



République Algérienne Démocratique et Populaire



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Akli Mohand Oulhadj de Bouira

Faculté des Sciences et des Sciences Appliquées

Département d'Informatique

Laboratoire de recherches LIMPAF

Mémoire de Master

en Informatique

Spécialité(s) : Génie des Systèmes Informatiques

et

Ingénierie des Systèmes d'informations et Logicielles

Thème

Dépistage et classification des tumeurs pancréatiques
par apprentissage automatique et profond.

Encadré par

— Lyes BADIS

— Hocine CHEBI

Réalisé par

— Amine AIT SAHED

— Saida HAMIDI

2021/2022

Remerciements

Tout d'abord on tiens à remercier *Dieu* pour nous avoir donnée la volonté, la force et la patience pour accomplir ce travail tout au long de cette période.

Nous tenons à remercier particulièrement *nos parents* pour leurs sacrifices tout au long de notre parcours éducatif.

Nous voudrions dans un premier temps remercier, notre encadreur de thèse **Mr. Lyes *BADIS***, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribués à alimenter nos réflexions.

Nos remerciements à tous les gens qui nous ont soutenu durant ce travail :
amis, ami(es) et collègues de promotion.

Merci.

Dédicaces

Je dédie ce travail à
mes *chers parents*,
ma soeur et son mari,
mes neveux
ainsi qu'à
toute ma précieuse famille...

Sans oublier
ma binôme Saida
et
mes chers ami(s) et ami(es)...

Amine AIT SAHED.

Dédicaces

Je dédie ce mémoire tout d'abord

A mes parents qui m'ont apportés tout le confort et les moyens pour avoir le meilleur parcours possible tout en me soutenant tout au long de mon cursus universitaire en me poussant à donner le maximum de moi-même.

A l'homme de ma vie, mon cher mari Ismail, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, que Dieu le tout puissant vous garde pour moi.

A celui qui m'a fait de moi une heureuse ambitieuse maman, à celui qui a illuminé notre vie de sa présence et lui a donné un autre sens, mon petit poussin Rassim.

A mon cher frère Mohsaid, mon idole, à qui je lui souhaite bonne santé et réussite, ainsi qu'à sa femme, ma chère Amira, et à leur cher fils, mon bien-aimé Iyad.

A mon cher frère mohamed.

Il va trouver ici l'expression de mes sentiments de respect et de reconnaissance pour le soutien qu'il n'a cessé de me porter, qui a partagé avec moi tous les moments d'émotions lors de la réalisation de ce travail.

A mon âme soeur Koukou au nom de l'amitié qui nous a réunis et au nom de nos souvenirs inoubliables, sa générosité et son soutien m'oblige à témoigner mon profond amour et ma loyale considération .

A ma famille, mes cousines, mes proches et à ceux qui me donnent de l'amour et de la vivacité . A tous ceux qui, par un mot, m'ont donné la force de continuer.

A ma chère douce amoureuse Hajoura, a mes amies Sara, Faiza, Saliha, Sihem, Hassiba .

A ma chère Khadidja pour son soutien.

Sans oublier mon binôme Amine pour son soutien moral, sa patience, sa compréhension tout au long de ce projet, à qui je souhaite de rester toujours brillant et d'avoir plus de succès.

A mes collègues de l'université Imen, Sara, Riane que j'ai trouvés à mes côtés lors des moments difficiles et qui ont essayés de m'aider de diverses manières et à qui je souhaite plus de succès.

Saida HAMIDI

Table des matières

Table des matières	i
Table des figures	v
Liste des tableaux	ix
Liste des abréviations	xi
Introduction générale	1
1 Contexte médical	4
1.1 Introduction	4
1.2 Terminologies médicales	4
1.2.1 Anatomie de l'appareil digestif	4
1.2.2 Anatomie du Pancréas	5
1.3 Processus de fonctionnement	7
1.3.1 Le Pancréas	7
1.3.2 Les glandes exocrines	7
1.3.3 Les glandes endocrines	8
1.4 Tumeurs du Pancréas	9
1.4.1 Définition de la tumeur du Pancréas	9
1.4.2 Catégories de tumeurs pancréatiques existantes	10
1.4.3 Développement et Symptômes du cancer du Pancréas	12
1.4.4 Classification des tumeurs pancréatiques	14

1.5	Techniques de dépistage médicales des lésions	
	Pancréatiques	16
1.5.1	L'imagerie médicale	16
1.5.2	L'Imagerie par Résonance Magnétique – (IRM)	19
1.5.3	Le Scanner Toraco - Abdomino - Pélvien – (TAP)	21
1.6	Examen histologique et cytologique	23
1.6.1	La Biopsie	23
1.7	Diagnostic et pronostic du cancer du Pancréas	24
1.7.1	Diagnostic	24
1.7.2	Pronostic	27
1.8	Traitement contre le cancer du Pancréas	28
1.8.1	La chirurgie	28
1.8.2	La chimiothérapie et radiothérapie	31
1.9	Causes et conseils	31
1.9.1	Causes liées au cancer du Pancréas	31
1.9.2	Conseils et préventions	32
1.10	Conclusion	32
2	État de l'art	33
2.1	Introduction	33
2.2	Le Machine Learning	33
2.2.1	Définition	33
2.2.2	Principe du Machine Learning	34
2.2.3	Types de machine learning	34
2.2.4	Les algorithmes du machine learning	35
2.3	Le Deep learning	45
2.3.1	Définition	45
2.3.2	Fontionnement du Deep learning	45
2.3.3	Algorithmes de Deep Learning	48
2.3.4	Les couches du Réseau de Neurones Convolutif - (CNN)	49
2.3.5	Les différentes Architectures du Réseaux de Neurones Convolutifs - (CNN)	53
2.4	La Matrice de Confusion	56

2.4.1	Definition	56
2.4.2	Calculer une Matrice de Confusion	57
2.4.3	Évaluation d’une Matrice de Confusion	57
2.4.4	Problème d’overfitting	59
2.5	L’image	63
2.5.1	Définition	63
2.5.2	Les types d’images	63
2.6	La différence entre le Machine Learning et le Deep Learning	64
2.6.1	Domaines d’applications de l’intelligence artificielle	65
2.7	L’imagerie médicale et l’intelligence artificielle	66
2.8	Travaux de références sur le cancer du Pancréas	67
2.9	Conclusion	67
3	Approche et solution proposée	68
3.1	Introduction	68
3.2	Techniques de dépistages Proposées - (Par Analyses médicales & Imageries médicales)	68
3.2.1	Dépistage par Analyses médicales	68
3.2.2	Dépistage par Imageries médicales	80
3.3	Plateforme de dépistage et diagnostic des tumeurs pancréatiques par CT-SCAN	91
3.4	Conclusion	91
4	Résultats et analyse	92
4.1	Introduction	92
4.2	Environnement de Développement	92
4.2.1	Google-Colaboratory	92
4.3	Librairies et Outils d’implémentation	93
4.3.1	Keras	93
4.3.2	Matplotlib	93
4.3.3	Seaborn	93
4.3.4	Tensorflow	93
4.3.5	Numpy	94

4.3.6	Pandas	94
4.4	Langages de développement informatique utilisés	94
4.4.1	Python	94
4.4.2	HTML & CSS	94
4.4.3	JavaScript	95
4.5	Frameworks et Serveurs	95
4.5.1	Flask	95
4.5.2	Ajax	95
4.6	Évaluation des techniques de dépistages utilisées	96
4.6.1	Dépistage par Analyses Médicales	96
4.6.2	Dépistage par Imagerie Médicales CT-SCAN	108
4.6.3	Résultats d'entraînement et Tests	108
4.7	Simulation de Prédiction	113
4.7.1	Dépistage par Analyses médicales	113
4.7.2	Plateforme de dépistage par imagerie médicale - (CT-SCAN)	114
4.8	Conclusion	120
 Conclusion générale		 121
 Perspective		 123

Table des figures

1.1	Anatomie de l'appareil digestif.	5
1.2	Anatomie du Pancréas.	7
1.3	Histologie du Pancréas.	9
1.4	Tumeurs du Pancréas.	11
1.5	Développement du cancer du Pancréas.	12
1.6	Symptômes du cancer du Pancréas.	13
1.7	Dépistage par échographie.	17
1.8	Fonctionnement d'une échographie.	17
1.9	Dépistage par endoscopie.	18
1.10	Déroulement d'une endoscopie.	19
1.11	Technique d'imagerie par dépistage à IRM.	20
1.12	Fonctionnement d'une IRM.	21
1.13	Technique d'imagerie par dépistage de Scanner.	22
1.14	Fonctionnement d'un Scanner.	22
1.15	Déroulement d'un examen de biopsie.	23
1.16	Les stades de la maladie.	27
1.17	Dilatation du canal biliaire comprimé par une tumeur au niveau de la tête du pancréas par une prothèse biliaire.	28
1.18	Intervention chirurgicale permettant de retirer une tumeur localisée au ni- veau de la tête du pancréas.	29
1.19	Intervention chirurgicale permettant de retirer une tumeur localisée au ni- veau de la partie libre (corps) du pancréas.	29

1.20	Intervention chirurgicale permettant de retirer une tumeur localisée au niveau de la queue du pancréas.	30
2.1	Exemple de figure montrant le principe du Decision Tree.	38
2.2	Exemple de figure montrant le principe du Random Forest.	41
2.3	Exemple de figure montrant le principe du K-Nearest Neighbors.	44
2.4	Architecture d'un réseau de neurones.	45
2.5	Principe d'un réseau de neurones.	47
2.6	Les sous domaines de l'intelligence Artificielle.	47
2.7	Concept de filtrage d'image au niveau de la couche convolution.	49
2.8	Concept du pooling au niveau de la couche Pooling.	51
2.9	Architecture d'un réseau de neurones convolutifs.	52
2.10	Architecture du Réseau de neurones Pré-entraînés [<i>Inception-V3</i>].	53
2.11	Architecture du Réseau de neurones Pré-entraînés [<i>Exception</i>].	54
2.12	Architecture du Réseau de neurones Pré-entraînés [<i>Densenet-121</i>].	55
2.13	Algorithmes des réseaux de neurones.	56
2.14	Principe de l'overfitting – (<i>Cas possibles</i>).	59
2.15	Le principe de la méthode du <i>K-fold</i>	61
2.16	Le principe de la méthode du <i>Stratified K-fold</i>	62
2.17	Le principe de la méthode du <i>Leave One out</i>	62
2.18	Image Matricielle VS Image Vectorielle.	63
2.19	Mécanisme de fonctionnement du Machine Learning et du Deep Learning.	65
2.20	L'intelligence artificielle au service du diagnostic médical.	66
3.1	Solution Proposée pour la première approche de dépistage - (<i>Dépistage par analyses médicales</i>).	69
3.2	Visualisation des données labélisées des trois – (03) classes de diagnostics se trouvant dans le Dataset au Format numérique.	71
3.3	Aperçu du Dataset avant l'augmentation de données.	74
3.4	Aperçu du Dataset après l'augmentation de données.	75
3.5	Solution Proposée pour la deuxième approche de dépistage - (<i>Dépistage par imageries médicales</i>).	80

3.6	Visualisation des images labélisées des deux – (02) classes de diagnostics se trouvant dans le Dataset des CT-SCAN.	82
3.7	Processus de fonctionnement du Model-Checkpoint durant la phase d'apprentissage.	85
3.8	Architecture du modèle Xception entraîné avec le classificateur Softmax	90
3.9	Architecture du modèle DenseNet121 entraîné avec le classificateur Softmax	90
4.1	Performance des deux Modèles utilisés.	97
4.2	Matrices de Confusion des deux Modèles - <i>Random Forest Classifier</i> et <i>K-Nearest Neighbors</i> Avant <i>l'augmentation des données</i>	98
4.3	Performance des deux Modèles utilisés.	100
4.4	Matrices de Confusion des deux Modèles - <i>Random Forest Classifier</i> et <i>K-Nearest Neighbors</i> Après <i>l'augmentation de données</i>	101
4.5	Matrice de Corrélation Après <i>l'augmentation de données</i>	102
4.6	Performance des deux Modèles utilisés.	104
4.7	Matrices de Confusion des deux Modèles - <i>Random Forest Classifier</i> et <i>K-Nearest Neighbors</i> après optimisation avec <i>la Cross-Validation • (K-fold)</i>	105
4.8	Comparaison entre les Performances des modèles utilisés - <i>Random Forest Classifier</i> et <i>K-Nearest Neighbors</i>	106
4.9	Performance du Modèle - (<i>Xception + Softmax</i>).	109
4.10	Performance du Modèle (<i>Densenet121 + Softmax</i>).	110
4.11	Matrices de Confusion des deux Modèles - <i>Xception + Softmax</i> et <i>Densenet-121 + Softmax</i>	110
4.12	Comparaison entre les Performances des modèles utilisés - <i>Xception + Softmax</i> et <i>Densenet121 + Softmax</i>	111
4.13	Performance de prédiction du Modèle <i>Random Forest Classifier</i> après Optimisation avec <i>Data-augmentation</i> et <i>Cross-Validation • K-fold</i>	113
4.14	Aperçu général de la plate-forme Medi-Predict.	114
4.15	Simulation avec l'objet inspect sur un CT-SCAN non-cancéreux.	115
4.16	Simulation avec l'objet inspect sur un CT-SCAN cancéreux.	115
4.17	Simulation avec l'objet filter sur un CT-SCAN non-cancéreux.	116

4.18 Simulation avec l'objet filter sur un CT-SCAN cancreux.	116
4.19 Simulation avec l'objet crop pour évaluer la taille de la tumeur.	117
4.20 Simulation de prédiction sur un CT-SCAN Non-cancereux.	118
4.21 Simulation de prédiction sur un CT-SCAN cancreux.	119

Liste des tableaux

1.1	Les types de cellules cancéreuses exocrines.	14
1.2	Les types de cellules cancéreuses endocrines.	15
1.3	Comportement biologique des tumeurs pancréatiques.	16
2.1	Les avantages et inconvénients de l'arbre de décision - (Decision Tree). . .	37
2.2	Les avantages et inconvénients de la forêt aléatoire - (Random forest classifier).	40
2.3	Les avantages et inconvénients du K-Plus Proches Voisins - (K-Nearest Neighbors).	43
2.4	Métriques de calcul d'une matrice de confusion.	58
2.5	La différence entre le Machine Learning et le Deep Learning.	64
2.6	Travail de références sur le cancer du Pancréas.	67
3.1	Répartition des données – (Entraînement et Test) du Dataset au Format Numérique utilisé dans le Machine Learning.	70
3.2	Répartition des données – (Entraînement, Validation et Test) du Dataset des images médicales utilisé dans le Deep Learning.	81
4.1	Résultats Primitifs du <i>Random Forest Classifier</i>	96
4.2	Résultats Primitifs du <i>K-Nearest Neighbors</i>	96
4.3	Résultats du Random Forest Classifier après <i>l'augmentation de données</i>	99
4.4	Résultats du K-Nearest Neighbors après <i>l'augmentation de données</i>	99
4.5	Résultats du Random Forest Classifier après optimisation avec <i>la Cross-Validation • (K-fold)</i>	103

4.6	Résultats du K-Nearest Neighbors après optimisation avec <i>la Cross-Validation • (K-fold)</i>	103
4.7	Modèle final sollicité pour effectuer des prédictions et diagnostics.	105
4.8	Comparaison de performances de nos modèles <i>Random Forest Classifier</i> et <i>K-Nearest Neighbors</i> avec les performances d'autres études menées sur la même thématique.	107
4.9	Résultats du Modèle <i>Xception</i> combiné avec la fonction d'activation <i>Softmax</i>	108
4.10	Résultats du Modèle <i>Densenet121</i> combiné avec la fonction d'activation <i>Softmax</i>	108
4.11	Modèle final sollicité pour effectuer des prédictions et diagnostics.	111

Liste des abréviations

API	Application Programming Interface.
IRM	Imagerie à Résonance Magnétique.
CT-SCAN	Computer Tomodensometrie Scanographie.
TAP	Scanner Toraco Abdomino Pélvien
TNM	<i>Tumor - (tumeur) Node – (ganglion) Metastasis – (métastase).</i>
TKS/TKM	Tumeur Kystique Séreuse / Tumeur Kystique Mucineuse.
TIPM	Tumeur Intra-canaulaire Papillaire et Mucineuse.
LR	Logistic Regression - (Regression Logistique).
DT	Decision Tree - (Arbre de Décision).
RFC	Random Forest Classifier - (Classificateur de Forêts Aléatoires).
K-NN	K-Nearest Neighbors - (K-Plus Proches Voisins).
NB	Naive Bayes.
SVM	Support Vector Machine - (Machine à vecteurs de support).
CNN	Convolutional Neuron Network - (Réseaux de Neurones Convolutifs).
TP	True Positive - (Vrai Positif).
TN	True Negative - (Vrai Négatif).
FP	False Positive - (Faux Positif).
FN	False Negative - (Faux Négatif).
GPU	Graphics Processing Unit - (Processeur Graphique).
CPU	Center Processing Unit - (Unité Centrale de Traitement).
TPU	Tensor Processing Unit - (Unité de Traitement du Tenseur).
HTML	Hyper Text Markup Language.
CSS	Cascading Style Sheet.

Abstract

The aim of this project project is to link artificial intelligence technologies *Machine Learning* and *Deep Learning* to the field of medicine and specifically to the screening of pancreatic cancer. The study was conducted on the basis of (02) two distinct approaches which are screening by medical analysis and screening by medical imaging. The first approach consists in making predictions on the basis of urine tests performed on patients and linking its data to learning models optimized by two methods (Data Augmentation and Cross Validation with the K-fold technique) so that they can make accurate and reliable predictions on other external data to be provided. In a comparative study between the two algorithms used (*Random Forest Classifier*) and (*K-Nearest Neighbors*), it was found that the *Random Forest Classifier* model provided the best results, i.e. 98.87 % in the training phase and 99.33 % in the test phase. Regarding the second approach, we used the technique of *Transfer-Learning* which proposes models that are already trained on the Imagenet database such as : *Xception* and *Densenet121* which are known to provide high accuracies on a large number of classes. Comparing the two models used, we selected the *Xcpetition* model for the high performances it provided i.e. : 98.14 % - (*Accuracy*) and 3.92 % - (*Loss*). By adapting its models to our dataset in the form of CT-SCAN, they provided very satisfactory results that we wanted to project on a platform that offers interesting tools that improve the handling aspect of CT-SCAN and offer a better diagnostic experience.

Key words : *Machine Learning, Deep Learning, Transfert-Learning, K-fold, Imagenet, CT-SCAN.*

Résumé

Le but de ce projet est de lier les technologies de l'intelligence artificielle *Machine Learning* et *Deep Learning* au domaine de la médecine et précisément au dépistage du cancer du Pancréas. L'étude s'est menée sur la base de (02) deux approches distinctes qui sont *le dépistage par analyses médicales* et *le dépistage par imageries médicales*. La première approche consiste à faire des prédictions sur la base des tests urinaires effectués sur des patients et de lier ses données à des modèles d'apprentissages optimisés par deux méthodes (Augmentation de données et Cross validation avec la technique du *K-fold*) pour qu'ils puissent faire des prédictions précises et fiables sur d'autres données externes qui seront fournies. En effectuant une étude comparative entre les deux algorithmes qu'on a utilisé (*Random Forest Classifier*) et (*K-Nearest Neighbors*), on a constaté que le modèle *Random Forest Classifier* a fourni les meilleurs résultats c-à-d : 98.87 % - dans la phase d'entraînement et 99.33 % - dans la phase de tests. En ce qui concerne la deuxième approche, on a utilisé la technique du *Transfert-Learning* qui propose des modèles qui sont déjà entraînés sur la base de données Imagenet tels que : *Xception* et *Densenet121* qui sont réputés pour fournir des précisions élevées sur un nombre important de classe. En comparant les deux modèles utilisés, on a sélectionné le modèle *Xception* pour les performances élevées qu'il a fourni c-à-d : 98.14 % - (*Accuracy*) et 3.92 % - (*Loss*). En adaptant ses modèles à notre dataset sous forme de CT-SCAN, ils ont fourni des résultats très satisfaisants qu'on a voulu projeter sur une plate-forme qui propose des outils intéressants qui améliorent l'aspect manipulation des CT-SCAN et offrent une meilleure expérience de diagnostic.

Mots clés : *Machine Learning, Deep Learning, Transfert-Learning, K-fold, Imagenet, CT-SCAN.*

Introduction générale

La santé d'un individu peut être décrite comme un atout en qui il dépend tout au long de sa vie. Ceci dit, avec l'avancée de l'âge et l'exposition à une mauvaise hygiène de vie, quelques soucis peuvent entraîner sa dégradation et donner naissance à des maladies délicates tels que le cancer.

Le Pancréas est l'un des organes qui assurent des fonctions indispensables pour le bon fonctionnement du corps humain. Cet organe est anatomiquement enfoui entre les autres organes de l'abdomen est constitué de trois (03) parties : la tête, la corp et la queue et mesure environs 15cm de long et à l'aspect biologie d'une feuille, ce qui rend son diagnostic un peu complexe vis-à-vis des médecins. Enfin, il assure deux (02) fonctions : exocrines et endocrines avec lesquelles aident à la digestion des aliments ainsi qu'à la régularisation de l'indice glycémique dans le sang.

Le cancer est une maladie qui surgit lorsque les cellules dont est constitué chaque organe du corps humain subissent un dérèglement biologique, ce qui cause la dégénérescence de tissus qui tapissent ses organes et donc leurs dysfonctionnements. Cependant, certains cancers ont la caractéristique de se manifester de manière silencieuse et ne peuvent malheureusement pas être dépistés à un stade précoce. Parmi les cancers qui sont répertoriés avec cette caractéristique, on cite le cancer du Pancréas. En effet, le cancer du Pancréas est très rare de nos jours, d'après les statistiques seulement (1,7 % des cas en Algérie) qui sont dépistés par an d'après l'article publié par l'organisme "*International Agency for Reasearch on Cancer, World Health organization*" - Algeria • Source : *Globocan 2020* en mars 2021.

Toutefois, le type de cancer du Pancréas qui est souvent diagnostiqué c'est l'*Adénocarcinome-canalair*e qui affecte les cellules exocrines du parenchyme et qui peut se manifester dans les trois (03) régions de l'organe : la tête, le corps et la queue. Lorsque, le cancer se développe au niveau de la tête du Pancréas, quelques signes cliniques peuvent apparaître et donner allusion à la suspicion d'un cancer ce qui permet de donner une piste aux médecins pour effectuer un diagnostic, entamer un traitement adéquat et ainsi avoir une idée sur la nature du pronostic vital du patient. Par contre, si le cancer se développe dans les deux (02) autres parties du Pancréas, des signes cliniques peuvent apparaître mais tardivement ce qui rend le diagnostic plus délicat et le pronostic vital sombre car des métastases peuvent se former au sein d'autres organes voisins.

L'intelligence artificielle en alliance avec la médecine, a mis en place des techniques de dépistages tels que le *Machine Learning* et le *Deep Learning* très performantes qui assistent les médecins lors de la phase de diagnostics, ce qui a apporté une solution radicale à la contrainte du facteur temps du dépistage et une meilleure procédure vis-à-vis du traitement de la maladie. Ces techniques fonctionnent avec des algorithmes qui permettent à la machine d'apprendre à faire des diagnostics sur des *données structurées - (numériques)* et *non-structurées - (Images)* avec une complexité accrue afin d'atteindre une précision fiable au point d'égaliser l'interprétation humaine. Enfin, toute cette technologie à permis de mieux cerner ce type de maladie et de sauver des milliers de vie qui auraient été induites en péril.

- Dans ce rapport, on va aborder le processus de fonctionnement de ces techniques de manière approfondie et montrer leurs apports sur le dépistage du cancer du Pancréas.
- L'objectif de ce travail est de réaliser une étude théorique et pratique sur les techniques de dépistage et classification des tumeurs du pancréas et en extraire les failles sur le plan déductif du diagnostic médicale et la prononciation du pronostic vital, pour pouvoir les améliorer et offrir une meilleure précision lors du traitement des datasets.

- Notre contribution dans ce travail s'est focalisée sur la conception et implémentation d'une plate-forme qui propose des outils performants qui aident dans la phase de diagnostic du cancer du pancréas.
- La structure de notre rapport se présente selon la chronologie suivante :

Chapitre 01 : Contexte médical

Dans ce chapitre on sera amenés à faire une étude approfondie sur la maladie pour mieux cerner son comportement, ces différentes manières d'évolution à travers des techniques de diagnostic modernes utilisés de nos jours pour le dépistage précisément le CT-SCAN.

Chapitre 02 : État de l'art

Ce chapitre à été consacré pour l'étude des techniques de l'intelligence artificielle telles que : *le machine learning et deep learning* sollicités dans la partie traitements et analyses des données existantes pour fournir des éléments concrets nécessaire au dépistage et diagnostic de la maladie et localiser les failles qui rendent l'étude moins performante et archaïque pour pouvoir proposer une solution plus rigide qui va remédier au problème posé.

Chapitre 03 : Approche et solution proposée

Dans ce chapitre nous allons présenter les approches à implémenter proposées, intitulées : ***dépistage par analyses médicaux et dépistage par imagerie médicale*** sur le plan implémentation ainsi qu'un avant-projet sur la contribution (plate-forme) à concerver pour offrir une meilleure expérience de dépistage sur la base d'image CT-SCAN.

Chapitre 04 : Résultats et analyse

Ce dernier chapitre résume la phase de tests après implémentation, pour avoir un aperçu sur l'algorithme conçu sur le point performance et temps de traitement sur une grande base de datasets, et le comparer avec les algorithmes modernes. Cette phase permet de valider l'efficacité de la solution proposée ou de l'ajuster dans le cas contraire pour obtenir le résultat attendu.

Contexte médical

1.1 Introduction

L'appareil digestif comporte plusieurs organes vitaux qui assurent des fonctions indispensables pour le bon fonctionnement du corps humain. Ces organes s'occupent de la décomposition alimentaire par des enzymes, hormones naturellement présentes. Parmi ces organes on en cite un qui occupe l'une des fonctions les plus importantes dans l'appareil digestif, le pancréas. Cet organe anatomiquement enfoui entre les autres organes dans l'abdomen aussi décrit par sa concavité par rapport au rachis reste très compliqué à dépister et à diagnostiquer précisément quand il s'agit de pathologies sévères tels que le cancer.

- Dans ce chapitre nous allons parler du pancréas du point de vue anatomique, morphologique et histologique ainsi que des types de cancers qui peuvent se développer sur ce dernier.

1.2 Terminologies médicales

1.2.1 Anatomie de l'appareil digestif

L'appareil digestif est constitué d'un ensemble d'organes qui s'occupent de la digestion des aliments. C'est un système qui commence de la cavité buccale à l'anus. Arrivé à l'œsophage la paroi du système prend une autre structure qui comportera quatre (04) couches qui se s'imbriquent depuis la partie inférieure à l'extérieure : *la muqueuse, la*

sous-muqueuse, la musculuse et l'adventice (séreuse). La morphologie de l'appareil digestif comporte deux – (02) parties principales [1] :

- **Le tube digestif** : Il est composé de la cavité buccale, le pharynx, l'œsophage, l'estomac, l'intestin grêle (petit intestin) et gros intestin [1].
- **Les glandes digestives** : Les glandes salivaires, glandes gastriques, le foie, le pancréas et les glandes intestinales [1].

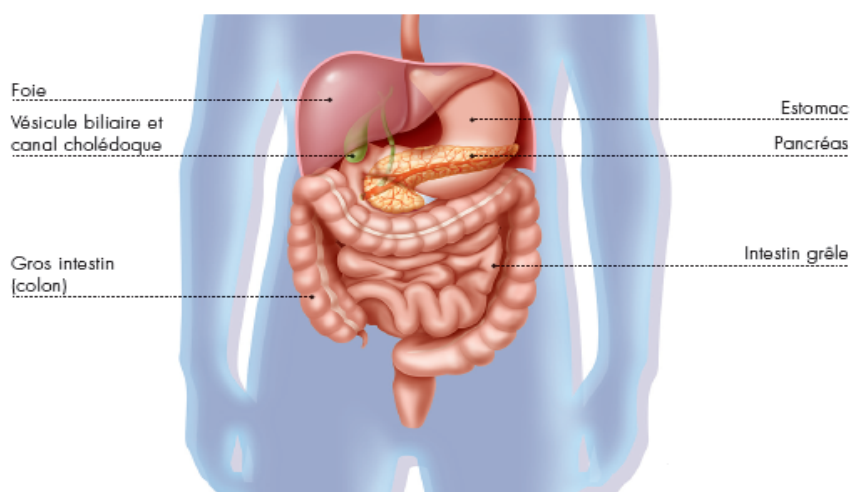


FIGURE 1.1 – Anatomie de l'appareil digestif.

1

1.2.2 Anatomie du Pancréas

Le pancréas est un organe qui se situe dans *l'abdomen*, entre *l'estomac* et les *reins*. Il mesure environ 15 centimètres de long pour un poids de 70 à 100 grammes, ce qui le place vu son gabarit en deuxième place après le foie parmi les organes les plus gros du corps humain. Son aspect visuel se présente sous forme de feuille. Il comprend quatre – (04) parties dont une qui est médiane [2] :

1. Anatomie de l'appareil digestif :

Image extraite de l'article référencié sous le nom :

[*Fondation ARC pour la recherche sur le cancer, Le cancer du Pancréas, décembre 2010*].

- **La tête du pancréas** se trouve sous le foie et entourée par le *duodénum* auquel elle y est *moulée*. Elle sert de pont dans lequel les voies biliaires principales ou le canal cholédoque y traversent pour transporter la bile qui est secrétée par le foie au duodénum qui servira d'élément principale lors du processus de la digestion de graisses [2].
- **L'isthme** est la partie médiane qui se situe entre la *tête* et le *corps* du pancréas. Les vaisseaux de l'intestin grêle y traversent au-dessous qui se charge d'acheminer le sang vers l'intestin ainsi que vers la veine mésentérique supérieure en direction de la veine de la rate pour former la veine porte qui se chargera de transporter le sang qui y circule entre les vaisseaux de l'intestin grêle et la rate vers le foie [2].
- **Le corps du pancréas** s'étant vers la *gauche* et le *haut* de *l'abdomen* de manière *oblique* qui rejoint la partie supérieure du rein et de la glande surrénale se situant à la tête de ce dernier [2].
- **La queue du pancréas** quant à elle, se situe au *pôle gauche* du parenchyme pancréatique, à parenté des vaisseaux ressortants de la rate [2].
- La langueur du parenchyme pancréatique est infiltrée par le canal pancréatique aussi appelé – (*canal de Wirsung*) qui se charge d'emmagasiner les *sucs digestifs* qui sont secrétés par le pancréas pour les évacuer dans le *duodénum* ultérieurement par le biais d'un orifice commun avec les voies biliaires auxquels ils se rejoignent, cet orifice est appelé – (*ampoule de Vater*). Ensuite, le suc pancréatique passera par un autre orifice intitulé – (*papille*) qui quant à lui se situe sur la paroi du duodénum [2].

2. Anatomie du Pancréas :

Image extraite de l'article référencié sous le nom :

[*Le groupe suisse d'entraide pour les maladies du pancréas GSP, Les maladies du Pancréas, 2007*].

3. Morphologie du Pancréas :

Image extraite de l'article référencié sous le nom : [[https ://fr.saleonline2022.ru/](https://fr.saleonline2022.ru/)].

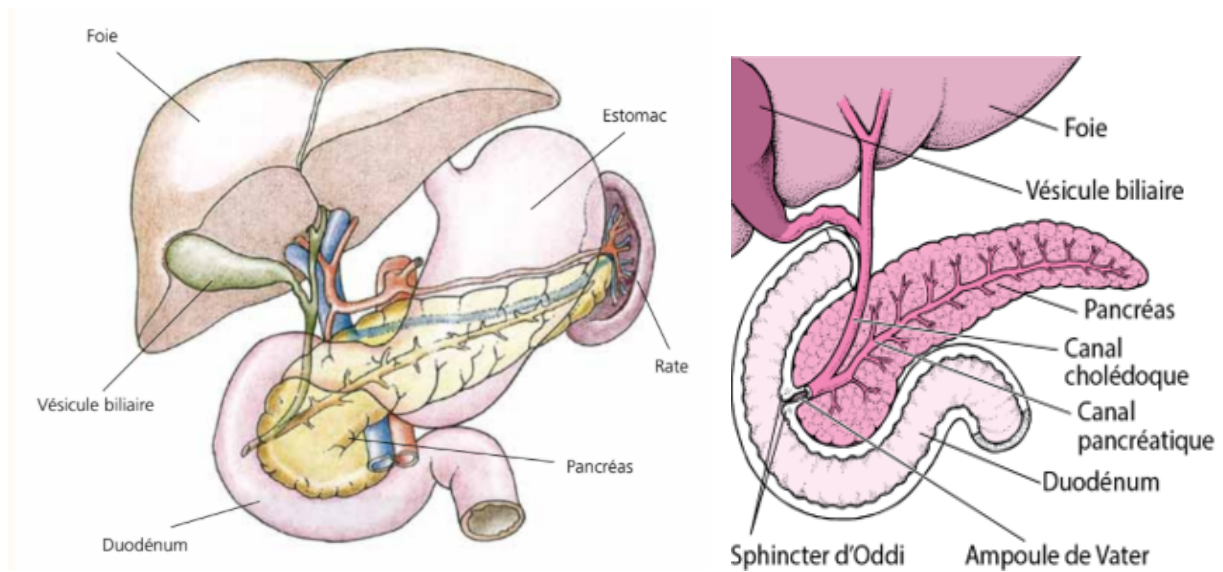


FIGURE 1.2 – Anatomie du Pancréas.

23

1.3 Processus de fonctionnement

1.3.1 Le Pancréas

Le parenchyme pancréatique histologiquement abordé est une *glande* dite *amphicrine* qui est constituée de deux (02) types de tissus [3] :

- Tissus constitués de *glandes exocrines*.
- Tissus constitués de *glandes endocrines*.

1.3.2 Les glandes exocrines

Les glandes exocrines ont pour rôle de synthétiser des enzymes pancréatiques qui sont essentiels au bon fonctionnement du *système digestif*. Ces enzymes seront par la suite déversés dans le *duodénum* infiltrant le *canal de Wirsung*. Les proenzymes du suc pancréatiques telles que (le protéase, ribonucléases, désoxyribonucléases, lipases et amylases) sont de nature inactives, c'est la présence du suc pancréatique dans la paroi intestinale qui permettra leurs activations. Toutefois, le suc pancréatique possédant une

propriété de pH-*basique* sera capable de maîtriser les acides qui seront déversés par l'estomac [3].

1.3.3 Les glandes endocrines

Les glandes endocrines sont arborées par les *îlots de Langerhans* du parenchyme pancréatique. Ces glandes sont éparpillées sur tout le amas pancréatique. Cependant, les îlots de Langerhans englobent quatre (04) types de cellules : (*cellules- α* , *cellules- β* , *cellules- δ* , *cellules-F*) [3].

- [**Cellules- α**] : Les cellules- synthétisent le glucagon [3].
- [**Cellules- β**] : Les cellules- synthétisent l'insuline [3].
- **Les cellules- α** et **cellules- β** secrètent des hormones de régulation de la glycémie dans le sang [3].

- [**Cellules- δ**] : Les cellules- synthétisent la somatostatine pancréatique [3].
- [**Cellules-F**] : Les cellules-F synthétisent le polypeptide pancréatique [3].

- **La somatostatine pancréatique** regroupe plusieurs fonctions qui assurent le bon fonctionnement de certains organes digestif telles que : *stabiliser la sécrétion de la bile par la vésicule biliaire, accroître la motricité de l'intestin et bloquer la synthèse du glucagon et de l'insuline* [3].

- **Le polypeptide pancréatique** quant à lui aurait la fonction de *bloquer* les sécrétions pancréatiques exocrines [3].

4. Histologie du Pancréas :

Image extraite de l'article référencié sous le nom : [<https://www.medindia.net/>].

