainsi, les annotations restent valides même après distorsion de l'image, ce qui garantit un apprentissage stable. Ce processus permet d'introduire une grande diversité dans les données

sans avoir à collecter de nouvelles images, tout en renforçant la robustesse du modèle face aux variations réelles telles que les changements d' angle, de luminosité ou de bruit. Il contribue également à limiter le risque de surapprentissage, en évitant que le modèle se spécialise trop sur des cas trop spécifiques. Grâce à cette stratégie d' augmentation avancée, YOLOv11s a été exposé à un large éventail de situations réalistes, ce qui a favorisé une meilleure stabilité durant l' apprentissage et permis d' obtenir des performances globales solides, sans dépendre d' un grand volume de données réelles. Finalement, YOLOv11s s' est imposé comme un modèle léger, rapide et suffisamment précis, parfaitement adapté à la tâche ciblée de détection automatique des signes pathologiques sur la langue.

V.6.2 Étape 2 : Simulation de la cartographie avec les caractéristiques

Au-delà de la simple détection, chaque anomalie identifiée est projetée sur une cartographie de la langue dans la MTC (présentée dans la figure Figure IV.15). Cette étape permet de dépasser la logique "langue + pathologie", en offrant une lecture organique et fonctionnelle des signes observés.

Chaque détection est localisée à partir de ses coordonnées dans l' image, puis automatiquement rattachée à l' une des zones fonctionnelles de la langue définies par la cartographie MTC, à l' aide d' une fonction de projection géométrique intégrée dans le système. Cette étape permet de relier les signes visibles à des organes internes précis, comme par exemple des ecchymoses concentrées dans la zone du Cœur, ou un enduit jaune apparaissant dans la zone de la rate-estomac.

Cette contextualisation est essentielle dans le raisonnement médical selon la MTC, car elle transforme une simple observation visuelle en hypothèse diagnostique ciblée sur les organes internes.

Exemple : 'lorsqu'une ecchymose est localisée dans la zone correspondant au foie, cela peut suggérer un déséquilibre ou une atteinte du foie lui-même. La visualisation devient alors un outil puissant, non seulement pour identifier les caractéristiques visibles, mais aussi pour simuler les organes potentiellement concernés, selon la cartographie énergétique propre à la MTC.'

V.6.3 Étape 3 : Classification avec les algorithmes de machine learning

Après l' extraction des caractéristiques pathologiques et leur projection sur la cartographie MTC, chaque image est traduite en un vecteur numérique résumant les signes visibles : type d'anomalies, nombre d'occurrences, zones concernées, niveaux de confiance, etc.
Ce vecteur est transmis à plusieurs classificateurs supervisés, dont l'objectif est de prédire automatiquement le stade clinique du patient : sain, stade précoce, ou stade avancé.
Pour cela, nous avons mis en uvre et comparé les performances de plusieurs algorithmes d' apprentissage machine, parmi les plus couramment utilisés dans la littérature :

- **XGBoost** : modèle d' ensemble très performant basé sur des arbres de décision

optimisés. Il s' est révélé particulièrement adapté aux données tabulaires issues des détections YOLO.

- **Random Forest** : ensemble d' arbres construits de manière aléatoire, robuste face au bruit et facile à interpréter. Idéal pour des prédictions fiables sans réglages complexes.
- **SVM (Support Vector Machine)** : algorithme efficace pour séparer les classes avec des frontières nettes. Particulièrement utile sur des datasets de taille moyenne et bien étiquetés.
- **MLP (Multi-Layer Perceptron)** : réseau de neurones à couches multiples, capable de modéliser des relations non linéaires complexes entre les signes cliniques détectés.
- **Gradient Boosting** : méthode d' apprentissage ensembliste proche de XGBoost, qui corrige progressivement les erreurs des modèles précédents pour améliorer la précision finale.

L'entraînement de ces modèles a été automatisé via une cross-validation en 5 plis [6], afin de garantir la fiabilité des scores. Ces modèles ont été comparés selon leur précision, stabilité et temps d'entraînement afin de sélectionner le plus adapté à notre tâche de diagnostic. Le meilleur classificateur est sélectionné selon la moyenne de précision obtenue. Le diagnostic est restitué sous forme visuelle avec :

- Le score de confiance associé à chaque caractéristique détectée sur la langue,
- La cartographie des zones MTC, annotée avec les signes cliniques identifiés dans chaque région (Foie, Rate, Cœur, etc.),
- Le niveau de confiance pour chaque stade (sain, précoce, avancé),
- La classe finale prédite pour le patient.

Cette hybridation de diagnostic médical automatisée, combinant détection d' objet et algorithmes d' apprentissage automatique de la classification, permet d' obtenir un diagnostic interprétable, cohérent avec la logique MTC, et applicable à grande échelle pour un dépistage intelligent. Il ouvre la voie à une analyse combinée entre médecine traditionnelle et intelligence artificielle, tout en mettant en valeur les données visuelles cliniques du patient.

V.7 Modèle étendu

Ce modèle propose une architecture modulaire orientée vers l' automatisation du diagnostic par la langue, selon les principes de la Médecine Traditionnelle Chinoise (MTC), tout en intégrant des avancées en intelligence artificielle. Il repose sur trois composantes

6. **Cross-Validation** : une technique d' évaluation robuste utilisée en ML pour estimer les performances d' un modèle.

fonctionnelles : la création automatique de données cliniques structurées, la détection des symptômes et zones affectées, et la génération de recommandations personnalisées via un agent conversationnel intelligent.

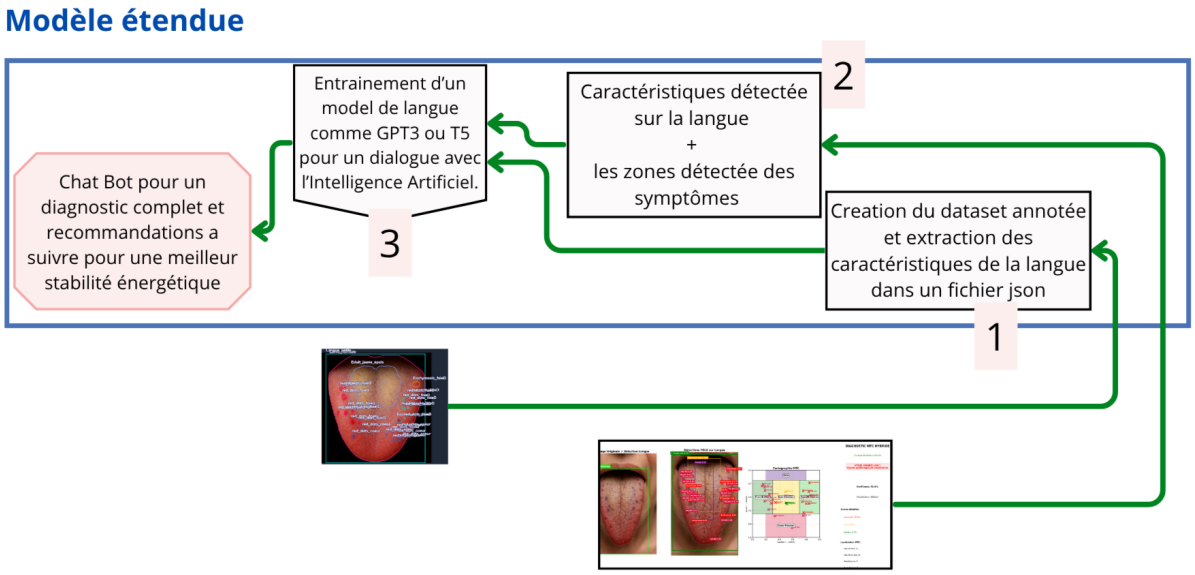


FIGURE V.17 – Modèle étendue

V.7.1 Étape 1 : Création du dataset structuré

Cette étape repose sur une extraction automatisée des différentes caractéristiques de la langue avec les syndromes MTC associée. Chaque image est analysée afin d' en extraire les attributs visuels pertinents. Ces attributs sont encodés dans un format structuré JSON, facilitant leur exploitation dans les phases ultérieures. Ce dataset enrichi constitue la base de connaissances du système.

V.7.2 Étape 2 : Détection des zones symptomatiques et interprétation clinique

La cartographie énergétique permet de relier chaque caractéristique détectée à une zone spécifique de la langue. Par exemple, une fissure médiane peut être associée à un déséquilibre du Cœur. Ce couplage repose sur un référentiel codé dans le système, inspiré de la littérature MTC. Les zones affectées sont ensuite visualisées sur une image annotée pour faciliter l'interprétation.

V.7.3 Étape 3 :Dialogue médical personnalisé via modèle de langage

Le modèle de langage exploite les résultats des étapes précédentes pour formuler un retour intelligent. Ce retour prend la forme d' un chatbot médical, capable d' engager un dialogue orienté vers l' explication des symptômes, la contextualisation énergétique (vide,

excès, stagnation⋯) et la proposition de pistes thérapeutiques. L' agent conversationnel s' adapte aux réponses de l' utilisateur et peut affiner son analyse.

Ce modèle étendu rend possible une médecine personnalisée, non invasive et interactive, combinant expertise traditionnelle et capacités modernes d' analyse. Il constitue une passerelle entre le raisonnement clinique MTC et les technologies d' IA conversationnelles.

V.8 Construction d' une plateforme de diagnostic du cancer du sein selon MTC

Afin de rendre accessible et opérationnel le système de détection des signes pathologiques de la langue selon la MTC, nous avons conçu et déployé une plateforme interactive nommée **MTC Hybrid Pro**. Celle-ci intègre l' ensemble de notre architecture de traitement et s' adresse principalement aux professionnels de la santé.

V.8.1 Besoins Fonctionnels

Dans cette section, nous présentons de manière détaillée les acteurs impliqués dans le système ainsi que les fonctionnalités principales qu'ils peuvent exercer.

Acteurs	**Fonctionnalité**	**Description**
Professionnel de santé	Téléverser une image de langue	Upload d'image de langue dans le système pour analyse
	lancer le diagnostic	lancement d'analyse diagnostique de l'image téléversées
	Consulter les résultats	Visualisation des résultats diagnostiques et détections pathologiques
	Consulter les statistiques	Accès aux statistiques détaillées et recommandations
	Consulter la cartographie MTC	Visualisation de la cartographie MTC interactive par zones
	Exporter le rapport	Export des résultats en format PDF
Administrateur Système	Gérer les utilisateurs	Gestion des comptes, rôles et droits d'accès
	Administrer le système	Configuration, Mise à jour, maintenance et consultation des statistiques globales

TABLE V.1 – Tableau des Besoins Fonctionnels - Système MTC Diagnostic Pro

V.8.2 Besoins Non-Fonctionnels

En complément des besoins fonctionnels précédemment décrits, les besoins non fonctionnels mesurent les performances qualitatives et quantitatives du système. En mettant

l' accent sur des critères de performance, de fiabilité, d' ergonomie. Ces exigences permettent de garantir que le système offre une expérience utilisateur satisfaisante.

Catégorie	Sous-critère	Exigence Détaillée	Métriques Cibles
Performance	Temps de traitement	Pipeline complet d'analyse : détection YOLO + extraction features + classification ML + affichage résultats	< 3 secondes
	Throughput	Capacité de traitement simultané pour usage clinique en environnement hospitalier	100 diagnostics/-heure
	Utilisation mémoire	Consommation RAM optimisée pour environnement standard (modèles YOLO + ML en mémoire)	< 2 GB
Disponibilité	Infrastructure	Hébergement sur plateforme cloud sécurisée	Multi-zone deployment
Fiabilité	Validation uploads	Contrôle automatique format, taille et intégrité des images téléchargées	Format whitelist + validation < 1s
Utilisabilité	Facilité d'usage	Interface intuitive pour professionnels de santé sans formation technique préalable	Zéro formation requise
	Responsive design	Adaptation automatique tablettes/ordinateurs pour usage en cabinet médical	Mobile + desktop

TABLE V.2 – Analyse des Besoins Non-Fonctionnels - Système MTC Diagnostic Pro

V.8.3 Developpement agile et Conception UML

Pour la réalisation de cette application, une méthode agile a été adoptée, malgré le fait que le développement ait été mené en binôme (pas de groupe). Afin de pallier la distance géographique entre les membres de l' équipe, nous avons mis en place une organisation inspirée de la méthode Scrum : réunions de coordination régulières en visioconférence, répartition claire des tâches à travers des sprints hebdomadaires, revue de code partagée et validation continue des fonctionnalités.

Afin de structurer et modéliser le fonctionnement de la plateforme, plusieurs diagrammes UML ont été élaborés.
Dans l'Annexe A.**la figure A.1** présente le diagramme de cas d'utilisation, qui décrit les principales interactions entre l' utilisateur et le système (ajout d'image, diagnostic, téléchargement de rapport, etc.). Et en Annexe A.
La Figure A.2 on trouve le diagramme de séquence, qui montre les interactions entre les différents objets intervenant dans le diagnostic de la langue.

La Figure A.3 présente le diagramme de classes. Celui-ci permet de visualiser la structure interne du système, en modélisant les classe impliquées dans le diagnostic de langue, leurs attributs, ainsi que les relations entre elles —de l'envoi de l'image jusqu'à l'affichage du résultat final.

V.8.4 Architecture modulaire de la platform

La plateforme repose sur une architecture modulaire structurée en quatre phases principales présenté dans la figure V.18, qui correspondent au flux complet de traitement des données depuis le téléversement jusqu'à l'affichage du diagnostic final. Ce flux, représenté en rouge dans la figure d'architecture YOLO_MTCBC V.4, est détaillé comme suit :

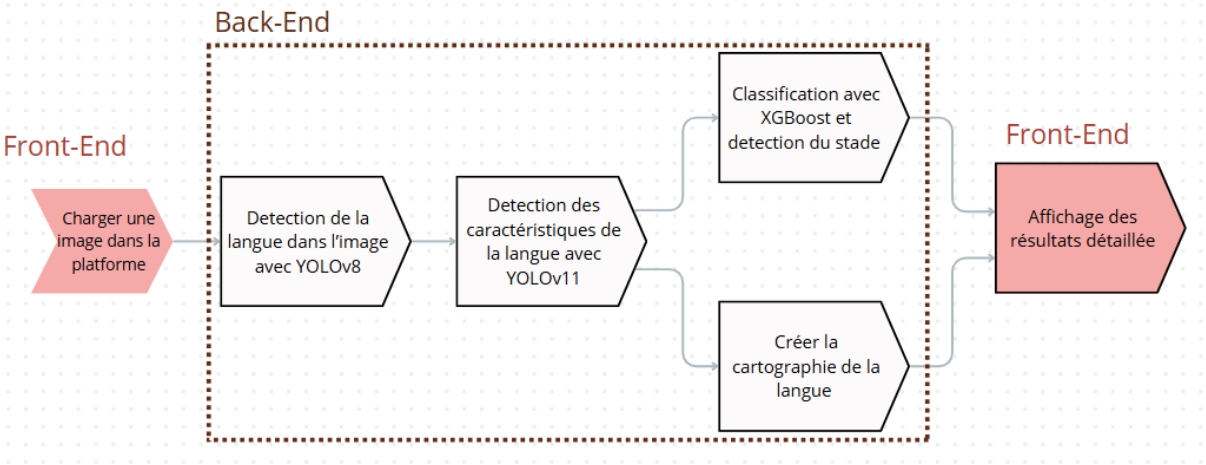


FIGURE V.18 – Schématisation MTC Hybrid Pro

V.8.4.1 Phase 1 –Téléversement et détection de la langue

- L'utilisateur téléverse une image de langue via l'interface (insertion ou glisser-déposer).
- L'image est transmise au modèle YOLOv8, entraîné pour détecter la langue.
- Une boîte englobante est générée autour de la langue détectée.
- La langue est extraite, redimensionnée, puis transmise à l'étape suivante.

V.8.4.2 Phase 2 –Détection des caractéristiques pathologiques

- L'image recadrée est analysée par YOLOv11, spécialisé dans la détection des signes pathologiques : fissures, ecchymoses, points rouges, enduits, anomalies de salive, etc.
- Les détections sont converties en un vecteur de caractéristiques.

V.8.4.3 Phase 3 –Prédiction du stade clinique

- Le vecteur de caractéristiques est transmis à un classificateur (SVM, XGBoost, etc.).
- Le modèle prédit le stade clinique : sain, précoce ou avancé.
- Les résultats sont projetés sur une carte des zones MTC (Foie, Rate, Cœur⋯).

V.8.4.4 Phase 4 –Restitution du diagnostic

- L' image est annotée avec les zones pathologiques détectées.
- Le stade clinique est affiché avec un code couleur : vert (sain), orange (précoce), rouge (avancé).
- Les scores de prédiction sont présentés en pourcentages pour chaque stade.

Intérêt médical de la plateforme:Elle s' inscrit dans une logique de médecine préventive intelligente, alliant les principes de la Médecine Traditionnelle Chinoise aux technologies de l' intelligence artificielle.

V.9 Conclusion

Dans ce chapitre nous avons présentées la conception et la mise en uvre d' un système intelligent de diagnostic précoce du cancer du sein basé sur l' analyse de la langue, en s' appuyant sur les fondements de la MTC. Construit en réponse à la rareté des données cliniques réelles et à l'absence d' outils préexistants, notre système repose sur une architecture hybride articulée autour de plusieurs innovations majeures.

Tout d' abord, un jeu de données réaliste et cohérent a été généré à l' aide de techniques d' intelligence artificielle (IA), puis annoté manuellement avec rigueur sur Roboflow. Cette base, nommée Tongue_MTCBC_Hope, constitue un socle essentiel pour entraîner et valider nos modèles.
Dans ce travail, nous avons détaillé l' architecture YOLO_BCTCM, articulée autour de quatre modules.Comme la détection automatique de la langue à l' aide de YOLOv8, et l' entraînement de YOLOv11 pour détecter les caractéristiques de la langue, selon les critères de la MTC.
Un vecteur de caractéristiques est généré pour alimenter un classificateur chargé d' évaluer le stade clinique (sain, précoce, ou avancé). Plusieurs algorithmes d' apprentissage automatique ont été testés et comparés (SVM, Random Forest, XGBoost, MLP et Gradient Boosting).
Enfin, l' ensemble de cette architecture a été intégré dans une plateforme numérique, MTC Hybrid Pro, pensée pour simplifier l' usage de ce système auprès des praticiens, des chercheurs, et potentiellement du grand public.

Dans le chapitre suivant, nous présenterons la suite des étapes (9 et 10) de notre système IV.3.3, avec pour objectif de finaliser le passage du prototype à un outil exploitable en pratique.