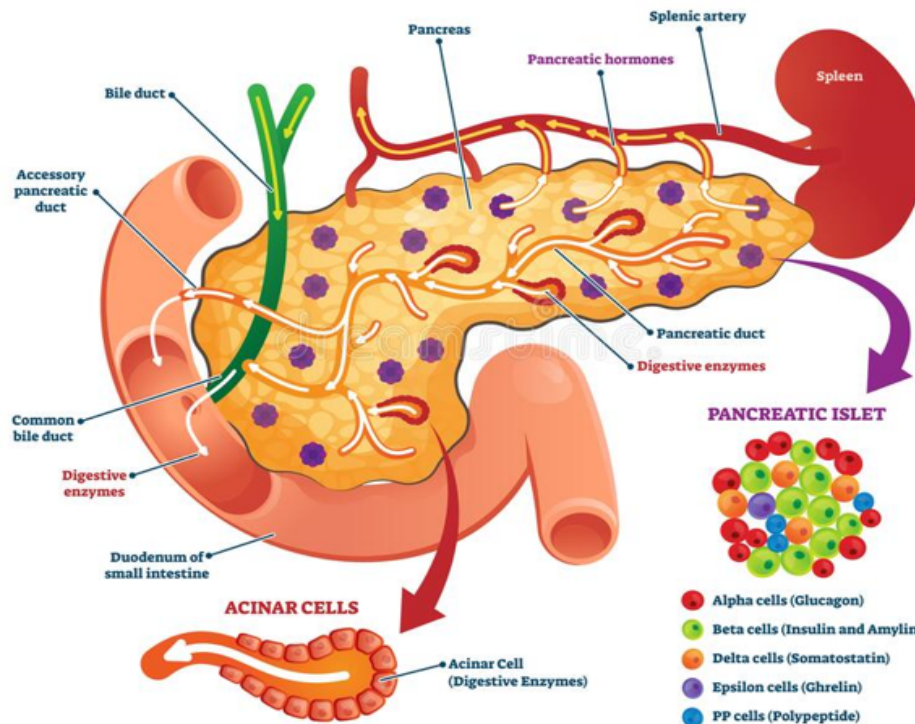


FIGURE 1.3 – Histologie du Pancréas.

4

1.4 Tumeurs du Pancréas

1.4.1 Définition de la tumeur du Pancréas

La tumeur du pancréas peut être décrite comme une *difformité de croissances* des cellules biologiques ce qui *perturbe* leur *fonctionnement*. Ces cellules étant incapable de fonctionner correctement, peuvent impacter les tissus voisins en causant leurs dégénérescences. Ce mécanisme incite le métabolisme à rejeter ces cellules en formant des amas cellulaire *symétriques* ou *asymétriques*. Les altérations qu'ont subies ces cellules peuvent susciter des *tumeurs non-cancéreuses* – (tumeurs bénignes) telle que : *les pseudo-kyste pancréatique* ainsi que *la tumeur kystique séreuse* – (TKS). Dans certains cas, elles peuvent occasionner des *états précancéreux*, autrement dit, les cellules altérées n'ont pas encore atteint un stade cancéreux mais qu'elles peuvent *converger* vers cet état si elles ne sont pas extraites chirurgicalement comme *la tumeur kystique mucineuse* – (TKM), *la tumeur intra-canaulaire papillaire et mucineuse* – (TIPM) ainsi que *la tumeur pseudo-papillaire et solide*. Toutefois, les cellules altérées qui se développent sur la paroi du parenchyme pancréatique

peuvent causer un cancer *grave* – (tumeur maligne), généralement le cancer du pancréas le plus répandu est l'*adénocarcinome canalaire* [4].

1.4.2 Catégories de tumeurs pancréatiques existantes

1.4.2.1 Tumeurs primitives et tumeurs secondaires

[Tumeurs primitives]

- Les tumeurs primitives sont des tumeurs qui se développent dans un ou *plusieurs tissus de l'organisme* sans que l'origine de celles-ci soit une *extension* d'autres pathologies d'ordre tumorales distinctes [5].

[Tumeurs secondaires]

- Les tumeurs secondaires, sont des tumeurs qui *ressortent* des tumeurs primitives, celles-ci sont décrite comme tumeurs d'ordre *métastatiques* et donc elles peuvent *migrer* dans d'autres région de l'organisme au contraire des tumeurs primitives [5].

1.4.2.2 Tumeurs bénignes, intermédiaires et malignes

[Tumeurs bénignes – (B)]

- Les tumeurs bénignes par définition, sont des masses qui se *développent* dans une partie d'un organe *sans causer* de *dégénérescence* tissulaire. La particularité de ces amas de cellules altérées est qu'elles ne se propagent pas vers d'autres parties du corps – (*pas de métastases*), ce qui ne les rends pas nocives. Les tumeurs bénignes peuvent être retirées et ne sont pas récidivantes [6].

[Tumeurs intermédiaires – (I)]

- Les tumeurs intermédiaires sont en partie des cellules cancéreuses à potentiel malin – (*tumeurs malignes*) mais *pas métastatique*. Donc elles peuvent êtres extraites, mais aussi localement récidiver [6].

[Tumeurs malignes – (M)]

- Les tumeurs malignes sont considérées comme cellules cancéreuses à *potentiel malin accru et métastatiques*. Ces cellules envahissent les tissus d'un ou plusieurs organes et engendrent des dégénérescences tissulaires irréversibles [6].

1.4.2.3 Tumeurs résécables, borderlines et non-résécables

[Tumeurs résécables]

- Les tumeurs résécables sont des cellules cancéreuses qui peuvent être *extraites par chirurgie*, ce qui permettra au tissu de l'organe touché de se *régénérer* [7].

[Tumeurs borderline]

- Les tumeurs borderline sont des cellules cancéreuses dite à *faible potentiel malin*. Comme les cellules résécables, celles-ci aussi ont la possibilité d'être *extraites chirurgicalement* et permettre une *régénérescence* des tissus [7].

[Tumeurs non-résécables]

- Les tumeurs non-résécables sont des cellules cancéreuses à *potentiel malin* et inéluctablement *métastatiques*. Ce genre de cellules *ne peuvent pas* être extraites chirurgicalement [7].

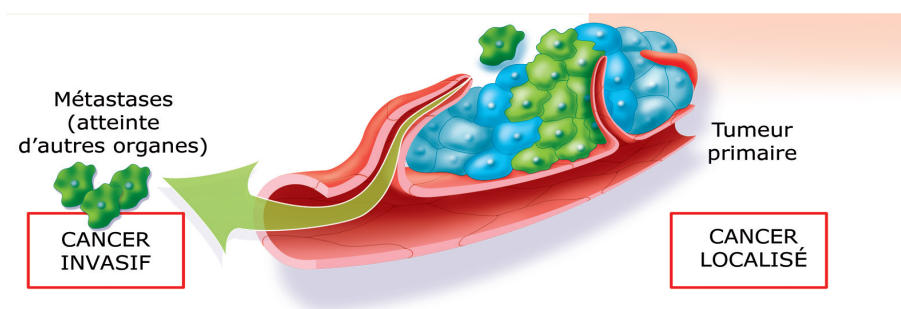


FIGURE 1.4 – Tumeurs du Pancréas.

5

5. Tumeurs du Pancréas :

Image extraite de l'article référencié sous le nom : [Fondation ARC pour la recherche sur le cancer, *Le cancer du Pancréas, L'évolution d'un cancer au sein de l'organisme, décembre 2010*].

1.4.3 Développement et Symptômes du cancer du Pancréas

1.4.3.1 Développement des cellules cancéreuses pancréatiques

Des recherches scientifiques précisément dans le domaine de la *biologie moléculaire*, ont connu une poussée de croissance en ce qui concerne le dépistage du *cancer du pancréas* qui s'avérer être l'un des *cancers* les plus *rare*s. Les évaluations fournis par cette étude ont donnée des résultats très satisfaisant sur les différents mécanismes de développement du cancer du pancréas. Le développement du cancer du pancréas parvient par la dégénérescence des tissus de la glande pancréatique – (*glande amphicrine*) qui renferme deux – (02) types de glandes dites : *exocrines* et *endocrines*. Un facteur très important est impliqué dans le développement des cellules cancéreuses qui est la *mutation génétique*. Cette dernière renferme des *gènes* qui peuvent *perturber* la progression des *cellules* et ainsi causer leurs *dégénérescences*. L'impact de la mutation génétiques sur les cellules infectées leurs donnera la caractéristique de se *multiplier rapidement* plus que le processus de multiplication des cellules saines et donc elles seront *plus résistantes* aux traitements tels que la chimiothérapie et la radiothérapie vu leurs progression accélérée [8].

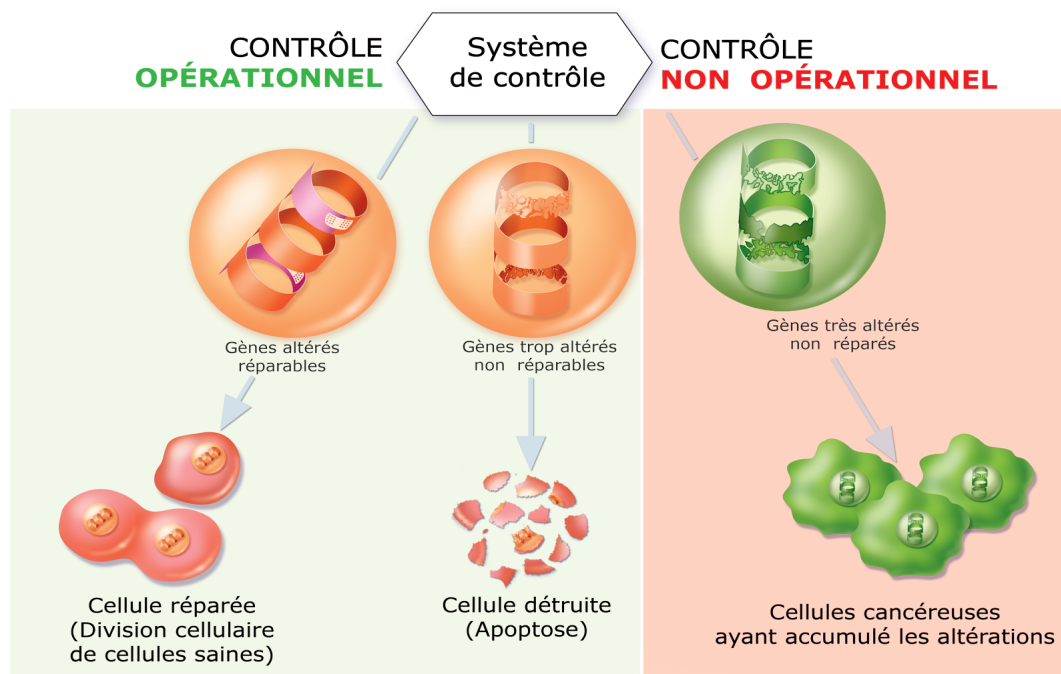


FIGURE 1.5 – Développement du cancer du Pancréas.

1.4.3.2 Symptômes apparents du cancer du Pancréas

Généralement le cancer du pancréas *ne manifeste* aucun *symptôme* apparent qui permettra de diriger le diagnostic vers cette hypothèse au *début* de l'*évolution*. Avec la progression de la maladie, certains patients présentent des symptômes qui peuvent laisser paraître une *suspicion* de la *maladie* en question. Ces symptômes se caractérisent en général par : ***une perte d'appétit et de poids, des douleurs abdominales localisées au niveau de la région du pancréas avec des projection dans le dos***. Lorsque l'évolution du cancer se manifeste dans la *zone* dédiée à la *tête* du pancréas, un *dérèglement* de l'épanchement biliaire sera diagnostiqué pour cause de *compression* de la tumeur sur le canal cholédoque qui fait office de *pont conducteur* entre la vésicule biliaire et le canal pancréatique – (*canal de Wirsung*). Cette compression tumorale peut engendrer une décoloration des tissus hépatiques appelée "ictère". Les urines auront un aspect foncé, les selles claires et des démangeaisons peuvent se manifester également [9].

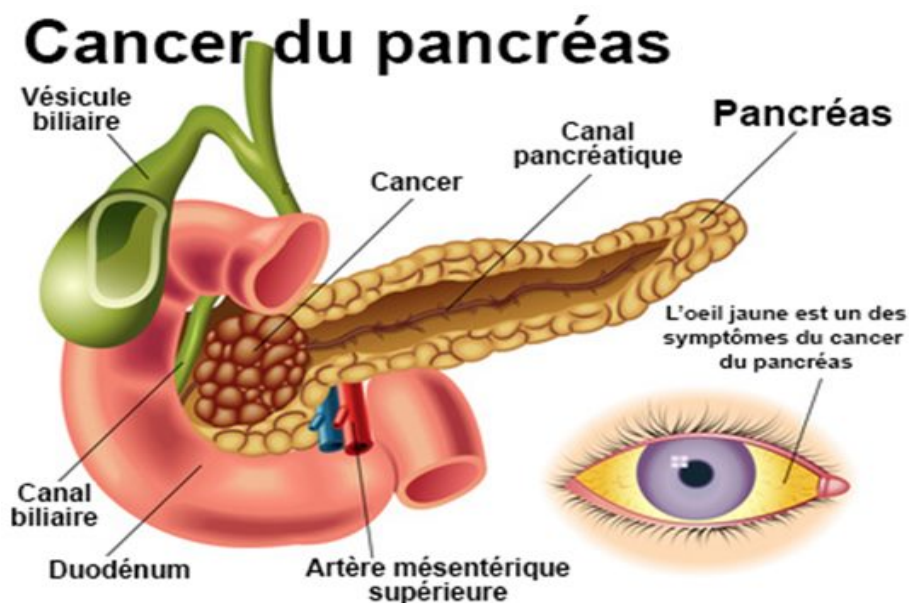


FIGURE 1.6 – Symptômes du cancer du Pancréas.

7

6. Développement du cancer du Pancréas :

Image extraite de l'article référencié sous le nom :

[*Fondation ARC pour la recherche sur le cancer, Le cancer du Pancréas, qu'est ce que le Cancer ?, décembre 2010*].

1.4.4 Classification des tumeurs pancréatiques

Les tumeurs du Pancréas sont organisées en deux (02) catégories : Les tumeurs qui se développent sur des *cellules exocrines* et celles qui se développent sur des *cellules endocrines*. Les tableaux ci-dessous résument ces catégories [10] :

Tissu d'origine	Type de tumeur	Définition
Cellules exocrines	<i>Adénocarcinome canalaire</i>	Il se développe sur la surface du parenchyme depuis les cellules qui sont logées dans ces canaux pancréatiques (90%-cas) [11].
	<i>Cystadéome séreux</i>	Se développe à partir des cellules acineuses qui se trouvent sur la surface du pancréas [11].
	<i>Cystadénome mucineux</i>	Ce sont des tumeurs kystiques qui encapsulent le mucus, le sang ainsi que quelques débris. Elles se développent sur les canaux pancréatiques qui traversent le parenchyme [11].
	<i>Cystadénome séreux microkystique</i>	Ce sont des souches micro-kystiques très rare qui dérivent du Cystadénome séreux –(seulement 1 – 2% des cas) [12].
	<i>Carcinome des cellules acineuses</i>	Ce sont des tumeurs rares, elles sont diagnostiquées lorsque les enzymes digestifs secrétées par le pancréas sont excessivement produits – (1 – 2% des cas) [13].
	<i>Tumeurs intracanales papillaires et mucineuses</i>	Ce sont des tumeurs qui se développent à partir des canaux pancréatiques. Elles ont comme caractéristique leur manière de se manifester imprévisiblement – (canaux pancréatiques principal ou secondaire) [13].

TABLE 1.1 – Les types de cellules cancéreuses exocrines.

7. Symptômes du cancer du Pancréas :

Image extraite de l'article référencié sous le nom : [<https://www.ndiasante.com/>].

Tissu d'origine	Type de tumeur	Définition
Cellules endocrines	<i>Insulinome</i>	Ce sont des tumeurs qui se manifestent au niveau des cellules- β des îlots de Langerhans du pancréas qui sont responsables de la sécrétion excessive de l'insuline. Ces tumeurs sont très rares [14].
	<i>Gastrinome</i>	Ce sont des tumeurs qui affectent la paroi duodénale du parenchyme pancréatique, elles produisent excessivement une hormone baptisée – (gastrine) [14].
	<i>Glucagonome</i>	Ce sont des tumeurs qui touchent les cellules- α qui sont logées au niveau des îlots de Langerhans qui secrètent un excès de glucagon [14].
	<i>Vipome</i>	Elles sont connues pour être des tumeurs non- β des îlots de Langerhans qui produisent un excès d'hormone baptisée – (vasoactive intestinal peptide – <i>vip</i>) [14].
	<i>Somatostatine</i>	Ce sont des tumeurs qui se manifestent par l'intermédiaire des cellules- δ qui subissent des anomalies de fonctionnement. Ces cellules englobent les îlots de Langerhans et les cellules endocrines [14].
	<i>Tumeurs neuro-endocrines</i>	Elles sont décrites comme des cellules qui se développent au niveau des glandes endocrines renfermant des cellules neuroendocrines qui ont pour rôle de concevoir des hormones [15].

TABLE 1.2 – Les types de cellules cancéreuses endocrines.

Comportement biologique des tumeurs		
Bénignes	Intermédiaires	Malignes
<ul style="list-style-type: none"> • <i>Cystadénome mucineux.</i> • <i>Cystadénome séreux microkystique.</i> 	<ul style="list-style-type: none"> • <i>Cystdénome séreux.</i> • <i>Tumeurs intracanalaires papillaires et mucineuses.</i> • <i>Insulinome.</i> • <i>Gastrinome.</i> • <i>Glucagonome.</i> • <i>Vipome.</i> • <i>Somatostatine.</i> 	<ul style="list-style-type: none"> • <i>Adénocarcinome canalaire.</i> • <i>Cystadénocarcinome séreux.</i> • <i>Carcinome des cellules acineuses.</i> • <i>Tumeurs neuroendocrines.</i>

TABLE 1.3 – Comportement biologique des tumeurs pancréatiques.

1.5 Techniques de dépistage médicales des lésions Pancréatiques

1.5.1 L'imagerie médicale

1.5.1.1 L'échographie

[Définition]

L'échographie est une technique de dépistage par imagerie qui utilise des ondes ultrasonores de haute fréquence. Ces ondes sont émises par l'appareil lui-même à travers une sonde (sonde) qui se déplace sur la peau au niveau de la zone à inspecter avec l'application d'un gel cutané. Un ordinateur dans la machine va s'occuper de convertir les ultrasons en images visibles [16].



FIGURE 1.7 – Dépistage par échographie.

[Fonctionnement]

L'échographie envoie des ondes à ultrasons avec des longueurs d'ondes inférieures au millimètre sur la zone du corps orienté vers l'organe à inspecter pour avoir la meilleure coupe possible qui sera réfléchié par l'organe qui va renvoyer les ondes qu'il perçoit. Les ondes qui ressortent par l'organe seront interprétées par la machine qui va les convertir sous forme d'image visible à l'écran. L'application d'un gel, aide à amplifier les ondes [17].

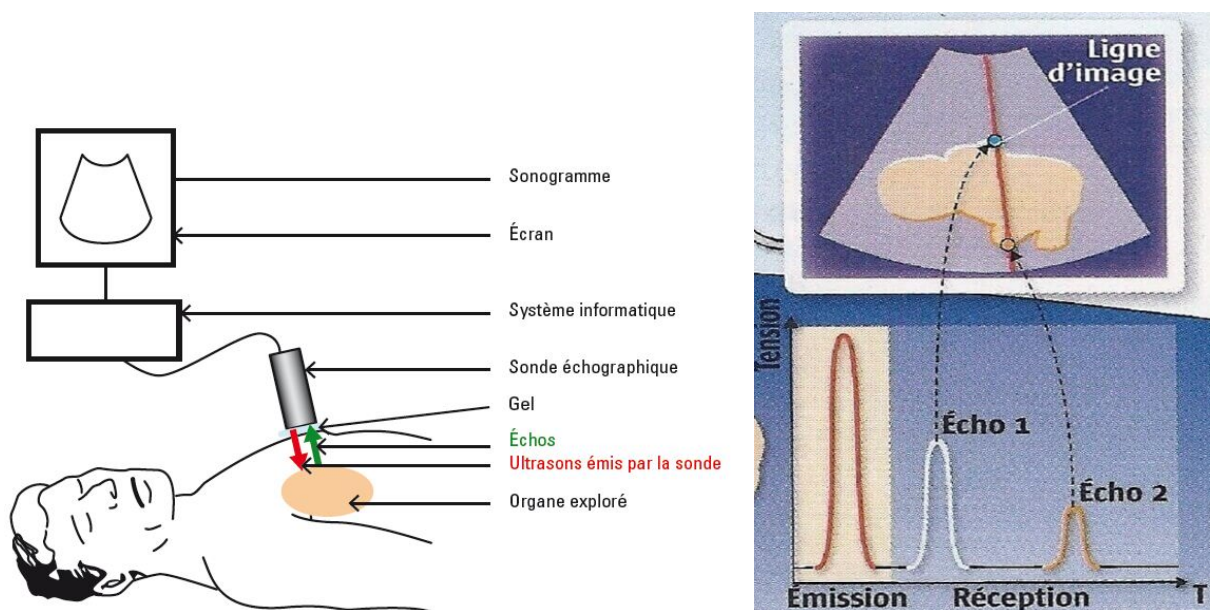


FIGURE 1.8 – Fonctionnement d'une échographie.

[Importance]

L'échographie est un examen primitif qui permet de dépister les anomalies qui se développent sur un organe. Son utilisation reste indispensable car elle permet au médecin de faire un diagnostic et même d'avoir des informations précises sur l'anomalie au cas où une biopsie sera envisageable [18].

1.5.1.2 L'endoscopie**[Définition]**

L'endoscopie aussi appelée "*fibroscopie*" est une technique de dépistage par imagerie qui consiste à introduire un tube fin et souple dans l'œsophage munie de fibres optiques à son extrémité qui regroupe une lampe ainsi qu'une caméra. Cet examen permet au médecin d'inspecter l'organe de plus près et d'en analyser les tissus avec précision [19].

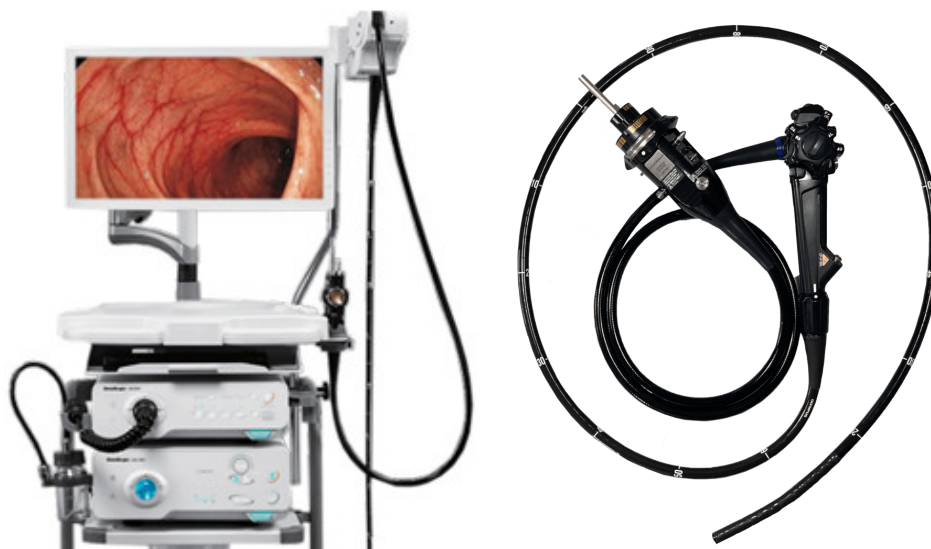


FIGURE 1.9 – Dépistage par endoscopie.

8. Fonctionnement d'une échographie :

Image extraite de l'article référencié sous le nom : [<https://sites.crdp-aquitaine.fr/>].

9. Aspect théorique d'une échographie :

Image extraite de l'article référencié sous le nom : [<https://www.pdfprof.com/>].

10. Déroulement d'une endoscopie :

Image extraite de l'article référencié sous le nom : [https://paediatricgastro.com.sg].

[Fonctionnement]

L'endoscopie permet d'inspecter les anomalies de plus près décelées à l'échographie sur un organe. L'examen se déroule à l'aide d'un tube souple muni d'une caméra et d'instruments chirurgicaux qui sera introduit par l'œsophage en direction de l'organe cible afin d'analyser l'anomalie suspecte et d'en collecter le maximum d'information à l'écran pour effectuer un diagnostic fiable [20].

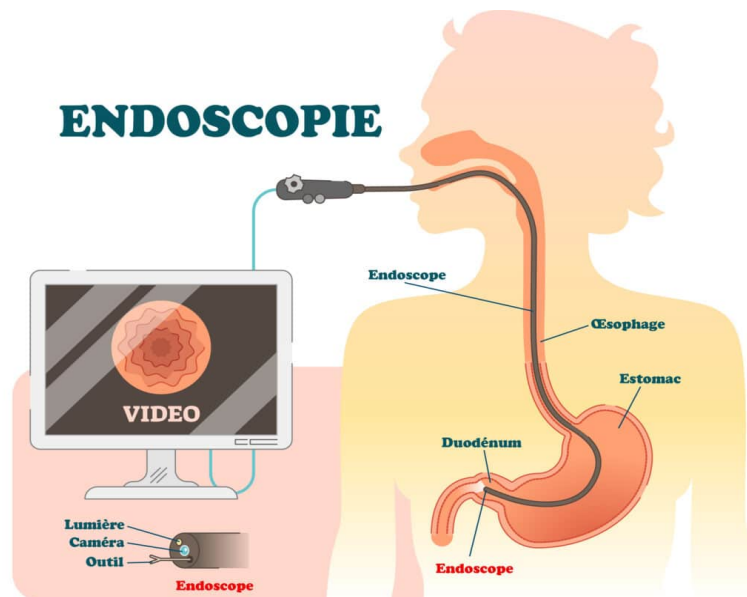


FIGURE 1.10 – Déroulement d'une endoscopie.

10

[Importance]

L'importance d'une endoscopie est la possibilité d'inspecter les anomalies décelées de près, ce qui permet d'avoir une bonne vision de la lésion et d'en extraire les informations nécessaires au diagnostic et dans certains cas faire des prélèvements à l'aide d'instruments chirurgicaux [21].

1.5.2 L'Imagerie par Résonance Magnétique – (IRM)

[Définition]

L'IRM ou Imagerie par résonance magnétique est l'une des techniques de dépistages les plus utilisées de nos jours. Cette technique permet de visualiser les tissus mous avec

une précision élevée et ce dans différents plans de l'espace. De ce fait, les lésions qui n'étaient pas bien visibles autrefois seront flagrantes [22].



FIGURE 1.11 – Technique d'imagerie par dépistage à IRM.

[Fonctionnement]

L'IRM est une technique qui utilise le mécanisme de l'effet du champ magnétique sur les molécules d'eau dont notre organisme en est structuré. Les atomes d'hydrogènes qui composent les molécules d'eau se déplacent dans toutes les directions avant d'émettre des signaux électriques en reprenant leur aspect initial. Cependant, la densité des organes et tissus qui sont élaborés dans notre organisme déterminera la vitesse avec laquelle ces molécules reprendront leur aspect naturel [23].

- Les émissions des champs électriques par les molécules d'hydrogène seront délicatement analysées par des détecteurs qui collectent des signaux électriques régulièrement. Ces signaux seront ensuite interprétés par un ordinateur qui les convertira en images 2D ou 3D [23].
- Généralement, les examens d'IRM sont stimulés par un *produit de contraste* – (*gadolinium*) qui a pour rôle de parfaire la résolution de l'image produite [23].

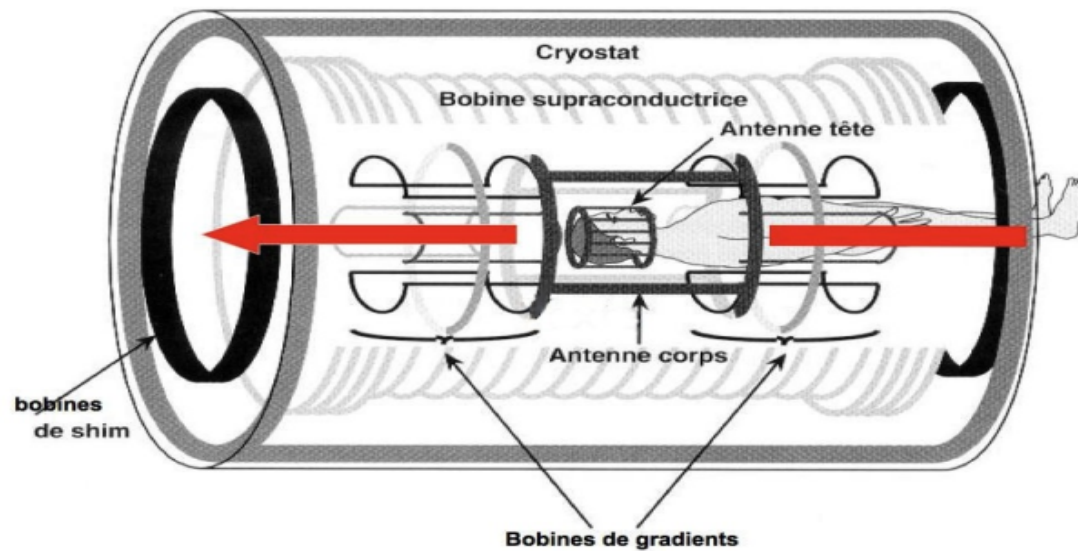


FIGURE 1.12 – Fonctionnement d'une IRM.

11

[Importance]

L'IRM peut être pratique quand elle est utilisée sur des organes *mou* tels que *le cerveau, moelle épinière, etc...* Dans ce genre de cas, elle fournit des explorations de hautes qualités [24].

1.5.3 Le Scanner Toraco - Abdomino - Pélvien – (TAP)**[Définition]**

Le scanner est une technique de dépistage par imagerie qui utilise les *rayons X*. Le principe de cette technique consiste à effectuer plusieurs plans de découpes du corps humain pour analyser les tissus et les organes. Les informations fournies par cet examen sont très persuasives en vue d'apporter un diagnostic précis [25].

11. Fonctionnement d'une IRM :

Image extraite de l'article référencié sous le nom : [<https://tpedotirm.wordpress.com/>].



FIGURE 1.13 – Technique d'imagerie par dépistage de Scanner.

[Fonctionnement]

Le scanner est un appareil qui dispose d'une table d'examen coulissante qui se déplace dans le champ d'un anneau fixe. Le patient est allié sur la table d'examen en présentant son thorax et abdomen dans l'axe de l'anneau à rayons X qui effectuera des rotations redondantes en émettant des rayons X. Cette opération permettra de réaliser des successions d'images de la partie interne du corps humain. Le scanner utilise aussi un *produit de contraste* – (*iode*) pour améliorer ses clichés [26].

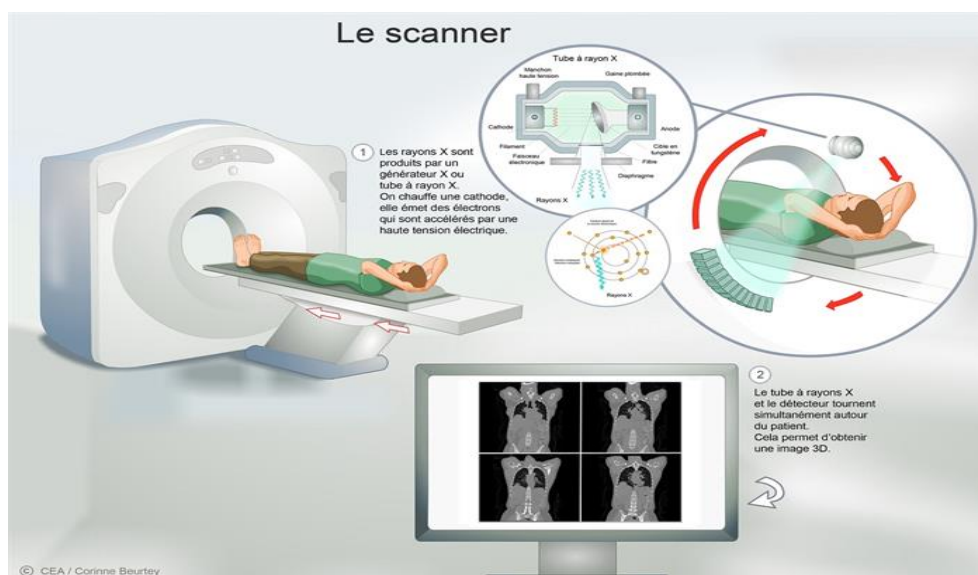


FIGURE 1.14 – Fonctionnement d'un Scanner.

[Importance]

Le scanner aide à visualiser les *variations de volumes* d'une ou plusieurs *lésions* telles que : *tumeurs, embolies, etc...* se développant sur un ou plusieurs organes [27].

1.6 Examen histologique et cytologique

1.6.1 La Biopsie

[Définition]

La biopsie se déroule de deux – (02) manières différentes, soit par le biais d'un *examen histologique* ou bien par le biais d'un *examen cytologique*. L'examen histologique permet de *ponctionner des échantillons* de tissus ou d'organes tels que : (*le foie, poumons, ganglions, etc...*) pour les visualiser au microscope et d'en extraire des informations sensibles au diagnostic [28].

- L'examen cytologique quant à lui, consiste à *ponctionner des échantillons* de *cellules isolées* telles que les *globules blancs* pour les analyser avec des produits de contraste et en extraire des informations sensibles au diagnostic. [28]

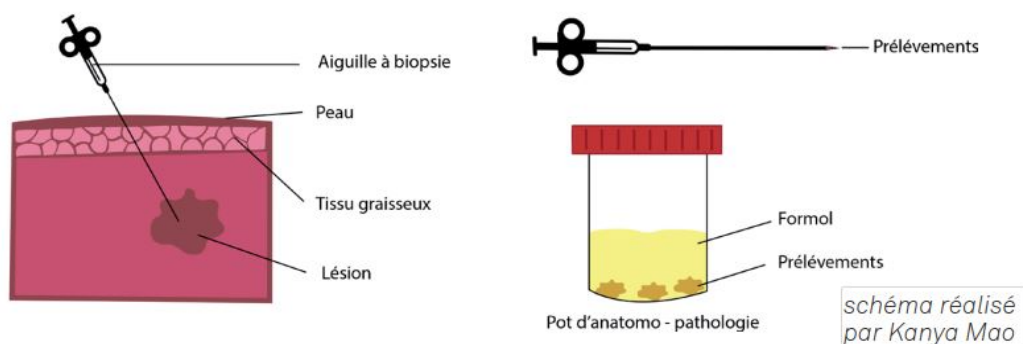


FIGURE 1.15 – Déroulement d'un examen de biopsie.

[Déroulement]

La biopsie est un examen qui se succède sous anesthésie locale ou générale. Cet examen est réalisé par un médecin qui se muni de plusieurs instruments chirurgicaux tels que : *une aiguille creuse ou une pince* pour prélever un ou plusieurs *échantillons* d'une zone suspecte d'un organe pour les anatomiser de plus près. L'échantillon prélevé sera envoyé dans un laboratoire spécialisé – (*l'anatomo-pathologie*) pour réaliser un rapport médical sur la nature de la pathologie [29].

[Importance]

La biopsie joue un rôle impératif lors du *dépistage des tumeurs*, elle permet de collecter des données très sensibles qui seront importantes dans le diagnostic tels que : *la taille, forme, degré d'agressivité*.

- Toutefois, la biopsie peut même localiser l'origine de la tumeur. L'une des importances secondaires de la biopsie, c'est sa capacité de spécifier un traitement adéquat à un type de cancer donnée [29].

1.7 Diagnostic et pronostic du cancer du Pancréas

1.7.1 Diagnostic

Pour diagnostiquer un cancer du pancréas, il faut suivre une démarche atypique qui permettra au médecin de bien préciser leur interprétation médicale sur l'état de santé du patient. Cette démarche se déroule de la manière suivante :

- **Suspicion de cancer :**

Les patients atteints d'un cancer du pancréas se présentent souvent avec les symptômes suivants [30] :

[Signes cliniques]

- Une jaunisse accompagnée d'un prurit - (démangeaisons) [30].
- Des vomissements après les repas [30].
- Des douleurs dorsales ainsi qu'au niveau du creux de l'estomac [30].

[Examens Primitifs]

Une échographie abdominale est préconisée en cas de jaunisse ou de douleurs initiales prononcées. Cet examen est toutefois complété par un scanner thoraco-abdomino-pelvien qui permettra de dépister la présence de tumeurs en cas de suspicion [30].

- **Confirmation diagnostique : déterminer le type de cancer (tumeur primitive) :**

Lors de la phase de confirmation du diagnostic, les médecins procèdent en trois (03) étapes :

[Examens paracliniques]

- Un scanner du pancréas doit être concrétisé en premier lieu [30].
- Une endoscopie sera ensuite réalisée sous anesthésie générale pour bien visualiser le pancréas et d'en analyser les tissus [30].
- L'examen sera finalisé par une biopsie pour prélever un échantillon des lésions apparentes [30].

[Examens Cytologique/Histologique]

- Dans le cas où les cellules cancéreuses sont résécables, la chirurgie sera envisageable pour disséquer les tissus du pancréas touché - (*tumeurs, ganglions, etc...*) ces tissus seront ultérieurement inspectés médicalement pour déterminer le type de traitement à administrer [30].

[Examens immunohistochimiques]

Toutefois, dans certains cas une études immunohistochimique (chromogranine A) sera réalisée pour déterminer la nature de la tumeur qui se développe sur le pancréas (*glandes exocrine ou endocrine*) car ces dernières ne se traitement pas de la même manière [30].

[Déterminer le stade - (bilan d'extension)]**— Examens paracliniques pour déterminer une extension**

Le bilan d'extension est un examen qui va permettre d'identifier s'il existe des contacts entre les tumeurs pancréatiques et les artères digestives pouvant empêcher la chirurgie. Cet examen s'effectue par le biais d'un scanner thoraco-abdomino-pelvien avec produit de contraste dans le but de déceler de potentielles métastases logeant dans cette partie du corps [30].

12. Fonctionnement d'un Scanner :

Image extraite de l'article référencié sous le nom : [<https://www.cea.fr/>].

13. Déroulement d'un examen de biopsie :

Image extraite de l'article référencié sous le nom : [*La ponction biopsie • www.chu-nantes.fr*].

14. Les stades de la maladie :

Image extraite de l'article référencié sous le nom :

[*Fondation ARC pour la recherche sur le cancer, Le cancer du Pancréas, les stades de la maladie, décembre 2010*].

STADE UICC	TNM			EXPLICATION
stade 0	Tis	N0	M0	On trouve des cellules anormales seulement dans le revêtement du pancréas (aussi appelé carcinome in situ).
stade IA	T1	N0	M0	La tumeur se trouve seulement dans le pancréas et mesure 2 cm ou moins.
stade IB	T2	N0	M0	La tumeur se trouve seulement dans le pancréas et mesure plus de 2 cm.
stade IIA	T3	N0	M0	La tumeur s'est propagée aux tissus voisins mais pas aux gros vaisseaux sanguins à proximité.
stade IIB	T1 à T3	N1	M0	La tumeur se trouve dans le pancréas ou s'est propagée aux tissus voisins mais pas aux gros vaisseaux sanguins à proximité. La tumeur s'est propagée aux ganglions lymphatiques régionaux.
stade III	T4	Tout N	M0	La tumeur s'est propagée au-delà du pancréas jusqu'aux gros vaisseaux sanguins à proximité. La tumeur peut s'être propagée aux ganglions lymphatiques régionaux.
stade IV	Tout T	Tout N	M1	Le cancer s'est propagé vers des emplacements éloignés, comme le foie, un poumon ou la cavité abdominale (péritonéale).

FIGURE 1.16 – Les stades de la maladie.

14

1.7.2 Pronostic

Le pronostic du cancer du pancréas reste très sombre en vue de son dépistage tardif. Cependant l'évaluation du taux de survie des patients dépends du type de tumeurs décelées (résécables ou non-résécables) ainsi que de leurs localisations (locales, régionales ou à distance). Avec l'avancée de la médecine, plusieurs techniques de dépistage et de traitement ont vu le jour, ce qui a permis de diagnostiquer plus vite la maladie et aussi améliorer la qualité du traitement et donc une augmentation de l'espérance de vie des patients. Mais dans tous les cas, la chirurgie demeure le traitement de référence [31].

1.8 Traitement contre le cancer du Pancréas

1.8.1 La chirurgie

La chirurgie reste le traitement de référence pour extraire une tumeur quelconque. Cependant, l'intervention suscite quelques contraintes pour assurer son déroulement ainsi que son efficacité telles que : **la taille de la tumeur, sa localisation** – (*pas de métastases*) ainsi que l'état de santé du patient – (*physique et mental*). Il existe différentes manières de réaliser l'intervention chirurgicale pour retirer une tumeur du pancréas selon sa localisation [32].

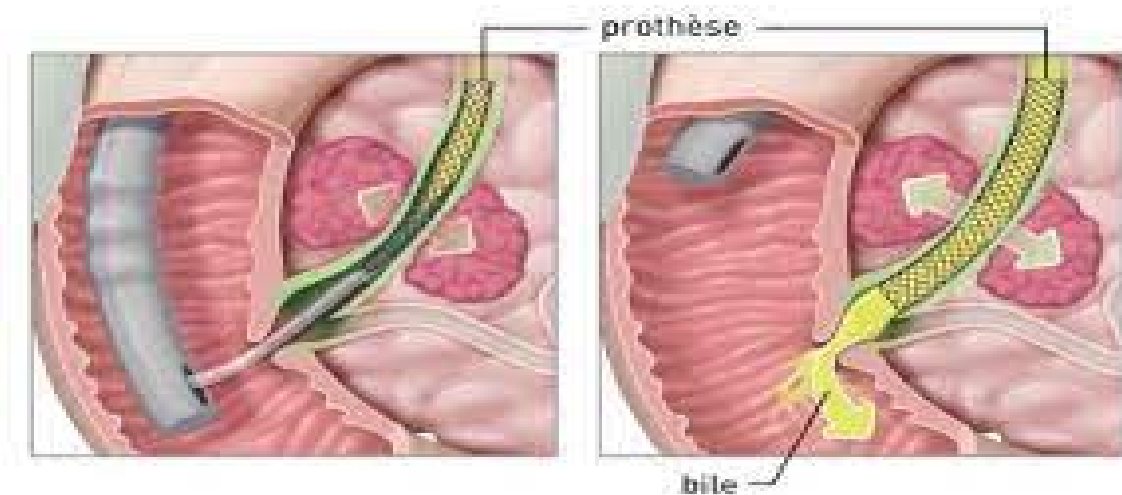


FIGURE 1.17 – Dilatation du canal biliaire comprimé par une tumeur au niveau de la tête du pancréas par une prothèse biliaire.

15

15. Dilatation du canal biliaire comprimé par une tumeur au niveau de la tête du pancréas par une prothèse biliaire :

Image extraite de l'article référencié sous le nom : [[https ://meldu834.wixsite.com/](https://meldu834.wixsite.com/)].

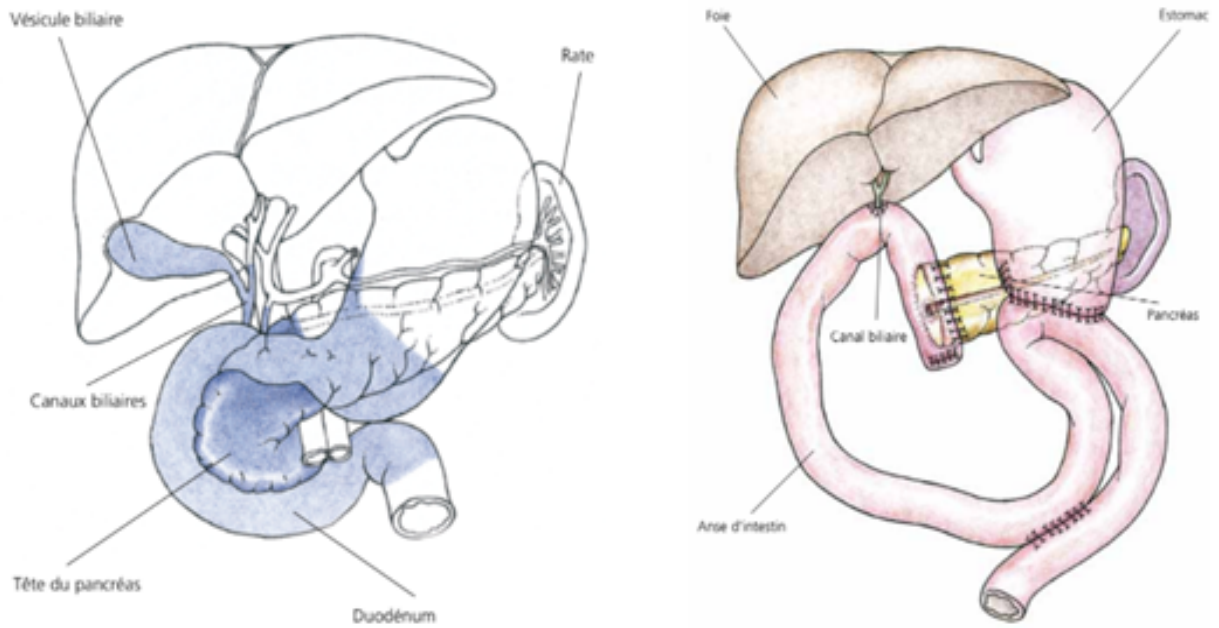


FIGURE 1.18 – Intervention chirurgicale permettant de retirer une tumeur localisée au niveau de la tête du pancréas.

16

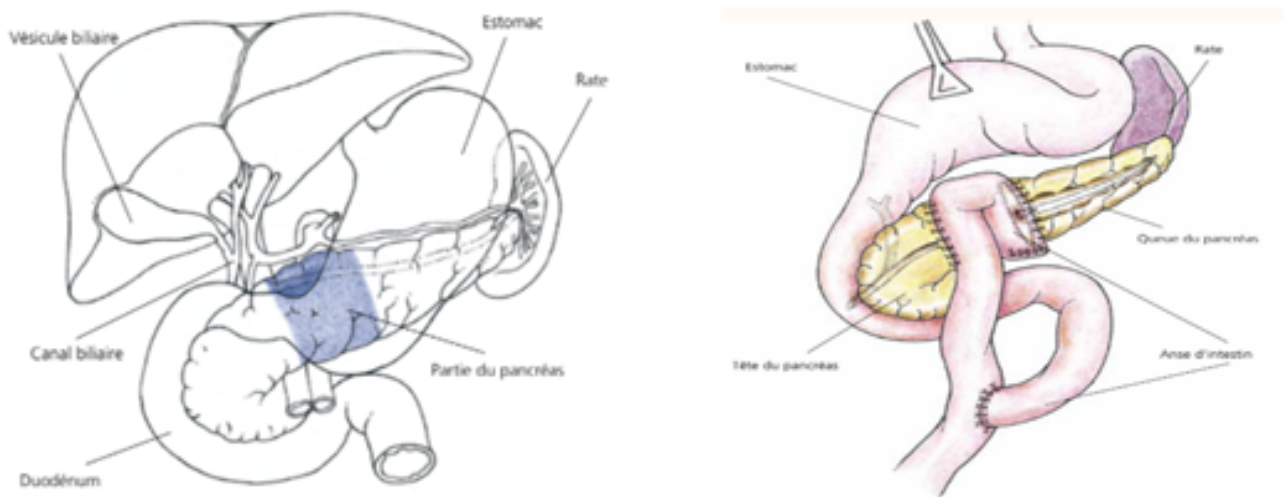


FIGURE 1.19 – Intervention chirurgicale permettant de retirer une tumeur localisée au niveau de la partie libre (corps) du pancréas.

17

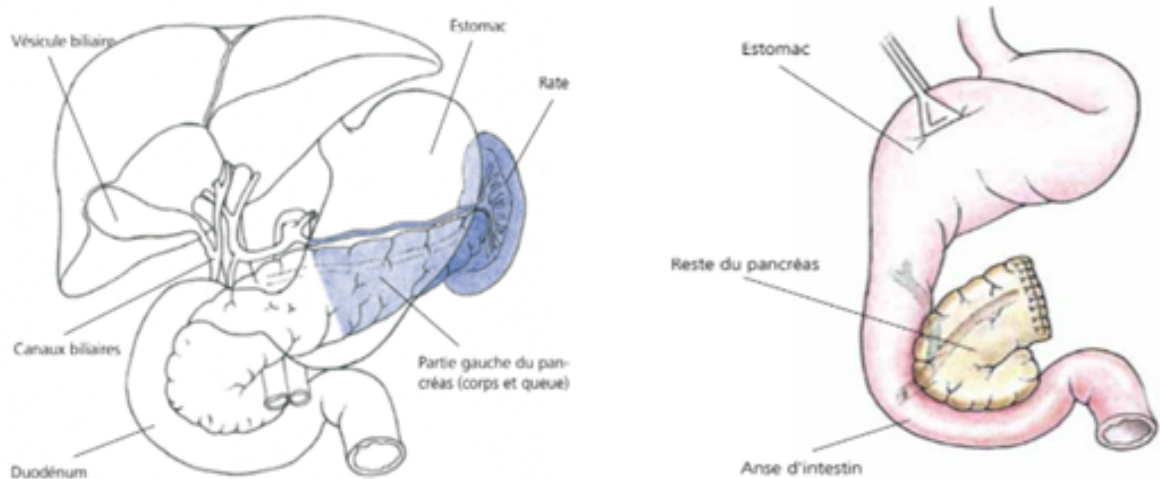


FIGURE 1.20 – Intervention chirurgicale permettant de retirer une tumeur localisée au niveau de la queue du pancréas.

18

16. Intervention chirurgicale permettant de retirer une tumeur localisée au niveau de la tête du pancréas :

Image extraite de l'article référencié sous le nom :

[Le groupe suisse d'entraide pour les maladies du pancréas GSP, maladies du Pancréas, Opérations du Pancréas en cas de cancer au niveau de la tête du paranchyme pancréatique, 2007].

17. Intervention chirurgicale permettant de retirer une tumeur localisée au niveau de la partie libre (corps) du pancréas :

Image extraite de l'article référencié sous le nom :

[Le groupe suisse d'entraide pour les maladies du pancréas GSP, maladies du Pancréas, Opérations du Pancréas en cas de cancer au niveau du corps du paranchyme pancréatique, 2007].

18. Intervention chirurgicale permettant de retirer une tumeur localisée au niveau de la queue du pancréas :

Image extraite de l'article référencié sous le nom :

[Le groupe suisse d'entraide pour les maladies du pancréas GSP, maladies du Pancréas, Opérations du Pancréas en cas de cancer au niveau de la queue du paranchyme pancréatique, 2007].

1.8.2 La chimiothérapie et radiothérapie

1.8.2.1 Chimiothérapie :

La chimiothérapie consiste à détruire les cellules altérées par le cancer avec l'administration d'un ou plusieurs médicaments. Ce traitement vise à désintégrer les cellules touchées qui se génèrent plus vite que les cellules saines. La prescription de ce traitement peut se faire de différentes manière – (*oral ou injectable*) qui sont administrés en suivant un protocole bien précis. Dans le cas du cancer du pancréas, la chimiothérapie demeure indispensable vis-à-vis du patient qu'il soit apte à la chirurgie ou pas [32].

1.8.2.2 Radiothérapie :

La radiothérapie est une autre technique de traitement contre le cancer. Le processus de traitement de cette technique consiste en une irradiation complète des cellules cancéreuses par des rayons à haut potentiel. Elle suscite une étape préalable qui permet de cibler les rayonnements sur les cellules tumorales afin qu'elles n'envahissent pas les autres tissus sains [32].

1.9 Causes et conseils

1.9.1 Causes liées au cancer du Pancréas

Il est malheureusement très rare de nos jours de diagnostiquer un cancer du pancréas à un stade précoce. Cependant, les principales causes de ce type de cancer restent méconnues. Finalement, les seuls facteurs de risque redondants sont : **le tabagisme, l'abus d'alcool, le diabète** parfois même par le biais d'une pathologie telle que la pancréatite aiguë qui converge vers un cancer ultérieurement [33].

1.9.2 Conseils et préventions

Pour réduire son exposition à ce type de maladie, une bonne hygiène de vie et habitudes sont à adopter, on cite :

- Éviter de fumer et limiter son exposition à d'autres fumées dérivées [34].
- Maintenir un poids équilibré [34].
- Arrêter de consommer de l'alcool [34].
- Faire de l'exercice physique [34].
- Réduire son apport de consommation de viandes – (rouges ou ses dérivées) [34].
- Adopter une hygiène de santé et de vie adéquates [34].

1.10 Conclusion

Dans ce chapitre nous avons parlé de l'anatomie du pancréas ainsi que de son processus de fonctionnement. On a aussi voulu cerner la maladie (le cancer) en nous basant sur son comportement et répercussions futurs sur la santé. Une partie a été consacré aux techniques de dépistages tels que : l'imagerie médicale, la biopsie et de traitements tels que : la chirurgie, la chimiothérapie et la radiothérapie que la médecine moderne offre de nos jours pour permettre de dépister la maladie et la traiter à temps.

- Cependant, dans les prochains chapitres, nous allons aborder la partie théorique sur le plan informatique, et essayer d'exploiter les techniques de l'intelligence artificielle tels que : *Machine Learning* et *Deep Learning* pour faire des prédictions sur les types de cancer qui peuvent se développer sur le pancréas par des calculs et traitements d'images à assistance artificielle.

Chapitre 2

État de l'art

2.1 Introduction

L'intelligence artificielle a atteint des degrés de développement très avancés et ne cesse d'évoluer de nos jours. Elle est utilisée dans différents domaines, et là où on se rends compte de sa performance en particulier c'est dans le domaine médical car cette dernière aide à diagnostiquer tous types de pathologies qu'elle soit simple - (bénignes) ou complexes - (cancer) à appréhender, ce qui permet de les dépister très tôt pour donner aux patients la chance d'être traités avant que la maladie atteigne un stade avancé de gravité et ainsi améliorer leur pronostic vital.

2.2 Le Machine Learning

2.2.1 Définition

Le Machine Learning aussi baptisé « apprentissage automatique » est un domaine de recherches scientifiques orientées côté informatique et mathématique qui est dérivée de l'intelligence artificielle. Ce domaine consiste à donner à des algorithmes implémentés la possibilité d'apprendre sur des modèles qui se forment à partir d'un ensemble de données. Ces données peuvent prendre divers aspects : des nombres, des mots, des images, des statistiques, etc... [35]

2.2.2 Principe du Machine Learning

Le Machine Learning est une technique de programmation informatique qui est fondée sur les calculs de probabilités statistiques ce qui donnera à la machine une certaine autonomie pour apprendre sur un ensemble de données qu'on appelle « Dataset ». Le principe est de donner à la machine une quantité de données - (datasets) sous format numérique ou autres et diviser cette quantité en deux parties (training datas - entraînement) et (testing datas - Tests) dans le but d'évaluer la performance via un algorithme qui fait des calculs complexes sur ces données et nous donne des résultats en sortie sous forme de valeur qui par la suite sera considérée comme précision de prédiction de cet algorithme sur les (testing datas - Tests) [36].

2.2.3 Types de machine learning

2.2.3.1 Apprentissage Supervisé

L'apprentissage supervisé est une méthode d'apprentissage automatique qui consiste à fournir des données en entrées et en sorties. Les données d'entrées seront par la suite classées et étiquetées pour fonder une base d'apprentissage qui servira à faire des traitements d'autres données fournis ultérieurement [37].

2.2.3.2 Apprentissage Semi-Supervisé

L'apprentissage semi-supervisé est une méthode d'apprentissage automatique qui consiste à fusionner une petite quantité de données étiquetées avec une grande quantité de données non-étiquetées durant la phase d'entraînement. Cette méthode d'apprentissage se situe entre l'apprentissage supervisé et l'apprentissage non-supervisé [38].

2.2.3.3 Apprentissage Non-Supervisé

L'apprentissage non-supervisé est une méthode d'apprentissage qui se déroule de manière opposée à l'apprentissage supervisé c-à-d que cette méthode est constituée d'un algorithme qui s'entraîne sur des données non-étiquetées : ce qu'on entend par apprentis-

sage autonome. Ainsi, les données sont par la suite fournies à la machine sans modèles de résultats en sortie [39].

2.2.3.4 Apprentissage Par Renforcement

L'apprentissage par renforcement est une méthode d'apprentissage automatique qui consiste à apprendre à partir d'un mécanisme particulier qui se déroule de manière continue, en d'autres termes c'est un apprentissage qui permet de trouver la meilleure solution [40].

2.2.4 Les algorithmes du machine learning

2.2.4.1 Decision Tree • (DT)

[Définition]

Les arbres de décisions sont des modèles d'apprentissage qui divergent de la classification et de la régression. Leur mécanisme de fonctionnement consiste à créer des modèles capables de prédire les valeurs des variables cibles qui seront fournies en sortie à partir du principe des règles de décision qui seront extraites des caractéristiques des données fournies en entrées [41].

[Usages]

— La classification

La classification est une méthode qui permet d'implémenter une fonction qui soit capable de faire des classifications sur diverses catégories de données [42].

— La regression

La regression est une méthode qui permet d'implémenter une fonction qui soit capable de faire la distinction entre les données à valeurs réelles [42].

[Crière de sélection]— **Entropie - (*Entropy*)**

L'entropie permet d'évaluer l'incertitude des variables aléatoires dans le but de diminuer l'impureté d'une classe [42].

— **Indice de Gini - (*Gini-index*)**

L'indice de gini permet de mesurer le degré de probabilité d'une variable lorsqu'elle est mal étiquetée [42].

— **Information Gain - (*Gain d'informations*)**

Le gain d'information permet de sélectionner les meilleures caractéristiques qui fournissent le maximum de données sur les classes [42].

[Déroulement de l'algorithme]

L'algorithme de l'arbre de décision se déroule comme suit [42] :

- [**Étape-01**] : Spécifier les meilleures caractéristiques parmi les données du dataset.
- [**Étape-02**] : Fractionner les données en sous-ensemble de manière à ce qu'ils contiennent les meilleures valeurs de caractéristiques pour la phase d'entraînement.
- [**Étape-03**] : Affecter à chaque sous-ensemble un nouvel arbre de décision.
- [**Étape-04**] : Arrêter le processus lorsque les données seront complètement classifiés.

[Les avantages et inconvénients de cet algorithme]

- Dans ce tableau nous allons présenter les avantages et inconvénients de l'algorithme du *Decision Tree* [42] :

<i>Avantages</i>	<i>Inconvénients</i>
Simple à interpréter	Risque d'overfitting
Temps de réponse moyen	Performances moyennes
Meilleur avec les données de catégorie	Apprentissage lent
Possibilité d'évaluer plusieurs résultats en parallèle	Sensible à l'ajustement de données
Fiabilité	Capacité des noeuds limitées

TABLE 2.1 – Les avantages et inconvénients de l'arbre de décision - (Decision Tree).

[Équations mathématiques]

- L'entropie se calcule comme suit [42].

$$H(S) = \sum_{c_i \in C} -P(c_i) \log_2 P(c_i)$$

- Fractionner l'ensemble de *classes* sur le *nombre de donnée total* du dataset [42].

$$P(c_i) = \frac{|c_i|}{|S|}$$

- Le *gain d'information* se calcule comme suit [42].

$$IG(S, f_j) = H(S) - \sum_{S_{jk} \in S_j} P(S_{jk}) H(S_{jk})$$

- Fractionner un *sous-ensemble* S_j sur le nombre de donnée S [42].

$$P(S_{jk}) = \frac{|S_{jk}|}{|S|}$$

— Le *Gini-index* se calcule comme suit [42].

$$E(S) = 1 - \sum_{c \in C} P^2(c_i)$$

$$P(c_i) = \frac{|c_i|}{|S|}$$

— $[S]$: Les données du dataset.

— $[c_i]$: Le nombre d'échantillon d'une classe C.

— $[C]$: L'ensemble de classes.

— $[P_i]$: La probabilité qu'un élément S se trouve dans la classe c_i .

— $[f_i]$: Vecteur de caractéristiques

— $[S_{jk}]$: Le nombre de sous-ensemble de donnée se trouvant dans l'ensemble j .

— $[k]$: L'indice de chaque donné.

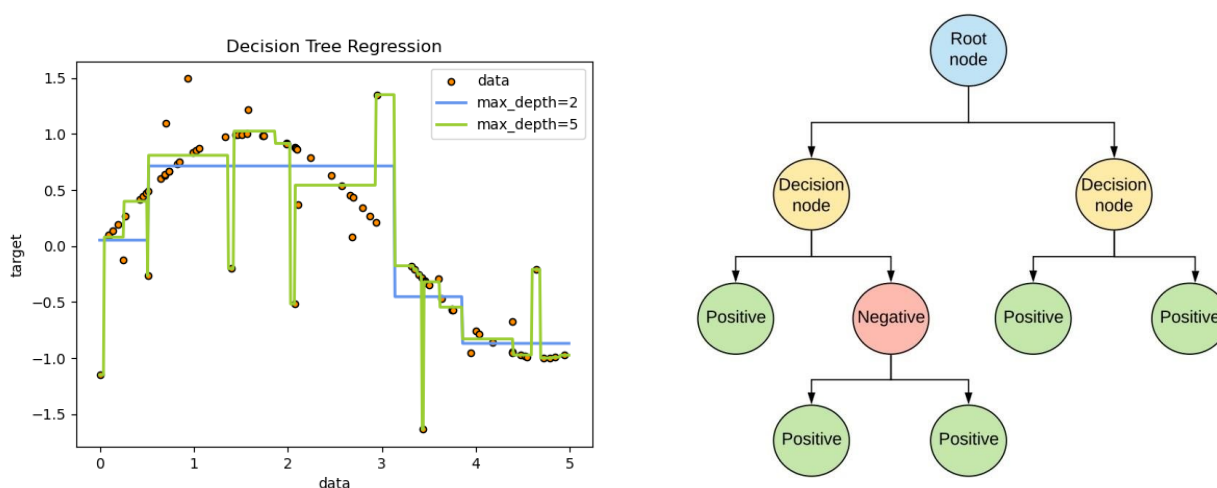


FIGURE 2.1 – Exemple de figure montrant le principe du Decision Tree.

1 2

1. Decision Tree - (représentation graphique) :

Image extraite de l'article référencié sous le nom : [<https://scikit-learn.org>].

2. Decision Tree - (forme théorique) :

Image extraite de l'article référencié sous le nom : [<https://www.researchgate.net>].

2.2.4.2 Random Forest • (RF)

Les forêts aléatoires sont décrites comme étant des métras-estimateurs qui consistent à faire des ajustements sur plusieurs classificateurs d'arbre de décision au niveau de plusieurs sous-échantillons de données. Puis, effectuer une moyenne sur les résultats fournies en sortie de chaque sous-échantillons pour contrôler et éviter le phénomène de surcharge-ment et ainsi pouvoir améliorer les résultats des prédictions fournies en sortie. Les forêts aléatoires sont une extension des arbres de décisions [43].

[Déroutement de l'algorithme]

L'algorithme de la forêt aléatoire se déroule comme suit [43] :

- [**Étape-01**] : Extraction d'un nombre d'enregistrement n appelés échantillons aléatoirement d'un ensemble de données incluant k enregistrement.
- [**Étape-02**] : Construction d'un ensemble d'arbres de décisions individuels à partir des échantillons extraits au préalable.
- [**Étape-03**] : Les arbres mis en place généreront des résultats en sortie.
- [**Étape-04**] : La sélection du meilleur résultat final qui sera retenu par un algorithme de trie calculant la moyenne des valeurs retournées par tous les arbres.

[Les avantages et inconvénients de cet algorithme]

- Dans ce tableau nous allons présenter les avantages et inconvénients de l'algorithme du **Random Forest Classifier** [43] :

<i>Avantages</i>	<i>Inconvénients</i>
Prévient l'overfitting	Complexité de calculs
Performances élevées	Temps d'exécution accru
Ajustement de paramètres simple	Apprentissage lent
Gère bien les cas de multi-classes	Optimisation complexe
Usage de données catégorielles	Instable

TABLE 2.2 – Les avantages et inconvénients de la forêt aléatoire - (Random forest classifier).

[Équations mathématiques]

- L'importance d'un *noeud* se calcule comme suit [43] :

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

- L'importance d'une *caractéristique* i se calcule comme suit [43] :

$$fi_i = \frac{\sum_{j: \text{node } j \text{ splits on feature } i} ni_j}{\sum_{k \in \text{all nodes}} ni_k}$$

- Les *caractéristiques* i importantes seront normalisées à travers la formule suivante [43] :

$$\text{norm } fi_i = \frac{fi_i}{\sum_{j \in \text{all features}} fi_j}$$

- Calculer la performance de chaque arbre à travers la formule suivante [43] :

$$RF fi_i = \frac{\sum_{j \in \text{all trees}} \text{norm } fi_{ij}}{T}$$

- $[ni_j]$: L'importance d'un noeud j .
- $[i]$: L'indice d'une caractéristique.
- $[j]$: L'indice d'un noeud.
- $[w]$: Le poids d'un noeud j .
- $[C]$: La valeur d'impureté d'un noeud j .
- $[fi_i]$: L'importance d'une caractéristique i .
- $[norm\ fi_i]$: La normalisation des caractéristiques importantes.
- $[RF\ fi_i]$: La moyenne des caractéristiques importantes normalisées à travers tous les arbres.

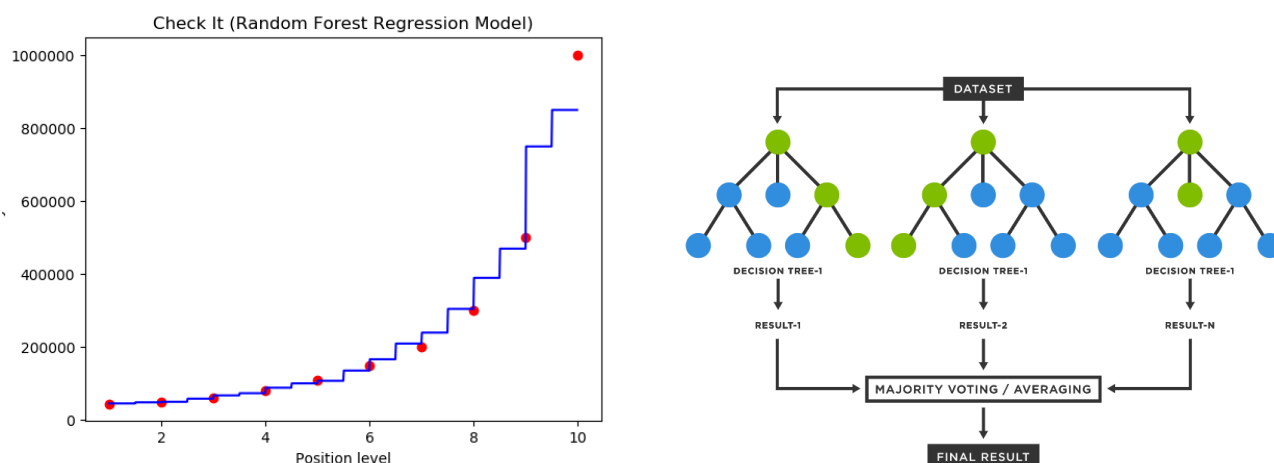


FIGURE 2.2 – Exemple de figure montrant le principe du Random Forest.

3 4

3. Random Forest - (représentation graphique) :

Image extraite de l'article référencié sous le nom : [<https://medium.com>].

4. Random Forest - (forme théorique) :

Image extraite de l'article référencié sous le nom : [<https://dou.ua>].

2.2.4.3 K-Nearest Neighbors • (K-NN)

Les K-Plus Proches Voisins – (K-NN) sont des algorithmes qui consistent à trouver un nombre d'échantillons d'apprentissages prédéfini qui se rapprochent le plus d'un nouveau point et d'en faire la prédiction sur la base de leurs étiquettes. Le nombre d'échantillon est une constante qui peut être définie soit par l'utilisateur soit varier en fonction de la densité locale des points. Cependant, la distance calculée qui sépare les points qui se trouvent dans le jeu de données et le nouveau point est juste une mesure métrique connu sous le nom de distance euclidienne. Ensuite, la distance la plus petite sera retenue pour faire la classification du point fixé aléatoirement avec le calcul d'une probabilité [44].

[Déroulement de l'algorithme]

L'algorithme du K-Plus Proches Voisins se déroule comme suit [44] :

- [**Étape-01**] : Déterminer le nombre de voisins k à vérifier.
- [**Étape-02**] : Appliquer la *distance-Euclidienne* pour calculer la distance entre les *k-valeurs*.
- [**Étape-03**] : Sélectionner les *k-voisins* les plus proches en terme de distances calculées.
- [**Étape-04**] : Évaluer le nombre de *classe* se trouvant dans la *perimètre* des *k-proches voisins*.
- [**Étape-05**] : Classifier le point fixé au départ dans la classe de *catégorie* la plus proche à son *egart*.

[Les avantages et inconvénients de cet algorithme]

- Dans ce tableau nous allons présenter les avantages et inconvénients de l'algorithme du *K-Nearest Neighbors* [44] :

<i>Avantages</i>	<i>Inconvénients</i>
Simple à implémenter	Données limitées
Facile à interpréter	Sensible au bruit
Performance accrue	Consommation de mémoire
Polyvalence	Temps de prédiction lent
Diversité de donnée	Sensible à la dimension

TABLE 2.3 – Les avantages et inconvénients du K-Plus Proches Voisins - (K-Nearest Neighbors).

[Équations mathématiques]

- La distance Eulidienne se calcule comme suit [44] :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

- Calculer la *probabilité* d'appartenance d'un *point* aléatoirement fixé à une classe donnée se formule comme suit [44] :

$$P(y = j|X = x) = \frac{1}{k} \sum_{i \in A} I(y^{(i)} = j)$$

- $[d(\mathbf{x}, \mathbf{y})]$: Le distance Enclidienne entre deux - (02) point.
- $[\mathbf{x}]$: Une donnée du dataset.
- $[\mathbf{y}]$: Le point fixé aléatoirement.
- $[\mathbf{j}]$: L'indice de classification
- $[\mathbf{k}]$: Le nombre de voisins.
- $[\mathbf{I}]$: L'ensemble de valeurs d'une classe.
- $[\mathbf{i}]$: L'indice d'une caractéristique.
- $[\mathbf{A}]$: L'ensemble des caractéristiques.
- $[\mathbf{X}]$: Le point de référence d'une classe.

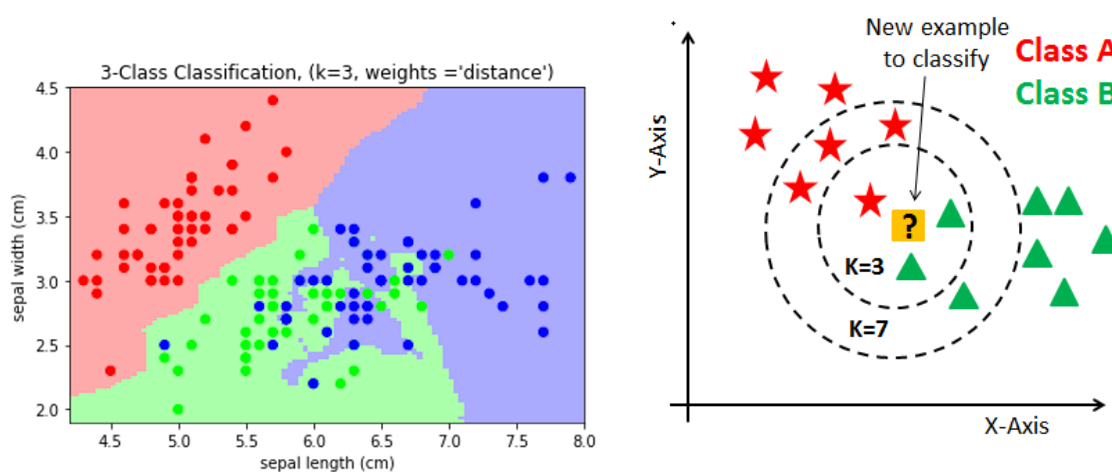


FIGURE 2.3 – Exemple de figure montrant le principe du K-Nearest Neighbors.

5 6

5. K-Nearest Neighbors - (représentation graphique) :

Image extraite de l'article référencié sous le nom : [<https://medium.com>].

6. K-Nearest Neighbors - (forme théorique) :

Image extraite de l'article référencié sous le nom : [<https://techvaluetrends.com>].

2.3 Le Deep learning

2.3.1 Définition

L'apprentissage en profondeur est un autre domaine de l'intelligence artificielle qui est une autre méthode dérivée de l'apprentissage automatique. Cette méthode d'apprentissage profond se focalise sur les réseaux de neurones artificiels. Le but est d'entraîner des réseaux de neurones artificiels pour qu'ils soient en capacité de simuler l'intelligence humaine et ainsi pouvoir faire des interprétations correctes qui par la suite seront incarnées dans des machines qui serviront à assister l'intelligence humaine [45].