Chapitre VI

Implémentation et tests

Sommaire

VI.1 Introduction **131**

VI.2 Entraînement de YOLOv8 pour la détection de la langue **131**

VI.3 Entraînement de YOLOv11 pour la détection des caractéristiques linguales **135**

VI.4 Adaptation du modèle YOLOv11 préentraîné via réentraînement . . . **140**

VI.5 Système de diagnostic précoce du cancer du sein **145**

VI.6 Plateforme interactive pour le diagnostic **151**

VI.7 Conclusion **154**

Résumé : *Ce chapitre met à l'épreuve l'architecture présentée précédemment en analysant les performances des modèles YOLOv8 et YOLOv11, avant et après réentraînement. Les résultats expérimentaux montrent une nette amélioration des scores (mAP, précision, rappel), validant l'approche hybride combinant détection par vision artificielle et classification par apprentissage automatique. Le modèle XGBoost s'est distingué comme le classificateur le plus performant pour prédire les stades cliniques. L'intégration finale dans une plateforme interactive confirme la faisabilité technique et clinique du système proposé.*

VI.1 Introduction

Après avoir conçu et expérimenté une architecture complète combinant vision par ordinateur, médecine traditionnelle chinoise et classification supervisée, il est désormais temps de passer à l'étape suivante : la mise à l'épreuve du système en conditions réelles. Dans ce chapitre, nous abordons les trois (03)ndernières étapes clés du pipeline de notre système de diagnostic intelligent (IV.3.3). Il s'agit tout d'abord d'entraîner et valider finement les modèles de l'architecture, en exploitant les métriques d'évaluation : la mAP, précision, recall et le F1-score. Nous terminons ce chapitre avec la présentation de la plateforme de diagnostic du cancer du sein qui intègre le système d'aide au diagnostic implémenté et testé.

VI.2 Entraînement de YOLOv8 pour la détection de la langue

Le module de détection de la langue repose sur YOLOv8. Le modèle `YOLOv8s-seg` a été chargé et entraîné sur un dataset de 130 images.

Ce dataset est composé de 50 images de langue annotées, réparties en **80%** pour l'entraînement, **10%** pour la validation et **10%** pour les tests. Il a été augmenté trois fois à l'aide de la plateforme Roboflow. Les transformations d'augmentation appliquées sont les suivantes :

- Retournements horizontaux et verticaux pour simuler des prises de vue inversées.
- Rotations à 90° (gauche et droite), ainsi qu'entre −15° et +15° pour reproduire des angles plus naturels.
- Ajout de bruit aléatoire : jusqu'à **0,1 %** des pixels.

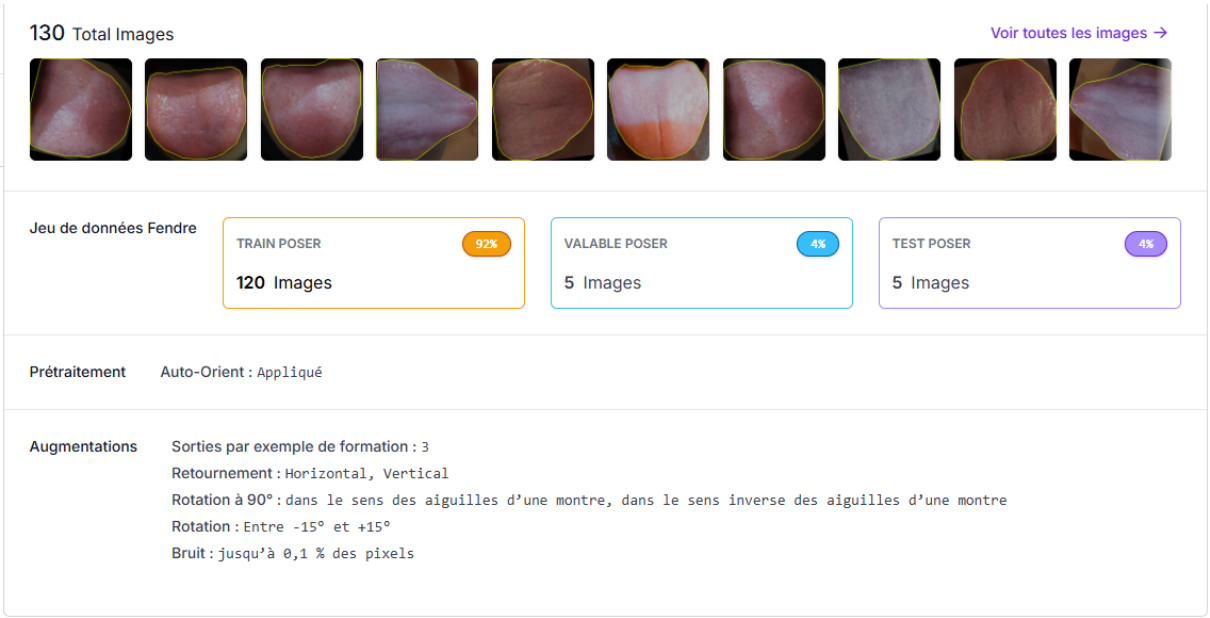


FIGURE VI.1 – Prétraitement du dataset de la détection de la langue sur Roboflow.

VI.2.1 Environement de développement

Le développement a été réalisé sur une machine Dell Latitude 7390, équipée d'un processeur Intel Core i7 vPro 8 génération et de 16 Go de RAM. L'environnement de programmation utilisé est Jupyter Lab, exécuté sous Windows 11 via le sous-système Linux (WSL). Nous avons également utilisé la version gratuite de Google Colab, avec un GPU Tesla T4 doté de 16 Go de mémoire GDDR6.

VI.2.2 Les hyperparamètres

- **20 epochs** : le modèle a été entraîné pendant 20 itérations complètes sur le jeu de données pour affiner ses prédictions.
- **Batch size de 16** : les images ont été traitées par lots de 16 pour équilibrer la mémoire RAM et la stabilité de l'apprentissage.
- **Format 640×640** : les images ont été redimensionnées à cette taille pour standardiser les entrées du modèle et garantir une bonne compatibilité avec YOLOv8.

VI.2.3 Évolution des pertes (loss) pendant l'entraînement

La figure VI.2 illustre l'évolution de chaque fonction de perte pendant l'entraînement de 20 epochs, tant que chaque fonction diminue avec les epochs, cela signifie que le modèle fait moins d'erreurs :

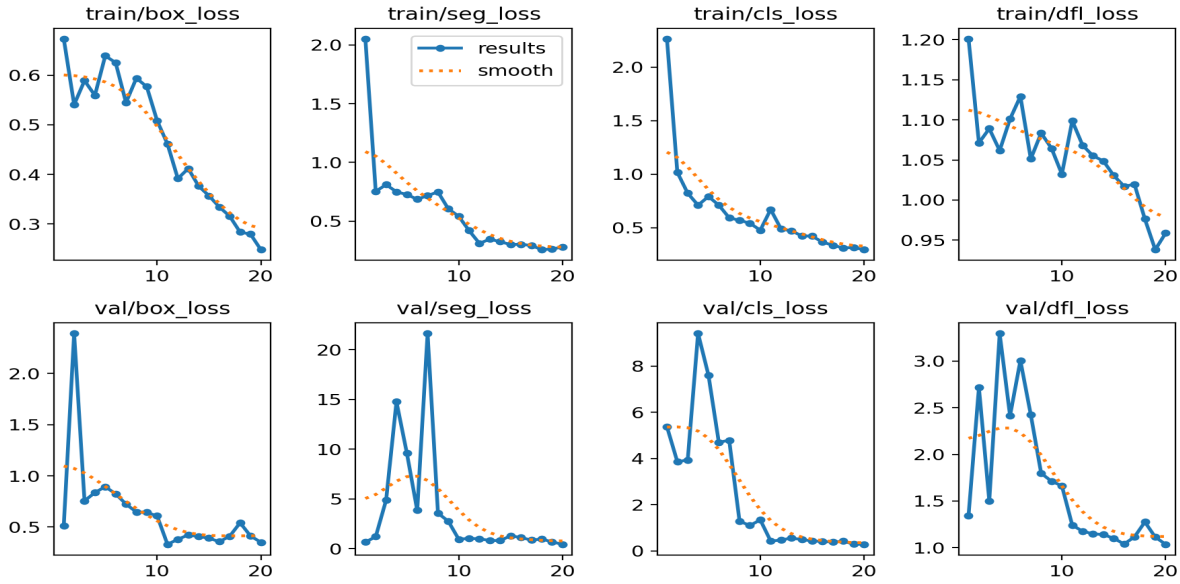


FIGURE VI.2 – Évolution de chaque fonction de perte par epoch YOLO8.

Les résultats des différentes fonctions de perte sont résumés dans le tableau ci-dessous, accompagnés d'une interprétation de chaque valeur afin d'évaluer la qualité d'apprentissage et de généralisation du modèle.

Métrique	Valeur finale	Interprétation
`train/box_loss`	≈ 0.26	Erreur de détection des boîtes : diminution régulière → bonne capacité à localiser les Bbox des langues.
`train/seg_loss`	≈ 0.22	Erreur de segmentation : perte chutant rapidement → précision élevée sur les zones contenant les langues.
`train/cls_loss`	≈ 0.30	Erreur de classification : stable et faible → les classes sont correctement apprises.
`train/dfl_loss`	≈ 0.94	Distance Focal Loss : qualité fine des boîtes. Diminue progressivement → les boîtes deviennent plus précises spatialement.
`val/box_loss`	≈ 0.50	Données de validation : instabilité initiale puis convergence → bonne généralisation.
`val/seg_loss`	≈ 0.30	Bonne performance sur la segmentation, même sur les images non vues. Le modèle généralise bien.
`val/cls_loss`	≈ 0.40	Aucune dérive observée → absence de surapprentissage.
`val/dfl_loss`	≈ 1.10	Légèrement plus élevée, mais reste dans une zone acceptable → prédictions encore fiables.

TABLE VI.1 – Analyse des pertes (loss) du modèle YOLOv8 sur les phases d'entraînement et de validation

VI.2.4 Évolution des métriques de performance pendant l'entraînement

On distingue ici les résultats obtenus pour la performance du modèle sur la détection (B : Bbox) et la segmentation de masque (M : Mask) pour yolo8s-seg.

Métrique	Valeur obtenue	Interprétation
Précision (B/M)	99.137 %	Taux de vrais positifs élevé : le modèle détecte bien la langue sans générer de fausses alertes.
Rappel / Recall (B/M)	100 %	Le modèle n'a manqué aucune instance d'objet (langue) à détecter.
mAP@50 (B/M)	100 %	Très bon score indiquant que, dès un seuil d' IoU $\geq$ 50 %, les prédictions sont très précises.
mAP@50:95 (B/M)	94.98 % / 94.49 %	Excellente précision même avec des seuils plus stricts (IoU de 0.5 à 0.95). Le modèle reste robuste sur des cas complexes et variés.

TABLE VI.2 – Valeurs finales des métriques de YOLOv8 avec leur interprétation.

Les résultats des métriques de performances des modèles de détection d' objet sont résumés dans le tableau VI.2, accompagnés d' une interprétation de chaque valeur afin d' évaluer la qualité d' apprentissage et de généralisation du modèle. La Figure VI.3 suivante illustre l'évolution de chaque métrique pendant l'entraînement de 20 epochs.

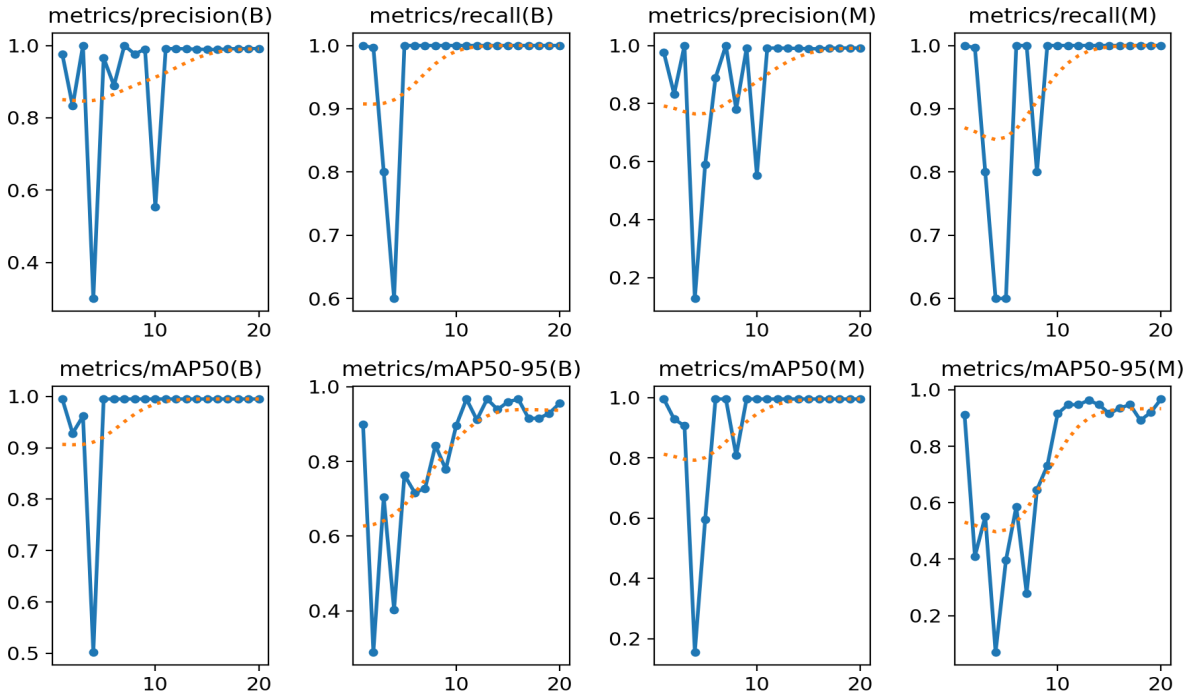


FIGURE VI.3 – Evolution des métriques de performance par epoch Yolo8.

La figure VI.4 ci dessous présente deux résultats de validation du modèle YOLOv8 : La matrice de confusion montre une précision de 100% pour la reconnaissance de la langue, sans erreur de classification (a). Les masques de segmentation sur les données de validation illustrent une détection précise et généralisable, confirmant la robustesse du modèle sur des images variées (b).

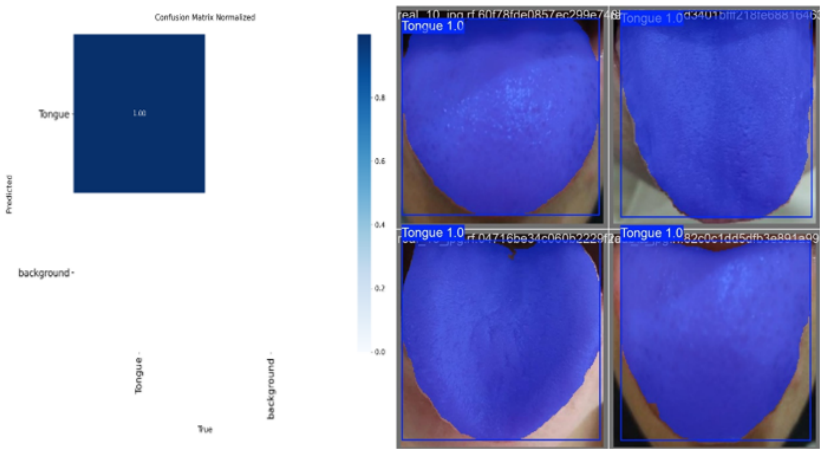


FIGURE VI.4 – Résultat de validation de YOLOv8. (a) présente la matrice de confusion, figure (b) présente les masque de segmentation sur les données de validation.

VI.3 Entraînement de YOLOv11 pour la détection des caractéristiques linguales

Le module Détection de la langue repose sur YOLOv11. Le modèle YOLOv11s a été chargé et entraîné sur notre dataset (MTC_BC_Hope), et en utilisnat le meme environnement de developpement que YOLOv8 (Google collab).

VI.3.1 Prétraitement automatisé du dataset

Une analyse intelligente du dataset a été réalisée, avec une répartition guidée par le nommage des fichiers.

Les images ont été automatiquement classées selon leur nom (par exemple : `ABC`, `EBC`, `healthy`, `real`). Les images nommées `real` ont été fusionnées avec la classe `healthy`, car elles correspondent à des langues de personnes saines.

Au total, 80 images ont été détectées : 38 images saines et 42 images présentant un cancer du sein, réparties entre trois classes :

- **Healthy** : 38 images
- **Early** : 23 images
- **Advanced** : 19 images

La répartition entre entraînement et validation a été effectuée automatiquement selon une proportion de **85% , 15%** :

- 67 images (85%) pour l'entraînement
- 13 images (15%) pour la validation : `Healthy` (6), `Early` (4), `Advanced` (3)

- **Création automatique du fichier `data.yaml`** : nécessaire pour l'entraînement avec YOLO, il contient les chemins des images et les classes.
- **Augmentation avancée des données (×8)** : comme détaillé dans le chapitre précédent, chaque image originale a généré 8 variantes (rotation, retournement, bruit, etc.), portant le total d' images d' entraînement à 603.

VI.3.2 Hyperparamètres d' entraînement YOLOv11 optimisé

Le tableau suivant présente les hyperparamètres d'apprentissage utilisée sur YOLO11 entrainé sur un GPU Tesla T4 (Google Colab).

L' hyperparamètre	La valeur mise	Rôle et justification
`epochs`	200	Nombre de passages sur le dataset.
`batch_size`	16	Taille des lots d' images —équilibre mémoire/temps.
`imgsz`	640	Taille standard des images, adaptée à YOLO.
`lr0`	0.0008	Taux d' apprentissage initial.
`lrf`	0.001	Taux final d' apprentissage (8×10^{-7}).
`momentum`	0.937	Inertie des poids, stabilise l' apprentissage.
`weight_decay`	0.001	Pénalité sur les poids pour éviter le surapprentissage.
`optimizer`	`AdamW`	Optimiseur adapté aux modèles profonds.
`Patience`	75	Nombre d' époques sans amélioration avant arrêt.

TABLE VI.3 – Les hyperparamètres utilisés pour le modèle `YOLOv11s`.

Les hyperparamètres ont été finement réglés pour assurer un équilibre entre vitesse, précision et robustesse.