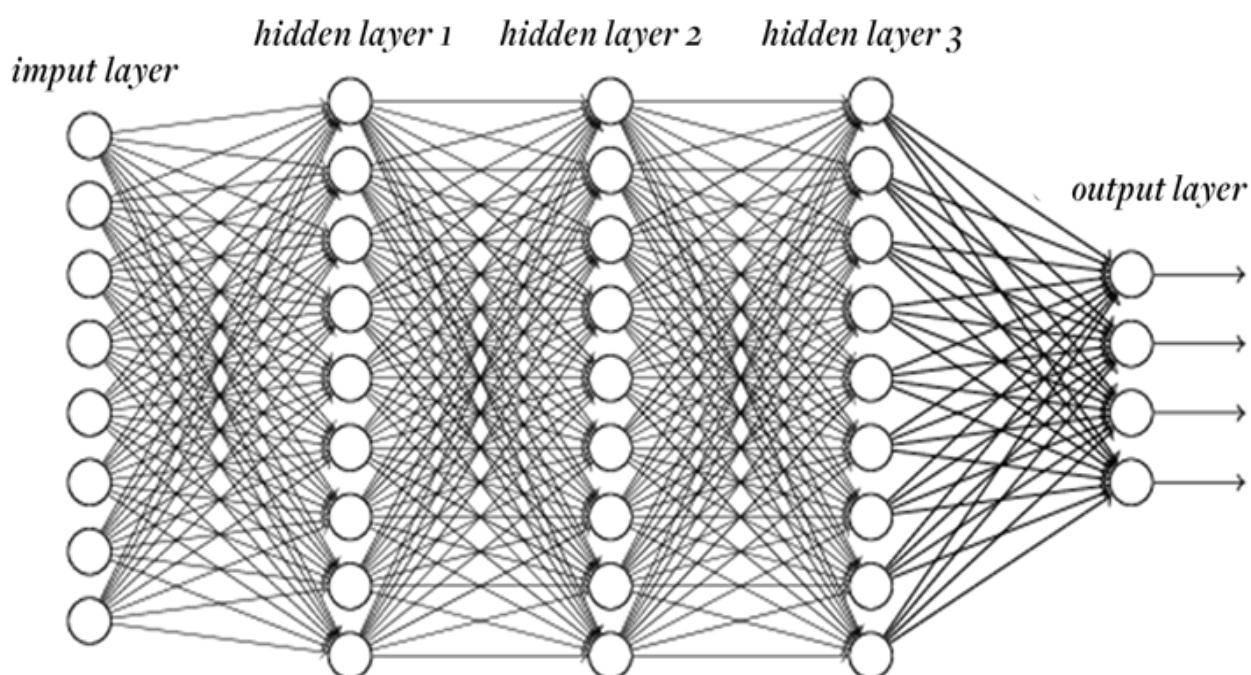


FIGURE 2.4 – Architecture d'un réseau de neurones.

7

2.3.2 Fonctionnement du Deep learning

Les réseaux de neurones artificiels sont développés sur la base de données structurées complexes. Ils génèrent des modèles d'entraînement constitués de plusieurs couches de traitement dans le but d'avoir plusieurs niveaux d'entités pour représenter les données [46].

- Les neurones sont regroupés en trois (03) types de couches différentes :

2.3.2.1 La couche d'entrée - Input Layer

La couche d'entrée s'occupe de recevoir les données d'entrées puis les transmettre à la première couche cachée [46].

2.3.2.2 La ou les couche(s) cachée(s) - Hidden Layer(s)

Les couches cachées quant à elles, effectuent des calculs sur les entrées. La plus grande contrainte qui puisse être poser c'est la détermination du nombre de couches cachées ainsi que le nombre de neurones se trouvant dans chaque couche [46].

2.3.2.3 La couche de sortie - Output Layer

La couche de sortie retourne les données de sortie en résultat qui sont fournies par les couches cachées [46].

- A chaque connexion entre ces neurones est associées un poids - $w(i, j)$. Il consiste à déterminer l'importance d'une valeur d'entrée. Les poids initiaux entre chaque neurone sont fixés aléatoirement [46].
- La principale idée derrière toutes ces étapes, est de « normaliser » les données qui circulent à travers ces réseaux de neurones pour pouvoir retourner des résultats concrets en sortie [46].
- Les réseaux de neurones artificiels ont la réputation d'être souvent sollicités dans l'intelligence artificielles et ce par leurs capacités à faire des traitements sur divers concepts ayants des quantités de données qui demandent énormément de calculs pour promouvoir à fournir des résultats précis [46].

7. Architecture d'un réseau de neurones :

Image extraite de l'article référencié sous le nom : [<https://www.researchgate.net>].

8. Principe d'un réseau de neurones :

Image extraite de l'article référencié sous le nom : [<https://passiondiscoverydot.wordpress.com>].

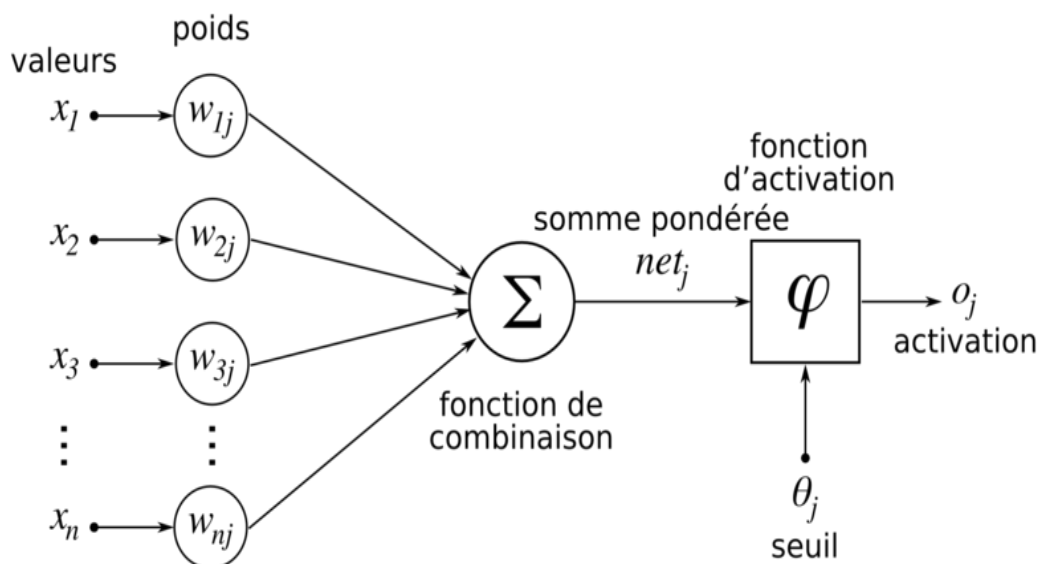


FIGURE 2.5 – Principe d'un réseau de neurones.

8

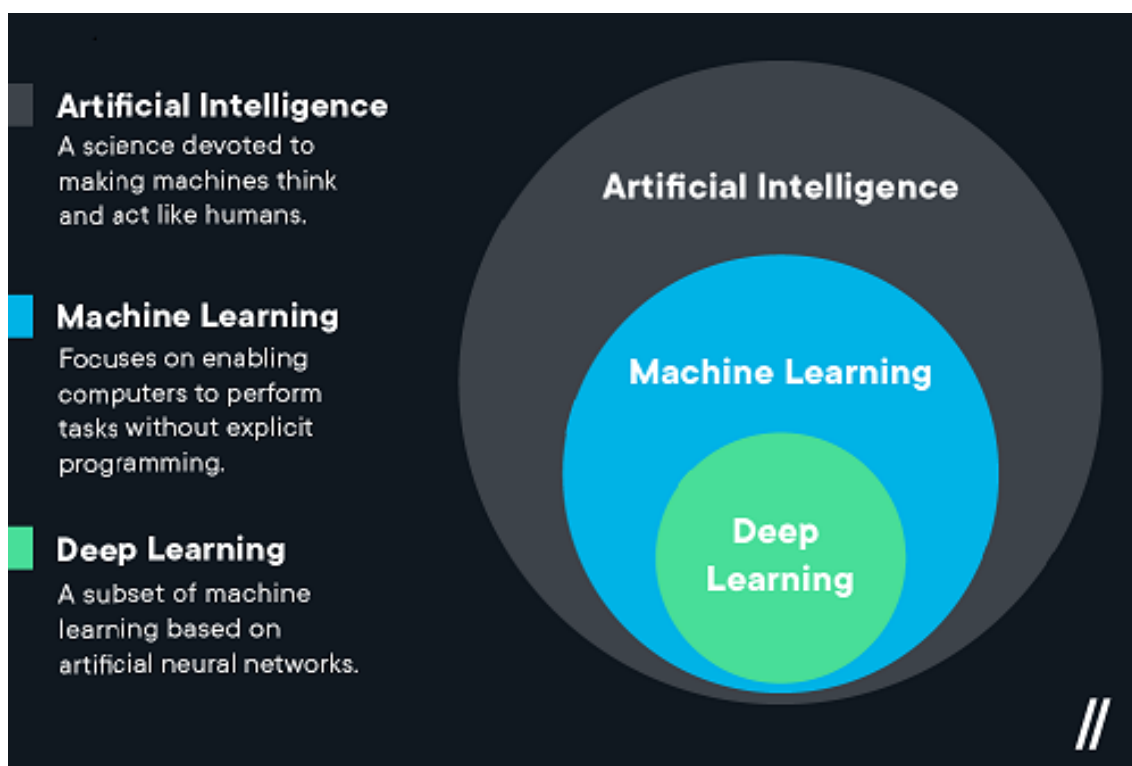


FIGURE 2.6 – Les sous domaines de l'intelligence Artificielle.

2.3.3 Algorithmes de Deep Learning

2.3.3.1 Réseaux de Neurones Convolutifs - (CNN)

[Définition]

Ce sont des réseaux de neurones particulier aux réseaux de neurones multicouches. Ces réseaux sont capables de créer des catégories sur les informations données qu'elles soient simples ou complexes. Le mécanisme de fonctionnement de ce type de réseaux de neurones consiste à un empilage multicouche de neurones ainsi que de fonctions mathématiques complexes avec des paramètres qui sont ajustables en permanence en vue d'effectuer des opérations de pré-traitement sur de petites quantités de données [47].

2.3.3.2 Architecture du Réseaux de Neurones Convolutifs - (CNN)

— Déroulement du Réseaux de Neurones Convolutifs - (CNN) :

[L'Extraction des caractéristiques]

Cette étape va permettre d'extraire les caractéristiques spécifiques de chaque image par un processus appelé la compression dans le but de réduire leur taille initiale. En réalité ce qui se passe, c'est que les images qui sont insérées en entrée passent par une série de filtre en créant à chaque fois des images multiples baptisées cartes d'entités. Ainsi les caractéristiques qui y seront extraites seront assemblées dans un vecteur de caractéristiques appelé le code CNN⁹ [47].

[La classification]

Dans cette étape, les caractéristiques qui ont été conservées dans le code - CNN seront combinées entres elles pour pouvoir effectuer le processus de classification d'images [47].

9. Convolutional Neural Network.

2.3.4 Les couches du Réseau de Neurones Convolutif - (CNN)

2.3.4.1 La couche de convolution - (Convolution Layer)

Cette première couche a pour rôle d'effectuer une analyse précise sur les images qui sont insérées en entrée et de détecter toutes les caractéristiques se trouvant sur leurs supports. Les caractéristiques qui y seront extraites vont être fournies en sortie sous forme de cartes d'entités [48].

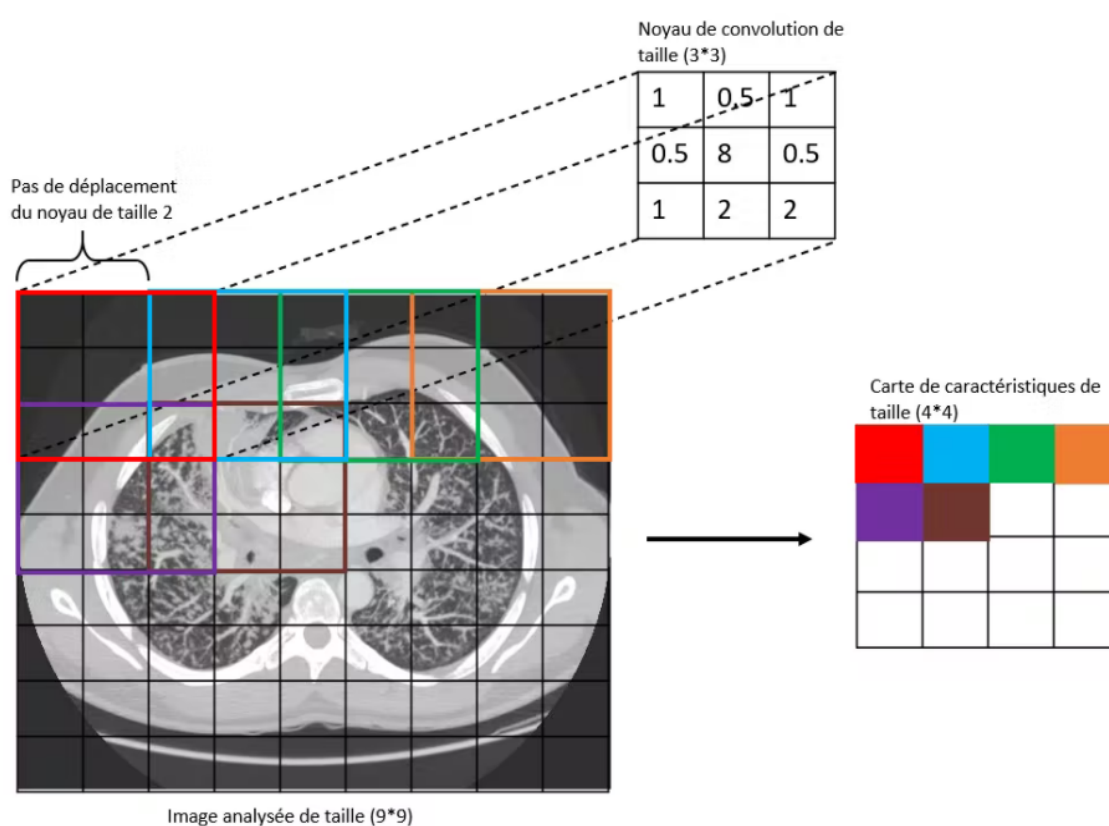


FIGURE 2.7 – Concept de filtrage d'image au niveau de la couche convolution.

10

10. Concept de filtrage d'image au niveau de la couche convolution :
Image extraite de l'article référencié sous le nom : [<https://www.imaios.com>].

2.3.4.2 La couche de Pooling

Le pooling se traduit par une opération de remplacement sur des valeurs. Ce principe est appliqué en remplaçant les valeurs de pixels carrés (2 x 2 ou 3 x 3) par des valeurs uniques. De ce fait, la dimension de l'image est réduite ainsi que son contenu simplifié [48].

- Pour appliquer le processus de Pooling, on doit d'abord avoir sélectionné un pixel carré de taille (2 x 2) autrement dit pour un pooling de 2 x 2, ensuite on calcule la valeur qui viendra remplacer ce carré. Après ça, on doit évaluer la taille du *pas* – (*stride*) de déplacement du carré, celle-ci varie entre (1 et 2). Si la valeur du stride est égale à 1 – (*stride* = 1), on déplace ce carré d'une case vers la droite, sinon on utilise d'autres stratégies dans le cas où la sélection sort du contexte de l'image, on cite par exemple : ne rien compter ou tout compter en ignorant les pixels manquants. Néanmoins ce genre de situation reste très rare car la sélection des strides sont souvent bien effectuées [48].
- Dans le cas où on arrive à la fin des déplacements à droite, on recommence le processus en repartant de la gauche après avoir effectué un (*stride* = 1) vers le bas, puis on se déplace à nouveau vers la droite. Il existe plusieurs opérations de pooling [48] :

- **Le Max pooling**, Ce qui équivaut à prendre la valeur maximale de la sélection. C'est le type d'opération qui est souvent utilisée, car elle est rapide et simplifie correctement les images [48].
- **Le Average pooling**, C'est-à dire la moyenne des pixels sélectionnés. Il faut procéder à un calcul en effectuant la somme de toutes les valeurs puis les diviser par le nombre de valeurs. Ce qui nous donnera une valeur intermédiaire en résultat qui va représenter le lot de pixels [48].
- **Le Sum pooling** Cette opération représente la moyenne (Addition) des valeurs sans effectuer de division.. [48]



FIGURE 2.8 – Concept du pooling au niveau de la couche Pooling.

11

11. Concept du pooling au niveau de la couche Pooling :

Image extraite de l'article référencié sous le nom : [<https://www.semanticscholar.org>].

2.3.4.3 La couche totalement connecté - (Fully Connected Layer)

La structure d'un réseau de neurones n'est rien d'autre qu'une succession de fonctions non-linéaires qui sont liées entres-elles. Chaque une de ces fonctions est constituée d'un neurone aussi dit perceptron. Au niveau de la couche totalement connectée, des transformations linéaires seront effectuées par les neurones sur le vecteur d'entrée par le biais de la matrice de poids. Ensuite une transformation non-linéaire suivra et sera appliquée sur le résultat donné en sortie à travers une fonction d'activation du même type – (non-linéaire) [49].

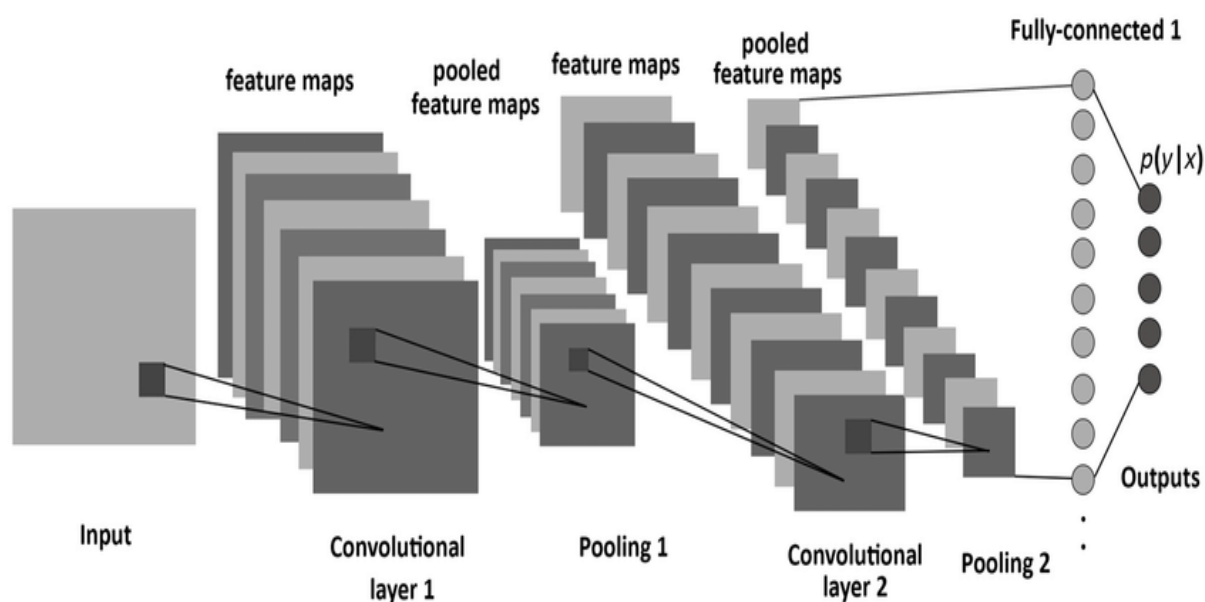


FIGURE 2.9 – Architecture d'un réseau de neurones convolutifs.

12

12. Architecture d'un réseau de neurones convolutifs :

Image extraite de l'article référencié sous le nom :

[*PLOS ONE* — <https://doi.org/10.1371/journal.pone.0237213> August 14, 2020].

2.3.5 Les différentes Architectures du Réseaux de Neurones Convolutifs - (CNN)

Dans l'apprentissage profond, il existe plusieurs modèles d'apprentissage qui sont *pré-entraînés* au préalable et qui sont à la disposition des développeurs pour les intégrer dans leurs travaux de recherches. Ce processus est connu sous le nom de *Transfer Learning*.

- Parmi les modèles les mieux évalués selon le classement de la plateforme - [*Keras*] on cite :

1. [*Inception-V3*]

L'*Inception-V3* est une architecture qui fait partie des réseaux de neurones convolutifs - (CNN) dont l'architecture contient 48 couches de convolutions. C'est un modèle pré-entraîné sur des millions d'images ce qui lui permet de pouvoir classifier des images dans 1000 catégories d'objets. Cependant, ce modèle est en mesure de faire des extractions de caractéristiques sur des images à complexité accrue vu sa qualité de pré-entraînement. L'*Inception-V3* ne traite que des images avec des dimension de (299 x 299 x 3) [50].

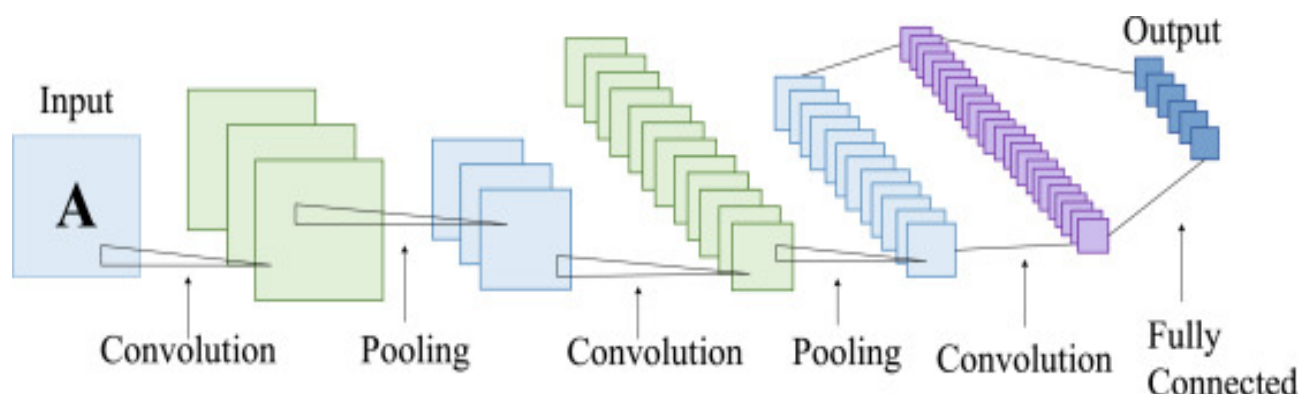


FIGURE 2.10 – Architecture du Réseau de neurones Pré-entraînés [*Inception-V3*].

13

13. Architecture du Réseau de neurones Pré-entraînés [*Inception-V3*] :

Image extraite de l'article référencié sous le nom : [<https://www.sciencedirect.com>].

2. [Xception]

L'Xception est aussi une architecture qui fait partie des réseaux de neurones convolutifs basée sur le principe des convolutions séparables en profondeur. Ce modèle à été développé par des chercheurs de Google. Le principe de convolution séparable en profondeur peut être décrit comme un module d'Inception avec un nombre de tours de convolution très élevé. Comme les modèle d'Inception, l'Xception ne support pas des images dont la taille est supérieure à $(299 \times 299 \times 3)$ [51].

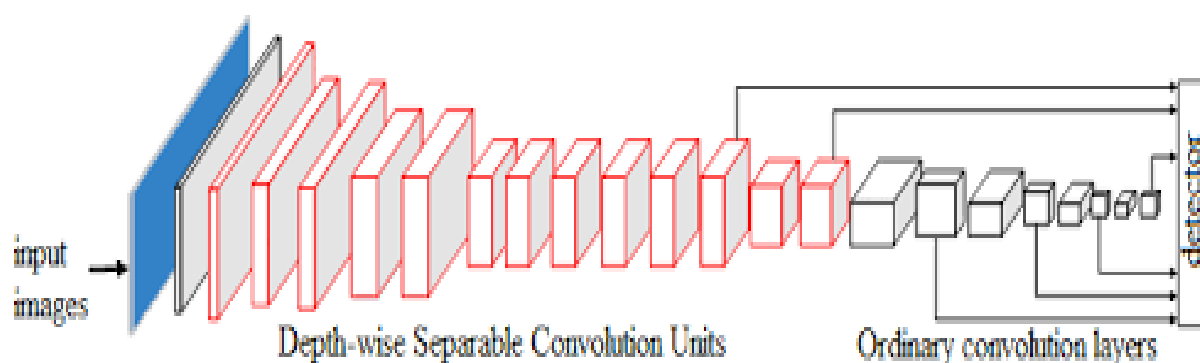


FIGURE 2.11 – Architecture du Réseau de neurones Pré-entraînés [*Xception*].

14

3. [DenseNet121]

Le DenseNet121 est une architecture qui vient apporter des améliorations au niveau des structures réseaux CNN en simplifiant le schéma de connectivité entre les couches de convolutions. On retrouve dans cette architecture des couches connectées directement à toutes les autres couches ce qui permet d'augmenter la densité du réseau CNN. L'architecture Densenet121 ne supporte pas des images à dimension supérieure à $(224 \times 224 \times 3)$ [52].

14. Architecture du Réseau de neurones Pré-entraînés [*Xception*] :

Image extraite de l'article référencié sous le nom : [*IEEE-ACCESS • 13 Décembre 2019*].

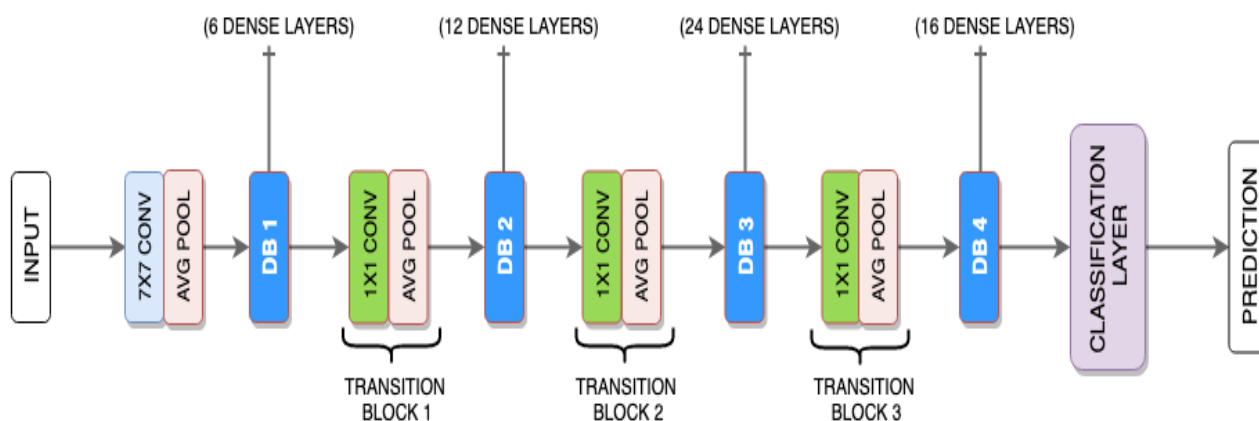


FIGURE 2.12 – Architecture du Réseau de neurones Pré-entraînés [*Densenet-121*].

15

- Toutefois, on note que tous ces modèles sont des modèles pré-entraînés sous la base de données [Imagenet]. Cette base de données contient 20 000 catégories dont plus de 14 millions d'images qui sont destinée à la vision par ordinateur [53].

15. Architecture du Réseau de neurones Pré-entraînés [*Densenet-121*] :

Image extraite de l'article référencié sous le nom : [<https://towardsdatascience.com/>].

16. Algorithmes des réseaux de neurones :

Image extraite de l'article référencié sous le nom : [<https://intelligence-artificielle.com/>].

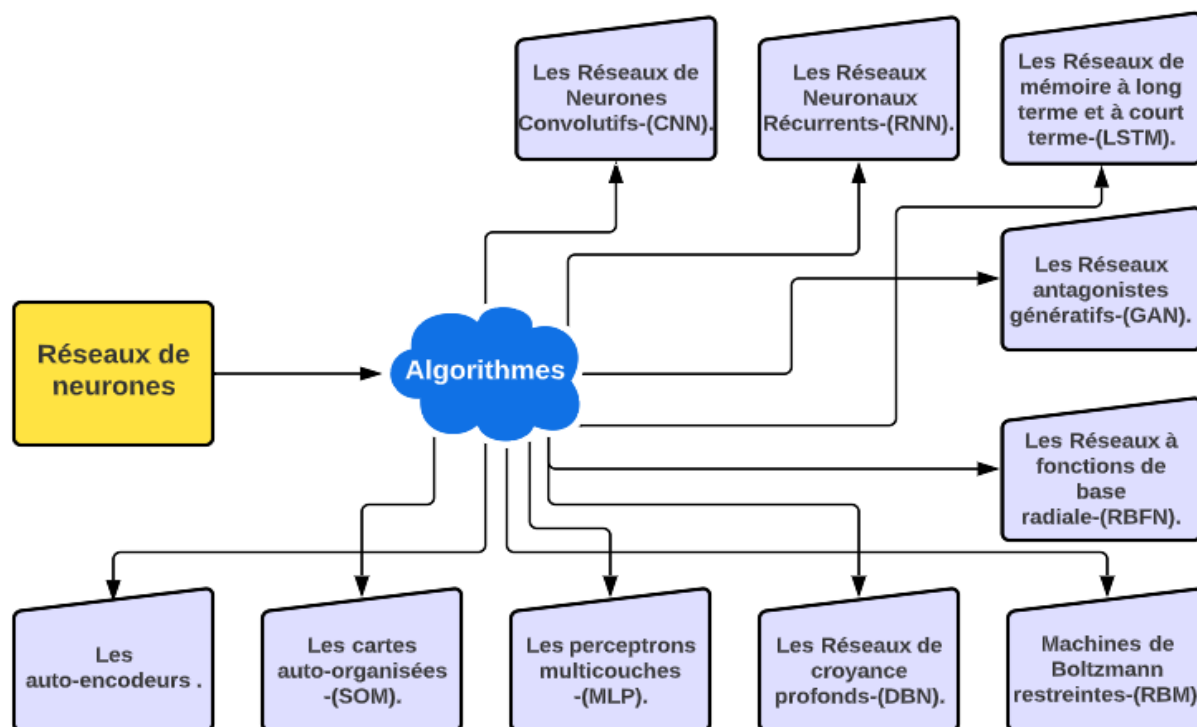


FIGURE 2.13 – Algorithmes des réseaux de neurones.

16

2.4 La Matrice de Confusion

2.4.1 Définition

Une matrice de confusion résume les prédictions qui ont été fournies après l'étude d'un problème de classification dans un processus d'apprentissage automatique. Le résultat de ces prédictions peut être correctes ou incorrectes et seront rangés par catégorie de classe.

- Ces résultats seront par la suite évalués et comparés avec les valeurs réelles attendus en sortie. L'objectif de cette matrice est de nous aider bien interpréter les performances des modèles qui ont effectués les classifications avant de fournir les prédictions.
- Toutefois, la matrice de confusion nous permet d'identifier les erreurs de classification mais pas seulement car, elle nous aide aussi à pointer le type d'erreurs..[54]

2.4.2 Calculer une Matrice de Confusion

Pour effectuer un calcul de matrice de confusion, il faut disposer d'une quantité de données de tests ou de validations qui incluent des valeurs de résultats qui seront fournies en sortie. Ensuite, on effectue des prédictions sur chaque ligne de l'ensemble de données de tests. Par la suite, les résultats attendus en sortie pour faire des prédictions, seront résumé dans une matrice qui fournira le nombre de prédictions correctes et incorrectes pour chaque catégorie de classe à prédire [55].

- Dans chaque ligne de cette matrice, on retrouve les classes à prédire, et dans chaque colonne, les classes réelles. On voit apparaître au niveau de chaque cellule de la matrice les résultats de prédictions qui ont été calculés. Le statut de ces résultats peut être *correctes* - [*Vrais positifs*], *correctes* - [*Vrais négatifs*], ou bien fournit de *fausses prédictions positives* - [*faux positifs*], ou des *prédictions négatives incorrectes* - [*faux négatifs*] [56].

2.4.3 Évaluation d'une Matrice de Confusion

Pour bien assimiler le processus de fonctionnement d'une matrice de confusion, il est nécessaire de bien comprendre la signification de ses paramètres. Ces paramètres sont : **TP**, **TN**, **FP** et **FN**. Ci-dessous, on va expliquer chaque paramètre [57] :

- [**TP** - (*vrais positifs*)] : Le cas où la prédiction est positive et la valeur réelle est positive [57].
- [**TN** - (*vrais négatifs*)] : Le cas où la prédiction est négative et la valeur réelle est négative [57].
- [**FP** - (*faux positifs*)] : Le cas où la prédiction est positive, mais la valeur réelle est négative [57].
- [**FN** - (*faux négatifs*)] : lorsque la prédiction est négative, mais que la valeur réelle est positive [57].

Métrique d'évaluation	Définition	Équation
<i>Precision</i>	C'est une métrique qui désigne le rapport entre les prédictions positives correctes et l'ensemble total de prédictions positives. [58].	$Precision = \frac{TP}{TP + FP}$ [58]
<i>Recall</i>	C'est une métrique qui mesure le nombre de prédictions positives correctes par rapport à l'ensemble total de données positives [58].	$Recall = \frac{TP}{TP + FN}$ [58]
<i>F1-Score</i>	C'est une métrique qui calcule la moyenne harmonique entre la précision et le rappel - (Recall) [58].	$F1 - Score = \frac{2 * (Recall * Precision)}{Recall + Precision}$ [58]
<i>Support</i>	Ce paramètre permet de faire la somme des toutes les données se trouvant dans chaque classe et les classer par ordre de classe puis donner le nombre de données se trouvant dans chacune [58].	$\sum_{i=1}^n \text{Datas on each class}$ [58]
<i>Accuracy</i>	C'est une métrique qui effectue la somme de tous les vrais positifs et vrais négatifs puis divise le résultat par l'ensemble total d'instances [58].	$Accuracy = \frac{TP + TN}{TP + PN + FN + TN}$ [58]

TABLE 2.4 – Métriques de calcul d'une matrice de confusion.

2.4.4 Problème d'overfitting

2.4.4.1 Principe de l'overfitting

L'overfitting est un problème qui est souvent croisé en Machine Learning. Il se manifeste lorsqu'un modèle essaye de s'ajuster excessivement aux données d'entraînement, ce phénomène est appelé - (*Bruit*).

- Ce qui se passe en réalité, c'est que l'apprentissage se focalise sur des données qui n'ont aucun impact direct avec les données d'entraînement. Ce qu'on traduit par bruit, ce sont les données d'entraînement qui sont représentées sous forme de points et qui présentent des variations et fluctuations de façon aléatoires. L'entraînement sur ce type de données rendent les modèles instables [59].

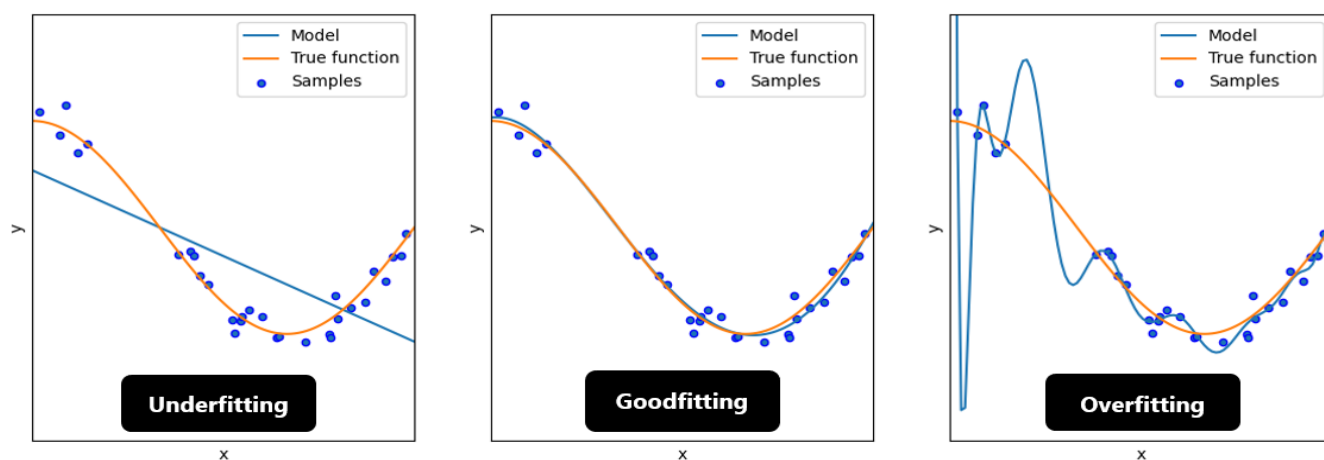


FIGURE 2.14 – Principe de l'overfitting – (*Cas possibles*).