VI.3.3 Évolution des pertes (loss) pendant l' entraînement

La figure suivante illistre l' evolution de chaque fonction de perte pendant l' entrainement de 200 epochs, tantque chaque fonction diminue avec les epochs, cela signifie que le modèle fait moins d' erreurs :

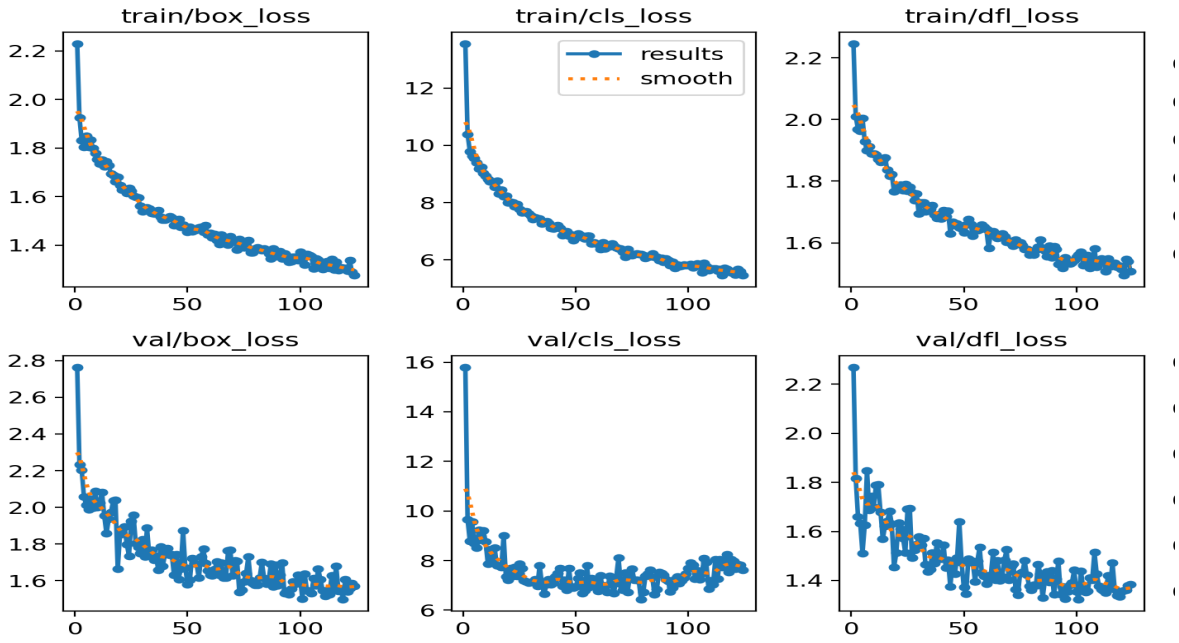


FIGURE VI.5 – Evolution de chaque fonction de perte par epoch yolo11.

Les résultats des différentes fonctions de perte sont résumés dans le tableau ci-dessous, accompagnés d' une interprétation de chaque valeur afin d' évaluer la qualité d' apprentissage et de généralisation du modèle.

VI.3.4 Évolution des métriques de performance pendant l'entraînement

En détection d'objets, on utilise plutôt des métriques comme la précision, le rappel, le F1-score et surtout le mAP, car elles se concentrent uniquement sur les objets détectés. L'accuracy, elle, prend en compte les vrais négatifs, ce qui fausse l'évaluation dans ce domaine. Elle peut donner une impression de bon résultat même si peu d'objets sont correctement détectés. C'est pourquoi elle est généralement écartée. On distingue ici les résultats obtenus pour la performance du modèle. La figure VI.6 illustre l'évolution de chaque métrique pendant l'entrainement de 124 epochs

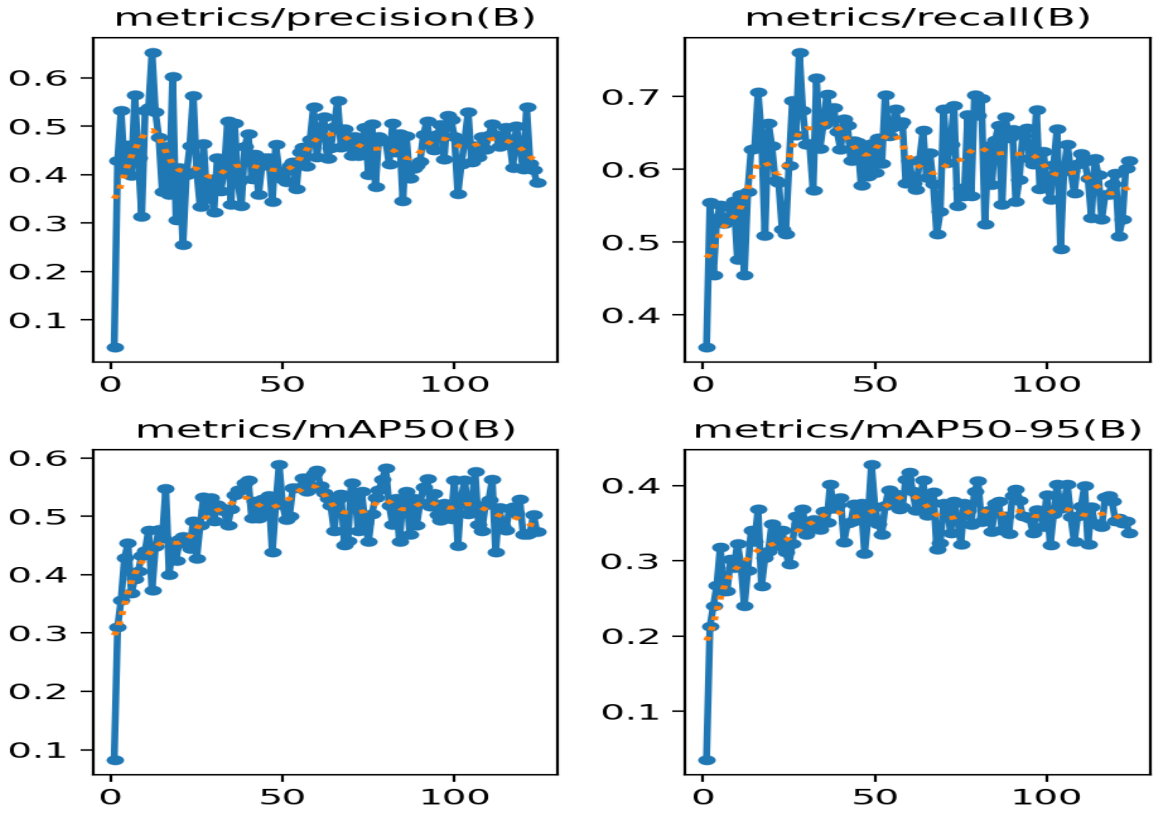


FIGURE VI.6 – Evolution des métriques de performance par epoch yolo11.

Les meilleures performances du modèle de détection d'objets ont été atteintes à l'epoch 49, correspondant au plus haut score de mAP, critère principal d'évaluation dans ce contexte. L'entraînement s'est poursuivi jusqu'à l'epoch 124, conformément au paramètre de patience fixé à 75 (49 + 75), afin de laisser au modèle le temps de progresser davantage.

Le tableau ci-dessous présente l'ensemble des métriques associées à cette meilleure itération, accompagnées d'une interprétation détaillée pour apprécier la qualité de l'apprentissage et la capacité de généralisation du modèle.

Métrique	Interprétation
mAP@50 58.88 %	Un score proche de 60 % indique que le modèle fait des prédictions précises dans leur localisation, même s' il peut encore se tromper sur la classe. Cela traduit une bonne compétence spatiale, surtout sur les objets bien visibles.
mAP@50:95 42.81 %	Un score supérieur à 40% montre que le modèle est **robuste**, même sur des cas complexes : petits objets, chevauchements, variations d'échelle. Cela prouve qu'il ne se contente pas de "juste détecter" mais qu' il apprend une détection *fine et généralisable*.
Précision (B) 40.64 %	Ce score est encore moyen : il signifie que **près de 6 prédictions sur 10 sont des faux positifs**. Le modèle détecte volontiers ···mais pas toujours à bon escient. Il a donc tendance à **sur-détecter**, ce qui affecte la confiance qu' on peut lui accorder.
Rappel / Recall 63.55 %	Un bon rappel montre que le modèle **voit large**, il ne rate pas beaucoup d' objets. Il est utile dans des situations critiques où "*mieux vaut trop que pas assez*" , comme dans les domaines médicaux ou la sécurité.
F1-score 49.6 %	Le modèle commence à **trouver un bon compromis** : il détecte la plupart des objets sans tomber complètement dans l' excès. Mais ce compromis est encore en *phase de stabilisation* : pour améliorer ce score, il faut **réduire les faux positifs** sans trop perdre en rappel.

TABLE VI.4 – Les valeurs de chaque métrique à l' epoch 49 avec leur interprétation.

VI.3.5 Performances du modèle préentraîné à détecter les classes dans la validation

Une évaluation globale ne suffit pas : il est essentiel d' analyser les performances par classe. La figure VI.7 offre un aperçu précis des capacités du modèle à détecter chaque caractéristique de langue durant la validation, on retrouve la précision (Box(P)), le rappel (R) et les métriques mAP à différents seuils d' IoU (mAP50 et mAP50-95), les scores montrent que le modèle détecte bien certaines caractéristiques fréquentes comme langue_normale ou salive_normale, mais échoue sur des classes plus rares comme fissure ou très petites comme red_dot, avec des scores très faibles. La figure VI.8, via la matrice de confusion normalisée, confirme ces difficultés : les classes proches visuellement (ex. langue_pale, langue_rose) sont souvent confondues. Ces résultats soulignent à la fois les points forts du modèle et les limites dues au déséquilibre du dataset ou à la complexité des cas à distinguer.

```
124 epochs completed in 0.498 hours.
Optimizer stripped from mtc_models/yolov11_mtc_optimized/weights/last.pt, 22.5MB
Optimizer stripped from mtc_models/yolov11_mtc_optimized/weights/best.pt, 22.5MB

Validating mtc_models/yolov11_mtc_optimized/weights/best.pt...
Ultralytics 8.3.156    Python-3.11.13 torch-2.6.0+cu124 CUDA:0 (Tesla T4, 15095MiB)
Model summary (fused): 72 layers, 11,131,776 parameters, 0 gradients, 28.5 GFLOPs
```

Class	Images	Instances	Box(P	R	mAP50	mAP50-95): 100%|████████|
all	13	291	0.406	0.636	0.589	0.428
Ecchymoses	4	12	0.35	0.667	0.703	0.351
Eduit_jaune_epais	2	2	1	0	0.745	0.27
Eduit_jaune_mince	3	3	0.435	1	0.995	0.392
Fissure	7	13	0.133	0.0769	0.0233	0.0037
Langue_normal	8	8	0.695	1	0.785	0.698
Langue_pale	2	2	0.585	1	0.663	0.663
Langue_petite	4	4	0.3	0.75	0.578	0.536
Langue_rose	6	6	0.451	0.96	0.686	0.634
Langue_rouge	2	2	0.208	0.5	0.662	0.596
enduit_blanc_epais	2	2	0	0	0.149	0.0573
enduit_blanc_mince	4	4	0.0893	0.25	0.108	0.0464
langue_ganfelee	2	2	0.251	1	0.695	0.637
red_dot	9	219	0.462	0.454	0.378	0.108
salive_humide	4	4	0.435	1	0.763	0.65
salive_normale	8	8	0.7	0.875	0.9	0.785

```
Speed: 0.2ms preprocess, 4.4ms inference, 0.0ms loss, 2.2ms postprocess per image
Saving mtc_models/yolov11_mtc_optimized/predictions.json...
Results saved to mtc_models/yolov11_mtc_optimized

Évaluation finale...
```

FIGURE VI.7 – Résultats de performance du modèle YOLOv11 par classe.

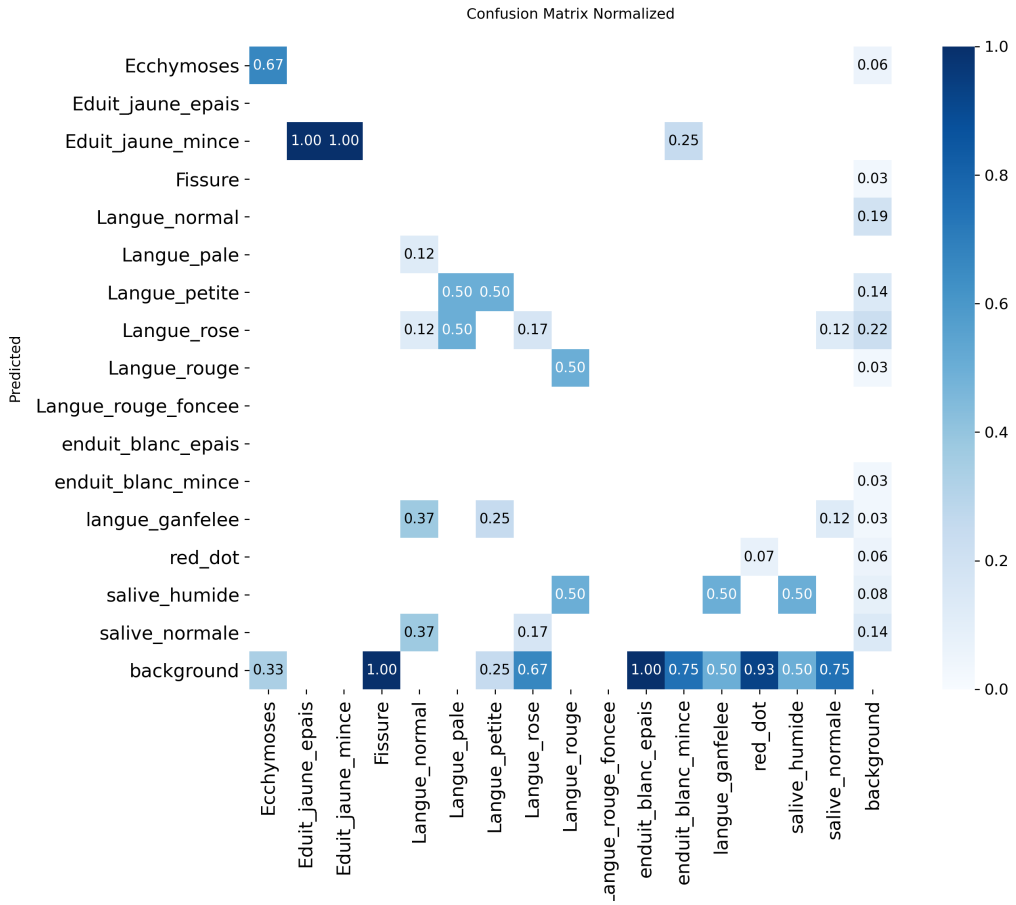


FIGURE VI.8 – Matrice de confusion normalisée sur les classes de caractéristiques linguales

VI.4 Adaptation du modèle YOLOv11 préentraîné via réentraînement

À partir du modèle préentraîné YOLOv11, un réentraînement (fine-tuning) a été effectué sur notre jeu de données annoté dans le but de spécialiser le modèle sur les classes spécifiques de notre étude (par exemple : fissure, enduit blanc, red_dots, etc.). Pour cette phase, un split 80/20 a été appliqué à la base de données (62 images de train et 18 images de validation), avec la même stratégie d' augmentation de données et les mêmes hyperparamètres, à l' exception de la patience fixée à 50 epoch.

VI.4.1 Performances globales du modèle réentraîné

Ce réentraînement a permis d' obtenir une amélioration notable des performances, avec un gain supérieur à 10% sur chaque métrique clé. La meilleure performance a été atteinte dès l' epoch 4. La figure suivante illustre les scores obtenus en précision, rappel et mAP à différents seuils, confirmant l'efficacité de cette optimisation ciblée.

```
============================================================
METRIQUES DE LA MEILLEUR PERFORMANCE RETROUVEE A L'EPOCH 4
============================================================
Epochs: 4
mAP50: 0.7153 (71.53%)
mAP50-95: 0.5573 (55.73%)
Precision: 0.5671 (56.71%)
Recall: 0.7753 (77.53%)
F1-Score: 0.6550 (65.50%)
Accuracy: 0.6712 (67.12%)
============================================================
```

FIGURE VI.9 – Les scores des métriques de performances du modèle YOLO11 réentraînée.

En comparant les deux modèles (Préentraîné, et Réentraîné) dans un tableau concret (le tableau suivant), on constate que le réentraînement n' a pas été inutile. Le modèle a gagné en robustesse et en fiabilité, ce qui se traduit par une meilleure reconnaissance des caractéristiques de la langue. Cette amélioration mesurable se confirmera dans les résultats du batch de validation présentés par la suite (figure VI.5).

Métrique	Préentraîné	Réentraîné (epoch 4)	Interprétation
mAP@50	58.88 %	71.53 %	Amélioration notable de la précision globale des détections.
mAP@50–95	42.81 %	55.73 %	Le modèle devient plus robuste sur des IoU variés (meilleure généralisation).
Précision	40.64 %	56.71 %	Moins de faux positifs : le modèle fait moins d'erreurs dans ce qu'il détecte.
Rappel	63.55 %	77.53 %	Le modèle capte davantage d'objets réels : meilleure couverture.
F1-score	49.6 %	65.50 %	Meilleur équilibre précision / rappel après réentraînement.

TABLE VI.5 – Comparaison entre les performances du modèle **YOLOv11** avant et après le réentraînement.

La figure VI.10 présente l'évolution des fonctions de perte et des métriques de performance du modèle réentraîné. On observe une dynamique similaire à celle du modèle préentraîné en ce qui concerne les fonctions de perte. Par ailleurs, on note une instabilité marquée au niveau des courbes de précision et de rappel. Cette perturbation peut s'expliquer par une confusion persistante entre faux positifs et faux négatifs, notamment pour certaines classes visuellement proches. Cela est cohérent avec le comportement d'un compromis entre précision et rappel : à mesure que le modèle augmente son rappel (détection plus large), la précision tend à diminuer.

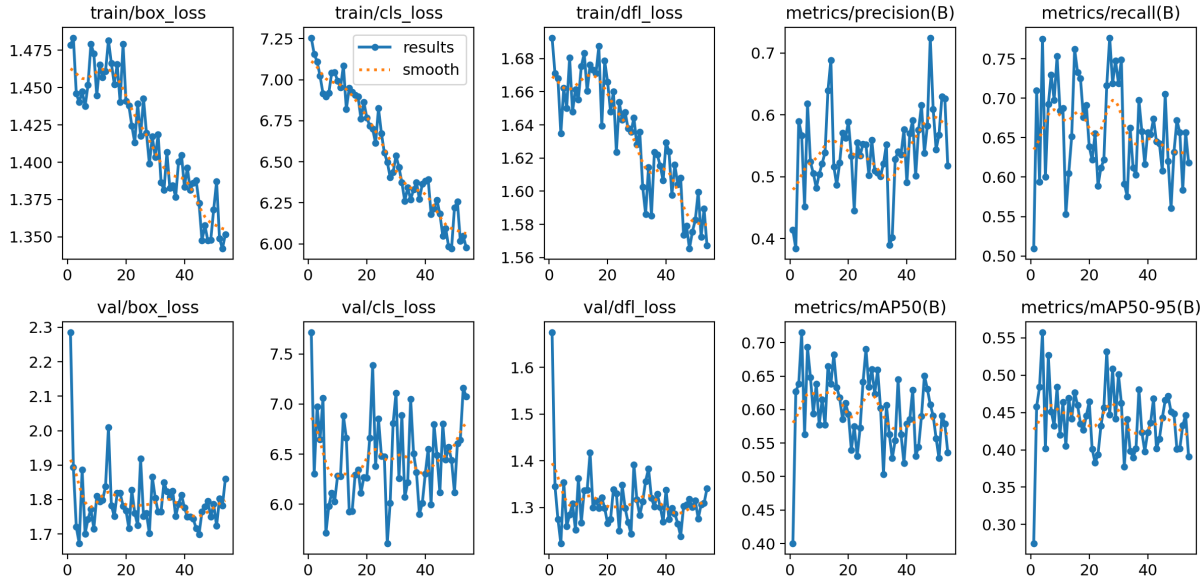


FIGURE VI.10 – Evolution de la performance du modèle yolo11 réentraînée.

VI.4.2 Performances du modèle réentraîné à détecter les classes dans la validation

Le modèle réentraîné améliore globalement la détection sur certaines classes clés, mais ne résout pas toutes les confusions visuelles, notamment pour les signes discrets ou très proches. Cela montre qu'un simple réentraînement ne suffit pas toujours : il faudra envisager d'autres stratégies (meilleures annotations et dataset, ou hybridation avec un classifieur secondaire).

Les figures VI.13 comparent les annotations réelles (val_batch1_labels.jpg), présentées en (a), aux prédictions générées par le modèle (val_batch1_pred.jpg), visibles en (b), sur un sous-ensemble d'images de validation. On y observe les boîtes de détection associées à différentes caractéristiques de la langue (par exemple : red_dot, fissure, langue_petite, etc.). Cette visualisation illustre la capacité du modèle à localiser et classer correctement plusieurs zones, tout en mettant en évidence certaines erreurs comme des surdétections ou des confusions entre classes proches. Ce type d'évaluation qualitative vient compléter les métriques globales et permet de mieux cerner les limites actuelles du modèle.

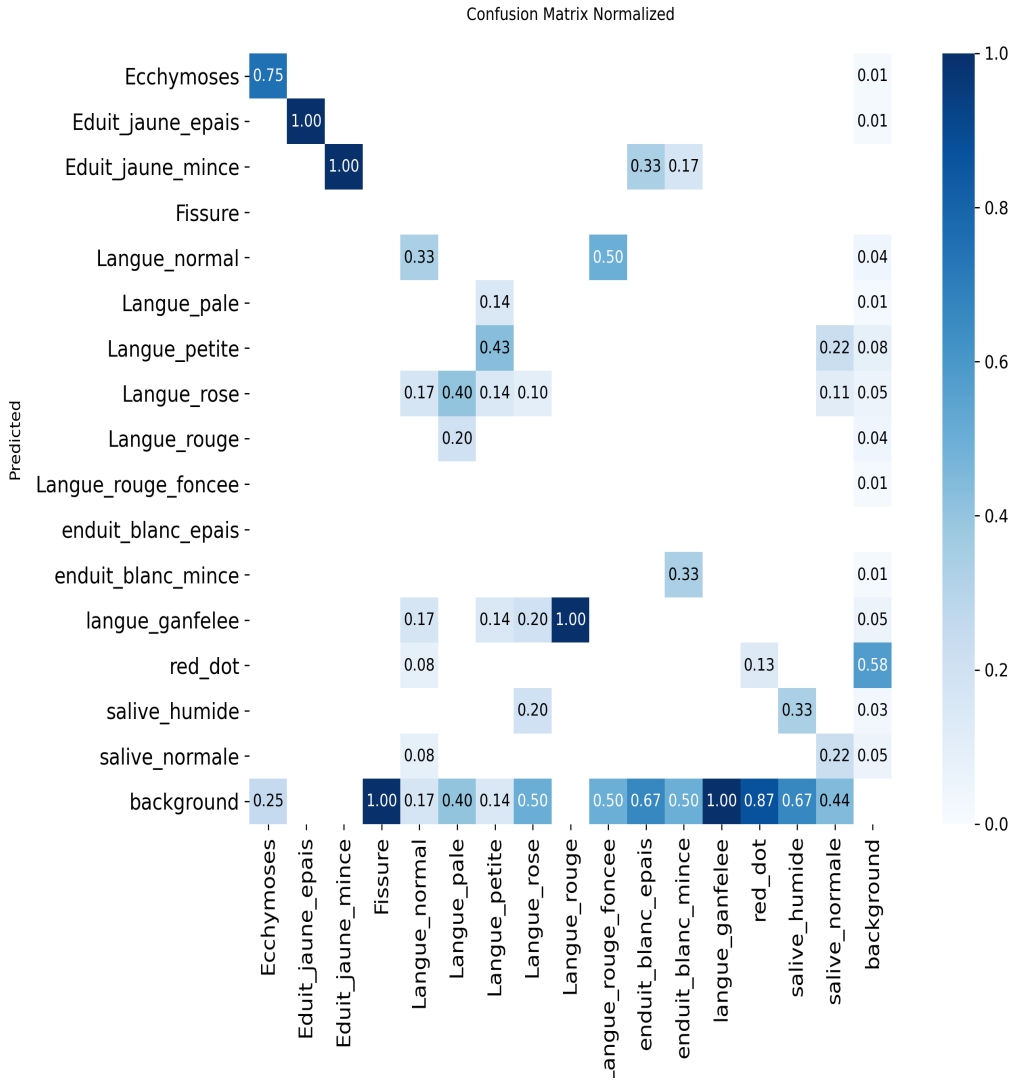


FIGURE VI.11 – Matrice de confusion du modèle yolo11 réentraînée.

```
Validating /content/mtc_models/yolov11_mtc_optimized/weights/best.pt...
Ultralytics 8.3.159    Python-3.11.13 torch-2.6.0+cu124 CUDA:0 (Tesla T4, 15095MiB)
Model summary (fused): 72 layers, 11,131,776 parameters, 0 gradients, 28.5 GFLOPs
```

Class	Images	Instances	Box(P	R	mAP50	mAP50-95): 100%|||||||||
all	18	576	0.566	0.777	0.715	0.554
Ecchymoses	5	12	0.826	0.75	0.79	0.47
Eduit_jaune_epais	1	1	0.339	1	0.497	0.298
Eduit_jaune_mince	4	4	0.596	1	0.828	0.44
Fissure	7	10	0.326	0.1	0.0745	0.0204
Langue_normal	12	12	0.845	0.908	0.945	0.844
Langue_pale	5	5	0.711	0.6	0.778	0.705
Langue_petite	6	7	0.506	0.857	0.77	0.746
Langue_rose	10	10	0.663	0.986	0.84	0.796
Langue_rouge	1	1	0.128	1	0.497	0.448
Langue_rouge_foncee	2	2	0.97	1	0.995	0.822
enduit_blanc_epais	3	3	1	0.446	0.806	0.437
enduit_blanc_mince	6	6	0.285	0.5	0.526	0.218
langue_ganfelee	1	1	0.0817	1	0.995	0.995
red_dot	13	487	0.548	0.388	0.35	0.128
salive_humide	6	6	0.599	1	0.829	0.648
salive_normale	9	9	0.634	0.889	0.919	0.853

```
Speed: 0.4ms preprocess, 4.2ms inference, 0.0ms loss, 3.0ms postprocess per image
Saving /content/mtc_models/yolov11_mtc_optimized/predictions.json...
Results saved to /content/mtc_models/yolov11_mtc_optimized
```

FIGURE VI.12 – les performances du modèle yolo11 réentraîner a détecter chaque classe.

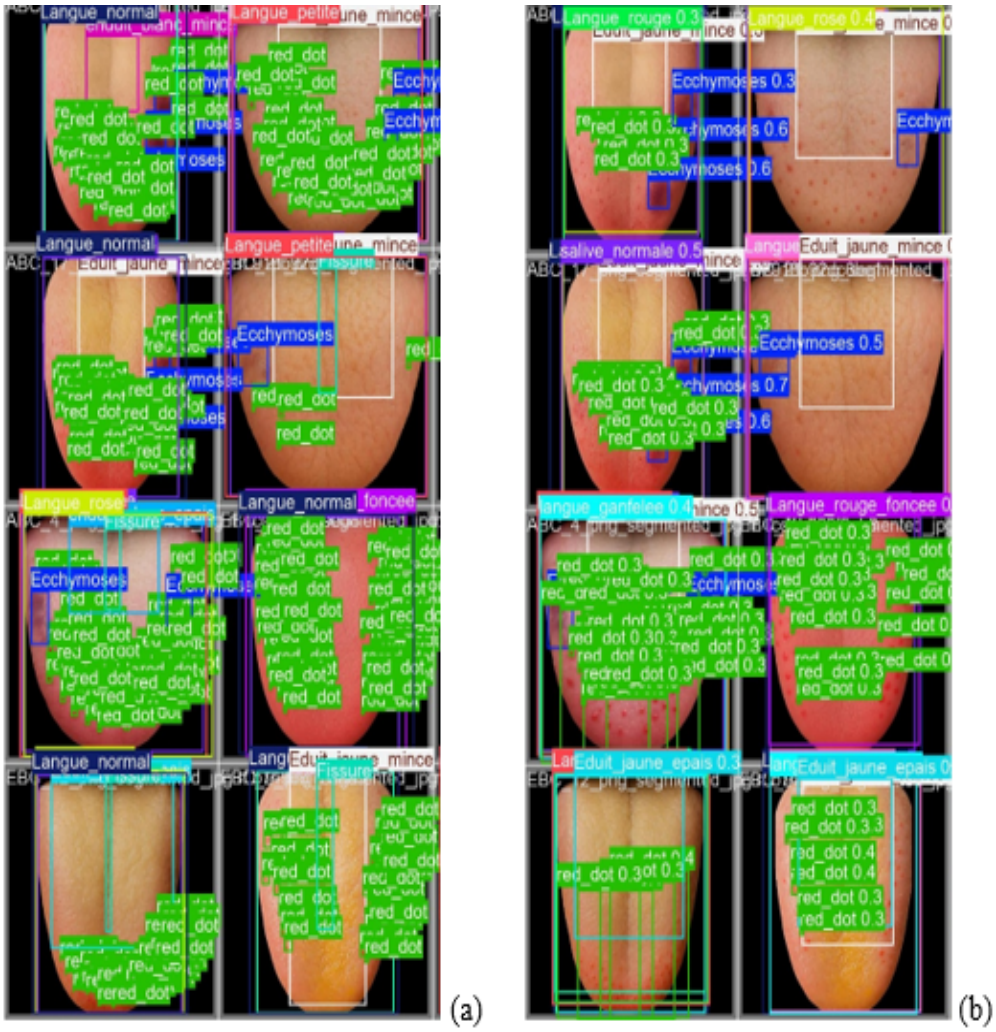


FIGURE VI.13 – Un exemplaire de validation du modèle yolo11 réentraînée.

VI.4.3 Modèle entraîné sur Roboflow

Roboflow propose un service en ligne permettant d'entraîner des modèles YOLO, avec une interface conviviale et des options de téléchargement via leur version premium. Dans le cadre de notre étude, plusieurs entraînements ont été effectués directement sur Roboflow en utilisant notre dataset (figure VI.14), sans ajustement des hyperparamètres, afin de tester leur efficacité par défaut et comparer les performances avec notre propre modèle réentraîné localement. Les modèles atteints une meilleure performance de mAP 61,6%.

Les résultats montrent que notre modèle réentraîné dépasse tous les modèles Roboflow en termes de mAP et de rappel, ce qui témoigne d'une meilleure couverture des caractéristiques de la langue. Cela signifie que notre modèle est plus généraliste et donc plus robuste.

Enfin, lors de tests sur des données totalement nouvelles, le modèle Roboflow a montré une dégradation significative de ses performances, détectant mal les caractéristiques voire la langue elle-même, contrairement à notre modèle qui a conservé une certaine cohérence dans ses prédictions. La figure VI.15 présente deux exemples testés sur ce modèle.

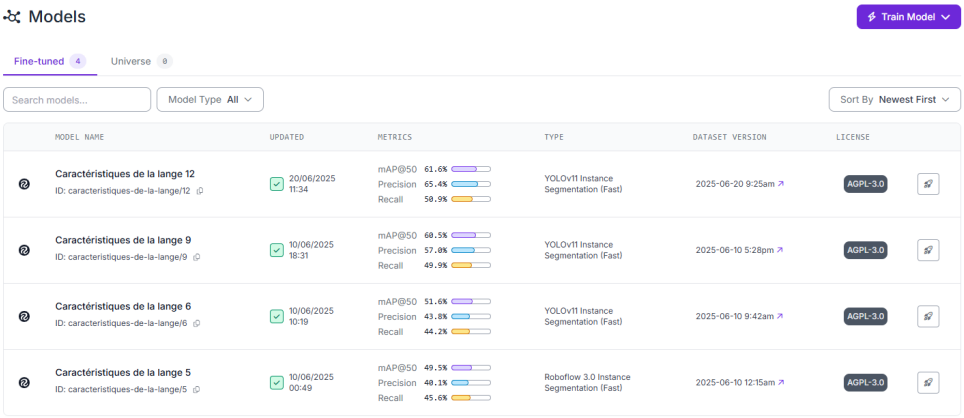


FIGURE VI.14 – Les différents modèles yolo11 entrainées dans Roboflow.

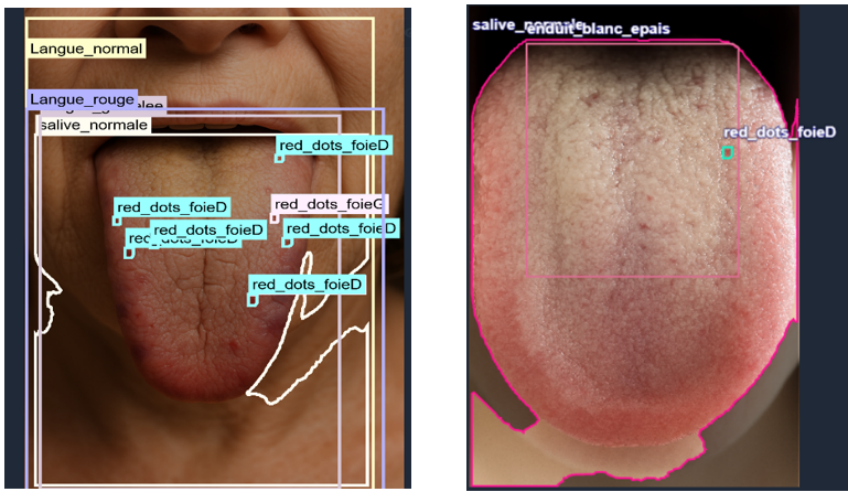


FIGURE VI.15 – Exemples d'échec partiel de détection du modèle de Roboflow, confusion entre classes et mauvaise localisation.

La figure suivante présente la détection de notre modèle sur les mêmes images : notre modèle détecte mieux les ecchymoses, et les autres caractéristiques.

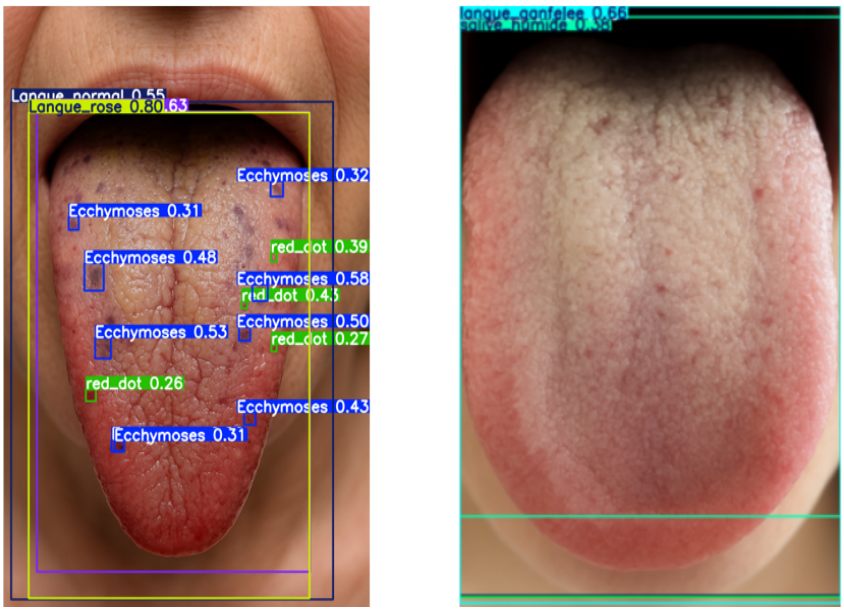


FIGURE VI.16 – Exemples de détection de notre modèle yolo11

VI.5 Système de diagnostic précoce du cancer du sein

YOLO est très performant pour détecter des objets visuels, mais il n'est pas suffisant seul pour un diagnostic médical complexe comme celui du cancer du sein via la langue. Ce type d'analyse nécessite de combiner plusieurs signes visuels et anatomiques. C'est pourquoi une solution hybride, associant YOLO à des algorithmes de machine learning, a été mise en place pour assurer un diagnostic plus complet et fiable.

VI.5.1 Pipeline général du système de diagnostic hybride (YOLO + ML)

- **Détection visuelle** : Le modèle YOLOv8 détecte la zone de la langue avec une Bbox et réduit les dimensions de l'image en fonction de cette Bbox pour éviter les bruits. Ensuite, le modèle YOLOv11, entraîné, est utilisé pour détecter automatiquement les signes visibles sur la langue (taches rouges, ecchymoses, fissures⋯), et génère un vecteur de caractéristiques contenant les classes détectées et leurs positions spatiales.

- **Projection anatomique** : Les caractéristiques détectées sont projetées sur une cartographie MTC en fonction de leurs coordonnées, afin de relier chaque signe à une zone (Rein, Cœur-Poumon, Foie-VB⋯). nous avons utilisé les dimensions de la Bbox de la langue pour définir les zones anatomiques sur la cartographie.