17

17. Principe de l'overfitting – (*Cas possibles*) :

Image extraite de l'article référencié sous le nom : [<https://scikit-learn.org/>].

2.4.4.2 Solutions contre le Problème d'overfitting

[Ajouter des données d'entraînement]

Lorsqu'un Dataset ne contient pas suffisamment de données, la machine ne sera pas en mesure de considérer tous les cas possibles lors des phases d'apprentissage et de test ce qui amènera le prédicteur à faire des erreurs et fournir des résultats soit *médiocres* ou bien *exagérés*. Dans ce genre de situation, il est recommandé de faire quelques ajustements au niveau du Dataset en y ajoutant des données - (*datas*) et refaire le traitement, ce qui permettra à la machine d'avoir plus de données à traiter et donc un meilleur apprentissage ce qui augmentera la performance du prédicteur lors de la phase du test [59].

[Enlever des données]

Dans certains cas, le recours à l'ajustements des données du Dataset c-à-d y ajouter des données peuvent améliorer la performance du prédicteur et dans un autre cas d'autres données - (*datas*) du Dataset vont impacter négativement sur la performance de ce dernier, ce genre de phénomène peut être détecté lors de l'analyse de la matrice de corrélation des données ce qui impose la nécessité de retirer certaines données - (*datas*) du Dataset ou bien programmer la machine à ignorer les colonnes contenant ces données pour qu'elle focalise son apprentissage que sur les données - (*datas*) qui permettront au prédicteur de fournir de meilleurs résultats [59].

[La Validation Croisée - (Cross-Validation)]

La validation croisée est une méthode qui permet de diviser un *dataset* en plusieurs parties pour évaluer la précision du modèle dans la phase d'apprentissage ainsi que celle du test. La principale tâche de cette méthode consiste à *améliorer les performances* des algorithmes d'apprentissage pour obtenir la meilleure précision possible que le prédicteur peut atteindre dans la phase d'*entraînement* et de *test* et ainsi limiter le *problème d'overfitting*. Il existe plusieurs *algorithmes* de la validation croisée, on cite : **le *k-fold*, stratified *k-fold*, Leave One Out** [59].

[K-fold]

La Méthode du **K-fold** consiste à effectuer des *découps symétriques* sur un dataset et réserver un ensemble de bloc pour l'entraînement et un bloc pour la validation puis répéter l'opération jusqu'à parcourir l'ensemble de ce dernier [59].

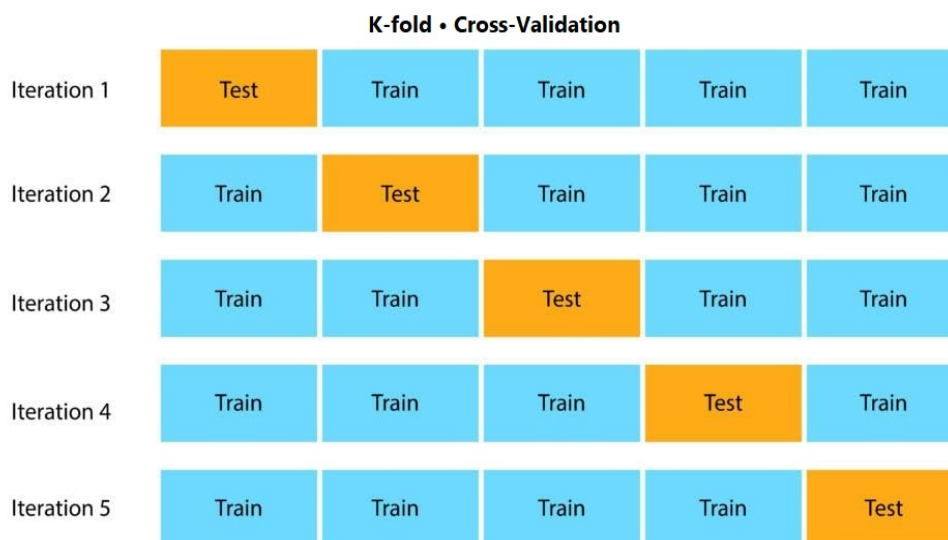


FIGURE 2.15 – Le principe de la méthode du *K-fold*.

18

[Stratified K-fold]

La Méthode du **Stratified K-fold** consiste à effectuer des *splits symétriques* sur un dataset de manière à ce que chaque split regroupe des échantillons de chaque classe se trouvant dans l'ensemble du dataset puis répéter l'opération jusqu'à parcourir l'ensemble de ce dernier [59].

18. Le principe du K-fold :

Image extraite de l'article référencié sous le nom :

[<https://www.section.io/engineering-education/how-to-implement-k-fold-cross-validation>].

19. Le principe du Stratified K-fold :

Image extraite de l'article référencié sous le nom :

[<https://www.askcodez.com/difference-entre-stratified-kfold-et-stratified-shuffle-split-dans-sklearn.html>].

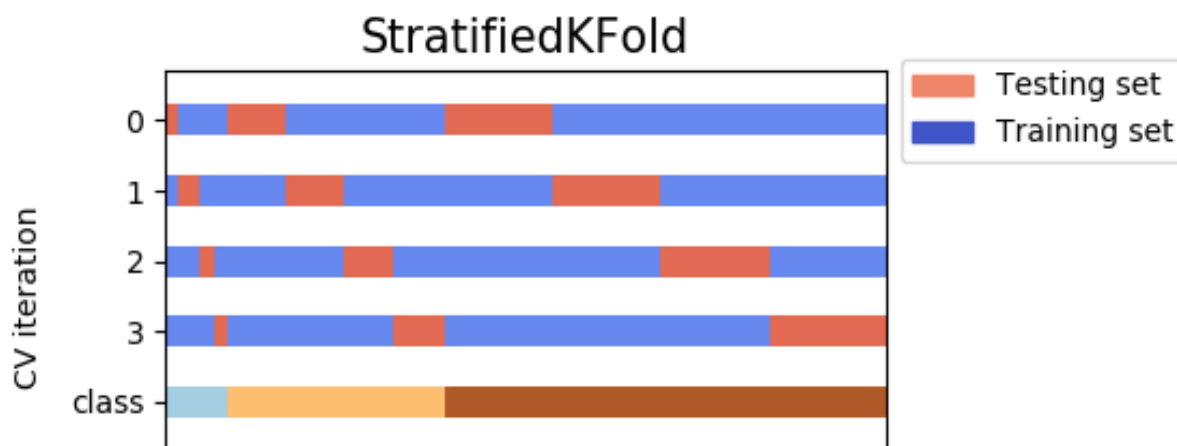


FIGURE 2.16 – Le principe de la méthode du *Stratified K-fold*.

19

[*Leave One Out*]

La Méthode du **Leave One Out** consiste à réserver *un échantillon* du dataset pour la validation et effectuer l'entraînement sur *le reste* de l'ensemble de données puis répéter l'opération jusqu'à parcourir l'ensemble de ce dernier [59].

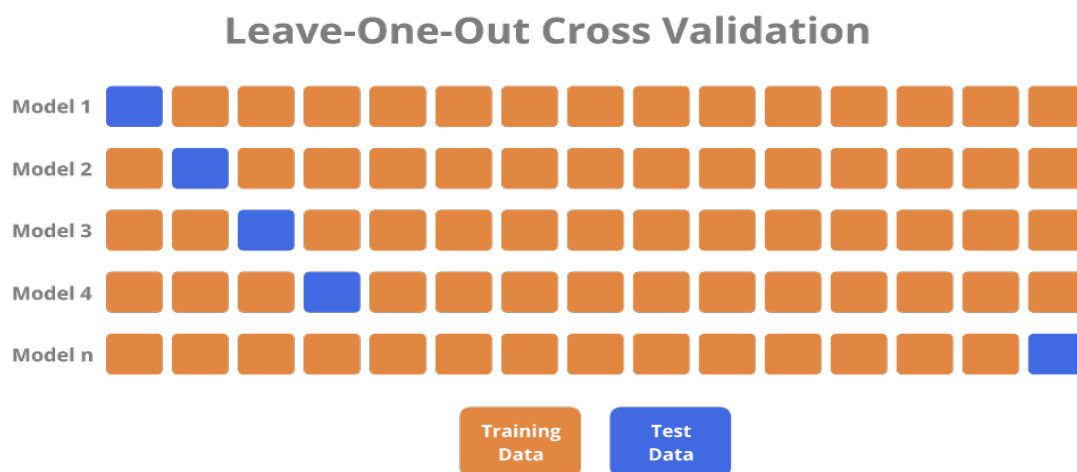


FIGURE 2.17 – Le principe de la méthode du *Leave One out*.

20

20. Le principe du Leave One Out :

Image extraite de l'article référencié sous le nom :

[<https://medium.datadriveninvestor.com/leave-one-out-cross-validation>].

2.5 L'image

2.5.1 Définition

L'image peut être définie comme une perception visuelle ou mentale d'un objet quelconque. Cette perception peut prendre un aspect naturel ou artificiel [60].

2.5.2 Les types d'images

2.5.2.1 L'image Matricielle

L'image matricielle ou aussi appelé – (*image raster*) se définit comme une image numérique qui est composée de points de couleurs connues sous le nom de – (*pixels*) qui sont représentés sur une grille à deux axes X et Y. Toutefois, la résolution d'une image matricielle dépend de sa composition [61].

2.5.2.2 L'image Vectorielle

L'image vectorielle est aussi décrite comme une image numérique qui est constituée par des représentations géométriques distincts – (*droites, carré, triangle, etc...*). Cette catégorie d'image utilise des équations mathématiques pour faire des calculs de positionnement des formes géométriques qui la constituent et ce en prenant en considération des paramètres – (*hauteur, largeur, diamètre, etc...*) qui seront fournis à des vecteurs qui représentent ces formes [62].

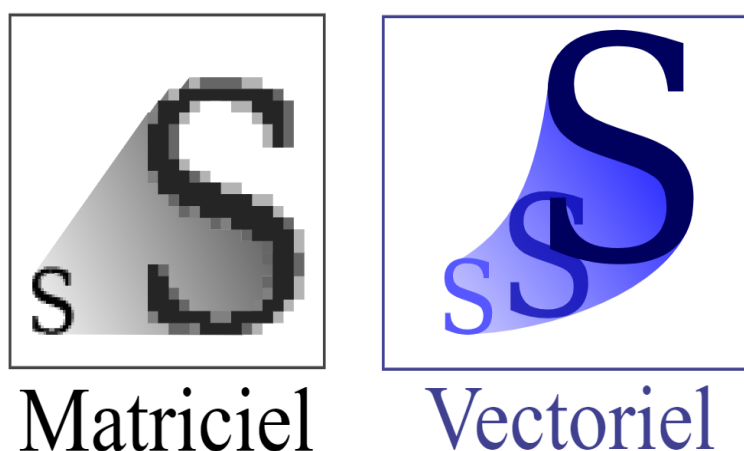


FIGURE 2.18 – Image Matricielle VS Image Vectorielle.

2.6 La différence entre le Machine Learning et le Deep Learning

Critères de Comparaison	<i>Machine Learning</i>	<i>Deep Learning</i>
Domaine d'application [63].	Principalement dans la Classification et Regression des données tabulaires - (numériques, chaînes de caractères, etc...) [63].	Utilisation diversifiée pour la Classification et Regression - (Données tabulaires, traitement de langage naturel et vision par ordinateur) [63].
Quantité de Données à utiliser [63].	Extraction d'une quantité importante d'information [63].	Extraction d'une quantité d'information encore plus importante avec complexités variées. [63].
Type de modélisation [63].	Déroulement et Optimisation par l'intermédiaire d'un modèle statique qui fournit des résultats stables. [63].	Déroulement et Optimisation par l'intermédiaire d'une architecture de réseau de neurones accompagné d'une fonction de perte d'optimisation pour évaluer les pertes de précisions. [63].
Processus d'entraînement [63].	Immédiat [63].	Des itérations - (<i>Batches</i> & <i>Epochs</i>) continus. [63].
Matériel Recommandé [63].	CPU - (Center Processing Unit) [63].	GPU - (Graphic Processing Unit) [63].
Bibliothèques python utilisées [63].	Scikit-learn, Scipy, Panda [63].	Keras, Tenserflow, pytorch [63].

TABLE 2.5 – La différence entre le Machine Learning et le Deep Learning.

2.6.1 Domaines d'applications de l'intelligence artificielle

- | | | |
|---|--|---|
| <ul style="list-style-type: none"> • La reconnaissance faciale. • Les robots industriels. • En médecine. • L'agriculture. | | <ul style="list-style-type: none"> • La détection d'objets. • Le langage naturel. [64] • La conduite autonome. |
|---|--|---|

Machine Learning



Deep Learning



FIGURE 2.19 – Mécanisme de fonctionnement du Machine Learning et du Deep Learning.

21. Mécanisme de fonctionnement du Machine Learning et du Deep Learning :
Image extraite de l'article référencié sous le nom : [<https://www.rotic.ir>].

2.7 L'imagerie médicale et l'intelligence artificielle

L'intelligence artificielle a connu une croissance révolutionnaire dans le domaine de la médecine et ce depuis les années 1950. Cette alliance a permis de développer des techniques sophistiquées qui ont apportées une valeur ajoutée et changer la conduite à tenir des professionnelles de la santé quant aux faites de faire des diagnostics et d'interpréter des données concernant l'état de santé des patients pour avoir une idée sur leurs pronostic vital. L'une de ces techniques est le *Deep Learning* – (Apprentissage profond) qui propose de nos jours divers algorithmes d'apprentissages qui font des prédictions sur des pathologies à gravité accrue tels que le *cancer* en se basant sur le principe de numérisation d'images médicales. L'utilisation de ce genre de technologie vise à simuler l'intelligence humaine dans le seul but de l'assister lors de la phase du diagnostic ce qui permettra un gain en temps et ainsi une prise en charge précoce vis-à-vis du patient et donc un meilleur pronostic vital. Au fil du temps ces techniques ont atteint des performances surprenantes dans le dépistage du cancer dont on en est devenu très dépendant [65].



FIGURE 2.20 – L'intelligence artificielle au service du diagnostic médical.

2.8 Travaux de références sur le cancer du Pancréas

- En 2021, les chercheurs [*Yevgeniy Bodyanskiy*] et [*Olha Chala*], ont utilisés le Dataset : "*A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer*" pour effectuer leur recherche intitulée : "*Generalized neo-fuzzy-neuron with membership functions of special type in medical diagnostics*" et ont obtenu lors de leurs tests une exactitude - (*accuracy*) qui s'élève à **94,78%** [66].

Année	Chercheurs	Dataset	Travail	Accuracy
2021	<i>Yevgeniy Bodyanskiy</i> & <i>Olha Chala</i>	A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer.	Generalized neo-fuzzy-neuron with membership functions of special type in medical diagnostics.	94,78 % (Tests)

TABLE 2.6 – Travail de références sur le cancer du Pancréas.

2.9 Conclusion

Dans ce chapitre, on s'est incrusté dans la partie théorique qui englobe le domaine de l'intelligence artificielle, et ses différents algorithmes.

- D'une part, on a montré l'apport que pouvait avoir l'intelligence artificielle en liaison avec la médecine en termes de traitement et interprétation de données pour faire des prédictions et diagnostics fiables en parlant de l'imagerie médicale qui est l'élément de références pour les médecins dans la phase de diagnostic des maladies à gravité accrue tels que le cancer.
- Prochainement, dans le chapitre qui suit, nous allons nous immiscer dans l'implémentation des deux (02) méthodes de l'intelligence artificielle *Machine Learning* et *Deep Learning* et évaluer les résultats fournis lors de la phase de test et choisir le meilleur modèle et suggérer une amélioration si possible.

Approche et solution proposée

3.1 Introduction

Dans ce chapitre nous avons impliqué le dépistage du cancer du Pancréas avec les techniques de l'intelligence artificielle. On a résumé notre étude sur la base des deux techniques invoquées précédemment en focalisant notre attention sur la phase d'entraînement des modèles tels que : *Random Forest Classifier* et *K-Nearest Neighbors* qui sont utilisés dans l'apprentissage automatique ainsi que les modèles d'apprentissage profond tels que : *Xception* et *DenseNet-121* afin qu'ils soient en mesure de faire des prédictions sur des données externes. Cependant, durant cette phase, on a eu *recours* à des ambiguïtés sur le plan évaluation des performances dont on y a remédié avec des solutions souvent utilisées en *Machine Learning* et *Deep Learning* qui sont détaillées ci-dessous.

3.2 Techniques de dépistages Proposées - (Par Analyses médicales & Imageries médicales)

3.2.1 Dépistage par Analyses médicales

3.2.1.1 Structure et Prototypage de la solution proposée

- Dans un premier temps, on va résumer sous forme de schéma, la structure de la solution proposée pour la première approche, ci-dessous :

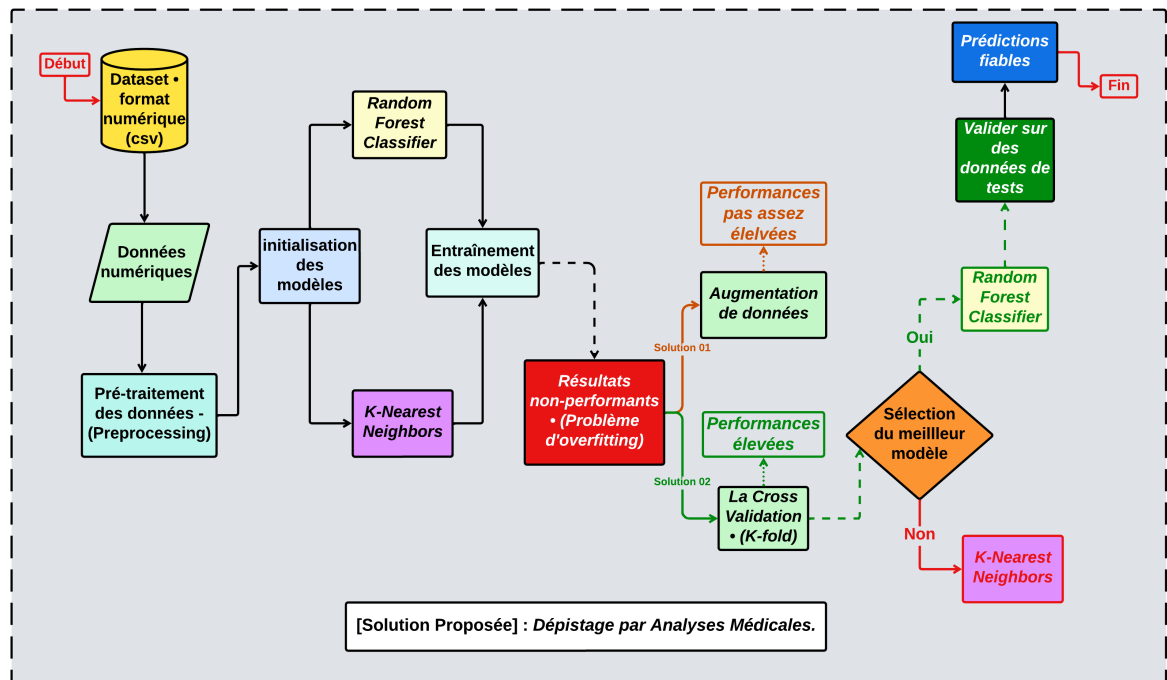


FIGURE 3.1 – Solution Proposée pour la première approche de dépistage - (*Dépistage par analyses médicales*).

3.2.1.2 Dataset

- Le Dataset qu'on a utilisé pour l'implémentation de l'algorithme du "*Dépistage par Analyses Médicales*", a été récupéré depuis la plate-forme [Kaggle] où on a trouvé un lien direct vers l'étude qui a été effectuée sur le Cancer du Pancréas intitulée : "*A combination of urinary biomarker panel and PancRISK score for earlier detection of pancreatic cancer : A case-control study*"- *En 2021* qui a fournit comme résultat ce *dataset* avec une dimension de [590-lignes] et [14-colonnes] contenant trois - (03) *classes* distinctes : [Healthy], [Benign] et [Malignant]. On cite aussi que ce Dataset a été utilisé dans une autre étude sur le *Machine Learning* intitulée : "*Generalized neo-fuzzy-neuron with membership functions of special type in medical diagnostics*" réalisée par : [Yevgeniy Bodyanskiy] et [Olha Chala] dans laquelle ils ont obtenus une Accuracy de 94,78 % durant la phase de tests.

3.2.1.3 Répartition du Dataset

- Le Dataset qu'on a utilisé dans l'implémentation de l'algorithme du "*Dépistage par Analyses Médicales*" est fourni sous *format CSV*.
- Ce Dataset contient [*590-lignes*] et [*14-colonnes*] qui sera redimensionné par la suite en [*590-lignes*] et [*12-colonnes*].
- Sur ce Dataset, on a effectué deux (02) découpes pour faire notre étude :

[75 % pour les données d'entraînement • (*Train-set*)]

&

[25 % pour les données de Test • (*Test-set*)].

- La répartition est détaillée dans le tableau ci-dessous :

Dataset	Données Fractionnées	Quantité de Données
<i>Données d'entraînement</i>	75 %	442
<i>Données de Test</i>	25 %	148

TABLE 3.1 – Répartition des données – (Entraînement et Test) du Dataset au Format Numérique utilisé dans le Machine Learning.

3.2.1.4 Visualisation des Données

- Pour avoir une meilleure vision des choses on a tenu à convertir les données de chaque Dataset sous forme de graphe qui résume les statistiques de chaque classes, c-à-d le nombre de données se trouvant dans chaque catégorie de classe :

Dataset • (Format numérique - [CSV]) :
[0] : *Healthy* = 183 | [1] : *Benign* = 208 | [2] : *Malignant* = 199.

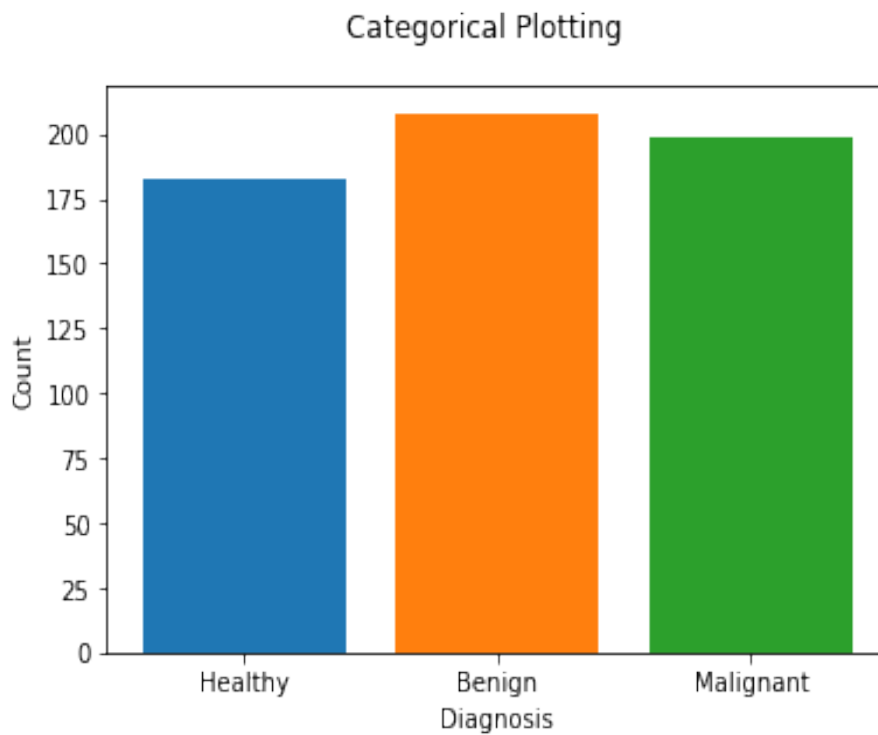


FIGURE 3.2 – Visualisation des données labelisées des trois – (03) classes de diagnostiques se trouvant dans le Dataset au Format numérique.

3.2.1.5 Encodage et Normalisation des Données

- Avant d'effectuer une étude sur un **dataset** et d'utiliser un modèle adéquat à cette étude, il faut procéder à une étape obligatoire qui permettra d'effectuer un filtrage et sélection de données dans ce Dataset, c'est le ***Preprocessing - (pré-traitement)***. Ce travail se déroule par le biais de deux (02) méthodes qui sont très souvent sollicitées en ***Machine Learning*** et ***Deep Learning*** qui sont : ***L'encodage*** et ***La normalisation*** de données.

[L'encodage]

- Lorsqu'on travaille sur un problème de machine learning, les données qui y sont traitées contiennent plusieurs *caractéristiques* - (features) qui peuvent se trouver dans une ou plusieurs colonnes du dataset. Ces données peuvent être interprétées sous différents format généralement numérique ou alphabétique. Elles sont à priori fournis sous ses formats-ci pour qu'elles soient interprétées facilement par l'homme.
- Dans le contexte de notre travail, on a procédé à un *encodage* de données pour que ces dernières pourront être interprétées par la machine. Pour ce faire il existe des méthodes qui sont implémentées tel que le ***LabelEncoder()*** et ***to_categorical()*** qui permettent de convertir ces données sous format numérique et qui seront considérées comme étiquettes d'un ou plusieurs ensemble de données [67].

[La Normalisation]

- Après l'étape de l'encodage de données, vient l'étape de la *normalisation*, qui quant à elle permet de réduire la complexité des modèles avant d'appliquer des algorithmes d'apprentissage. Pour ce faire, elle standardise la moyenne et l'*écart-type* de chaque type de données pour simplifier l'apprentissage. Toutes ces transformations de données s'effectuent par le biais de la formule suivante [68] :

$$X_{normalisé} = \frac{X - m}{\sigma}$$

[68]

- $[\mathbf{X}]$: Les données à normaliser.
- $[\mathbf{m}]$: La moyenne empirique des données.
- $[\sigma]$: L'écart-type entre les données.

- Comme pour l'encodage, il existe aussi des méthodes implémentées qui permettent de faire ce type de conversion tels que *StandardScaler()* et *normalize()* [68].

3.2.1.6 Augmentation de Données

- L'augmentation de données consiste à augmenter artificiellement la taille d'un jeu de données d'apprentissage en élaborant d'autres instances sur les mêmes données à partir de données fournies à la machine lors de la phase d'entraînement. Le but de cette manoeuvre est d'accroître le nombre de données se trouvant dans le dataset [69].
- L'application de cette technique varie selon le type de données se trouvant dans le dataset qu'on traite.
- Dans notre étude, on a eu recours à l'augmentation de données au niveau de la première approche "*Dépistage par analyses médicales*" comme première solution pour améliorer les performances des modèles après une première simulation. En effet, notre dataset contenant [590-lignes] et [14-colones] qui regroupent les données suivante : *sample_id*¹, *patient_cohort*², *sample_origin*³, *age*⁴, *sex*⁵, *diagnosis*⁶, *stage*⁷, *benign_sample_diagnosis*⁸, *plasma_CA19_9*⁹, *creatinine*¹⁰, *LYVE1*¹¹, *REG1B*¹², *TFF1*¹³, *REG1A*¹⁴.
- Après avoir effectué des opérations de filtrage de données, quatre (04) colones ont été tronquées du dataset qui sont : *stage*, *benign_sample_diagnosis*, *plasma_CA19_9*, *REG1A*, car ces dernière contenaient des valeurs nulles - (*NaN*).

- Dans un premier temps, on a essayé d’ajuster les colonnes telles que : **stage**, **benign_sample_diagnosis** qui contiennent des **cellules vides** par ces valeurs respectives : [”R.A.S”] pour **benign_sample_diagnosis** et [0] pour **stage**. En ce qui concerne les autres colonnes contenant des valeurs nulles telles que : **plasma_CA19_9**, **REG1A**, on n’a pas eu à les augmenter pour les utiliser dans notre études car elle contiennent un nombre important de cellules vides et pour les ajuster, il faut utiliser les valeurs d’autres cellules de la même colonne pour y effectuer la moyenne ce qui n’est pas ergonomique sur le plan pratique et théorique vu que les valeurs résultantes restent purement théoriques et non réelles, ce qui nous a contraint de les ignorer compte tenu de l’amélioration qu’il y a eu en ajustant seulement les colonnes : **stage**, **benign_sample_diagnosis**.
- Cependant, les colonnes suivantes : **sample_id** , **patient_cohort**, **sample_origin**, **age**, **sex** sont facultatives, elles contiennent des données qui ne sont pas obligatoires dans le diagnostic du cancer de pancréas mais seulement à faire des statistiques sur la maladie selon les professionnelles de la santé.
- Toutefois, les colonnes qui sont concernées par la phase d’apprentissages des modèles sont : **diagnosis**, **stage**, **benign_sample_diagnosis**, **creatinine**, **LYVE1**, **REG1B**, **TFF1**.

sample_id	patient_cohort	sample_origin	age	sex	diagnosis	stage	benign_sample_diagnosis	plasma_CA19_9	creatinine	LYVE1	REG1B	TFF1	REG1A	
0	S1	Cohort1	BPTB	33	F	1	NaN	NaN	11.7	1.83222	0.893219	52.948840	654.282174	1262.000
1	S10	Cohort1	BPTB	81	F	1	NaN	NaN	NaN	0.97266	2.037585	94.467030	209.488250	228.407
2	S100	Cohort2	BPTB	51	M	1	NaN	NaN	7.0	0.78039	0.145589	102.366000	461.141000	NaN
3	S101	Cohort2	BPTB	61	M	1	NaN	NaN	8.0	0.70122	0.002805	60.579000	142.950000	NaN
4	S102	Cohort2	BPTB	62	M	1	NaN	NaN	9.0	0.21489	0.000860	65.540000	41.088000	NaN

FIGURE 3.3 – Aperçu du Dataset avant l’augmentation de données.

	sample_id	patient_cohort	sample_origin	age	sex	diagnosis	stage	benign_sample_diagnosis	plasma_CA19_9	creatinine	LYVE1	REG1B	TFF1	REG1A
0	S1	Cohort1	BPTB	33	F	1	0	RAS	11.7	1.83222	0.893219	52.948840	654.282174	1262.000
1	S10	Cohort1	BPTB	81	F	1	0	RAS	NaN	0.97266	2.037585	94.467030	209.488250	228.407
2	S100	Cohort2	BPTB	51	M	1	0	RAS	7.0	0.78039	0.145589	102.366000	461.141000	NaN
3	S101	Cohort2	BPTB	61	M	1	0	RAS	8.0	0.70122	0.002805	60.579000	142.950000	NaN
4	S102	Cohort2	BPTB	62	M	1	0	RAS	9.0	0.21489	0.000860	65.540000	41.088000	NaN

FIGURE 3.4 – Aperçu du Dataset après l’augmentation de données.

- Enfin, l’augmentation de données bien qu’elle ait permis de mieux optimiser les performances des modèles, ces derniers ont montré qu’ils pouvaient atteindre de meilleures performances, ce qui nous a amené à proposer une autre solution pour mieux améliorer ces performances qui est la validation croisée - (Cross-Validation) dont les performances seront abordées par la suite.

-
1. L’identifiant des Patients.
 2. Population sur laquelle l’étude a été menée.
 3. L’origine des Patients.
 4. L’âge des Patients.
 5. Le sexe des Patients.
 6. Le diagnostic effectué sur les patients.
 7. Stade de la maladie - (Cancer du Pancréas : 0 | I | II | III | IV).
 8. Premières observations sur l’état des patients.
 9. Antigène Carbohydre, c’est un marqueur tumoral utilisé pour dépister les tumeurs du Pancréas.
 10. Produit éliminé par les urines après la filtration des reins.
 11. Niveaux urinaires du récepteur 1 de l’hyaluronane endothélial des vaisseaux lymphatiques, une protéine qui pourrait jouer un rôle dans les métastases tumorales.
 12. Taux urinaires d’une protéine qui pourrait être associée à la régénération du pancréas.
 13. Niveaux urinaires du Trefoil-Factor1, lié à la régénération et réparation des voies urinaires.
 14. Niveaux urinaires de protéines, qui peuvent être associés à la régénération du pancréas.

3.2.1.7 Entraînement des Modèles

[Résumé de l'étude]

- Dans cette technique de dépistage, notre étude s'est effectuée sur la comparaison de plusieurs algorithmes pour en évaluer la performance dans le côté entraînement et test pour en sélectionner par la suite le meilleur algorithme qui a pu s'adapter le mieux à notre modèle d'apprentissage. Les algorithmes qu'on a sollicités durant cette étude sont :
 1. Le Modèle • [*Random Forest Classifier* – (*RFC*)].
 2. Le Modèle • [*K-Nearest Neighbors* – (*K-NN*)].
- Cependant, durant ce parcours d'étude on a eu à faire à quelques obstacles durant les deux - (02) phases : **Entraînement** et **Test** qui a fait que certains algorithmes se sont bien adaptés au modèle d'apprentissage mais n'ont pas donnés de bon résultat dans le test de performance - (**Test-Accuracy**) et d'autres n'ont pas pu satisfaire les deux - (02) phases dans leurs performances. Ce phénomène s'est produit car il y avait des caractéristiques - (*features*) qui n'avaient pas assez de données et qui influaient négativement sur les données sensibles qui sont fortement sollicitées dans le calcul de performance, cette déduction nous ait parvenu par l'analyse de la corrélation entre les données.
- Cette Problématique nous a contraint à faire des modification - **Augmentation de données** au niveau de ces caractéristiques - (*features*) pour en évaluer le résultat qui y est avéré satisfaisant. Mais d'un autre côté en évaluant la limite que certains algorithmes pouvaient atteindre avec la fonction *learning_curve()*, on a constaté que leurs performances pouvaient atteint des résultats encore plus élevés.
- Ensuite en utilisant la fonction *validation_curve()* après l'implémentation de l'algorithme de la *cross-validation*, on a confirmé que le modèle à bien atteint la performance attendue.

- Dans cas où, on a eu recours à la **Cross-Validation**, on a utilisé trois (03) méthodes successivement **LeaveOneOut**, **K-fold** et **Stratified K-fold**. En effectuant une comparaison entre les résultats fournis par chaque méthode, on a constaté que la méthode **K-fold** avait fourni les meilleurs résultats et on l'a validé en faisant **la prédiction sur les 25 %** de données de tests et qui ont été très satisfaisante.

3.2.1.8 Préparation de la phase d'entraînement

[Initialisation des Modèles]

- Le modèle • [**Random Forest Classifier**] :

```
1          # Random Forest Classifier
2      from sklearn.ensemble import RandomForestClassifier
3      RFC = RandomForestClassifier(n_estimators=10,
4          criterion='entropy', random_state=0)
5      RFC.fit(X_train, Y_train)
```

Implémentation du Modèle - Random Forest Classifier avec l'augmentation de données.

- Le modèle • [**K-Nearest Neighbors**] :

```
1          # K-Nearest Neighbors Classifier
2      from sklearn.neighbors import KNeighborsClassifier
3      KNN = KNeighborsClassifier(n_neighbors=5)
4      KNN.fit(X_train, Y_train)
```

Implémentation du Modèle - K-Nearest Neighbors avec l'augmentation de données.

- Le modèle • [*Random Forest Classifier*] - optimisé avec la méthode • [*K-fold*] :

```
1           # Random Forest Classifier
2   # Importing the K-Fold Method and the Cross Val Score
3   # from Sklearn
4   from sklearn.model_selection import KFold
5   from sklearn.model_selection import cross_val_score
6
7   ## Fitting the Random Forest Classifier Model
8   rfc2 = RandomForestClassifier(n_estimators=10,
9   criterion='entropy', random_state=0)
10  rfc2.fit(X_train, Y_train)
11
12  # Creating the K-Fold Object with the number of folds
13  # parameter
14  Kfold_validation = KFold(10)
15
16  # Training and Testing the datas on differents models
17  # with the K-Fold Method
18  rfcCVTrainAccuracy= cross_val_score(rfc2, X_train,
19  Y_train, cv=Kfold_validation)
20
21  rfcCVTestAccuracy= cross_val_score(rfc2, X_test, Y_test,
22  cv=Kfold_validation)
```

Implémentation du Modèle - Random Forest Classifier avec la Cross-Validation.

Le modèle • [*K-Nearest Neighbors*] - optimisé avec la méthode • [*K-fold*] :

```
1           # K-Nearest Neighbors Classifier
2   # Importing the K-Fold Method and the Cross Val Score
3   # from Sklearn
4   from sklearn.model_selection import KFold
5   from sklearn.model_selection import cross_val_score
6
7   ## Fitting the K-Nearest Neighbors Classifier Model
8   knn2 = KNeighborsClassifier(n_neighbors=5)
9   knn2.fit(X_train, Y_train)
10
11  # Creating the K-Fold Object with the number of folds
12  # parameter
13  Kfold_validation = KFold(10)
14
15  # Training and Testing the datas on different models
16  # with the K-Fold Method
17
18  knnCVTrainAccuracy= cross_val_score(knn2, X_train, Y_train,
19  cv=Kfold_validation)
20
21  knnCVTestAccuracy= cross_val_score(knn2, X_test, Y_test,
22  cv=Kfold_validation)
```

Implémentation du Modèle - K-Nearest Neighbors avec la Cross-Validation.

[Lancement de la phase d'entraînement des Modèles]

- Après le lancement de l'apprentissage, les algorithmes vont essayer de s'adapter au maximum au modèle d'apprentissage puis fournir une *précision* d'entraînement par la suite qui nous permettra d'avoir une idée de l'algorithme final à utiliser dans notre étude.

3.2.2 Dépistage par Imageries médicales

3.2.2.1 Structure et Prototypage de la solution proposée

- Dans cette partie, on va résumer sous forme de schéma, la structure de la solution proposée pour la deuxième approche qui à été implémenté avec la technique du *Transfert Learning* comme suit :

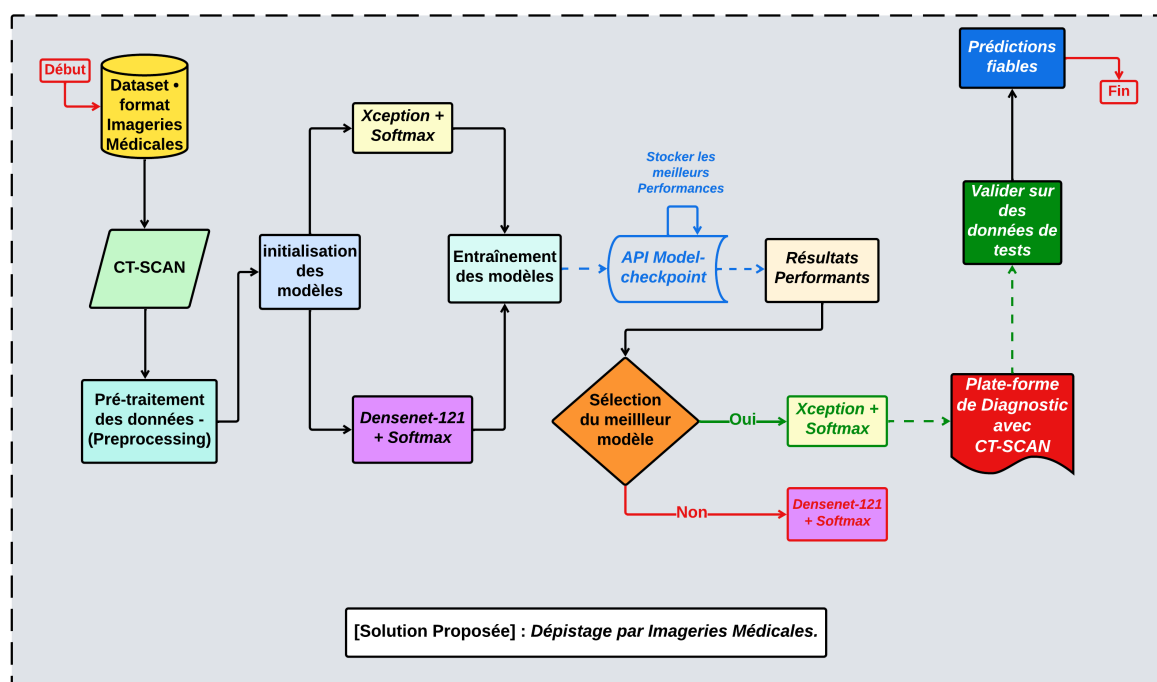


FIGURE 3.5 – Solution Proposée pour la deuxième approche de dépistage - (*Dépistage par imageries médicales*).

3.2.2.2 Dataset

- En ce qui concerne le Dataset des **CT-SCAN**¹⁵- (**Imagerie Médicale**) qu'on a utilisé pour l'implémentation de l'algorithme du "*Dépistage par Imagerie Médicales*". Il a été récupéré depuis la plate-forme [*Github*] qui a été publié récemment le : [28/07/2021]. Ce Dataset contient deux (02) classes distinctes [*Cancerous*] et [*Non-Cancerous*] avec une dimension de (908 x 722 x 3) qui seront redimensionnées par la suite en (299 x 299 x 3) et (224 x 224 x 3) selon le modèle pré-entraîné utilisé.

3.2.2.3 Repartition du Dataset

- Le Dataset qu'on a utilisé dans l'implémentation de l'algorithme "*Dépistage par Imagerie Médicales*", inclut **deux (02) classes** :

- *Cancerous* et *Non-cancerous* qui sont des images au *format JPG* qui font en tout un total de **1603-images** qu'on a par la suite découpés en deux (02) parties - (**1292-images**) :

- [75 % Données d'entraînement • (*Train-set*)] - ce qui fait **969 images**.
- &
- [25 % Données de Validation • (*Validation-set*)] - ce qui fait **323 images**.

- Ainsi que les **données de Test** • (*test-set*) qui font (**310-images**) fractionnées sur les deux – (02) classes.

- La répartition est détaillée dans le tableau ci-dessous :

Dataset	Quantité total de données	Classes	Données Fractionnées	Quantité de Données
<i>Données d'entraînement</i>	1292	<i>Cancerous</i>	75 %	969
		<i>Non-Cancerous</i>		
<i>Données de Validation</i>		<i>Cancerous</i>	25 %	323
		<i>Non-Cancerous</i>		
<i>Données de Test</i>	310	<i>Cancerous</i>	Non-fractionnées	310
		<i>Non-Cancerous</i>		

TABLE 3.2 – Répartition des données – (Entraînement, Validation et Test) du Dataset des images médicales utilisé dans le Deep Learning.

15. Computer-Tomodensometrie SCAN ou Scanographie, Tomographie axiale calculée par ordinateur.

3.2.2.4 Visualisation des Données

- Pour avoir une meilleure vision des choses on a tenu à convertir les données de chaque Dataset sous forme de graphe qui résume les statistiques de chaque classes, c-à-d le nombre de données se trouvant dans chaque catégorie de classe :

Dataset • (Image - [CT-SCAN]) :
[*Cancerous*] = 1003 | [*Non-cancerous*] = 288.

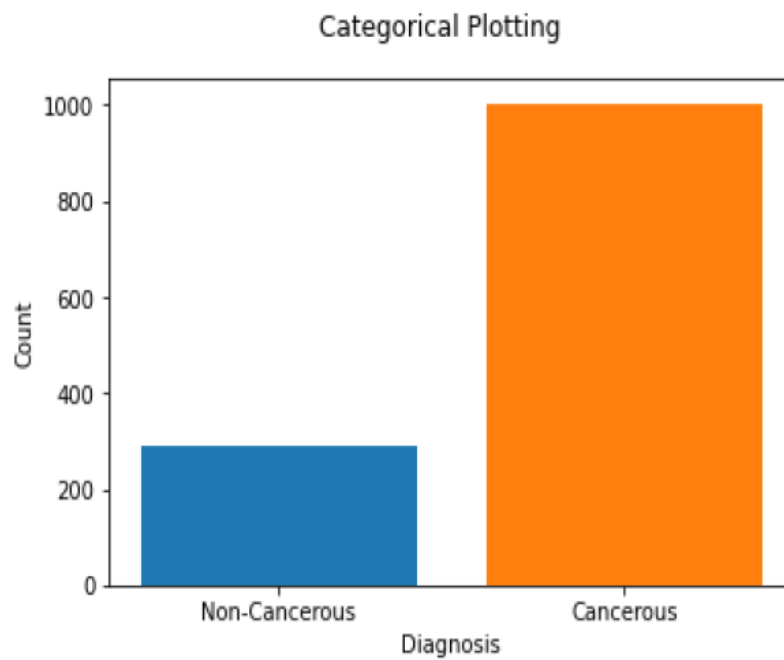


FIGURE 3.6 – Visualisation des images labelisées des deux – (02) classes de diagnostiques se trouvant dans le Dataset des CT-SCAN.

[Résumé de l'étude]

- Pour effectuer notre étude sur cette technique de dépistage, nous avons suivi la même logique que celle de la première technique c-à-d importer un Dataset contenant un nombre de données - (**Images CT-SCAN**) qu'on a mis à disposition de deux modèles d'apprentissages - (algorithmes pré-entraînés) implémentés sur la technique des réseaux de neurones convolutifs - (CNN), puis comparer les résultats qu'ils ont fournis durant leur apprentissage. Les algorithmes - (CNN) sollicités dans cette étude sont :

1. Le Modèle • [*Xception*].
2. Le Modèle • [*DenseNet-121*].

- Après avoir effectué cette étape avec succès, en comparant les données fournies en sortie par chaque modèle, on a constaté que l'apprentissage des modèles en question s'est très bien passé quant tenu des valeurs qu'ils ont fournis sur les données *entraînement* et en *validation*.

3.2.2.5 Les Fonctions d'activation utilisées pour la classification**[La Fonction - *Relu*¹⁶]**

La fonction *Relu* consiste à appliquer une **fonction linéaire rectifiée** sur les neurones artificiels. Cette fonction retourne par défaut une valeur d'entrée maximal aussi appelée "*seuil*" de "0" - $\max(x, 0)$. Toutefois, cette valeur peut être ajustée pour obtenir d'autres valeurs non nulles [70].

[La Fonction - *Sigmoid*]

La fonction *sigmoïde* connue sous sa formule mathématique :

$S(x) = 1 / (1 + \exp(-x))$ qui est **une fonction non-linéaire** utilisé pour fournir des résultats de prédiction en sortie dans un réseau de neurones artificiels. Cette fonction effectue des calculs complexes sur des données numériques fournies par chaque neurone

puis retourne un résultat qui sera évalué par la suite. Si les valeurs fournies sont inférieures à (< -5), la fonction retourne des valeurs *proches* de **zéro - (0)** sinon elle retourne des résultats *proches* de **1**. La fonction sigmoïde retourne des valeurs se trouvant dans un *intervalle* de $[0, 1]$ [70].

[La Fonction - *Softmax*]

La fonction *Softmax* effectue des conversions sur des vecteurs de valeurs en des distributions de probabilités. Cette opération s'effectue par le biais d'une équation mathématique formulée comme suit : $\exp(x)/\text{tf.reduce_sum}(\exp(x))$ qui est **une fonction non-linéaire**. Les valeurs représentées par les vecteurs qui sont fournies dont chacun est traité indépendamment, sont comprises entre $[0, 1]$ et leur somme donne toujours **la valeur 1** [70].

- Dans notre étude, on a utilisé la fonction d'activation **Softmax** car, notre Dataset est fourni avec deux - (02) classes **Cancerous** et **Non-cancerous**.

3.2.2.6 Optimisateur et Fonctions de perte d'optimisation

[La Fonction - *Adam*]

L'optimisateur **Adam** est décrit comme une méthode dérivée des méthodes de descente de gradient qui sont connues pour leurs applications sur l'évaluation des estimations d'un modèle [71].

[La Fonction - *Categorical-Crossentropy*]

La fonction *categorical_crossentropy()* permet de calculer la perte d'entropie croisée entre les étiquettes réelles et prédites. On utilise cette fonction généralement quand on dispose de plus de deux (02) classes d'étiquettes - (2, 3, 4 classes) [72].

[La Fonction - *Binary-Crossentropy*]

La fonction *binary_crossentropy()* permet de calculer la perte d'entropie croisée entre les étiquettes réelles et prédites. On utilise souvent cette fonction dans le cas d'une classification **binaires (0 ou 1)** - (deux (02) classes) [72].

- Dans notre étude, au niveau de l'étape de compilation des modèles, on a utilisé l'optimisateur *Adam* avec la fonction *categorical_crossentropy()* comme fonction qui calcule les pertes d'optimisation pendant la phase d'apprentissage des modèles.

3.2.2.7 Les API Callbacks

Les *API Callbacks*, sont des objets fournis par la plateforme *Keras*, qui permettent d'améliorer la phase d'apprentissage d'un modèle en gardant que les meilleures itérations – (*Epochs*). Parmi les API utilisés, on cite [73] :

[Model-Checkpoint]

Le *model-Checkpoint*, permet d'enregistrer les phases d'apprentissages – (*Epochs*) qui se sont bien déroulées et d'ignorer les phases qui ont fournis des résultats médiocres. Le but de cette opération est d'obtenir un bon modèle d'entraînement qui fournira des prédictions correctes et fiables [74].

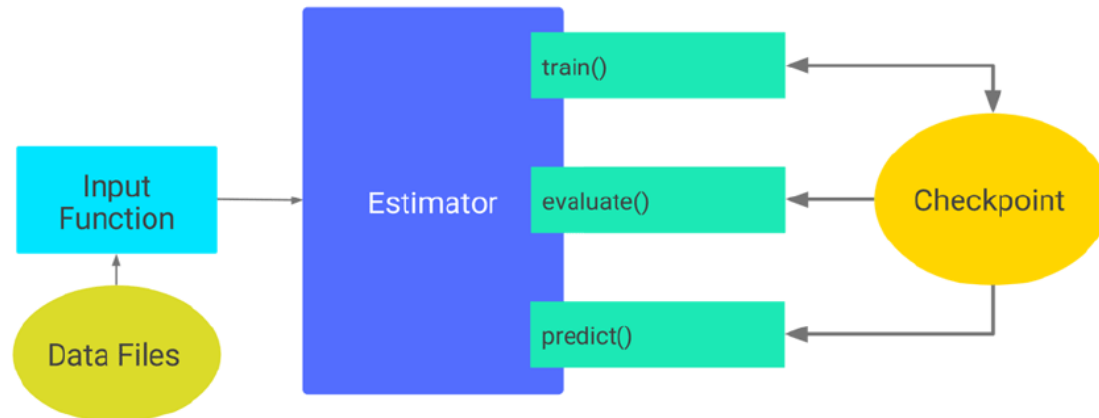


FIGURE 3.7 – Processus de fonctionnement du Model-Checkpoint durant la phase d'apprentissage.

17. Processus de fonctionnement du Model-Checkpoint :

Image extraite de l'article référencié sous le nom : [<https://camo.githubusercontent.com/>].

3.2.2.8 Préparation de la phase d'entraînement des modèles

[Initialisation des Modèles]

1. Le modèle • [Xception].

```
1           # The Xception Model
2
3     modelXception = Sequential(Xception(weights="imagenet",
4     include_top=False, input_shape=(299, 299, 3),
5     pooling='avg', classes=2))
6
7     for layer in modelXception.layers:
8         layer.trainable=False
9
10    modelXception.add(Flatten())
11
12    modelXception.add(Dense(512))
13    modelXception.add(Activation('relu'))
14
15    modelXception.add(Dropout(0.2))
16    modelXception.add(Dense(2))
17
18    modelXception.add(Activation('softmax'))
19
20    modelXception.summary()
```

Implémentation du Modèle (Xception + Softmax).

2. Le modèle • [DenseNet-121].

```
1          # The Densenet-121 Model
2
3      modelDenseNet121 = Sequential(densenet.DenseNet121(
4          weights="imagenet", include_top=False,
5          input_shape=(224, 224, 3),
6          pooling='avg', classes=2))
7
8      for layer in modelDenseNet121.layers:
9          layer.trainable=False
10
11     modelDenseNet121.add(Flatten())
12
13     modelDenseNet121.add(Dense(512))
14     modelDenseNet121.add(Activation('relu'))
15
16     modelDenseNet121.add(Dropout(0.2))
17     modelDenseNet121.add(Dense(2))
18
19     modelDenseNet121.add(Activation('softmax'))
20
21     modelDenseNet121.summary()
```

Implémentation du Modèle (Densenet-121 + Softmax).

[Compilation des Modèles]**1. Le modèle • [Xception]**

```
1      # Compiling the [Xception + Softmax] Model
2      modelXception.compile(optimizer = 'adam',
3                             loss = 'categorical_crossentropy',
4                             metrics=['accuracy'])
```

Compilation du Modèle (Xception + Softmax).

2. Le modèle • [DenseNet-121]

```
1      # Compiling the [Densenet-121 + Softmax] Model
2      modelDenseNet121.compile(optimizer = 'adam',
3                               loss = 'categorical_crossentropy',
4                               metrics=['accuracy'])
```

Compilation du Modèle (Densenet-121 + Softmax).

[Lancement de l'entraînement des Modèles]**1. Le modèle • [Xception]**

```
1      # Launching the Model [Xception + Softmax])
2      historyXception = modelXception.fit(
3          X_train, Y_train,
4          batch_size=16, verbose=1, epochs=30,
5          validation_data=(X_test, Y_test),
6          shuffle=False, callbacks=[Callbacks])
```

Lancement du Modèle (Xception + Softmax).

2. Le modèle • [DenseNet-121]

```

1  # Launching the Model [Densenet121 + Softmax]
2  historyDenseNet121 = modelDenseNet121.fit(
3  X_train, Y_train,
4  batch_size=16, verbose=1, epochs=30,
5  validation_data=(X_test, Y_test), shuffle=False,
6  callbacks=[Callbacks])

```

Lancement du Modèle (Densenet121 + Softmax).

3.2.2.9 Aperçu et résumé des modèles créés

1. Le Modèle • [*Xception*]

- Le modèle *Xception* va recevoir des images CT-SCAN en entrées qui sont redimensionnées en **(299 x 299 x 3)** puis procéder à un processus d'extraction de caractéristiques. Sachant que le modèle *Xception* est un modèle pré-entraîné la avec la base de données *Imagenet*, nous avons isolé la partie du modèle qui inclus les couches de convolutions et de Pooling - (Couches cachées - (*Hidden layers*) du CNN) et apporter des modifications que sur la couche totalement connectée - (*Fully conected Layer*) pour l'adapter à notre étude. Pour ce faire, on a procéder au départ avec une fonction *GlobalAveragePooling2D()* qui reçoit les caractéristiques des images extraites au préalable et en effectue la moyenne. Après cette étape, nous avons appliqué une autre fonction *Flatten()* qui permet de redimensionner ces caractéristiques pour leurs affecter une seule et même dimension.
- Ensuite, pour effectuer la classification, on a créé une couche *dense* qui contient **512-neurones** suivi d'une fonction d'activation *Relu*. On a appliqué par la suite un *dropout (0.2) = 20 %* pour prévenir *l'overfitting* et enfin ajouter une autre couche dense qui contient **2-neurones** suivi de la fonction d'activation *Softmax* pour fournir un résultat de classe en sortie.

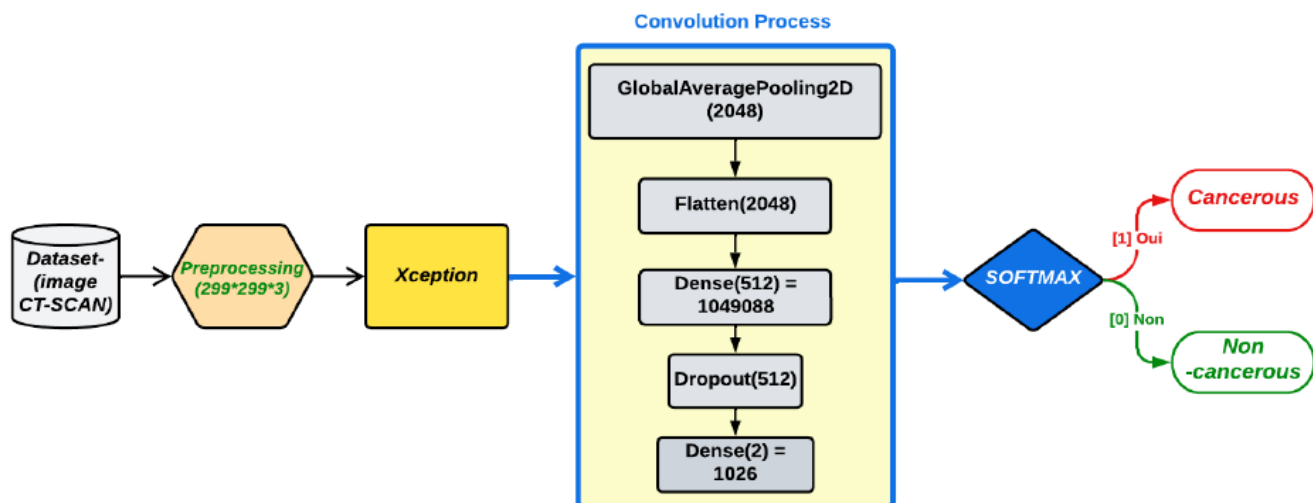


FIGURE 3.8 – Architecture du modèle **Xception** entraîné avec le classificateur **Softmax**.

2. Le Modèle • [*Densenet121*]

- Pour entraîner le modèle *Densenet121*, on a suivi la même logique que celle de [*Xception*] sauf que les images qui seront fournis en entrée doivent avoir une dimension de (224 x 224 x 3). Le reste de l'étude est identique.
- On note que le modèle *Densenet121* est aussi un modèle pré-entraîné sur la base de données *Imagenet*.

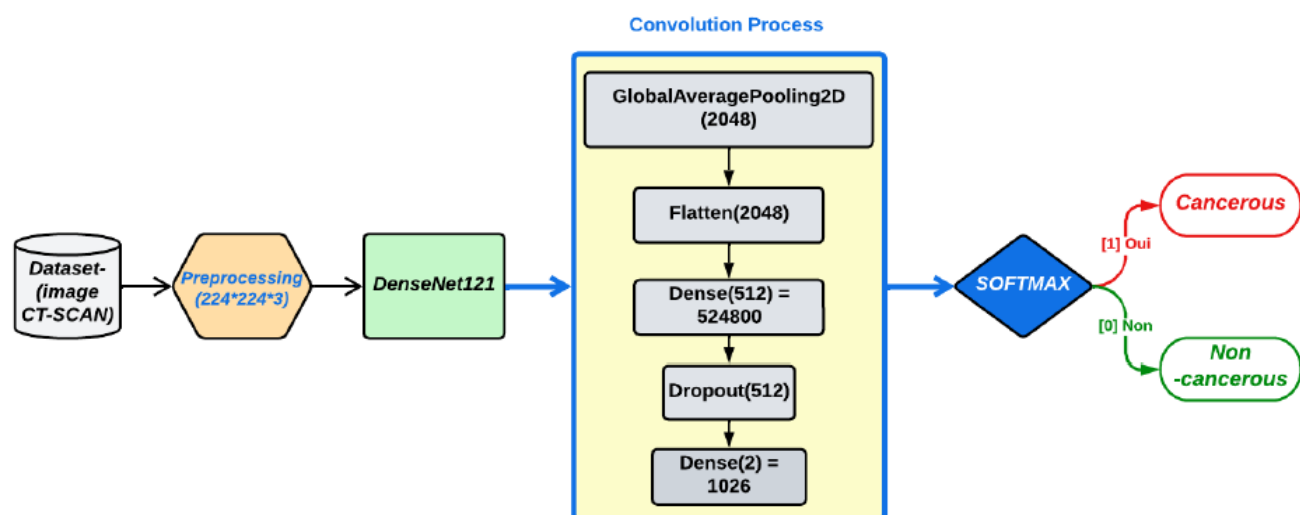


FIGURE 3.9 – Architecture du modèle **DenseNet121** entraîné avec le classificateur **Softmax**.

3.3 Plateforme de dépistage et diagnostic des tumeurs pancréatiques par CT-SCAN

L'interprétation d'un CT-SCAN pancréatique-abdominal pour effectuer des diagnostics précis est une étape très complexe quant à la disposition du Pancréas par rapport aux autres organes. Pour ce fait, on a conçu une plate-forme qui dispose de certains outils pratiques qui nous permettront de visualiser la paroi abdominale précisément pancréatiques pour faire valider les prédictions fournies par l'algorithme qu'on a implémenté et donc faire la distinction entre un Pancréas atteint du cancer et un autre sain.

3.4 Conclusion

- D'après l'étude préalable, on a pu évaluer nos modèles d'apprentissage et distinguer les modèles qui sont le plus adaptés à nos Datasets, en se basant sur la précision qu'ils ont fournis durant leur apprentissage.
- Dans le chapitre suivant nous allons présenter les résultats de performance des modèles sollicités pour effectuer des prédictions fiables ainsi qu'un aperçu de la plateforme développée.

Chapitre 4

Résultats et analyse

4.1 Introduction

Dans ce chapitre, nous allons présenter les résultats qu'on a obtenu durant l'étude préalable puis sélectionner les meilleures modèles pour établir des prédictions et diagnostics fiables. Nous allons d'une autre part fournir des échantillons de simulation de ses modèles pour mieux intercepter les choses.

4.2 Environnement de Développement

4.2.1 Google-Colaboratory



Google-Colaboratory est une plateforme qui fonctionne avec le principe du Cloud. Elle est basée sur la plateforme Jupyter Notebook et conçu pour des travaux de recherches orientées vers le Machine Learning. Cette plateforme permet d'entraîner des modèles de Machine Learning de façon immédiate sur le CCloud. Ce dernier propose trois-(03) modes d'exécutions pour entraîner ces modèles (**CPU**¹, **GPU**² et **TPU**³) [75].

1. Central Processing Unit.
2. Graphics Processing Unit.
3. Tensor Processing Unit.

4.3 Librairies et Outils d'implémentation

4.3.1 Keras



Keras est une API⁴ conçu pour manipuler les réseaux de neurones artificiels développé en python et interfaçable avec TensorFlow. Cette API à été mise en place pour permettre au développeurs de faire des expérimentations rapides [76].

4.3.2 Matplotlib



Matplotlib est une bibliothèque issue du langage de programmation Python qui permet de concevoir des graphes optimaux. L'idée ultime de cette bibliothèque est de rendre les données complexes simple à interpréter [77].

4.3.3 Seaborn



Seaborn est une bibliothèque issue du langage de programmation Python, qui permet de visualiser des données par le biais de la bibliothèque Matplotlib. Cette bibliothèque est alimentée par une interface qui permet de dissiner des graphiques statistiques [78].

4.3.4 Tensorflow



TensorFlow est une plate-forme Open Source dédiée au Machine Learning. Elle propose des outils et bibliothèques permettant aux développeurs de concevoir et améliorer des applications performantes [79].

4. Application Programming Interface.

4.3.5 Numpy



Numpy est une bibliothèque issue du langage de programmation Python qui permet de faire des calculs numériques sur des tableaux de nombres [80].

4.3.6 Pandas



Pandas est une bibliothèque issue du langage de programmation Python qui fournit des outils d'analyses et de manipulations de données de manière optimale et rapide [81].

4.4 Langages de développement informatique utilisés

4.4.1 Python



Python est un langage de programmation interprété c-à-d qu'il s'exécute sans être compilé ce qu'il lui donne un côté simple et pratique et fait de lui le langage le plus utilisé de nos jours dans différents domaines. Ce langage a été inventé par **Guido Van Rossum en 1991** [82].

4.4.2 HTML & CSS

4.4.2.1 HTML



HTML⁵ est langage de structuration de pages web. La particularité de ce langage est qu'il utilise des balises pour formuler une instruction et s'exécute sur un navigateur internet pour afficher son contenu [83].

5. HyperText Markup Language

4.4.2.2 CSS



*CSS*⁶ est un langage qui à été standardisé par *W3C* qui consiste à apporter une certaine allure et styles aux pages *HTML* pour leurs donner un côté visuel plus esthétique [84].

4.4.3 JavaScript



Javascript est un langage de programmation qui permet d'installer un mécanisme ou prototype de fonctionnement auxquelles les pages web oeuvrent. Ce mecanisme peut être des *animations*, *contenus multimédia*, *etc..* [85].

4.5 Frameworks et Serveurs

4.5.1 Flask



Flask est décrit comme un petit Framework⁷ léger qui encapsule est outils pratiques pour créer et alimenter des *applications-web* implémenté avec *Python*. Il est facile à utiliser, après quelques lignes de codes on peut l'exploiter [86].

4.5.2 Ajax



*Ajax*⁸ est une méthode qui permet de faire une inter-connexion entre un serveur-web et une page web dynamique de manière asynchrone autrement dit sans avoir à actualiser l'ensemble de la page pour echanger des données. Il fonctionne avec deux - (02) langages de programmation le *Xml* et *JavaScript* [87].

6. Cascading Style Sheet

7. Un ensemble d'outils ou composants qui permettent de concevoir des logiciels ou des applications dynamiques.

8. Asynchronous JavaScript And XML

4.6 Évaluation des techniques de dépistages utilisées

4.6.1 Dépistage par Analyses Médicales

4.6.1.1 Résultats d'entraînement et Tests - (Simulation Primitive des Modèles)

- Dans un premier temps, nous allons présenter les performances des modèles *Random Forest Classifier* et *K-Nearest Neighbors* avant d'avoir effectué une Optimisation quelconque :

[Performances du modèle • (*Random Forest Classifier*)]

Modèle	Train-Accuracy	Test-Accuracy	Precision	Recall	F1-Score
<i>Random Forest Classifier</i>	98.41 %	55.40 %	58 %	55 %	55 %

TABLE 4.1 – Résultats Primitifs du *Random Forest Classifier*.

[Performances du modèle • (*K-Nearest Neighbors*)]

Modèle	Train-Accuracy	Test-Accuracy	Precision	Recall	F1-Score
<i>K-Nearest Neighbors</i>	68.55 %	57.43 %	59 %	57 %	58 %

TABLE 4.2 – Résultats Primitifs du *K-Nearest Neighbors*.

4.6.1.2 Courbes de performances des Modèles

- Dans cette partie, on peut visualiser les courbes de performances primitives des modèles *Random Forest Classifier* et *K-Nearest Neighbors*.
- Les courbes d'entraînement des deux modèles sont élevées - (**98.41 %** pour le *Random Forest Classifier* et **68.55 %** pour le *K-Nearest Neighbors*) comparé aux courbes de tests - (**55.40 %** pour le *Random Forest Classifier* et **57.43 %** pour le *K-Nearest Neighbors*).
- D'après notre analyse, on constate un problème d'overfitting dans les performances des deux modèles.

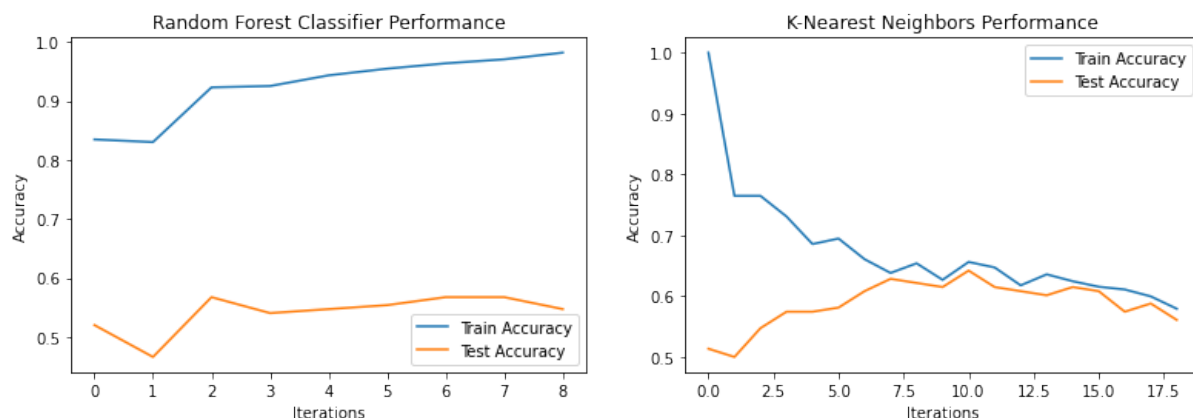


FIGURE 4.1 – Performance des deux Modèles utilisés.

4.6.1.3 Comparaison entre les Matrices de Confusion des deux Modèles

- D'après les matrices de confusion des deux - (02) modèles *Random Forest Classifier* et *K-Nearest Neighbors* on peut analyser la qualité de prédictions qu'ils peuvent fournir :
- La matrice de confusion nous permet d'évaluer les performances d'un modèle c-à-d le nombre de fois qu'il fournit des classifications correctes ou pas.

- Sachant qu’une matrice de confusion se lit en diagonal, les cellules qui forment cette diagonal révèlent le classement correcte du modèle. En ce qui concerne les autres cellules elles révèlent les erreurs de classement du modèle.
- Le principe expliqué ci-dessus s’applique sur le reste des matrices de confusions.
- On constate que les deux modèles font beaucoup d’erreurs de classification et se avant l’augmentation de données.

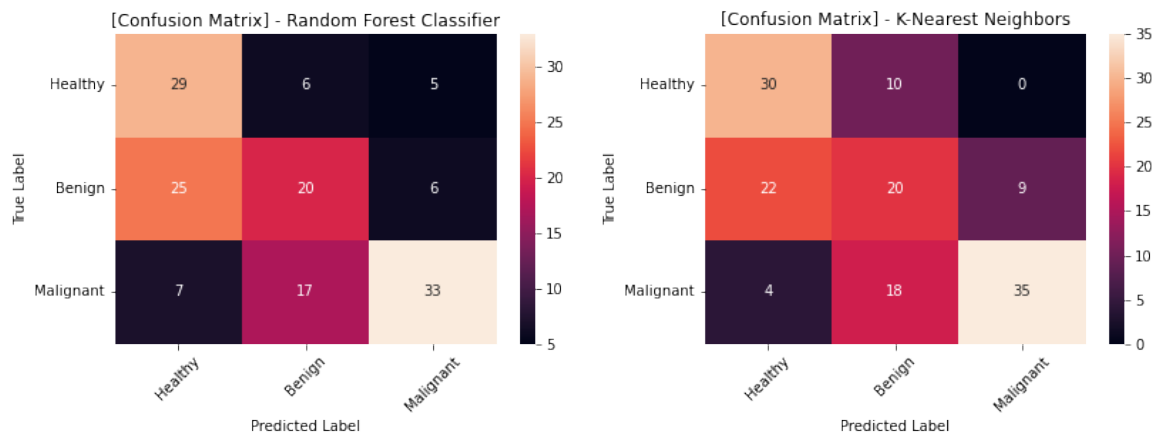


FIGURE 4.2 – Matrices de Confusion des deux Modèles - *Random Forest Classifier* et *K-Nearest Neighbors* Avant *l’augmentation des données*.

4.6.1.4 Résultats d'entraînement et Tests - (Optimisation des Modèles)

4.6.1.5 Augmentation de données • (Data - augmentation)

- Après avoir analysé les performances primitives des modèles *Random Forest Classifier* et *K-Nearest Neighbors* et constaté que ces dernières n'étaient pas suffisantes pour effectuer des prédictions fiables, on a procédé à l'augmentation de données, ce qui a fourni les résultats suivants :

[Performances du modèle • (*Random Forest Classifier + Data-augmentation*)]

Modèle	Train-Accuracy	Test-Accuracy	Précision	Recall	F1-Score
<i>Random Forest Classifier</i>	100 %	77.70 %	84 %	78 %	74 %

TABLE 4.3 – Résultats du Random Forest Classifier après *l'augmentation de données*.

[Performances du modèle • (*K-Nearest Neighbors + Data-augmentation*)]

Modèle	Train-Accuracy	Test-Accuracy	Précision	Recall	F1-Score
<i>K-Nearest Neighbors</i>	95.92 %	91.21 %	91 %	91 %	91 %

TABLE 4.4 – Résultats du K-Nearest Neighbors après *l'augmentation de données*.

4.6.1.6 Courbes de performances des Modèles

- Après avoir effectué l'optimisation avec *l'augmentation de données*, on peut visualiser les courbes de performances des modèles *Random Forest Classifier* et *K-Nearest Neighbors*.
- On note une amélioration de performances des deux modèles d'après les nouvelles courbes d'entraînement - (**100 %** pour le *Random Forest Classifier* et **95.62 %** pour le *K-Nearest Neighbors*) comparé aux courbes de tests - (**77.70 %** pour le *Random Forest Classifier* et **91.21 %** pour le *K-Nearest Neighbors*).
- On constate que le *problème d'overfitting persiste* dans les performances des deux modèles.