Modèle	Description	Avantages d' utilisation
Random Forest	Agrégation d' arbres de décision avec échantillonnage aléatoire.	Robuste au bruit, facile à interpréter, idéal pour évaluer les performances de départ sur données tabulaires.
Gradient Boosting	Arbres successifs corrigeant les erreurs précédentes (boosting séquentiel).	Améliore la généralisation par rapport à Random Forest, bon pour capter des patterns faibles sur données structurées.
XGBoost	Boosting extrême (optimisé et régularisé de boosting gradient).	Très performant sur petits datasets, haute robustesse, gestion des données déséquilibrées, très recommandé.
SVM	Cherche les hyperplans maximisant la séparation entre classes.	Recommandé car les classes sont bien séparables (ex. sain vs avancé), bonne performance en faible dimension.
MLP	Réseau à couches profondes apprenant des relations complexes.	Utile pour détecter des combinaisons non linéaires entre signes et zones anatomiques.

TABLE VI.6 – Descriptions de chaque algorithme de ML choisi.

- **Classification médicale** : Le vecteur est transmis à un classifieur automatique, entraîné à prédire le stade clinique du patient (sain, précoce, avancé).
- **Affichage des résultats** : Un rapport visuel synthétique est généré. Il montre la langue détectée, les prédictions YOLO, la cartographie MTC annotée, et le diagnostic final avec son score de confiance.

VI.5.2 Hybridation du modèle YOLOv11 avec des modèles de classification de ML

Afin d' assurer une bonne généralisation et une interprétation fiable, cinq algorithmes de ML ont été testés sur le vecteur enrichi. Ces modèles ont été choisis pour leur efficacité sur de petits jeux de données et leur capacité à traiter des données tabulaires combinées, le tableau ci-dessus résume la description et l' avantage de chaque algorithme :

VI.5.2.1 Méthodologie d'entraînement des classificateurs

Dans ce système, les vecteurs générés par YOLOv11 ont servi à entraîner cinq modèles de classification supervisée : RandomForest, XGBoost, SVM, MLP et GradientBoosting. Chaque modèle a été optimisé avec des hyperparamètres adaptés et validé à l'aide de la validation croisée pour assurer sa fiabilité. Enfin, le modèle le plus performant a été sélectionné automatiquement en se basant sur le score moyen obtenu lors de la validation.

1. Préparation des données Avant d' entraîner un modèle de machine learning :

- Les étiquettes (sain, précoce, avancé) sont encodées en chiffres (ex. 0, 1, 2).

- Les données sont divisées : 80% pour l'entraînement (train), 20% pour la validation (valid).

- Les vecteurs sont normalisés (mis à la même échelle), ce qui est particulièrement utile pour les modèles sensibles aux écarts de valeurs comme SVM ou MLP.

VI.5.2.2 Paramètres utilisés pour chaque modèle ML

Modèle	Paramètres clés utilisés
Random Forest	200 arbres limités à 15 niveaux, avec minimum 5 exemples pour chaque séparation. Robuste et facile à interpréter.
XGBoost	200 arbres enchaînés (boosting), profondeur max 8, apprentissage lent (0,1), utilise 80 % des données à chaque étape.
SVM	Frontière courbe (RBF), erreur pondérée à 10, donne aussi les probabilités. Bon sur petits jeux.
MLP (réseau de neurones)	3 couches (128, 64, 32 neurones), activation ReLU, apprentissage adaptatif, jusqu'à 500 itérations.
Gradient Boosting	200 arbres successifs, profondeur 8, taux d'apprentissage 0,1. Bonne généralisation.

TABLE VI.7 – Les hyperparamètres utilisés pour l'entraînement des modèles de ML.

VI.5.2.3 Mesure de performance

Les modèles ont été évalués à l'aide de deux métriques :

- **Accuracy** : taux de bonnes prédictions.

- **CV Score** : score de validation croisée, pour évaluer la fiabilité sur de nouvelles données.

XGBoost a été retenu comme modèle final car il combine à la fois précision et stabilité, deux critères cruciaux dans le contexte médical.

```
...  Entraînement RandomForest...
       Accuracy: 0.9375
       CV Score: 0.7808 (+/- 0.1531)

     Rapport détaillé RandomForest:
                   precision    recall  f1-score   support

         advanced       1.00      1.00      1.00         4
            early       1.00      0.80      0.89         5
          healthy       0.88      1.00      0.93         7

         accuracy                           0.94        16
        macro avg       0.96      0.93      0.94        16
     weighted avg       0.95      0.94      0.94        16


     Entraînement XGBoost...
       Accuracy: 0.8125
       CV Score: 0.8269 (+/- 0.1240)

     Entraînement SVM...
       Accuracy: 0.8750
       CV Score: 0.7987 (+/- 0.1529)

     Entraînement MLP...
       Accuracy: 0.8750
       CV Score: 0.7500 (+/- 0.1141)

     Entraînement GradientBoosting...
       Accuracy: 0.8125
       CV Score: 0.7641 (+/- 0.1788)

     🏆 MEILLEUR MODELE: XGBoost
        Accuracy: 0.8125
```

FIGURE VI.17 – Performances des modèles ML entraînée pour la classification.

VI.5.3 Test et résultats

Nous avons testé notre modèle sur de nouvelles données, totalement absentes du jeu de données, afin d' évaluer sa capacité de généralisation. Les résultats obtenus confirment la robustesse de notre architecture, qui a su détecter correctement les caractéristiques attendues.

L' image est analysée en trois étapes : détection de la langue, identification des signes pathologiques par YOLOv11, puis projection sur la cartographie MTC. Ces données alimentent le classifieur XGBoost, qui prédit le stade clinique du patient.

Une personne en stade avancé:

Le modèle prédit un stade avancé avec une forte confiance (99,38%). Tous les classificateurs le confirment. Les signes détectés sont localisés sur différentes zones de la langue selon la cartographie MTC (ex. : Rate-Estomac, Foie). Ce retour complet combine fiabilité

du diagnostic et interprétabilité clinique.

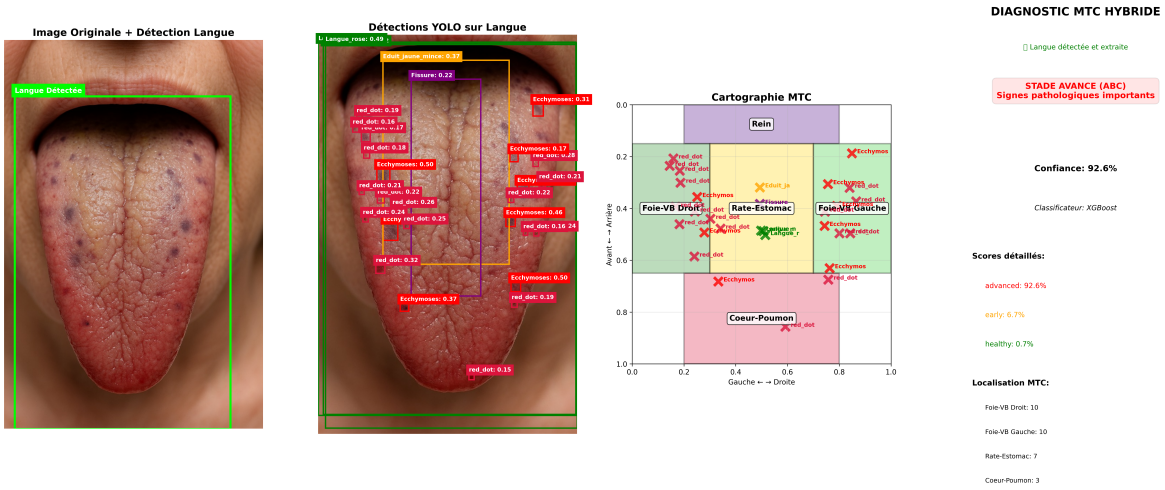


FIGURE VI.18 – Diagnostic d'un stade avancé avec YOLO_MTCBC.

```
Image originale: ABC_11_png_segmented_jpg.rf.bab6fe1fdae5b285cbe2cafcfecabddd.jpg
Langue détectée et extraite
Image analysée: tongue_ABC_11_png_segmented_jpg.rf.bab6fe1fdae5b285cbe2cafcfecabddd_20250622_003657.jpg

Prédiction: ADVANCED
Confiance: 99.38%
Classificateur: XGBoost

Probabilités détaillées:
 advanced: 0.9938 (99.38%)
 early: 0.0043 (0.43%)
 healthy: 0.0018 (0.18%)

Prédictions ensemble:
 RandomForest: advanced (0.949)
 XGBoost: advanced (0.994)
 SVM: advanced (0.871)
 MLP: advanced (1.000)
 GradientBoosting: advanced (1.000)

Localisation par zones MTC:
 Rate-Estomac: 10 détection(s)
 Foie-VB Droit: 8 détection(s)
 Foie-VB Gauche: 4 détection(s)
 Coeur-Poumon: 3 détection(s)
 Rein: 2 détection(s)

Visualisation complète: mtc_results/hybrid_mtc_ABC_11_png_segmented_jpg.rf.bab6fe1fdae5b285cbe2cafcfecabddd_
```

FIGURE VI.19 – Sortie du terminal détaille sur le diagnostic d'un stade avancée avec YOLO_MTCB

Une personne en stade précoce :

Grâce à ces informations spatiales et sémantiques, le classifieur XGBoost prédit un stade précoce du cancer du sein avec 95.5% de confiance, confirmé par l'ensemble des autres modèles testés (SVM, MLP, GradientBoosting⋯). Ce résultat met en lumière la capacité du système à détecter les premiers signes cliniques de manière localisée et cohérente avec les zones physiopathologiques de la Médecine Traditionnelle Chinoise (MTC).

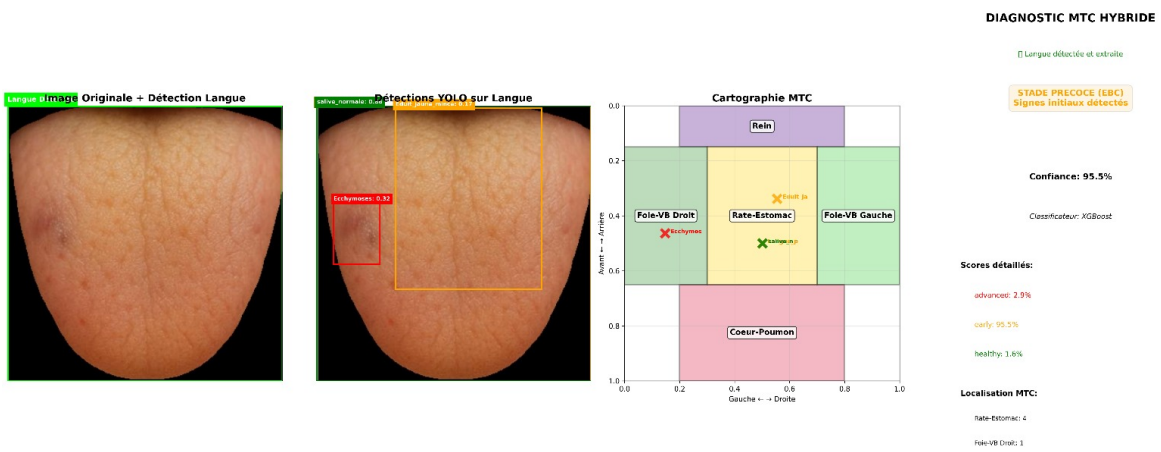


FIGURE VI.20 – Diagnostic d'un stade précoce avec YOLO_MTCBC.

```
Image originale: EBC_21_png_segmented_jpg.rf.3c1e580c135a3fedec4c20aff7b2c0bb.jpg
Langue détectée et extraite
Image analysée: tongue_EBC_21_png_segmented_jpg.rf.3c1e580c135a3fedec4c20aff7b2c0bb_20250622_002004.jpg

Prédiction: EARLY
Confiance: 95.51%
Classificateur: XGBoost

Probabilités détaillées:
 advanced: 0.0286 (2.86%)
 early: 0.9551 (95.51%)
 healthy: 0.0163 (1.63%)

Prédictions ensemble:
 RandomForest: early (0.615)
 XGBoost: early (0.955)
 SVM: early (0.850)
 MLP: early (0.997)
 GradientBoosting: early (1.000)

Localisation par zones MTC:
 Rate-Estomac: 4 détection(s)
 Foie-VB Droit: 1 détection(s)

Visualisation complète: mtc_results/hybrid_mtc_EBC_21_png_segmented_jpg.rf.3c1e580c135a3fedec4c20aff7b2c0bb_
```

FIGURE VI.21 – Sortie du terminal détaille sur le diagnostic d'un stade précoce avec YOLO_MTCB.

Une personne en bonne santé :

Le classifieur XGBoost prédit un état sain avec une confiance de 97.4%, une prédiction validée. Malgré une légère erreur de détection d'un bouton situé sur le visage (interprété comme un red_dot), le système reste robuste et n'interprète pas cette anomalie comme un signe pathologique. Ce résultat montre la capacité du modèle à ignorer les faux positifs isolés et à produire un diagnostic globalement fiable.

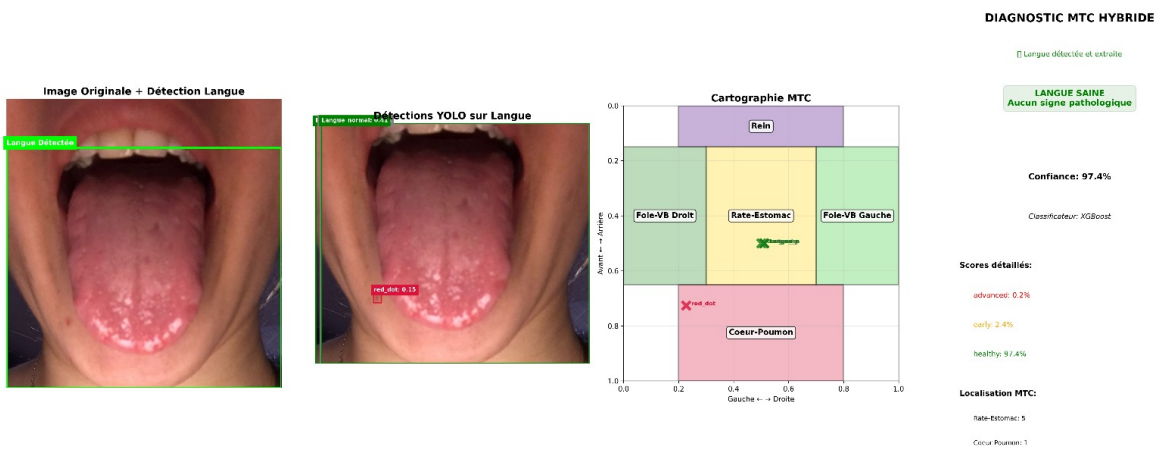


FIGURE VI.22 – Diagnostic d'une personne en bonne santé avec YOLO_MTCBC

```
Image originale: WhatsApp Image 2025-06-22 à 02.17.30_408db7b6.jpg
Langue détectée et extraite
Image analysée: tongue_WhatsApp Image 2025-06-22 à 02.17.30_408db7b6_20250622_003017.jpg

Prédiction: HEALTHY
Confiance: 97.40%
Classificateur: XGBoost

Probabilités détaillées:
 advanced: 0.0025 (0.25%)
 early: 0.0235 (2.35%)
 healthy: 0.9740 (97.40%)

Prédictions ensemble:
 RandomForest: healthy (0.837)
 XGBoost: healthy (0.974)
 SVM: healthy (0.919)
 MLP: healthy (0.995)
 GradientBoosting: healthy (0.635)

Localisation par zones MTC:
 Rate-Estomac: 5 détection(s)
 Coeur-Poumon: 1 détection(s)
```

FIGURE VI.23 – Sortie du terminal détaille sur le diagnostic d'une personne en bonne santé avec YOLO_MTCB.

VI.6 Plateforme interactive pour le diagnostic

Notre application web a été développée pour permettre un diagnostic assisté du cancer du sein à partir de l'analyse de la langue, selon les principes de la MTC.

VI.6.1 Environnement de développement

Le développement s'est effectué localement à l'aide de :

- **IDE utilisé** : Visual Studio Code (VS Code), un environnement de développement puissant et extensible. La figure suivante montre l'arborescence du projet.
- **Versionnement** : Git et GitHub ont été utilisés pour gérer l'historique du projet, collaborer et sauvegarder les différentes versions du code, avec des *commits* réguliers pour chaque étape du pipeline.
- **Langage** : Python, langage principal utilisé pour le développement du projet.
- **Interface web** : Streamlit, un framework Python minimaliste qui permet de créer rapidement une interface utilisateur interactive, sans compétences en développement frontend.

VI.6.2 Architecture de la plateforme

Le projet est structuré autour d'une architecture Python modulaire, avec un hébergement du code source sur GitHub (figure VI.24), facilitant la traçabilité et la collaboration.

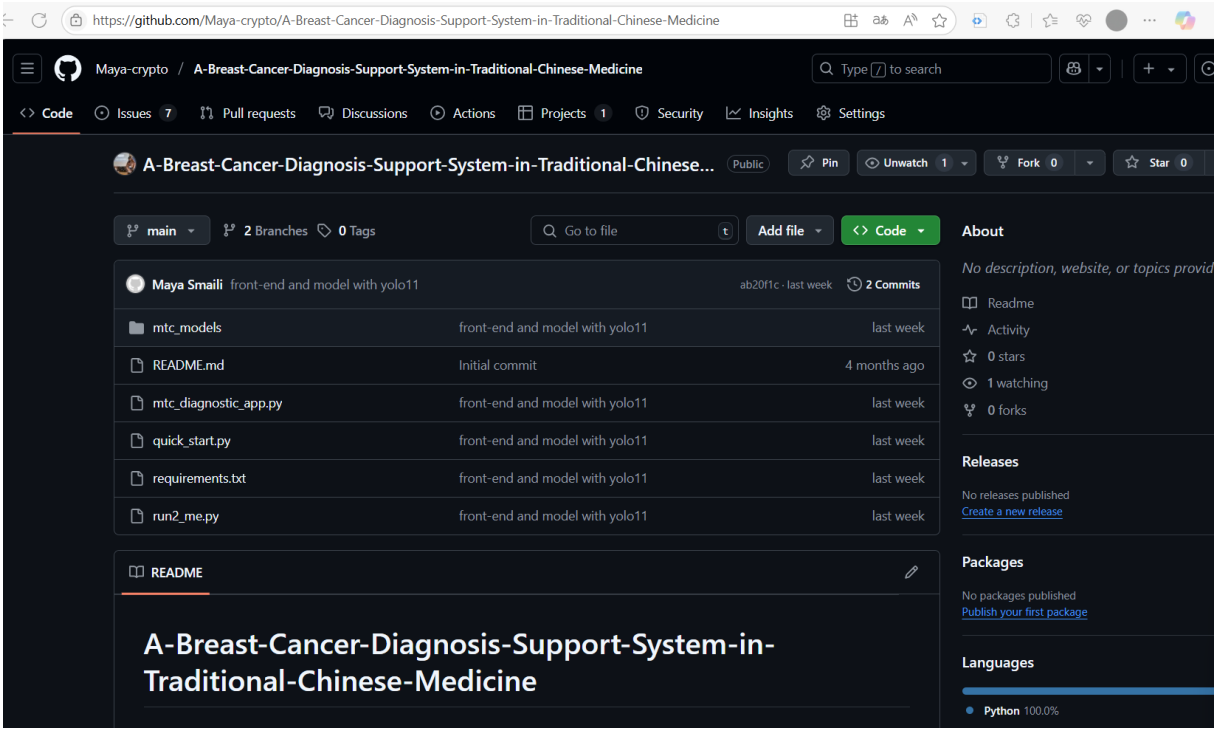


FIGURE VI.24 – L'interface GitHub pour le code accessible.