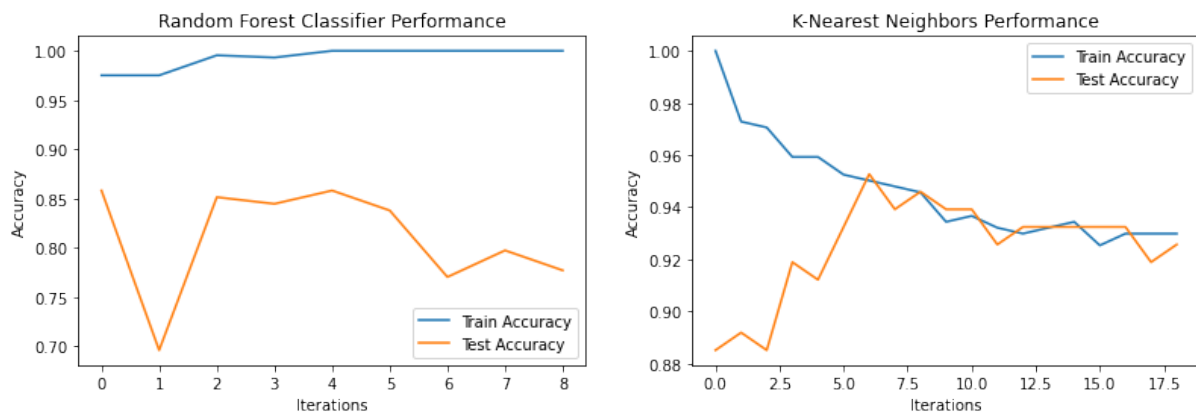


FIGURE 4.3 – Performance des deux Modèles utilisés.

4.6.1.7 Comparaison entre les Matrices de Confusion des deux Modèles

- Les performances des modèles *Random Forest Classifier* et *K-Nearest Neighbors* après avoir effectué l'optimisation avec *l'augmentation de données* à travers leurs matrice de confusion sont fournis ci-dessous.
- On constate que les deux modèles ont subi des améliorations sur le plan performances de la classification et se après l'augmentation de données.

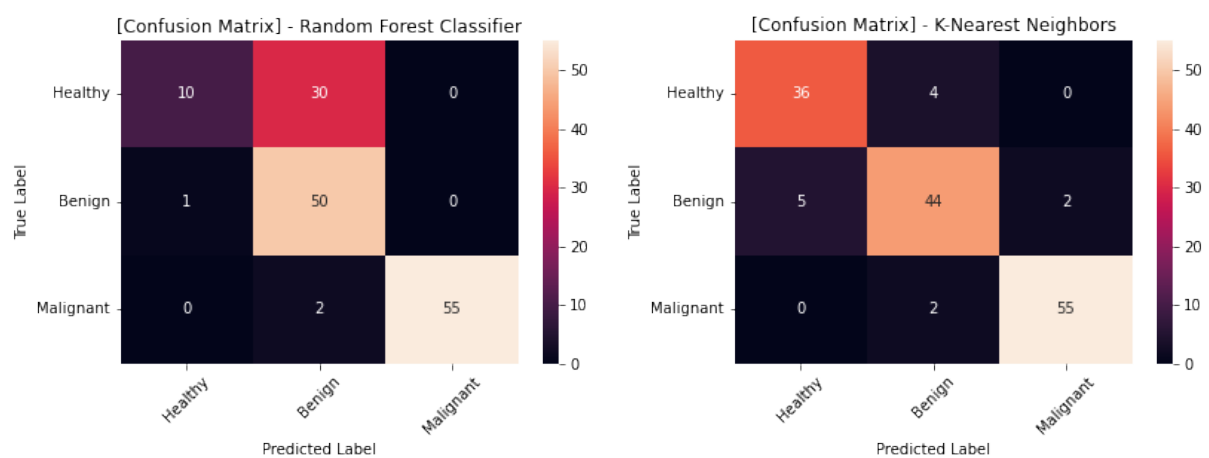


FIGURE 4.4 – Matrices de Confusion des deux Modèles - *Random Forest Classifier* et *K-Nearest Neighbors* Après *l'augmentation de données*.

4.6.1.8 Évaluation de la Matrice de Corrélacion

— Voici un aperçu de la *Matrice de Corrélacion* qu'on a obtenu lors de la phase d'analyse de données du DATASET fourni et après optimisation avec *augmentation de données* :

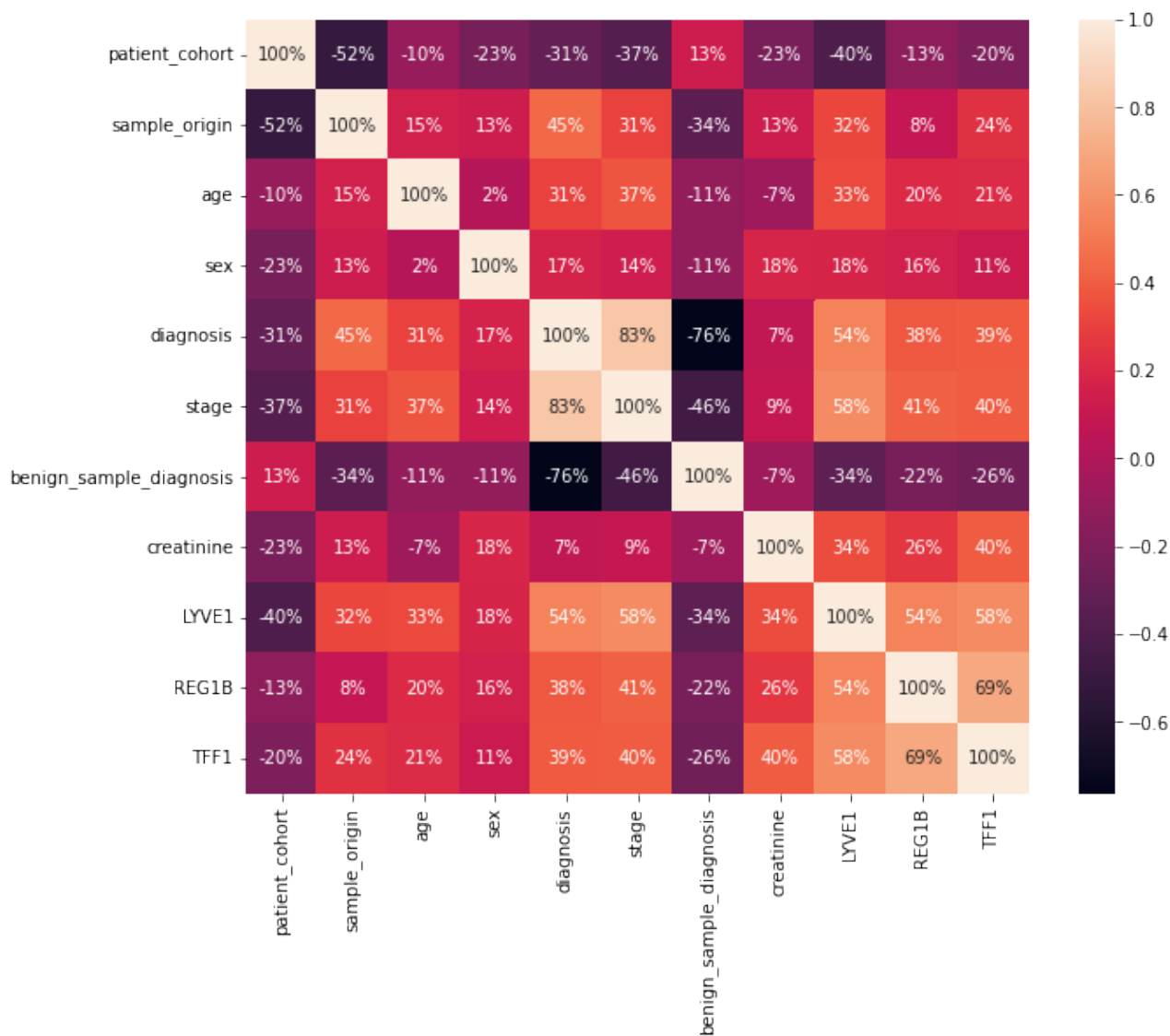


FIGURE 4.5 – Matrice de Corrélacion Après *l'augmentation de données*.

4.6.1.9 La Cross Validation • (Validation Croisée avec la technique K-Fold)

- L'augmentation de données a permis aux modèles *Random Forest Classifier* et *K-Nearest Neighbors* de fournir de bon résultats, mais on avait constaté que leurs performances pouvaient atteindre de meilleures valeurs en les optimisant avec la cross validation, ce qui a fourni les résultats suivants :

[Performances du modèle • (*Random Forest Classifier + K-fold*)]

Modèle	Train-Accuracy	Test-Accuracy	Précision	Recall	F1-Score
<i>Random Forest Classifier</i>	98.87 %	99.33 %	99 %	99 %	99 %

TABLE 4.5 – Résultats du Random Forest Classifier après optimisation avec *la Cross-Validation • (K-fold)*.

[Performances du modèle • (*K-Nearest Neighbors + K-fold*)]

Modèle	Train-Accuracy	Test-Accuracy	Précision	Recall	F1-Score
<i>K-Nearest Neighbors</i>	94.58 %	94.66 %	95 %	95 %	95 %

TABLE 4.6 – Résultats du K-Nearest Neighbors après optimisation avec *la Cross-Validation • (K-fold)*.

4.6.1.10 Courbes de performances des Modèles

- Après avoir effectué l'optimisation avec *la Cross-Validation • (K-fold)*, on peut visualiser les courbes de performances des modèles *Random Forest Classifier* et *K-Nearest Neighbors*.

- Dans cette partie, on note que le problème de l'overfitting s'est dissipé et que les courbes d'entraînement des modèles ont atteint les performances suivantes - (**98.87** % pour le *Random Forest Classifier* et **94.58** % pour le *K-Nearest Neighbors*) comparé aux courbes de tests - (**99.33** % pour le *Random Forest Classifier* et **94.66** % pour le *K-Nearest Neighbors*).

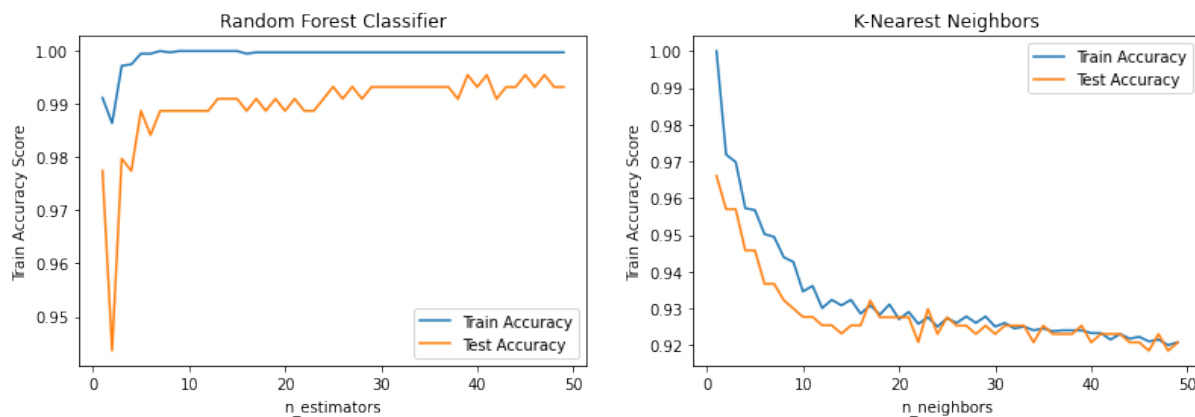


FIGURE 4.6 – Performance des deux Modèles utilisés.

4.6.1.11 Comparaison entre les Matrices de Confusion des deux Modèles

- Les performances des modèles *Random Forest Classifier* et *K-Nearest Neighbors* après avoir effectué l'optimisation avec **La Cross-Validation • (K-fold)** à travers leurs matrice de confusion sont fournis ci-dessous.
- Après l'usage de la technique de la cross validation avec la methode k-fold ($k = 10$), on constate une nette amélioration lors de la classification c-à-d moins d'erreurs.

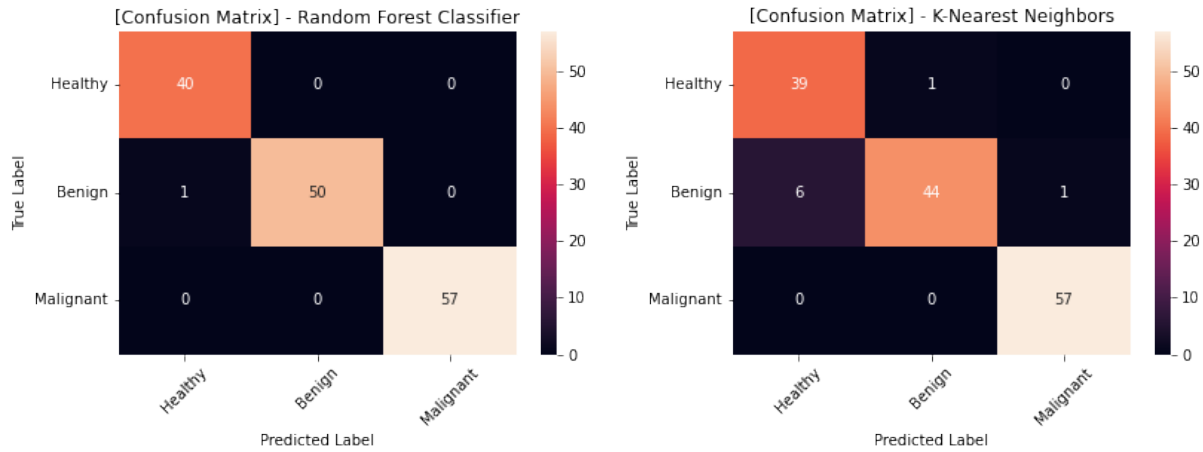


FIGURE 4.7 – Matrices de Confusion des deux Modèles - *Random Forest Classifier* et *K-Nearest Neighbors* après optimisation avec *la Cross-Validation • (K-fold)*.

4.6.1.12 Modèle final Sélectionné pour effectuer des Prédictionns

[Modèle final sélectionné : *Random Forest Classifier + K-fold*]

Modèle	Train-Accuracy	Test-Accuracy	Précision	Recall	F1-Score
<i>Random Forest Classifier</i>	98.87 %	99.33 %	99 %	99 %	99 %

TABLE 4.7 – Modèle final sollicité pour effectuer des prédictions et diagnostics.

4.6.1.13 Comparaison entre les Performances des modèles utilisés

- Dans ce modèle graphique, nous avons représenté les statistiques de performances fournis par chaque modèle durant leurs phases d'optimisation par les deux (02) méthodes d'optimisation : *Random Forest Classifier* et *K-Nearest Neighbors*.

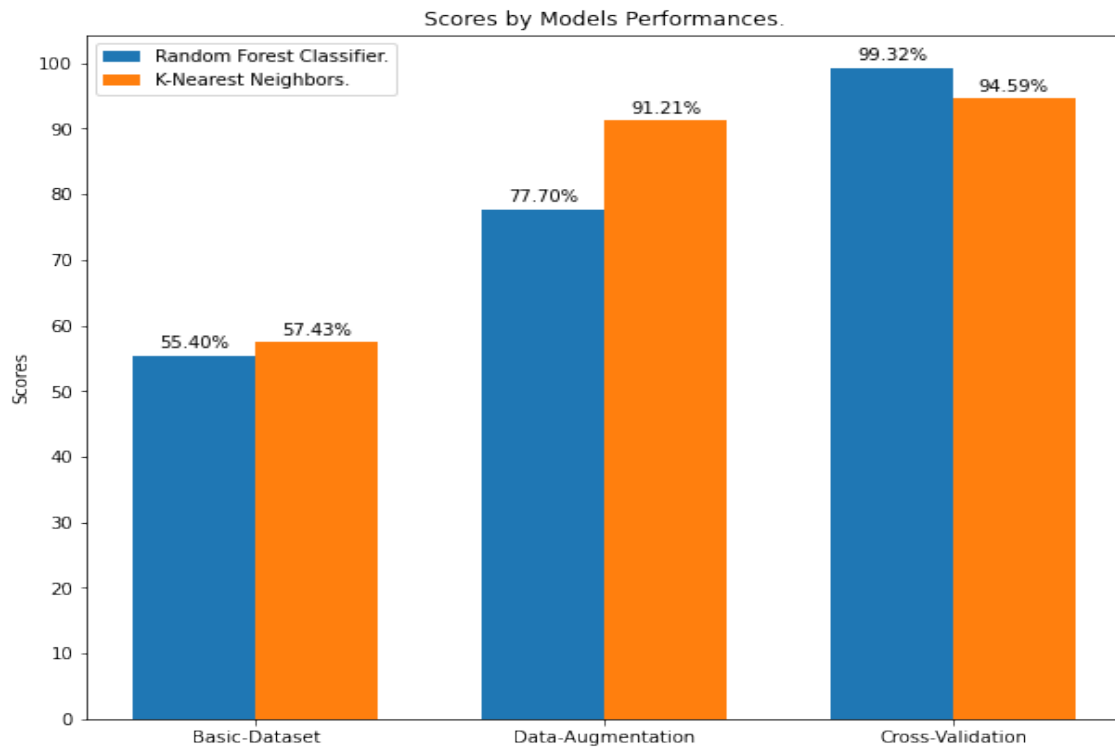


FIGURE 4.8 – Comparaison entre les Performances des modèles utilisés - *Random Forest Classifier* et *K-Nearest Neighbors*.

4.6.1.14 Comparaison entre les performances de nos modèles avec celles d'autres travaux connexes

- Pour prouver que nos modèles sollicités dans l'étude de la première approches "**Dépistage par analyses médicales**" sont performants, on les a comparés avec d'autres travaux connexes qui ont traité la même problématique et qui ont fournis les résultats de performances de leurs modèles. On présente ci-dessous, un tableau de comparaison entre les performances de nos modèles et celles d'autres recherches :

Travaux Connexes				
Modèle	Article	Année	Train-Accuracy	Test-Accuracy
SVM ⁹	[66] ¹⁰	2021	/	77.83%
MLP ¹¹			/	79.98%
GNFN ¹²			/	94.78%
Notre Étude				
Modèle	Article	Année	Train-Accuracy	Test-Accuracy
<i>K-Nearest Neighbors</i>	/	2022	94.58%	94.66%
<i>Random Forest Classifier</i>			98.87%	99.33%

TABLE 4.8 – Comparaison de performances de nos modèles *Random Forest Classifier* et *K-Nearest Neighbors* avec les performances d'autres études menées sur la même thématique.

9. Support Vector Machine.

10. *Bodyanskiy, Yevgeniy and Chala, Olha* - [Generalized neo-fuzzy-neuron with membership functions of special type in medical diagnostics-2021].

11. Multi-Layer Perceptron.

12. Generalized neo-fuzzy-neuron.

4.6.2 Dépistage par Imagerie Médicales CT-SCAN

4.6.3 Résultats d'entraînement et Tests

- Dans cette partie, on va présenter les résultats de performances des modèles de réseaux de neurones convolutifs : *Xception* et *Densenet121* combinés avec la fonction d'activation *Softmax*, puis sélectionner le modèle ayant fourni les meilleurs scores pour gérer les prédictions.

[Performances du modèle • (*Xception* + *Softmax*)]

Modèle	Accuracy	Loss	Précision	Recall	F1-Score
<i>Xception</i> + <i>Softmax</i>	98.14 %	3.92 %	98 %	98 %	98 %

TABLE 4.9 – Résultats du Modèle *Xception* combiné avec la fonction d'activation *Softmax*.

[Performances du modèle • (*Densenet121* + *Softmax*)]

Modèle	Accuracy	Loss	Précision	Recall	F1-Score
<i>Densenet-121</i> + <i>Softmax</i>	95.36 %	12.23 %	96 %	95 %	95 %

TABLE 4.10 – Résultats du Modèle *Densenet121* combiné avec la fonction d'activation *Softmax*.

4.6.3.1 Courbes de performances des Modèles

- Après la phase d'apprentissage, on peut visualiser les courbes de performances après optimisation des modèles *Xception* + *Softmax* et *Densenet121* + *Softmax* et effectuer une comparaison qui se base sur les paramètres suivants : *Accuracy-Xception* et *Loss-Xception*.

[Courbes de performances du Modèle • (*Xception + Softmax*)]

- Voici les courbes tracées durant la phase d'apprentissage qui résument l'historique de performance du modèle *Xception + Softmax*.
- Après avoir analysé les courbes de performances des deux modèles, on a constaté que le modèle *Xception + Softmax* a fourni les meilleures performances après l'application de l'api *modelCheckpoint*.
- L'utilisation du *modelCheckpoint* a permis de sélectionner les meilleures performances du modèle *Xception + Softmax* dans les métriques : (**accuracy = 98.14 %**) et (**loss = 3.92 %**) ce qui a fourni en sortie des courbes élevées dans la phase d'entraînement et test et donc des courbes qui tendant vers zéro dans la métrique loss.

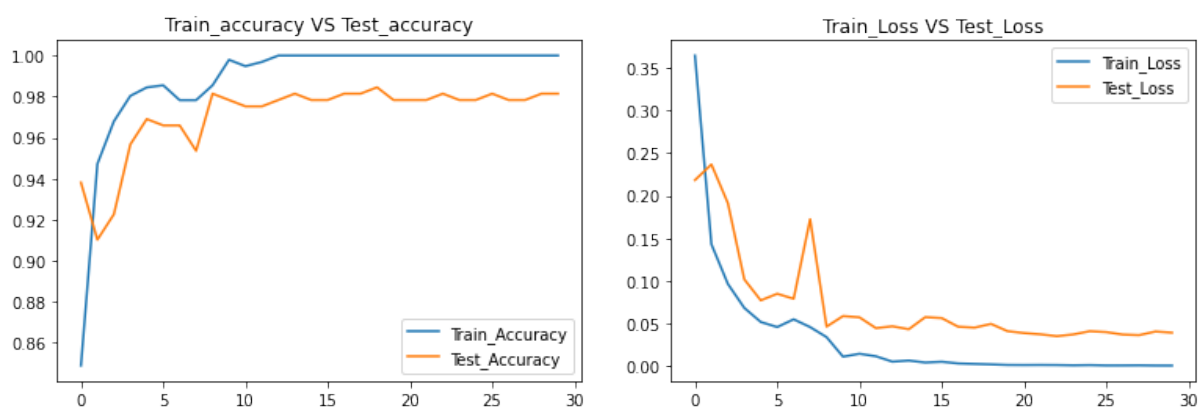


FIGURE 4.9 – Performance du Modèle - (*Xception + Softmax*).

[Courbes de performances du Modèle • (*Densenet121 + Softmax*)]

- Voici les courbes tracées durant la phase d'apprentissage qui résument l'historique de performance du modèle *Densenet121 + Softmax*.
- Les performances du modèle *Densenet121 + Softmax* ont fournis les résultats suivants : (**accuracy = 95.36 %**) et (**loss = 12.23 %**).

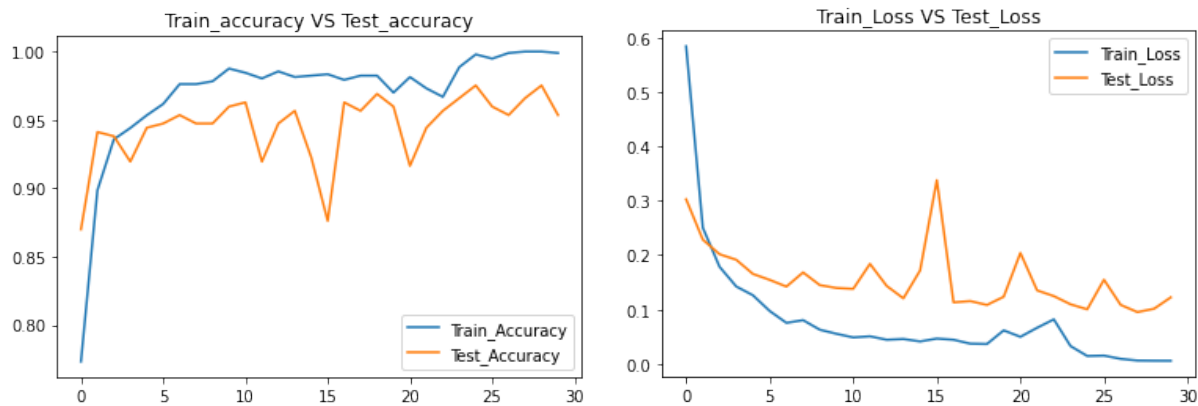


FIGURE 4.10 – Performance du Modèle (*Densenet121 + Softmax*).

4.6.3.2 Comparaison entre les Matrices de Confusion des deux Modèles

- Les performances des modèles *Xception + Softmax* et *Densenet121 + Softmax* à travers leurs matrice de confusion sont fournis ci-dessous.
- La matrice de confusion révèle que le modèle Xception fait le moins d’erreurs lors de la classification comparé au modèle densenet121.

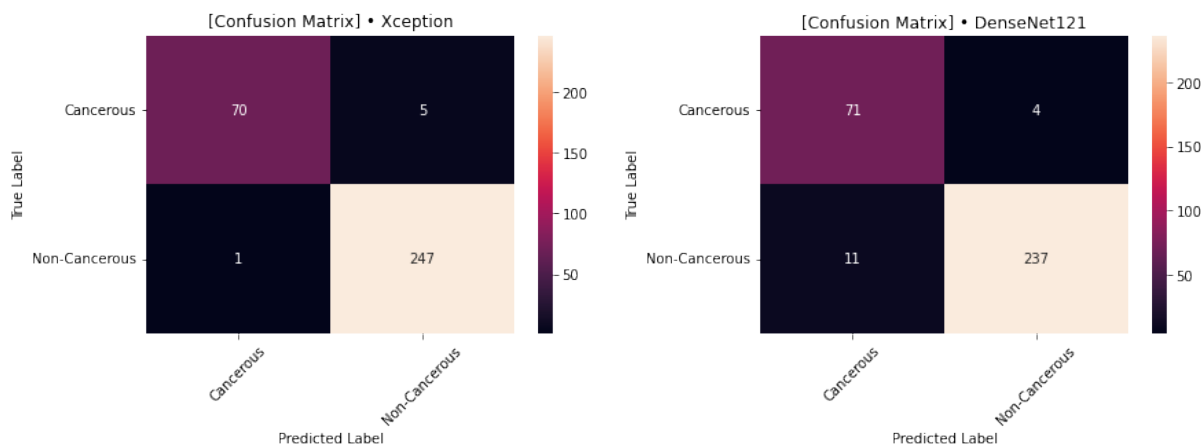


FIGURE 4.11 – Matrices de Confusion des deux Modèles - *Xception + Softmax* et *Densenet-121 + Softmax*.

4.6.3.3 Modèle final Sélectionné pour effectuer des Prédictions

[Modèle final sélectionné : *Xception + Softmax*]

Modèle	Accuracy	Loss	Précision	Recall	F1-Score
<i>Xception + Softmax</i>	98.14 %	3.92 %	98 %	98 %	98 %

TABLE 4.11 – Modèle final sollicité pour effectuer des prédictions et diagnostics.

4.6.3.4 Comparaison entre les Performances des modèles utilisés

- Dans ce modèle graphique, nous avons représenté les statistiques de performances fournis par chaque modèle durant leurs phases d'apprentissage sur les deux modèles de *Transfert Learning* combiné avec la fonction d'activation *Softmax* *Xception + Softmax* et *Densenet121 + Softmax*.

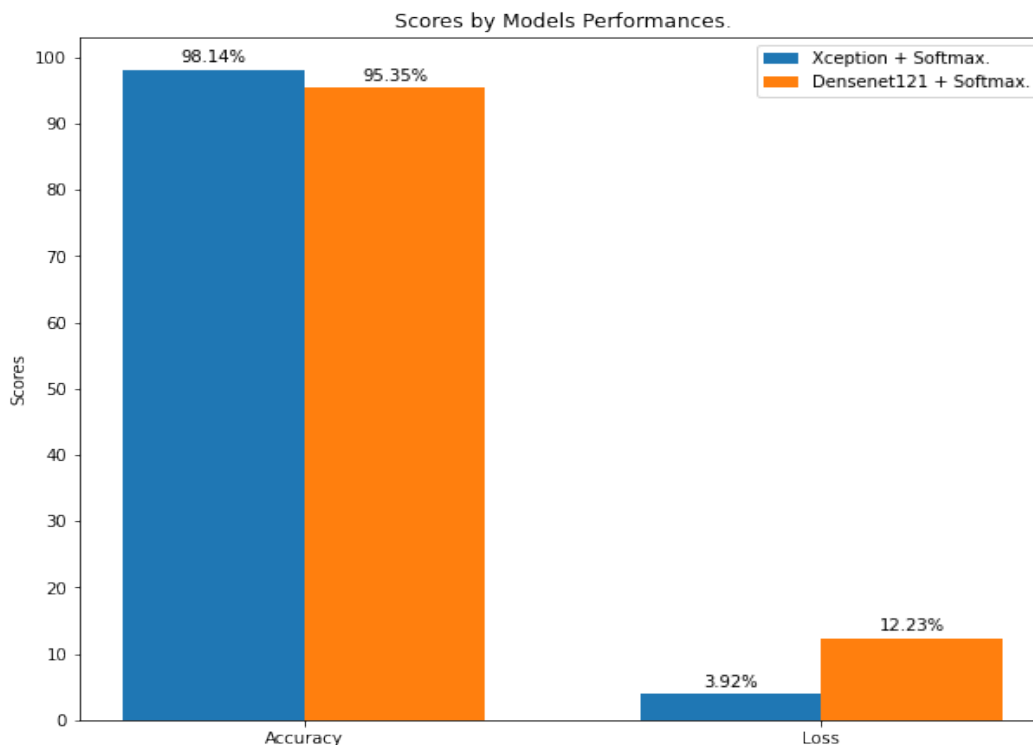


FIGURE 4.12 – Comparaison entre les Performances des modèles utilisés - *Xception + Softmax* et *Densenet121 + Softmax*.

4.6.3.5 Comparaison entre les performances de nos modèles avec celles d'autres travaux connexes

- En ce qui concerne la deuxième approche "*Dépistage par imagerie médicale*", on n'a pas trouvé d'autres articles qui ont traités la même problématique pour comparer les performances fournis par nos modèles durant leurs apprentissage et test.
- La raison est que les [CT-SCAN] du cancer du pancréas sont difficiles à trouver en vue de la *rareté de ce cancer*, ce qui fait que le dataset qu'on à utilisé n'a été ajouté sur la plate-forme • [Github] que récemment - [28/07/2021] ce qui explique le manque de son exploitation par les chercheurs.

4.7 Simulation de Prédictions

- Dans cette partie du chapitre, on va tester nos modèles *Random Forest Classifier* et *Xception + Softmax* sur de nouvelles données pour valider leurs performances.

4.7.1 Dépistage par Analyses médicales

- Voici un aperçu de performances fournis par le modèle *Random Forest Classifier* avec *Data-augmentation* et *Cross-Validation • K-fold*.

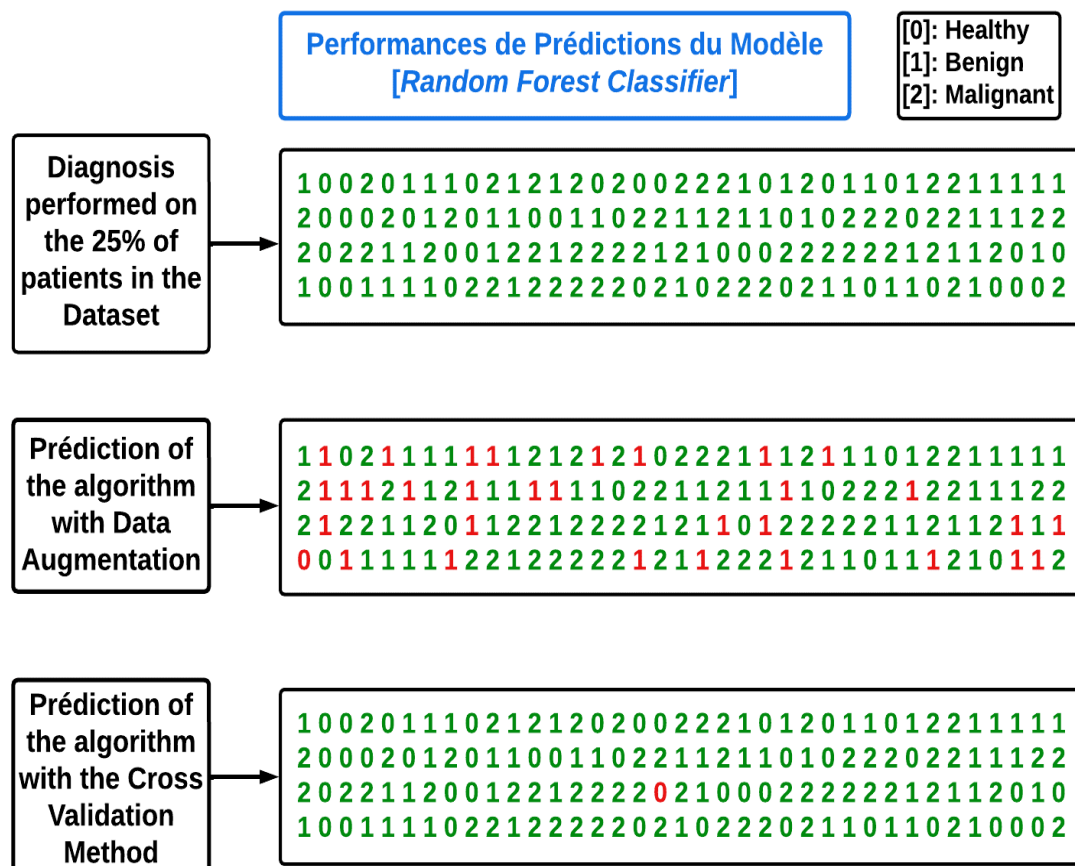


FIGURE 4.13 – Performance de prédiction du Modèle *Random Forest Classifier* après Optimisation avec *Data-augmentation* et *Cross-Validation • K-fold*.

4.7.2 Plateforme de dépistage par imagerie médicale - (CT-SCAN)

- Voici un aperçu de la plate-forme *déployée* avec le *framework Flask* qui permet de faire des prédictions et diagnostics par le biais du modèle *Xception + Softmax*.

4.7.2.1 Prototype de Fonctionnement

[Aperçu-général]

- La plate-forme *Medi-Predict* est une interface graphique qui propose des outils pratique qui permettent de manipuler les CT-SCAN pour effectuer des diagnostics sur l'état de santé des patients.
- Parmi ces outils on cite : *l'objet inspect, crop, filter.*, nous allons voir en détails l'usage de chaque outils par la suite.

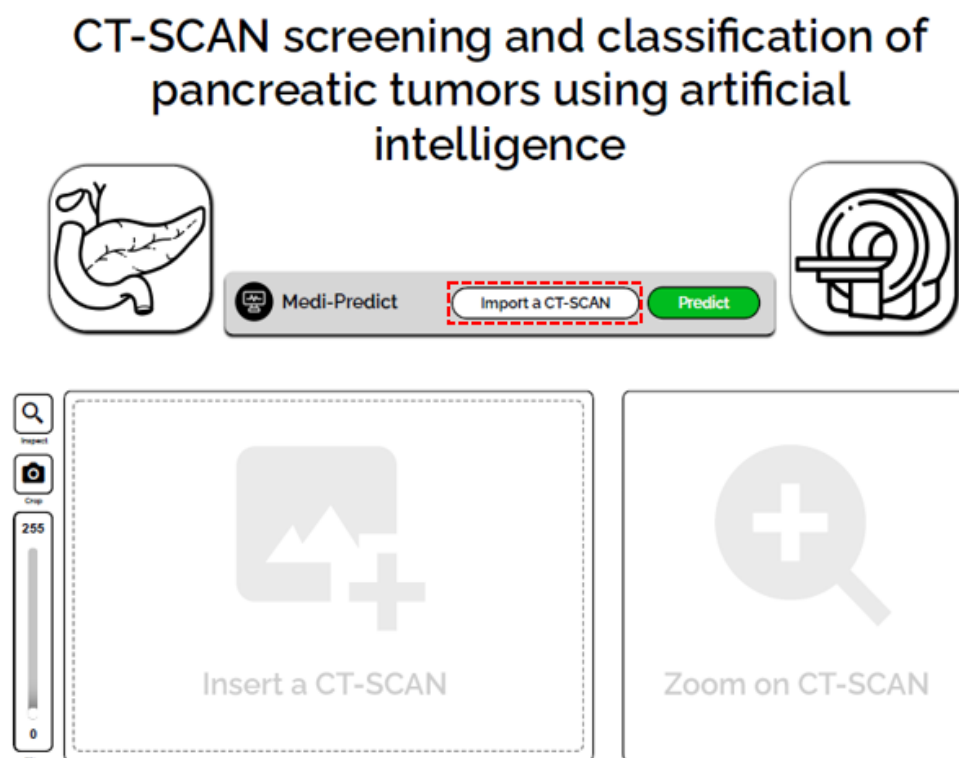


FIGURE 4.14 – Aperçu général de la plate-forme Medi-Predict.