- **Frontend** : composants `Streamlit`, HTML et CSS personnalisés pour offrir une expérience utilisateur fluide, pédagogique et esthétique.
- **Backend** :
  - **YOLOv8** pour la détection automatique de la langue.
  - **YOLOv11** pour l'extraction des caractéristiques pathologiques.

– **Classificateurs ML (XGBoost)** pour prédire le stade de développement du cancer (sain, précoce, avancé).
– **Cartographie MTC** de la langue, basée sur les caractéristiques détectées.

VI.6.3 Déploiement et accessibilité

L'application a été déployée en ligne via la plateforme Streamlit Cloud (présentée dans la figure VI.25) destinée à des médecins professionnels (praticiens chinois, médecins).



FIGURE VI.25 – Capture de l'Interface d'acceuil de la platform MTC diagnostique Pro

L'intégration de fonctionnalités avancées (détection des zones MTC, visualisation des scores, analyse probabiliste) démontre la maturité du système, tandis que l'architecture modulaire du code facilite les mises à jour, l'extension des modèles, ou le déploiement futur dans des contextes cliniques réels.

VI.6.4 Test de la plateforme

Afin de valider l'efficacité fonctionnelle de la plateforme MTC Diagnostic Pro, plusieurs scénarios utilisateurs ont été testés. Ces tests couvrent les principales fonctionnalités attendues.

En Annexe B, la Figure B.1 présente l'interface de téléversement d'image, première étape essentielle du processus diagnostique. Une fois l'image chargée, le système déclenche automatiquement l'analyse des caractéristiques visuelles de la langue.

La Figure B.2 illustre le cas d'un patient sain, avec une langue homogène et des statistiques compatibles avec un état de santé normal. À l'inverse, les Figures B.3 et B.4 montrent des cas de cancer du sein à un stade avancé. Ces interfaces affichent des résultats plus contrastés, accompagnés d'indicateurs cliniques révélateurs.

La Figure B.5 apporte une dimension supplémentaire grâce à la cartographie des zones affectées selon les principes de la Médecine Traditionnelle Chinoise, facilitant ainsi l'interprétation fonctionnelle du diagnostic.

Enfin, la Figure B.6 met en évidence la détection automatisée des caractéristiques pathologiques sur la langue à l'aide du modèle YOLOv11, tandis que la Figure B.7 détaille l'interprétation fine de ces anomalies selon les référentiels médicaux traditionnels.

VI.7 Conclusion

Ce chapitre a permis d'évaluer de manière rigoureuse la performance de l'architecture proposée pour le diagnostic du cancer du sein à partir de l'analyse de la langue, en combinant la détection par YOLO et la classification par apprentissage supervisé. Dans un premier temps, les entraînements des modèles YOLOv8 et YOLOv11 ont été analysés à travers les courbes de perte, les métriques clés (précision, rappel, mAP) et les matrices de confusion. Un modèle YOLOv11 réentraîné sur notre dataset a démontré une amélioration significative par rapport à la version préentraînée, avec un gain de plus de 10% sur toutes les métriques. Cette montée en performance confirme que le 2eme fine-tuning sur des mêmes données avec diffèrent hyperparamètres peut faire un changement. De plus, les sorties du modèle YOLO11 ont été utilisées pour afficher la cartographie avec les caractéristiques détectées, ainsi que par plusieurs modèles de ML pour effectuer un diagnostic par classification des stades saine, précoce et avancée. Les tests de classification confirment la robustesse de cette approche, XGBoost est le modèle optimal, grâce à sa capacité de généralisation et sa précision élevée. L'ensemble du pipeline a ensuite été intégré dans une plateforme web interactive, MTC Diagnostic Pro, accessible, intuitive, pour une évaluation rapide et sécurisée. Cependant, les résultats montrent aussi des limites sur certaines classes rares ou visuellement proches, mettant en lumière les difficultés liées à l'annotation, au déséquilibre des classes ou à la complexité des signes cliniques. Les tests sur des nouvelles données ont confirmé la capacité du système à produire des diagnostics fiables, interprétables et cohérents avec les principes de la MTC.

Chapitre VII

Conclusion générale et perspectives

Notre mémoire à pour objectif d'explorer les apports de l'intelligence artificielle dans l'automatisation du diagnostic précoce non invasive du cancer du sein à partir de l'analyse de la langue, en s'appuyant sur les fondements de la Médecine Traditionnelle Chinoise (MTC). Le cancer du sein, pathologie agressive féminin en Algérie, souffre d'un diagnostic souvent tardif qui limite considérablement les options thérapeutiques et affecte les taux de survie. Face aux limites des méthodes conventionnelles (mammographie, échographie, IRM) en termes d'accessibilité, de coût et de détection précoce, particulièrement dans les pays en développement, notre travail a exploré une voie alternative : l'objectivation par l'intelligence artificielle des observations linguales traditionnelles chinoises.

Notre mémoire a débuté par l'étude des fondements théoriques de la MTC, notamment le concept du Qi, le Yin et le Yang, les cinq éléments, les méridiens, ainsi que les techniques diagnostiques dont l'observation de la langue, outil central dans l'approche énergétique de la santé. Nous avons ensuite présenté les grandes familles d'algorithmes d'apprentissage automatique, des réseaux de neurones profonds aux techniques de vision par ordinateur, particulièrement les algorithmes YOLO pour la détection d'objets. L'état de l'art a révélé le potentiel élevé des approches d'intelligence artificielle pour renforcer la précision des analyses linguo-diagnostiques, tout en soulignant l'absence d'approches spécifiques au cancer du sein par la langue en MTC.

Pour relever ce défi, nous avons conçu et implémenté avec succès une architecture hybride originale, YOLO_BCTCM, combinant YOLOv8 pour la détection automatique de la langue (100% de mAP), YOLOv11 pour l'identification des caractéristiques pathologiques (amélioration de 10% après réentraînement, a atteindre 71.53% de mAP), des algorithmes de Machine Learning pour la classification des stades cliniques (XGBoost avec une validation croissee [82,69 %], Très bonne généralisation et Moins sensible aux variations de données), et une cartographie MTC pour la projection anatomique des signes détectés. Cette architecture constitue une première mondiale dans l'automatisation du diagnostic du cancer du sein par analyse linguale en MTC. Face à l'absence de datasets publics, nous avons développé une approche novatrice de génération de données synthétiques guidée par l'IA, créant le dataset *Tongue_MTCBC_Hope* (88 images, uniquement 80 fonctionnel réparties en trois classes : saine, précoce, avancée). Le développement de la plateforme web interactive *MTC Hybrid Pro* concrétise la transition de la recherche vers l'application pratique.

Les tests sur données nouvelles confirment la robustesse du système avec des performances remarquables : diagnostic correct du stade avancé avec 99,38% de confiance, identification du stade précoce avec 95,5% de confiance, et classification de l'état sain à 97,4%. Cependant, plusieurs limites méritent d'être soulignées. La principale limite réside dans l'absence de validation clinique à grande échelle, ce qui limite son intégration immédiate dans des environnements médicaux réels. La qualité et la diversité des données demeurent perfectibles, la base d'images reposant encore sur des données générées ou peu nombreuses, ce qui peut affecter la robustesse du modèle. L'interprétation fine des signes pathologiques reste dépendante d'une lecture contextualisée qui pourrait être améliorée par l'intégration de données patient plus complètes. De plus, l'annotation manuelle des 16 classes caractéristiques reste chronophage et dépendante de l'expertise humaine.

Ce mémoire démontre qu'il est possible de créer un système intelligent de diagnostic précoce du cancer du sein basé sur l'analyse de la langue selon la MTC, avec des performances encourageantes et une architecture évolutive. Ce travail pose les jalons d'une approche nouvelle de l'aide au diagnostic médical, respectueuse des savoirs traditionnels tout en répondant aux exigences de rigueur scientifique contemporaine. L'impact sociétal potentiel est considérable : dans un contexte où l'accès aux soins spécialisés reste limité dans de nombreuses régions du monde, un outil de dépistage précoce, non invasif et peu coûteux pourrait transformer la prise en charge du cancer du sein, particulièrement dans les pays émergents. Cette recherche illustre comment l'intelligence artificielle peut servir de pont entre les médecines traditionnelles et modernes, ouvrant la voie à une médecine véritablement intégrative et personnalisée où l'intuition humaine et l'intelligence artificielle coopèrent pour le bien-être des patientes.

Perspectives

Les perspectives de développement s'articulent autour de plusieurs axes complémentaires, tenant compte du fait que la plateforme *MTC Hybrid Pro* demeure en phase de développement actif avec plusieurs composantes critiques encore à implémenter.

Amélioration des performances et validation clinique : La priorité absolue demeure l'enrichissement du dataset avec des images cliniques réelles annotées par des experts MTC pour renforcer la robustesse et la validité clinique du système. Une collaboration avec des centres hospitaliers spécialisés en oncologie permettra de constituer une cohorte de 500-1000 patientes avec suivi à long terme de developpemet de la maladie, pour une comparaison rigoureuse avec les méthodes diagnostiques conventionnelles. L'optimisation algorithmique se concentrera sur l'exploration d'architectures plus récentes comme les modèles bitemporels, l'implémentation de techniques d'apprentissage fédéré, et le développement de mécanismes d'explicabilité pour renforcer la confiance clinique.

Développements techniques en cours : Plusieurs composantes critiques de la plateforme restent à finaliser, notamment l'implémentation d'un système de gestion des utilisateurs robuste avec authentification sécurisée, la mise en place d'une base de données relationnelle pour le stockage et la traçabilité des diagnostics, et le renforcement de la sécurité par l'intégration de protocoles de chiffrement conformément aux réglementations. L'optimisation des performances vise à réduire le temps de traitement du pipeline complet sous la barre des 2 secondes pour un usage clinique fluide. L'ajout d'un module d'auto-

apprentissage permettra au système de s'adapter continuellement grâce à des techniques de fédération de données ou d'apprentissage en ligne.

Modèle étendu et intelligence conversationnelle : L'implémentation du modèle V.7 étendu intégrera un module de traitement du langage naturel (NLP) pour l'analyse des symptômes textuels et la génération automatique de rapports personnalisés. Ce modèle incorporera un chatbot médical intelligent basé sur des modèles de langage (GPT-4, T5) capable d'engager un dialogue contextualisé avec les patients, d'expliquer les résultats en termes accessibles, et de proposer des recommandations préventives selon les principes de la MTC.

Extension multi-pathologique et déploiement : L'intégration mobile constituera un axe majeur avec le développement d'une application légère et ergonomique pour faciliter l'usage en télémédecine ou dépistage rural. L'extension à d'autres pathologies détectables par la langue (diabète, troubles digestifs, déséquilibres émotionnels) et l'intégration d'autres modalités MTC (analyse du pouls digital, reconnaissance faciale pour l'examen du teint) enrichiront l'écosystème diagnostic. L'innovation technologique se concrétisera par un dispositif de capture standardisée intégrant des capteurs IoT et une infrastructure cloud sécurisée capable de gérer des millions de diagnostics simultanés.

Vision à long terme : L'évolution vers une médecine personnalisée préventive s'articulera autour d'un assistant médical virtuel multilingue, de la corrélation avec des biomarqueurs génétiques, et du développement d'algorithmes prédictifs avancés. La recherche fondamentale explorera les corrélations neurobiologiques entre manifestations linguales et pathologies systémiques, contribuant à la validation scientifique des méridiens énergétiques et à la compréhension moderne des mécanismes de la MTC.

En somme, ce travail constitue une première pierre vers une médecine intégrative, intelligente et centrée sur la prévention, jetant les bases d'un système d'aide au diagnostic original à la croisée de la médecine traditionnelle et de l'intelligence artificielle moderne, s'inscrivant résolument dans une démarche d'innovation responsable au service de la santé publique algérienne et pourquoi pas mondiale.

Bibliographie

[1] S. M. Rabah, "Vision artificielle et traitement d'images." 2024.

[2] Ultralytics, "Yolo11 documentation." `https://docs.ultralytics.com/`.

[3] M. Guérin, "Cancer du sein : signes, stades, un cancer agressif?." consulter en juin.

[4] J. Ferlay, M. Ervik, F. Lam, M. Laversanne, M. Colombet, L. Mery, M. Piñeros, A. Znaor, I. Soerjomataram, and F. Bray, "Global cancer observatory: Cancer today (globocan2022)." `https://gco.iarc.who.int/today`, 2024.

[5] G. Maciocia, *L' EXAMEN DE LA LANGUE EN MÉDECINE CHINOISE*. PhD thesis, Journal of Chinese Medicine, 1989.

[6] M. ALLEK, "Un etat de l' art sur les systèmes d' aide au diagnostic médical dans la médecine traditionnelle chinoise," Master's thesis, Université de Mouloud Mammerie de Tizi-Ozou, 2024.

[7] J. Ljubelsek, *La Médecine traditionnelle chinoise et son application dans le Diabète*. PhD thesis, Université de Lille 2, 2016.

[8] J. Zhao, *Initiation au monde de la médecine traditionnelle chinoise : De la philosophie des cinq éléments et du Qi aux pratiques de guérison*. Avenet Edition, 2023.

[9] A. H. et al., "La médecine traditionnelle chinoise," *Esprit libre*, 02 2014.

[10] C. Despeux, "Les fondements théoriques de la médecine chinoise traditionnelle," in *Les fondements théoriques de la médecine chinoise traditionnelle*, 2006. Professeur à l'INALCO.

[11] medecine-traditionnelle-chinoise, "Médecine Traditionnelle Chinoise." `https://chinoistips.com/medecine-traditionnelle-chinoise/`. Consulté en mars 2025.

[12] santé global, "la theorie du yin yang." .

[13] david-plourde, "Les 5 Éléments médecine traditionnelle chinoise," 2024. Consulté en mars 2025.

[14] changeyourenergy, "Science proves meridians exist & how you can sense this energy." .

[15] Change Your Energy, "Science proves meridians exist and how you can sense this energy." .

[16] M. Gaillard, "Médecine chinoise et entretien de la santé," in *M-Chinese Corner*, 2025.

[17] T. M. Mitchell, *Machine Learning*. McGraw-Hill, 1997.

[18] Z. T. et Wissam Mouder ., "Systèmes de diagnostic médical de la fibrose pulmonaire, start-up," Master's thesis, Université de Mouloud Mammerie de Tizi-Ozou, 2024.

[19] Y. B. A. et Boussad Ait-Djoudi-Oufella, "Système de surveillance des malades basé sur l' iot," Master's thesis, Université de Mouloud Mammerie de Tizi-Ozou, 2024.

[20] D. O.-T. et Kociela Lemdani, "Reconnaissance et traduction de la langue des signes basée sur le deep learning pour les personnes sourdes muettes," Master's thesis, Université de Mouloud Mammerie de Tizi-Ozou, 2024.

[21] J. Kufel, K. Bargie-czek, S. Kocot, M. Kolik, W. Bartnikowska, M. Janik, ukasz Czogalik, P. Dudek, M. Magiera, A. Lis, K. Gruszczyska, and et al., "What is machine learning, artificial neural networks and deep learning?—examples of practical applications in medicine," *Diagnostics*, vol. 13, no. 15, p. 2582, 2023.

[22] L. Lazib, "Cours de module : Traitement automatique de la langue naturelle (taln)." Cours magistral, UMMTO - Département Informatique - Master 2 SII, 2024. Non publié.

[23] DataScientest, "Fonctionnement des réseaux de neurones." `https://datascientest.com/fonctionnement-des-reseaux-neurones`. Consulté en avril 2025.

[24] GeeksforGeeks, "Understanding vgg-16 cnn model." `https://www.geeksforgeeks.org/vgg-16-cnn-model/`.

[25] M. M. e. B. A. Farhad, Moomal et Masud, A. Ahmad, and L. Ahmed, "A review of medical diagnostic video analysis using deep learning techniques," *Applied Sciences (Switzerland)*, vol. 13, no. 11, p. 6582, 2023. Reçu 28 février 2023; révisé 11 mai 2023; accepté 19 mai 2023; publié 29 mai 2023.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, . Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[27] OpenAI, "Dall-e 3." `https://openai.com/index/dall-e-3`, 2025. Consulté le 11 avril 2025.

[28] DataCamp, "A comprehensive guide to the dall-e 3 api." `https://www.datacamp.com/tutorial/a-comprehensive-guide-to-the-dall-e-3-api`, 2025. Consulté le 11 avril 2025.

[29] D. A. Ray, "Transfer learning step-by-step guide." `https://amitray.com/transfer-learning-step-by-step-guide/`, 2025. Consulté le 12 avril 2025.

[30] L. J. et Li Mingyang, *Few-shot Learning: Methods and Applications*, vol. 70. EDP Sciences, 2025.

[31] R. M. et N. Malloul, "Segmentation intégrant les contours actifs hybrides pour les irm-3d du cerveau," Master's thesis, École nationale Supérieure d' Informatique, 2015.

[32] T.-Y. Lin, M. Maire, and S. e. a. Belongie, "Microsoft COCO: Common Objects in Context," 2015.

[33] scikit-learn developers, "Model Evaluation —scikit-learn 1.2.2 Documentation," 2023.

[34] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *arXiv preprint arXiv:1506.02640*, 2016.

[35] N. Rao, "Yolov11 explained: Next-level object detection with enhanced speed and accuracy," 2024.

[36] M. Guérin, "Cancer du sein : signes, stades, un cancer agressif ?." `https://sante.journaldesfemmes.fr/fiches-maladies/2505880-cancer-du-sein-depistage-stade-symptomes-traitement-survie/`, 2025. Consulté le 28 juin 2025.

[37] Z. Lezar, "Cancer du sein." ResearchGate, 2019. Contenu mis en ligne le 18 juillet 2019, consulté via ResearchGate.

[38] ESLAN, "Elsan centre-finisterien-radiotherapie-oncologie." En ligne. Consulté le [date de consultation].

[39] M. Tardy, *Apprentissage automatique pour l' aide au diagnostic précoce du cancer du sein*. Thèse de doctorat, École centrale de Nantes, Nantes, France, 2021. Soutenue le 12 octobre 2021.

[40] W. Junde, "Etude clinique des caractéristiques de l'image de la langue chez les patients atteints de cancer," *Université de médecine chinoise de Pékin*, 2011.

[41] Y. Wang, X. Wang, Z. Wang, L. Xie, and L. Cheng, "Dr. tongue: Sign-oriented multi-label tongue image analysis for remote tcm diagnosis," *arXiv preprint arXiv:2402.11808*, 2024.

[42] L. Jia and et al., "Modernizing tongue diagnosis: Ai integration with traditional chinese medicine for precise health evaluation," *IEEE Access*, 2024.

[43] J. Xie and et al., "Digital tongue image analyses for health assessment," *arXiv*, February 2022.

[44] S. Sun and et al., "Ammonia-net: A multi-task joint learning model for multi-class segmentation and classification in tooth-marked tongue diagnosis," *arXiv*, 2023.

[45] Y. Zhang and et al., "Weakly supervised object detection for automatic tooth-marked tongue recognition," *arXiv*, August 2024.

[46] S. Cao, Q. Ruan, Q. Wu, and W. Lin, "A non-invasive interpretable nafld diagnostic method combining tcm tongue features," *arXiv preprint arXiv:2309.02959v3 [eess.IV]*, 2023. Available online: `https://github.com/cshan-github/SelectorNet`.

[47] H. P. Ng, S. Y. Chong, Y. H. Li, T. H. Goh, K. Y. Pang, M. J. Pereira, and C.-M. Huang, "Objective analysis of traditional chinese medicine syndrome differentiation of patients with diabetes and prediabetes," *JMIR Research Protocols*, vol. 13, p. e56024, 2024.

[48] D. N. Lu, "A traditional chinese medicine view of breast cancer," in *A Traditional Chinese Medicine View of Breast Cancer*, 2024.

[49] L.-C. L. et al., "Breast cancer index: A perspective on tongue diagnosis in traditional chinese medicine," *Journal of Traditional and Complementary Medicine*, vol. 3, no. 3, pp. 194–203, 2013.

[50] L.-C. Lo, T.-L. Cheng, Y.-J. Chen, S. Natsagdorj, and J. Y. Chiang, "Tcm tongue diagnosis index of early-stage breast cancer," *Complementary Therapies in Medicine*, vol. 23, no. 5, pp. 705–713, 2015.

[51] F. Baochao, "Recherche sur le diagnostic de la langue mtc et la technologie de recommandation de la mtc pour le cancer du sein basée sur un graphe de connaissances et un réseau neuronal," *Université de médecine traditionnelle chinoise de Guangzhou*, 2021.

[52] S.-M. Deng, A.-F. Chiu, S.-C. Wu, Y.-C. Huang, S.-C. Huang, S.-Y. Chen, and M.-Y. Tsai, "Association between cancer-related fatigue and traditional chinese medicine body constitution in female patients with breast cancer," *Journal of Traditional and Complementary Medicine*, vol. 11, no. 1, pp. 62–67, 2021.

[53] R.-Q. Feng, D.-H. Li, X.-K. Liu, X.-H. Zhao, Q.-E. Wen, and Y. Yang, "Traditional chinese medicine for breast cancer: A review," *Breast Cancer: Targets and Therapy*, 2021.

[54] Y. H. et al., "Guideline of traditional chinese medicine diagnosis and treatment for breast cancer patients with emotional disorders (2023)," *Global Science of Traditional Chinese Medicine (GSCM)*, 2023.

[55] Y. W. et al., "Symmap: an integrative database of traditional chinese medicine enhanced by symptom mapping," *Nucleic Acids Research*, vol. 47, no. D1, pp. D1110–D1117, 2019.

[56] SymMap Project, "SymMap: A Comprehensive Platform Integrating Traditional Chinese Medicine with Modern Medicine." `https://www.symmap.org/`, 2019.

[57] Shen-Apps Team, "Shen-Acupuncture: Atlas of Acupuncture and TCM Diagnostics." `https://shen-apps.com/`, 2025.

[58] Shen-Apps Team, "Shen-acupuncture: Atlas of acupuncture and tcm diagnostics." `https://shen-apps.com/`.

[59] Kaggle, "Kaggle: Your machine learning and data science community," 2025. Accessed: 2025-06-28.

[60] Roboflow, "Roboflow: Annotate, train, deploy computer vision models," 2025. Accessed: 2025-06-28.

[61] OpenML, "Openml: The open machine learning platform," 2025. Accessed: 2025-06-28.

[62] Harvard Dataverse, "Harvard dataverse repository," 2025. Accessed: 2025-06-28.

[63] ScienceDirect, "Sciencedirect | elsevier," 2025. Accessed: 2025-06-28.

[64] IEEE DataPort, "Ieee dataport: Trusted data repository," 2025. Accessed: 2025-06-28.

[65] Baidu AI Studio, "Ai studio - powered by baidu," 2025. Accessed: 2025-06-28.

[66] Europe PMC, "Europe pmc - an open access life sciences literature database," 2025. Accessed: 2025-06-28.

[67] R. J. Chen, M. Y. Lu, T. Y. Chen, D. F. K. Williamson, F. Mahmood, *et al.*, "Synthetic data in machine learning for medicine and healthcare," *Nature Biomedical Engineering*, vol. 5, pp. 493–497, 2021.

[68] e. a. Hu, "A survey of synthetic medical image generation for improving disease classification," *Stanford University Press (ou arXiv/vierzon selon publication)*, 2024.

[69] e. a. Van Booven, "Synthetic genitourinary image synthesis via generative adversarial networks," *Journal of Personalized Medicine*, vol. 14, no. 7, p. 703, 2024.

[70] OpenAI, "Chatgpt: large language model," 2025. Modèle Consulté le 8 juin 2025 .

[71] GitHub and OpenAI, "Copilot," 2025. Consulté le 5 juin 2025.

Annexe A

Annexe - Diagramme UML

I.1 Diagramme de cas d'utilisation

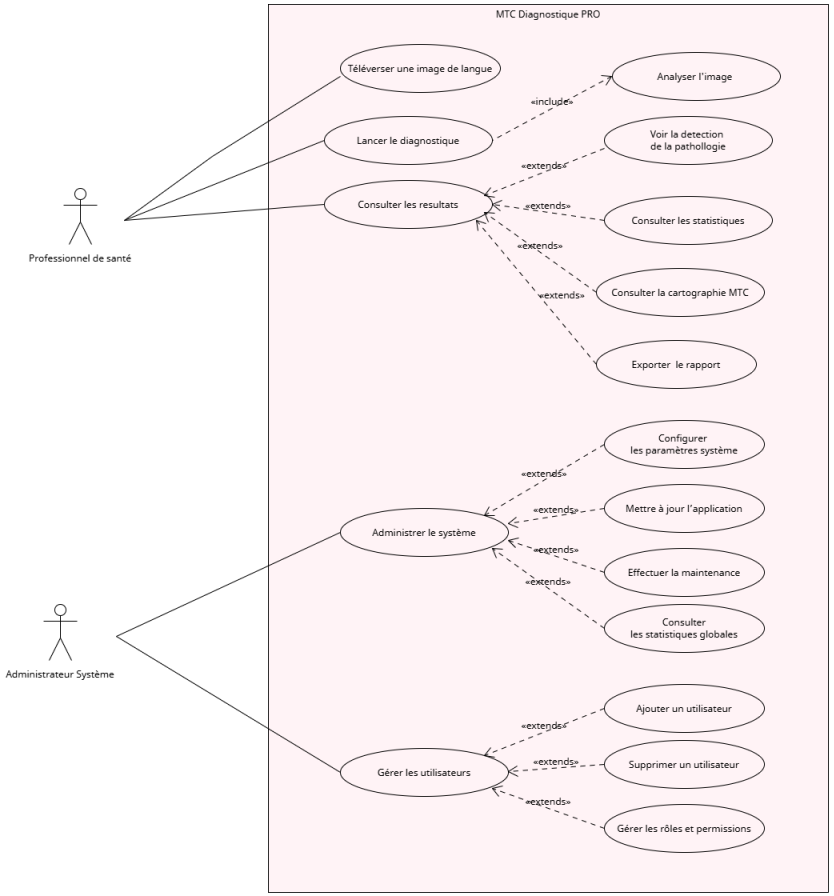


FIGURE A.1 – Annexe: Diagramme des cas d' utilisation de la plateforme MTC Hybrid Pro

I.2 Diagramme de séquences

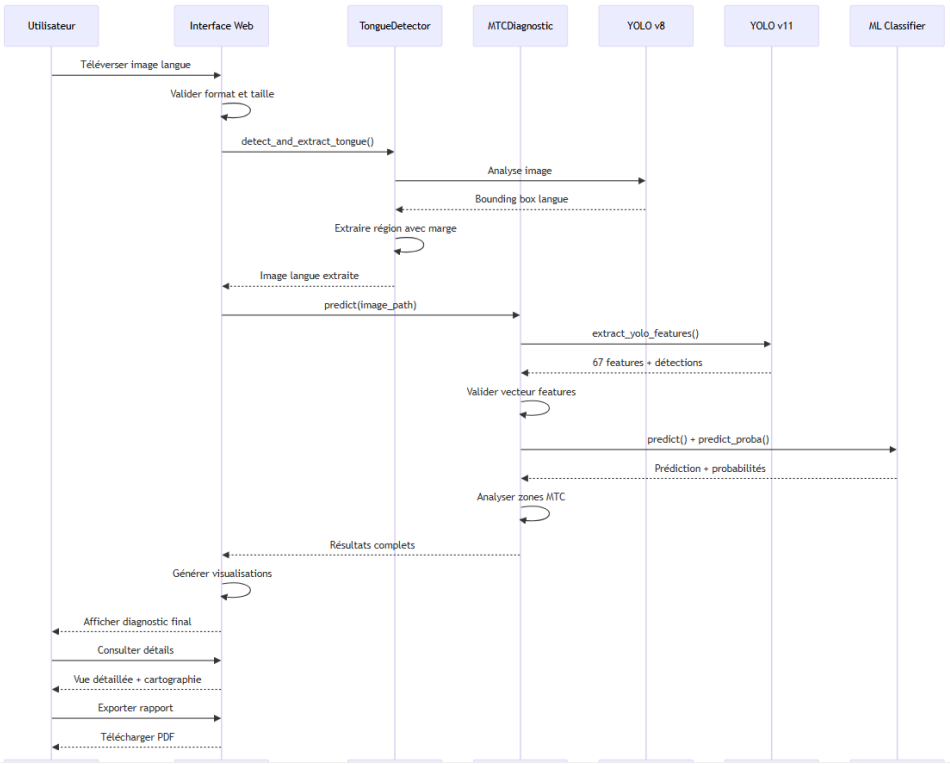


FIGURE A.2 – Annexe:Diagramme de sequence de la plateforme MTC Hybrid Pro

Ce diagramme de séquence présente le fonctionnement complet de l'application MTC Diagnostic Pro, depuis l'importation d'une image jusqu'à la génération du diagnostic final. **L'utilisateur** commence par téléverser une image de langue via **l'interface Web**. Celle-ci vérifie automatiquement le format et les dimensions avant d' appeler la méthode **detect_and_extract_tongue()**du module **TongueDetector**. Cette fonction applique **YOLOv8** pour détecter la région linguale, puis en extrait la zone pertinente. L' image extraite est transmise à **MTCDiagnostic**, qui utilise **YOLOv11** afin de détecter plusieurs signes visuels pathologiques (fissures, couleur, enduit, etc.). Chaque détection alimente un vecteur de **67 features** réparties comme suit :

- 48 features proviennent des 16 caractéristiques détectées par YOLO, chacune résumée par trois indicateurs : nombre d' occurrences, confiance maximale, confiance moyenne ;
- 10 features proviennent de la cartographie MTC, avec deux mesures (nombre et intensité des signes) pour chacune des cinq zones de la langue ;
- 9 features globales complètent l' ensemble, incluant : le nombre total de détections, la moyenne et l' écart-type des confiances, la densité des signes, leur dispersion horizontale et verticale, la surface estimée de la langue, le nombre de signes pathologiques et le nombre de signes sains.

Ce vecteur est ensuite transmis à un **classifieur ML Classifier** (ex. : SVM, MLP, XGBoost), qui retourne une probabilité de stade de la maladie : sain, précoce, avancé. Les

résultats sont ensuite visualisés via des courbes, une cartographie MTC annotée, et un rapport PDF généré automatiquement.

I.3 Diagramme de classe

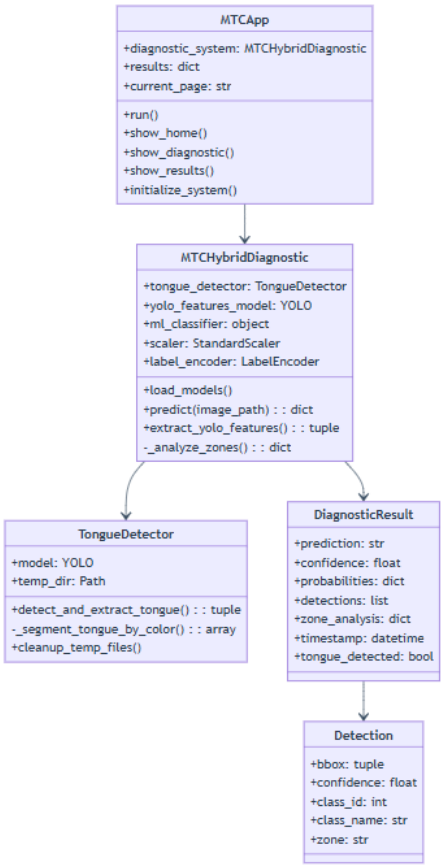


FIGURE A.3 – Annexe: Diagramme de classes de la plateforme MTC Hybrid Pro

Ce diagramme de classes illustrel' architecture objet de l' application MTC Diagnostic Pro, en mettant en évidence la séparation des responsabilités entre les différentes composantes.
La classe MTCApp joue le rôle de contrôleur principal, orchestrant l' interface utilisateur et le diagnostic. Elle repose sur **la classe MTCHybridDiagnostic**, qui regroupe l' ensemble de la logique métier du diagnostic, combinant détection de la langue, extraction de caractéristiques avec YOLO, classification par apprentissage automatique, et cartographie énergétique selon les zones MTC. Cette dernière fait appel à **la classe TongueDetector,** spécialisée dans l'extraction de la langue à partir des images. Bien que **la classe DiagnosticResult** ne soit pas explicitement implémentée dans le code, elle est utilement modélisée dans le diagramme pour structurer les résultats retournés sous forme de dictionnaire. De même, **la classe Detection** formalise les objets représentant les caractéristiquesădétectées, on peut imaginer cette classe comme une façon de structurer chaque "signe clinique visuel" identifié sur la langue, par exemple: Fissure avec confiance 0.87 à tel endroit.