[L'objet - *Inspect*]

- L'objet *Inspect* permet de faire un *zoom-détaillé* sur un CT-SCAN donné dans le but de cibler le pancréas et l'analyser de plus près.

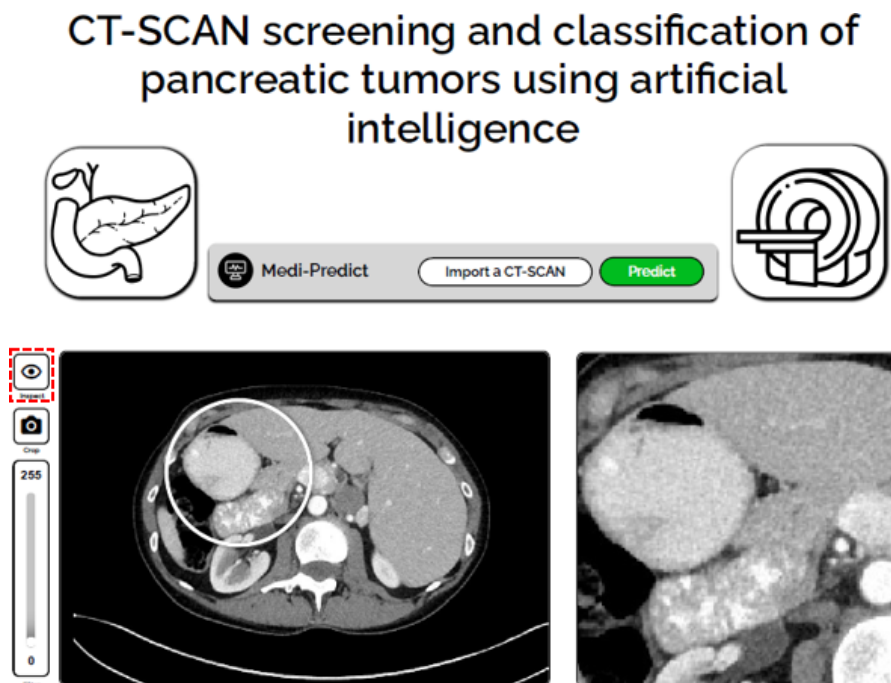


FIGURE 4.15 – Simulation avec l'objet inspect sur un CT-SCAN non-cancereux.

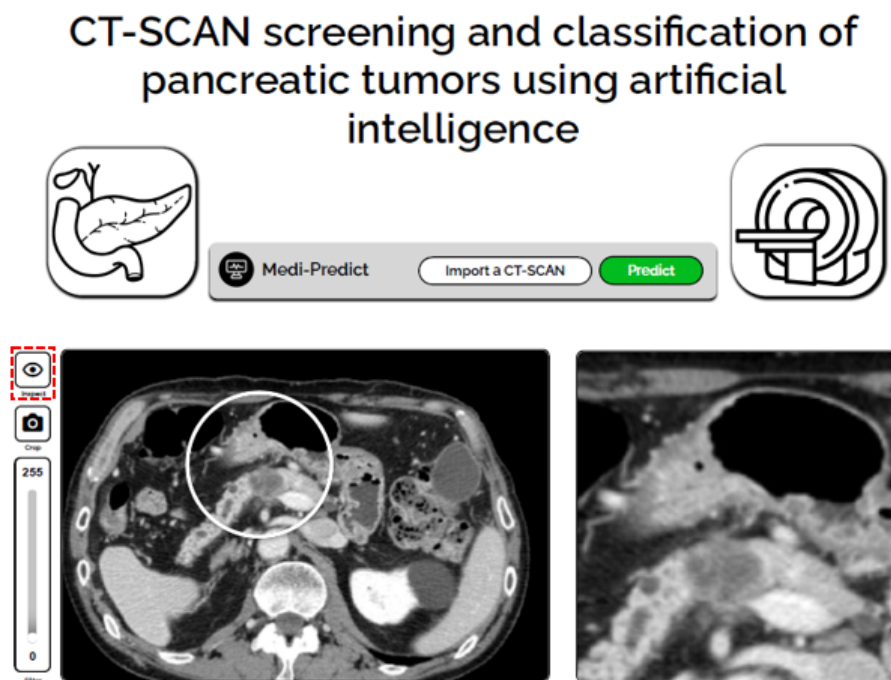


FIGURE 4.16 – Simulation avec l'objet inspect sur un CT-SCAN cancreux.

[L'objet - *Filter*]

- L'objet *Filter* permet d'appliquer trois - (03) *filtres* (*sepia*, *brightness et invert*) sur un CT-SCAN donné pour offrir un aspect visuel plus *dense* et *lumineux* ce qui permet de bien identifier les lésions cancéreuses.

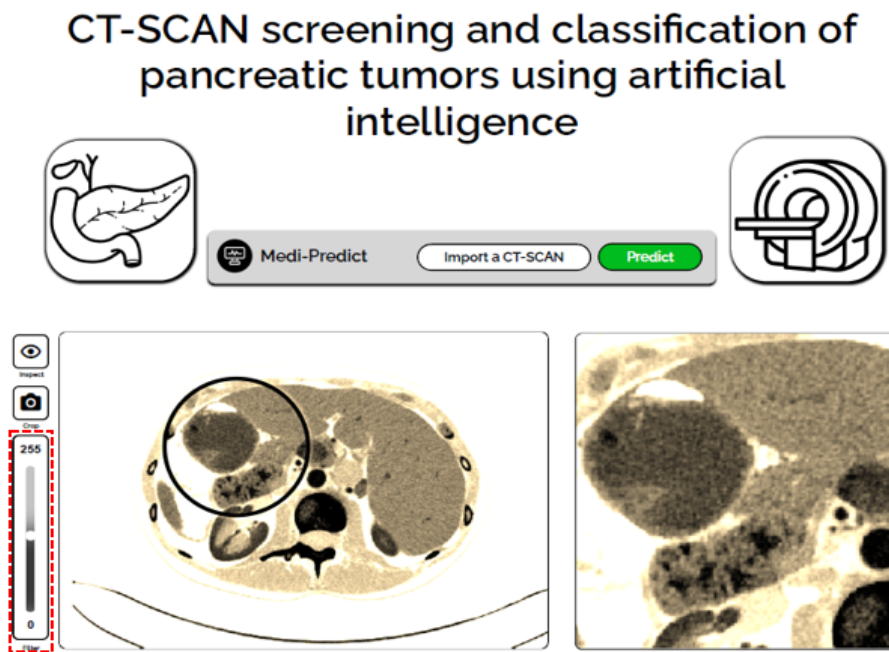


FIGURE 4.17 – Simulation avec l'objet filter sur un CT-SCAN non-cancereux.

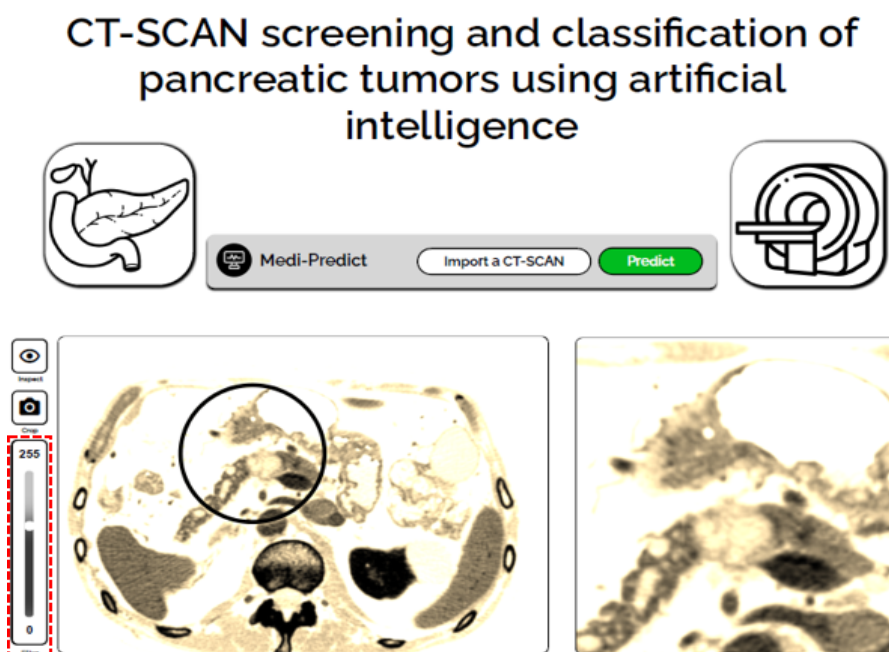


FIGURE 4.18 – Simulation avec l'objet filter sur un CT-SCAN cancreux.

[L'objet - Crop]

- L'objet **Crop** permet d'évaluer la taille de la tumeur sur un CT-SCAN révélant une masse confirmée sur le parenchyme pancréatique.

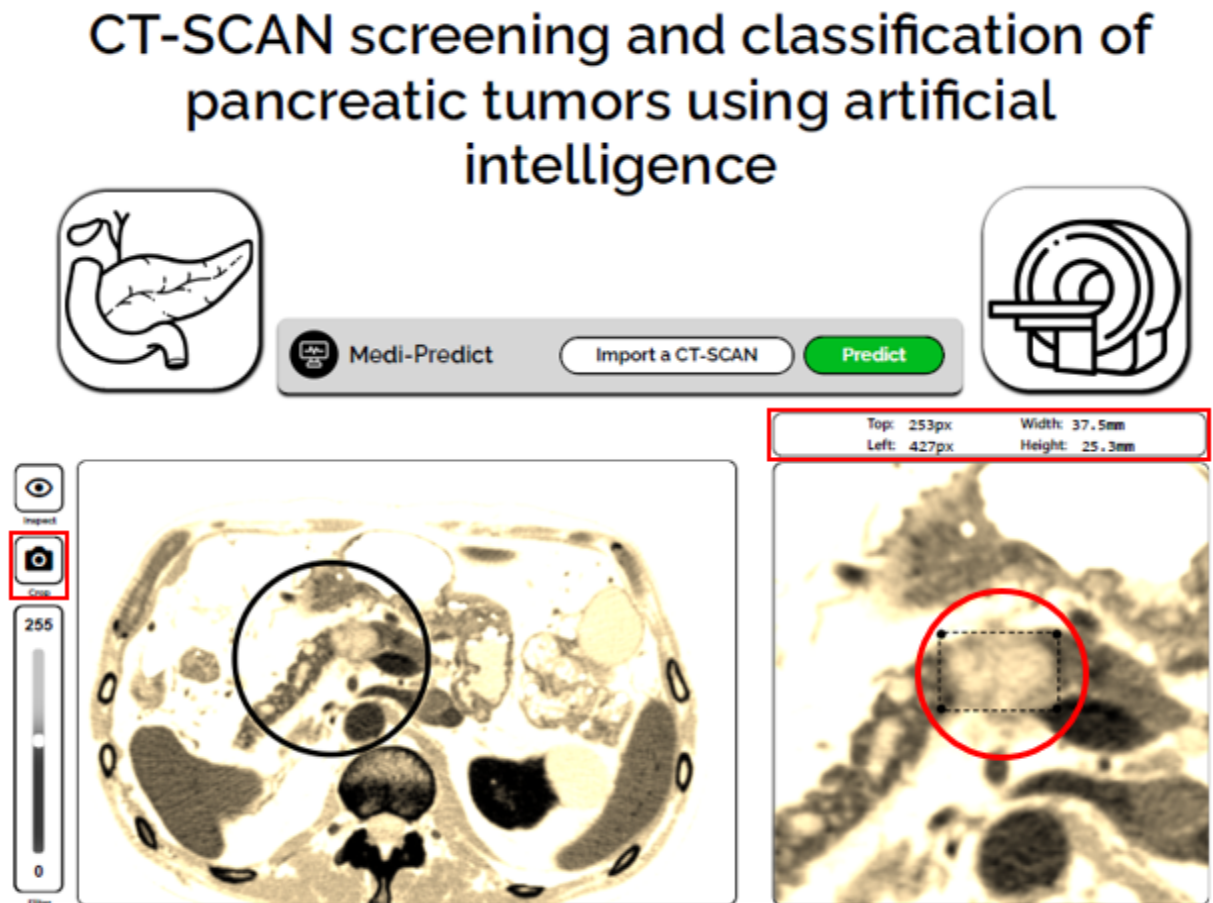


FIGURE 4.19 – Simulation avec l'objet crop pour évaluer la taille de la tumeur.

[Diagnostic - (*Healthy Pancreas*)]

- Voici un aperçu de la performance de l'algorithme *Xception* combiné à la fonction d'activation *Softmax* lors de la *prédiction* sur un CT-SCAN révélant l'état du pancréas avec la déduction "*Healthy Pancreas*" qui veut dire que l'organe est sain.

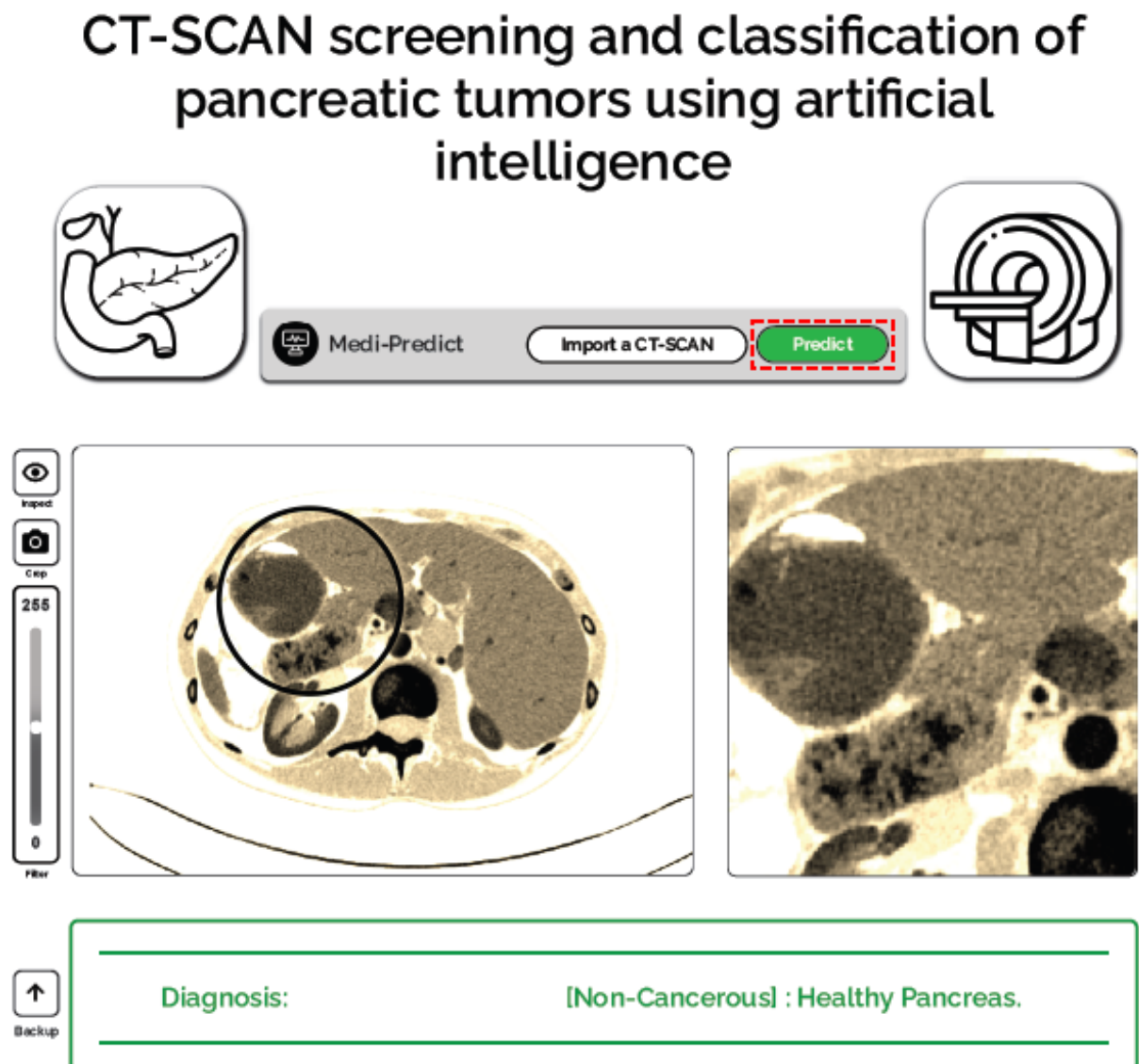


FIGURE 4.20 – Simulation de prédiction sur un CT-SCAN Non-cancereux.

[Diagnostic - (*Pancreas affected by cancer*)]

- Voici un aperçu de la performance de l'algorithme *Xception* combiné à la fonction d'activation *Softmax* lors de la *prédiction* sur un CT-SCAN révélant l'état du pancréas avec la déduction "*Pancreas affected by cancer*" qui veut dire que l'organe est atteint par le cancer.

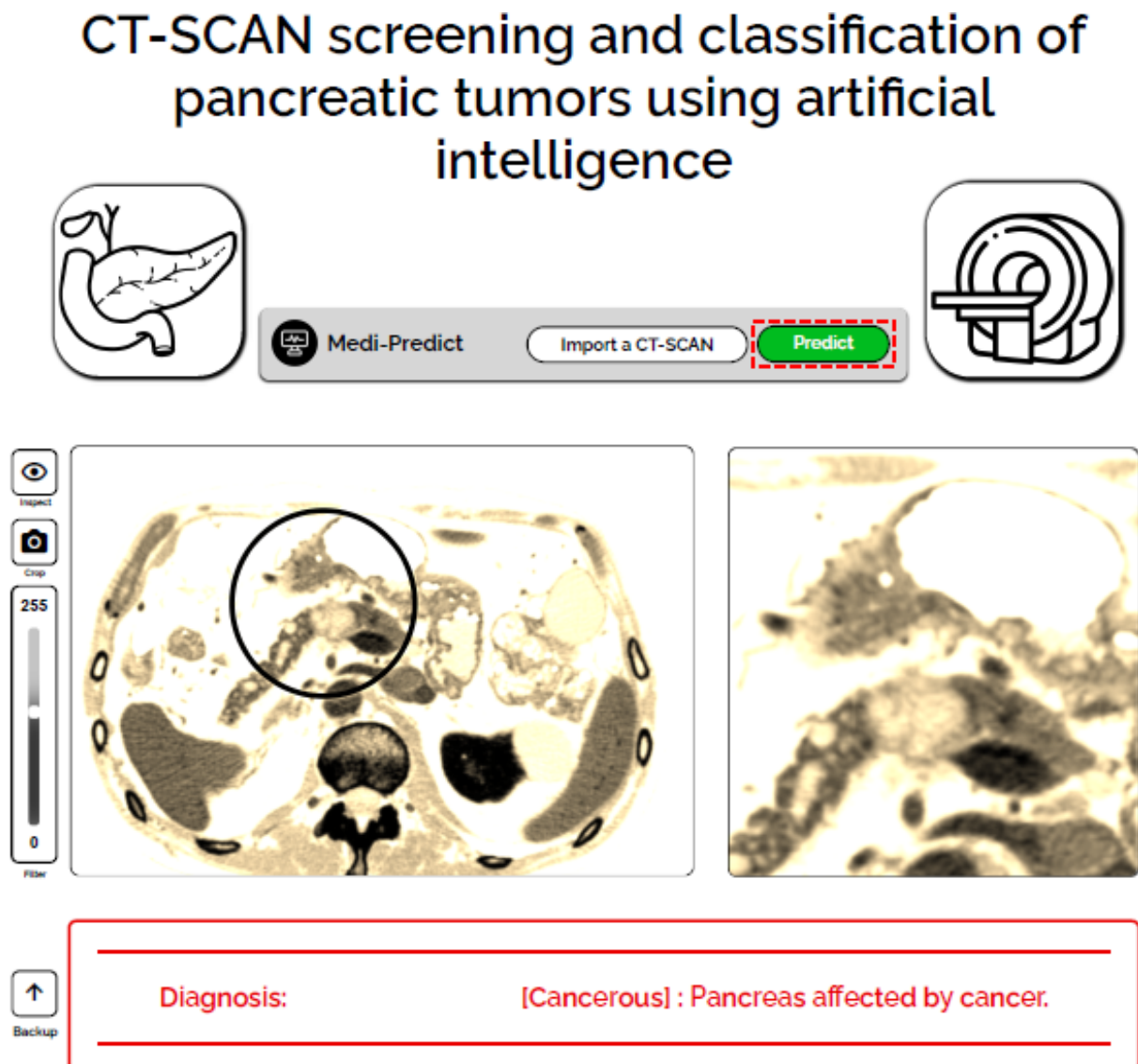


FIGURE 4.21 – Simulation de prédiction sur un CT-SCAN cancéreux.

4.8 Conclusion

D'après le compte rendu précédent, on a démontré que les modèles qu'on a sollicité pour notre étude étaient bien adaptés à nos données - (Datasets).

- Cependant, les résultats de performances qu'ils ont fournis durant leurs apprentissages et tests de validation sur des données quelconques étaient très satisfaisants.
- Bien que, les chiffres fournissent toujours des interprétations exactes, un modèle d'apprentissage n'est jamais parfait et ne peut jamais surpasser l'intelligence humaine.

Conclusion générale et perspective

Dans ce projet, on a lié les techniques de l'intelligence artificielle *Machine Learning* et *Deep Learning* à la problématique du dépistage du cancer du Pancréas. Pour ce fait, on a d'abord consacré une partie du rapport pour étudier l'aspect théorique sur le plan médical ainsi que sur le plan informatique puis nous sommes passés à la partie pratique qui aborde l'aspect implémentation des modèles qui serviront à faire des prédictions sur des données tierces. Dans notre étude, on a procédé à l'élaboration de deux (02) approches distinctes ***le dépistage par analyses médicales*** et ***le dépistage par imageries médicales***.

En ce qui concerne la première approche - (le dépistage par analyses médicales), nous avons sollicité les algorithmes suivants : *Random Forest Classifier* et le *K-Nearest Neighbors* et pour la deuxième approche - (le dépistage par imageries médicales), on a utilisé des modèles de *Transfert-Learning* : *Xception* et *Densenet-121* qui fournissent de bons résultats selon la plateforme *Keras* combinés avec la fonction d'activation *Softmax*.

Après la phase d'entraînement et de tests, les modèles qu'on a utilisés ont fournis de bons résultats sur les deux approches ce qui nous a contraint à faire une comparaison entre les modèles de chaque approche et en choisir le meilleur pour l'exploiter dans la phase de tests sur des données externes, ces modèles sont : *Random Forest Classifier* pour la première approche et *Xception + Softmax* pour la deuxième approche.

Enfin, en effectuant des tests de prédiction sur des données externes au niveau des deux (02) approches, on a constaté que ces modèles ont générés des prédictions qui sont rarement incorrectes ce qui a confirmé leurs performances.

En ce qui concerne la deuxième approche - ***Dépistage par imagerie médicale*** on a conçu une *plate-forme* qui propose des outils intéressants qui aident dans la phase de diagnostic pour donner un côté réaliste à notre étude.

Perspective

De nos jours, la médecine et l'informatique ne cessent d'évoluer, pour cette raison, on ne peut jamais concevoir une approche qui soit une référence statique pour les années à venir...

En effet, de nouvelles théories et techniques sont à l'heure actuelle en cours de développement sans que personne ne s'en aperçoive, jusqu'au jour où une nouvelle révolution scientifique naisse et vienne apporter sa contribution dans notre société.

Cependant, notre étude est de loin d'être la plus performante ni la meilleure, ce qui lui laisse un large volet de contributions qui pourront que l'améliorer dans le futur. Dans notre étude, on s'est focalisé sur la classification des tumeurs du Pancréas sur (03) trois classes pour la première approche : *Healthy*, *Benign* et *Malignant* et seulement deux (02) classes : *Cancerous* et *Non-cancerous* pour la deuxième approche. Particulièrement, la deuxième approche qui a été implémenté sur la base des réseaux de neurones convolutifs et de **Dataset** utilisé était limité à deux (02) classes d'origine, ce qui peut être améliorée en proposant diverses classes qui répertorient les types de cancer qui touchent le Pancréas et ainsi améliorer les résultats de prédictions en fournissant plus de précisions sur le diagnostic.

Bibliographie

- [1] <https://www.alloprof.qc.ca/fr/eleves/bv/sciences/le-systeme-digestif-et-son-anatomie-s1258> - [site web visité le : 15/02/2022].
- [2] <http://chirurgie-digestive-sat.aphp.fr/chirurgie/pancreatectomies/anatomie-du-pancreas/> - [site web visité le : 16/02/2022].
- [3] <https://www.futura-sciences.com/sante/definitions/biologie-pancreas-6962/> - [site web visité le : 16/02/2022].
- [4] <https://cancer.ca/fr/cancer-information/cancer-types/pancreatic/what-is-pancreatic-cancer> - [site web visité le : 16/02/2022].
- [5] <https://sante.journaldesfemmes.fr/fiches-maladies/> - [site web visité le : 17/02/2022].
- [6] <https://cancer.ca/fr/cancer-information/cancer-types/pancreatic/what-is-pancreatic-cancer/non-cancerous-tumours> - [site web visité le : 17/02/2022].
- [7] <https://www.cancer.be/lexique/r/r-s-cable> - [site web visité le : 17/02/2022].
- [8] Le groupe suisse d'entraide pour les maladies du pancréas GSP, *Les maladies du pancréas*, Information pour les patientes et patients, *Tumeurs du pancréas*, Développement, page-14, Année - 2007.
- [9] <https://www.fondation-arc.org/cancer/cancer-pancreas/symptomes-diagnostic-cancer> - [site web visité le : 23/02/2022].
- [10] Le groupe suisse d'entraide pour les maladies du pancréas GSP, *Les maladies du pancréas*, Information pour les patientes et patients, *Tumeurs du pancréas*, Classification et origine des divers types de tumeurs du pancréas, page-12, Année - 2007.

- [11] <https://www.info-radiologie.ch/menu1-pancreas.php> - [site web visité le : 23/02/2022].
- [12] Mélanie Legrand, Flore Delalande, Claire Blechet, Rémy Kerdraon, Olivier Saint-Marc, and Patrick Michenet. Cystadénome séreux microkystique : une tumeur rare du pancréas. à propos de deux observations. In *Annales de Pathologie*, volume 39, pages 292–296. Elsevier, 2019.
- [13] <https://www.msmanuals.com/fr/professional/troubles-gastro-intestinaux/tumeurs-du-tractus-gastro-intestinal/insulinome> - [site web visité le : 25/02/2022].
- [14] <https://www.academie-medecine.fr/dictionnaire/> - [site web visité le : 25/02/2022].
- [15] P Amsalhem, S Peschard, L Crickx, B Gaudin, R Lévy, and E Leroy-Terquem. Carcinome à cellules acineuses du pancréas occulte révélé par des métastases hépatiques. volume 14, pages 82–83, 2005.
- [16] <https://www.themedicalchambers.com/fr/chographie-comment-ca-marche?> - [site web visité le : 25/02/2022].
- [17] <https://www.lumni.fr/video/comment-fonctionne-une-echographie> - [site web visité le : 25/02/2022].
- [18] <https://www.passeportsante.net/fr/Maux/examens-medicaux-operations/Fiche.aspx?doc=examen-echographie> - [site web visité le : 26/02/2022].
- [19] <https://www.cancer.be/le-cancer/jeunes-et-cancer/les-examens/lendoscopie-fibroscopie> - [site web visité le : 26/02/2022].
- [20] <https://www.futura-sciences.com/sante/definitions/medecine-endoscopie-2677/> - [site web visité le : 26/02/2022].
- [21] <https://www.topsante.com/medecine/examens-medicaux/comment-se-deroule-une-endoscopie> - [site web visité le : 26/02/2022].
- [22] <https://www.futura-sciences.com/sante/definitions/medecine-irm-8096/> - [site web visité le : 27/02/2022].
- [23] <https://www.futura-sciences.com/sante/definitions/medecine-irm-8096/> - [site web visité le : 27/02/2022].

- [24] <https://www.futura-sciences.com/sante/definitions/medecine-irm-8096/> - [site web visité le : 27/02/2022].
- [25] <https://www.vidal.fr/sante/examens-tests-analyses-medicales/irm-imagerie-resonance-magnetique.html> - [site web visité le : 27/02/2022].
- [26] <http://institutcancerologieprive.fr/scanner-thoraco-abdomino-pelvien-tap/> - [site web visité le : 28/02/2022].
- [27] <https://www.chu-montpellier.fr/fr/offre-de-soins/consultations/scanner-abdomino-pelvien-et-ou-thoracique-avec-injection-de-produit-de-contrast-333> - [site web visité le : 28/02/2022].
- [28] <https://www.chu-montpellier.fr/fr/offre-de-soins/consultations/scanner-abdomino-pelvien-et-ou-thoracique-avec-injection-de-produit-de-contrast-333> - [site web visité le : 28/02/2022].
- [29] <https://www.cancer.be/le-cancer/jeunes-et-cancer/les-examens/la-biopsie> - [site web visité le : 28/02/2022].
- [30] Fondation ARC pour la recherche sur le cancer, *Le cancer du Pancréas*, Diagnostic, Collection *Comprendre et Agir*, 1^{ère} édition, décembre 2010.
- [31] Université Médicale Virtuelle Francophone, *Tumeurs du Pancréas*, Support du cours - [Item 155], 2008-2009.
- [32] Fondation ARC pour la recherche sur le cancer, *Le cancer du Pancréas*, Pronostic, Collection *Comprendre et Agir*, 1^{ère} édition, décembre 2010.
- [33] Le groupe suisse d'entraide pour les maladies du pancréas GSP, *Les maladies du Pancréas*, Informations pour les Patients et Patientes, *GSP*, 2007.
- [34] <https://cancer.ca/fr/cancer-information/cancer-types/pancreatic/risks/reducing-your-risk> - [site web visité le : 28/02/2022].
- [35] <https://datascientest.com/machine-learning-tout-savoir> - [site web visité le : 11/03/2022].
- [36] <https://www.talend.com/fr/resources/what-is-machine-learning/> - [site web visité le : 11/03/2022].
- [37] <https://www.lemagit.fr/definition/Apprentissage-supervise> - [site web visité le : 11/03/2022].

-
- [38] https://stringfixer.com/fr/Semi-supervised_learning - [site web visité le : 13/03/2022].
- [39] <https://datascientest.com/apprentissage-non-supervise> - [site web visité le : 13/03/2022].
- [40] <https://dataanalyticspost.com/Lexique/apprentissage-par-renforcement/> - [site web visité le : 14/03/2022].
- [41] <https://dhirajkumarblog.medium.com/top-5-advantages-and-disadvantages-of-decision-tree-algorithm> - [site web visité le : 14/03/2022].
- [42] <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html#sklearn.ensemble.RandomForestClassifier> - [site web visité le : 18/03/2022].
- [43] https://www.imo.universite-paris-saclay.fr/~goude/Materials/ProjetMLF/rf_web.html - [site web visité le : 18/03/2022].
- [44] <https://datascientest.com/knn> - [site web visité le : 20/03/2022].
- [45] <https://www.journaldunet.fr/web-tech/guide-de-l-intelligence-artificielle/1501333-deep-learning-definition-et-principes-de-l-apprentissage-profond/> - [site web visité le : 25/03/2022].
- [46] <https://moncoachdata.com/blog/comment-fonctionne-le-deep-learning/> - [site web visité le : 25/03/2022].
- [47] <https://dataanalyticspost.com/convolutional-neural-network> - [site web visité le : 26/03/2022].
- [48] <https://datascientest.com/convolutional-neural-network> - [site web visité le : 26/03/2022].
- [49] <https://towardsdatascience.com/convolutional-layers-vs-fully-connected-layers-364f05ab460b> - [site web visité le : 26/03/2022].
- [50] <https://www.mathworks.com/help/deeplearning/ref/inceptionv3.html;jsessionid=a061b6d2444bad035fbae9c74dad> - [site web visité le : 27/03/2022].
- [51] <https://maelfabien.github.io/deeplearning/xception/#> - [site web visité le : 27/03/2022].
- [52] <https://iq.opengenus.org/architecture-of-densenet121/> - [site web visité le : 28/03/2022].

- [53] <https://en.wikipedia.org/wiki/ImageNet> - [*site web visité le : 28/03/2022*].
- [54] <https://towardsdatascience.com/understanding-and-implementing-lenet-5-cnn-architecture-deep-learning-a2d531ebc342> - [*site web visité le : 29/03/2022*].
- [55] <https://towardsdatascience.com/deep-learning-googlenet-explained-de8861c82765> - [*site web visité le : 29/03/2022*].
- [56] <https://www.geeksforgeeks.org/introduction-to-residual-networks/?ref=lbp> - [*site web visité le : 2/04/2022*].
- [57] <https://www.lebigdata.fr/confusion-matrix-definition> - [*site web visité le : 2/04/2022*].
- [58] <https://www.jedha.co/blog/matrice-de-confusion-comment-la-lire-et-l-interpreter> - [*site web visité le : 2/04/2022*].
- [59] <https://www.jedha.co/blog/quest-ce-que-loverfitting>
<https://www.youtube.com/c/MachineLearnia/> - [*site web visité le : 6/04/2022*].
- [60] <https://www.techno-science.net/glossaire-definition/Image-matricielle.html> - [*site web visité le : 6/04/2022*].
- [61] <https://www.futura-sciences.com/tech/definitions/informatique-image-matricielle-13097/> - [*site web visité le : 10/04/2022*].
- [62] <https://www.macapflag.com/blog/image-vectorielle/> - [*site web visité le : 10/04/2022*].
- [63] <https://datascientest.com/quelle-difference-entre-le-machine-learning-et-deep-learning> - [*site web visité le : 10/04/2022*].
- [64] <https://medium.com/jedha/les-applications-du-deep-learning-d674483d44af> - [*site web visité le : 12/04/2022*].
- [65] <https://www.sciencedirect.com/science/article/abs/pii/S0001407919350745> - [*site web visité le : 12/04/2022*].
- [66] Bodyanskiy, Yevgeniy and Chala, Olha. Generalized neo-fuzzy-neuron with membership functions of special type in medical diagnostics. -2021.
- [67] <https://fr.acervolima.com/ml-encodage-detiquettes-densembles-de-donnees-en-python/> - [*site web visité le : 15/04/2022*].
- [68] <https://dataanalyticspost.com/Lexique/normalisation/> - [*site web visité le : 15/04/2022*].

- [69] <https://www.invivoo.com/augmentation-images-modeles-machine-learning-\partie-1/> - [site web visité le : 15/04/2022].
- [70] <https://keras.io/api/layers/activations/> - [site web visité le : 20/04/2022].
- [71] <https://keras.io/api/optimizers/adam/> - [site web visité le : 20/04/2022].
- [72] https://keras.io/api/losses/probabilistic_losses/ - [site web visité le : 20/04/2022].
- [73] <https://keras.io/api/callbacks/> - [site web visité le : 20/04/2022].
- [74] https://keras.io/api/callbacks/model_checkpoint/ - [site web visité le : 23/04/2022].
- [75] <https://ledatascientist.com/google-colab-le-guide-ultime/> - [site web visité le : 23/04/2022].
- [76] <https://www.actuia.com/keras/> - [site web visité le : 27/04/2022].
- [77] <https://he-arc.github.io/livre-python/matplotlib/index.html> - [site web visité le : 28/04/2022].
- [78] <https://seaborn.pydata.org/> - [site web visité le : 2/05/2022].
- [79] <https://www.tensorflow.org/?hl=fr> - [site web visité le : 2/05/2022].
- [80] <https://courspython.com/apprendre-numpy.html> - [site web visité le : 2/05/2022].
- [81] <https://pandas.pydata.org/> - [site web visité le : 2/05/2022].
- [82] <https://python.doctor/> - [site web visité le : 8/05/2022].
- [83] <https://developer.mozilla.org/fr/docs/Web/HTML> - [site web visité le : 8/05/2022].
- [84] <https://developer.mozilla.org/fr/docs/Web/CSS> - [site web visité le : 8/05/2022].
- [85] https://developer.mozilla.org/fr/docs/Learn/JavaScript/First_steps/What_is_JavaScript - [site web visité le : 12/05/2022].
- [86] <https://www.digitalocean.com/community/tutorials/how-to-make-a-web-application-using-flask-in-python-3-fr> - [site web visité le : 12/05/2022].

[87] <https://www.pierre-giraud.com/javascript-apprendre-coder-cours/introduction-ajax/> - [*site web visité le : 15/05/2022*].