Annexe B

Annexe - Captures de la plateforme

II.1 Annexe:Tests de chargement d'images dans la plateforme

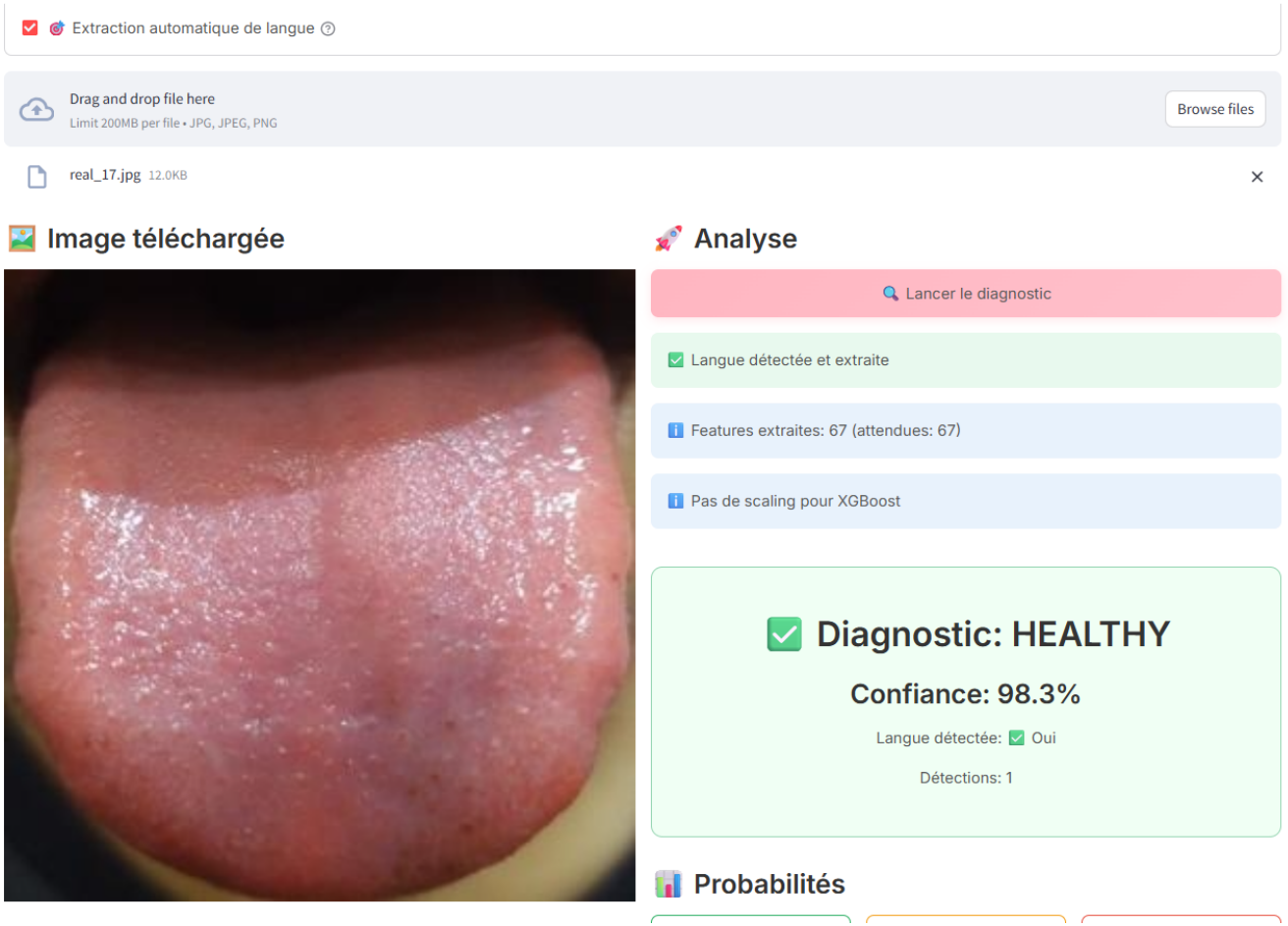


FIGURE B.1 – Annexe: Image healthy détectée avec sucés

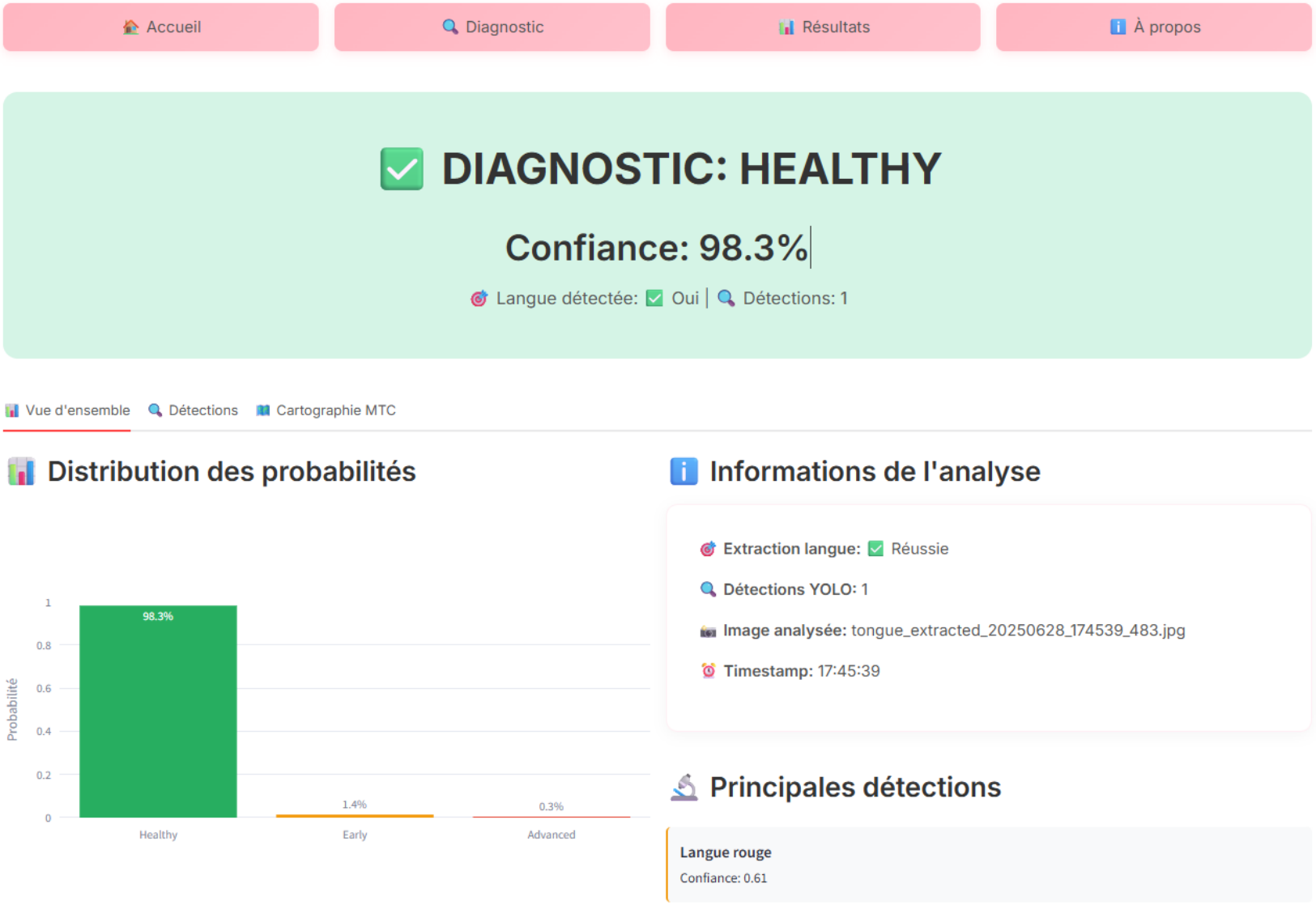


FIGURE B.2 – Annexe:Image healthy et statistiques

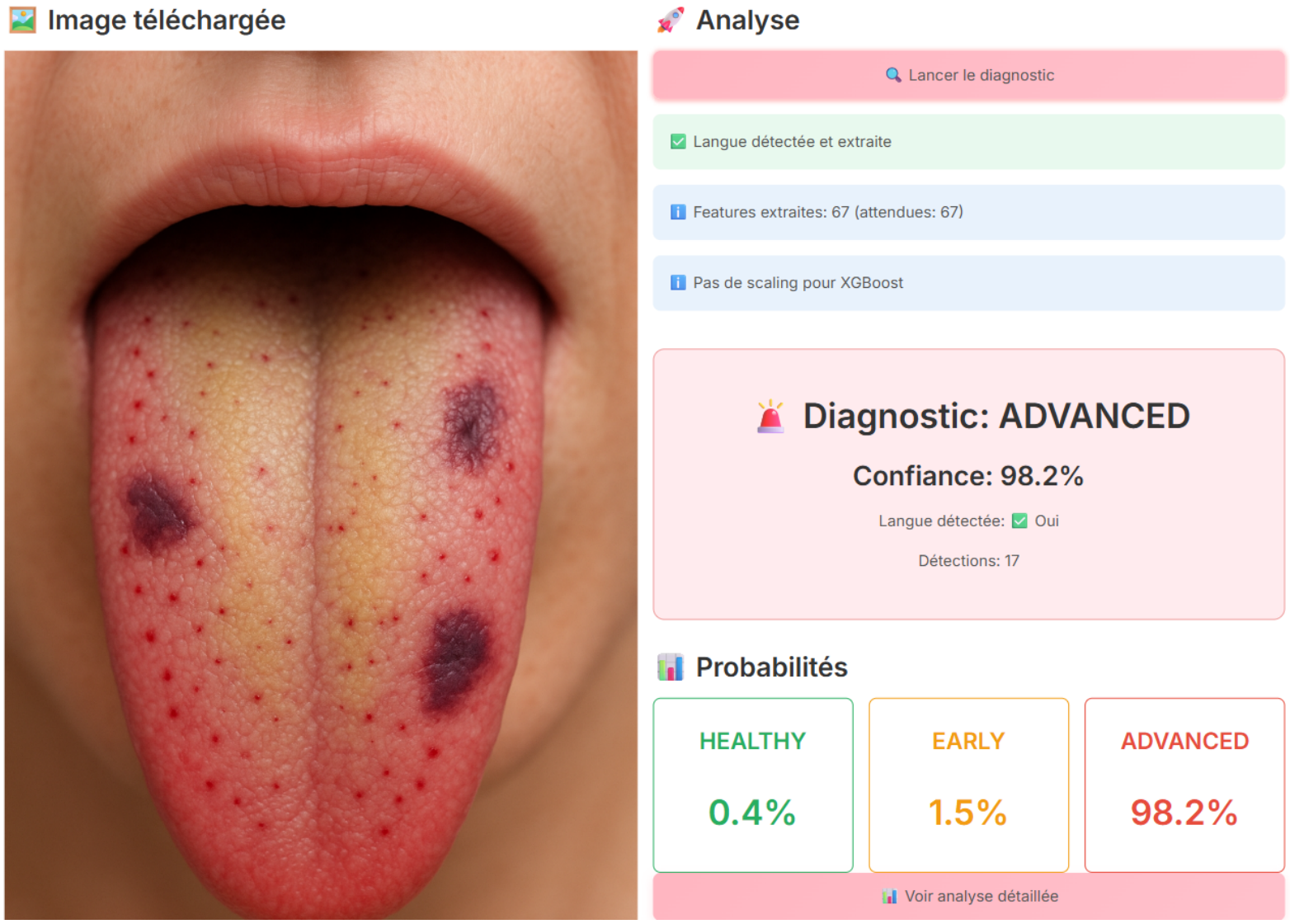


FIGURE B.3 – Annexe: Image stade avancé détecté avec succès avec les statistiques

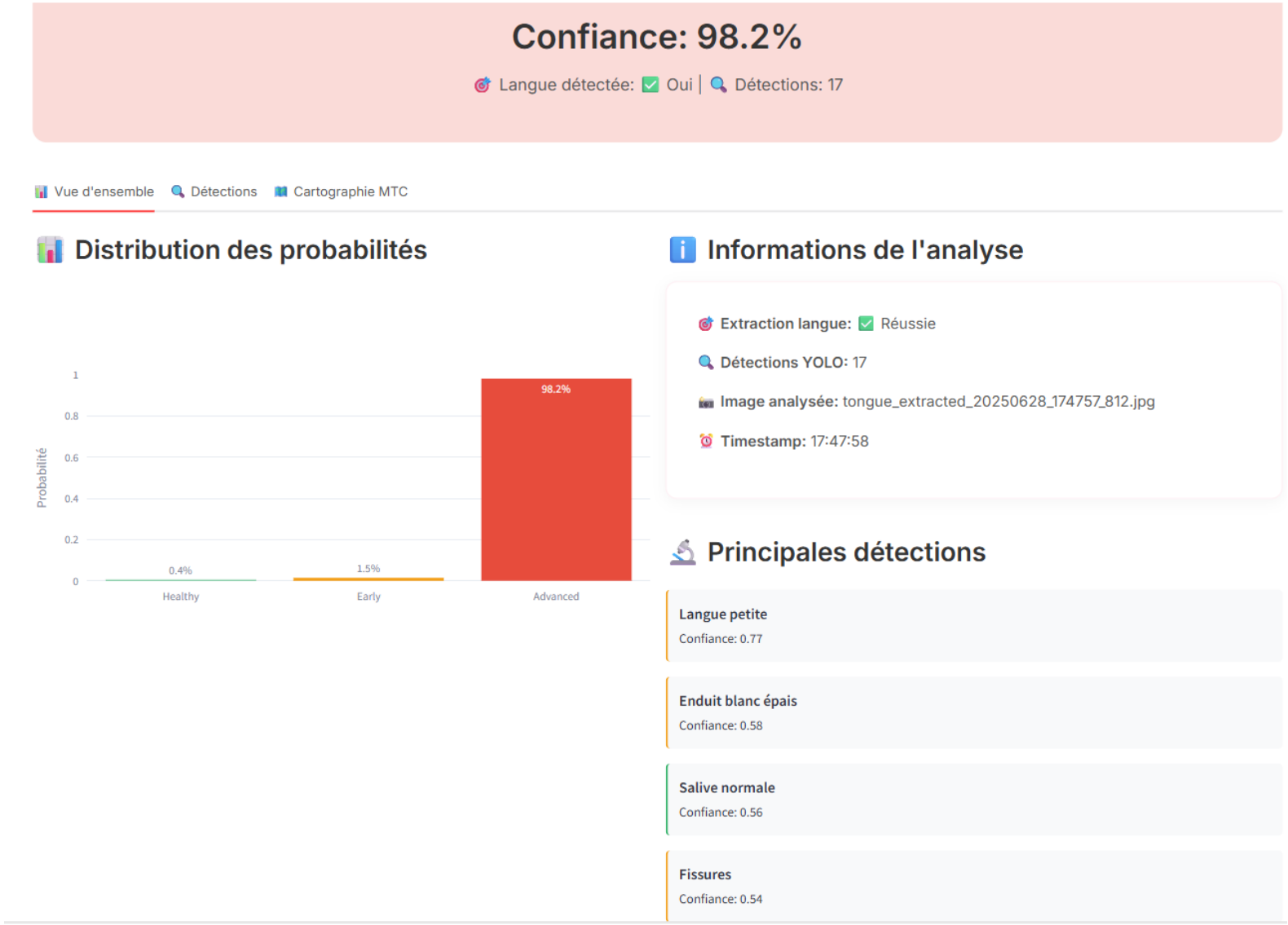


FIGURE B.4 – Annexe:Image stade avancé détecté avec succès avec les statistiques

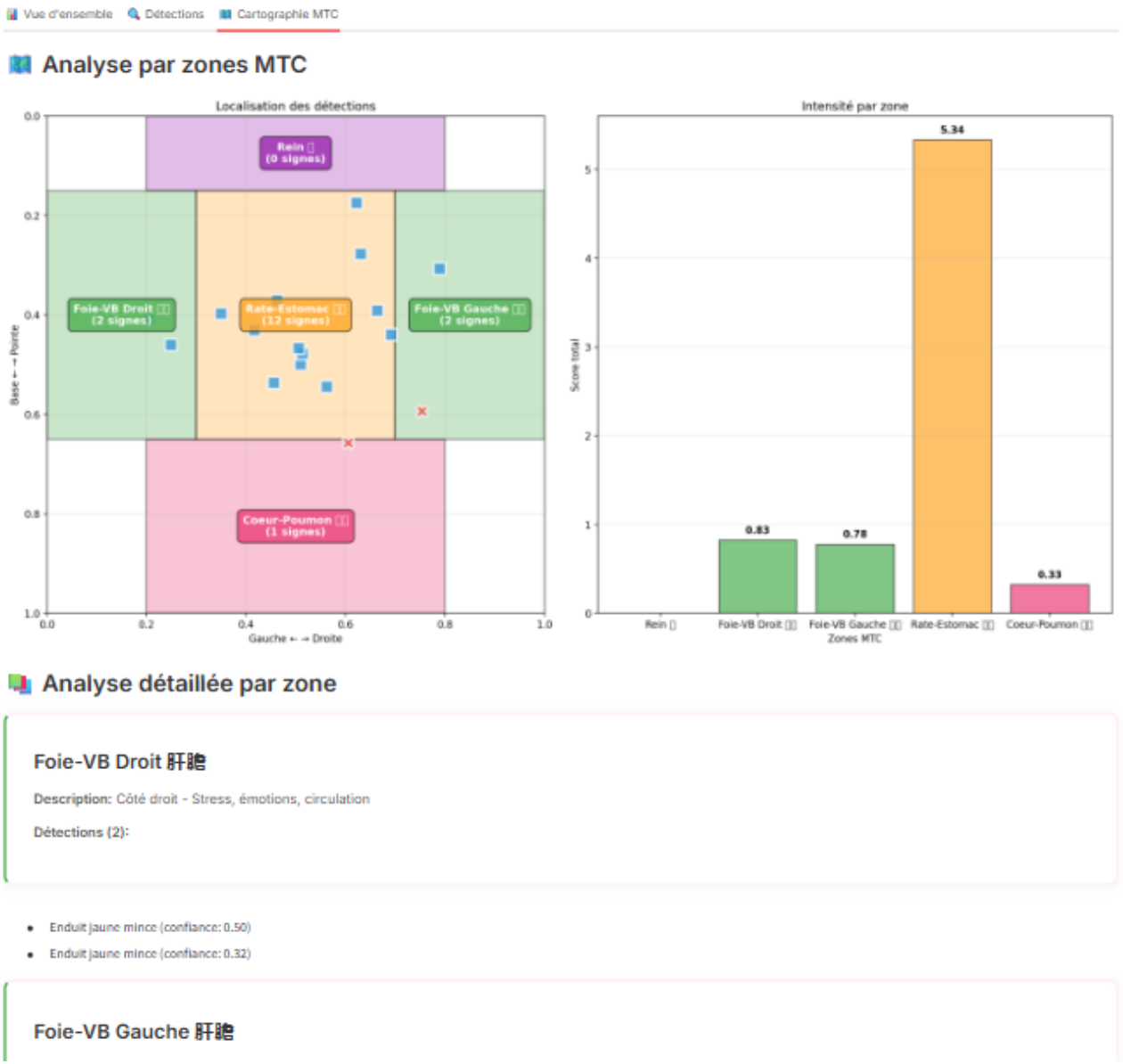
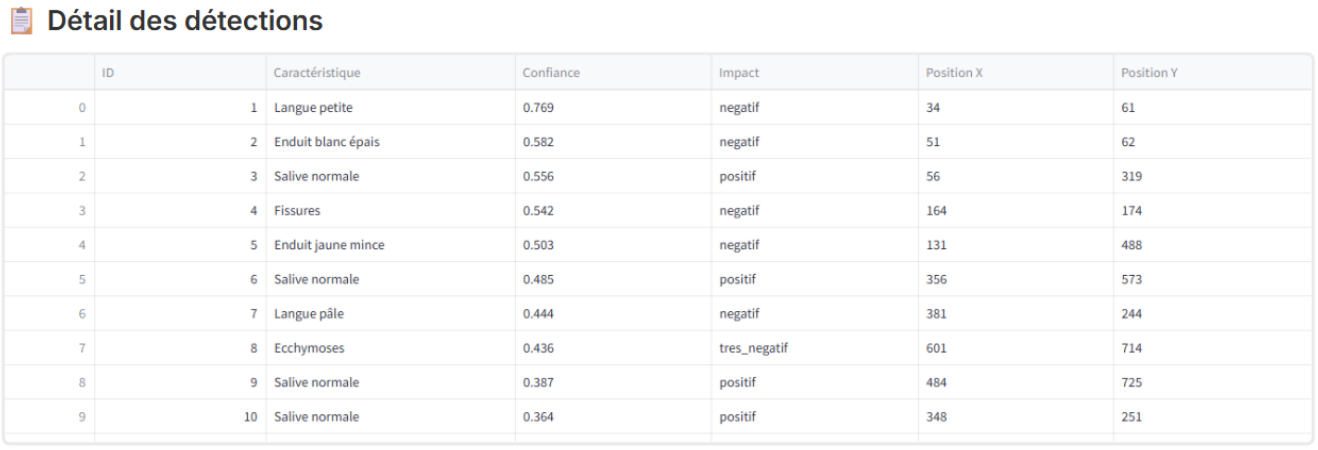


FIGURE B.5 – Annexe:Image stade avancé avec cartographie



FIGURE B.6 – Annexe: Image stade avancé avec détection des caractéristiques sur la langue



Détail des détections

	ID	Caractéristique	Confiance	Impact	Position X	Position Y
0	1	Langue petite	0.769	negatif	34	61
1	2	Enduit blanc épais	0.582	negatif	51	62
2	3	Salive normale	0.556	positif	56	319
3	4	Fissures	0.542	negatif	164	174
4	5	Enduit jaune mince	0.503	negatif	131	488
5	6	Salive normale	0.485	positif	356	573
6	7	Langue pâle	0.444	negatif	381	244
7	8	Ecchymoses	0.436	tres_negatif	601	714
8	9	Salive normale	0.387	positif	484	725
9	10	Salive normale	0.364	positif	348	251

FIGURE B.7 – Annexe: Image stade avancé avec détails sur les caractéristiques détectées

Département d'Informatique
Faculté des Genie Eectrique et Informatique
Université Mouloud Mammeri
Tizi-Ouzou