



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA
Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie
Département d'informatique

N° d'ordre : **SIOD21/M2/2019**

Mémoire

présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : **SIOD**

Une Approche intelligente pour la santé

(Aide au diagnostic du cancer de la peau)

Par :

HEMILI MOHAMED LAMINE

Soutenu le **07** juillet 2019, devant le jury composé de :

Houadjli Hadia

MAA

Président

Terrissa Sadek Labib

MCA

Rapporteur

Zouiouche Amina

MAA

Examineur

Dédicace

Je dédie ce modeste travail à :

mon père et ma Mère

Toute ma famille

Et tous mes amis.

A mon Défunt Grand Père (abi).

Remerciements

Je tiens à remercier mon encadreur le Dr TERISSA Labib Sadek pour ses conseils, ses orientations et son aide si Précieuse.

Je remercie également les membres de jury, Dr.Houadjli Hadia et Dr.Zouioueche Amina d'avoir accepté de juger ce travail.

Enfin, je présente mes remerciements à toutes les personnes qui m'ont encouragé à construire et élaborer ce mémoire.

Résumé

Le Machine Learning est devenu depuis quelque années un outil très puissant grâce à l'évolution des calculateurs et des capteurs.

Cet outil permet de présenter des solutions performantes et intelligentes dans plusieurs domaines et en particulier celui de la santé.

Dans ce travail de master, nous avons développé une solution basée sur les techniques sur le Machine Learning (Arbre de Décision, Random Forest) pour l'aide au diagnostic du cancer de la peau. Cette étude a été effectuée sur une base de données réelles fournis par le département d'oncologie de l'hôpital de Franche – Comté, et les résultats ont été comparés et discutés.

Mots-clés:

Machine Learning, Clasification, ,Arbre de Décision, Random Forest,Cancer de la peau.

Table des Matières :

Dédicace	1
Remerciements	2
Résumé	3
Table des Matières	4
Liste des Figures	7
Liste des Tableaux	8
Introduction Générale	9
Chapitre 1 :	10
1) Introduction	10
2) Définition du cancer	10
3) Cause du cancer	11
4) Développement du cancer	12
5) Caractéristique des cellules cancéreuses	14
6) Stade du cancer	15
7) Type de cancer	15
8) Traitement du cancer	20
9) Conclusion	24
Chapitre 2 : (Machine Learning)	
1. Introduction	25
2. Définition du Machine Learning	26
3. Principe du Machine Learning	26
4. Domaine d'application	27
5. Modèle d'apprentissage	28
6. Type d'apprentissage	29
6.1. Apprentissage supervisé	29
6.1.1. Régression	30

6.1.1.1. Type de régression	30
6.1.2. Classification	31
6.1.2.1. Support Vector Machine (SVM)	32
6.1.2.2. Classification Naïve Bayésienne	36
6.1.2.3. Arbre de décision	36
6.1.2.4. K –plus proche voisin (K-PPV)	38
6.1.2.5. Réseaux de neurones artificielles (RNA)	39
6.2. Apprentissage non supervisé	41
6.2.1. Clustering	42
7. Conclusion	45

Chapitre 3 : (Conception)

1. Introduction	46
2. Schéma Synoptique	47
3. Objectif	48
4. Description de la base de données	48
5. Diagramme de cas d'utilisation	49
6. Prétraitement	50
7. Traitement	52
7.1. Arbre de décision	52
7.2. Random Forest	55
7.3. Evaluation des Performance	56
8. Conclusion	59

Chapitre 4 : (Implémentation et discussion)

1. Introduction	60
2. Outils de programmation	60
2.1. Langage de programmation	60
2.2. Outils de développement	60
2.3. Les Packages	60
3. Implémentation	61
3.1. Les interfaces	61
4. Résultat de l'application	62

4.1. Arbre de Décision	62
4.2. Random Forest	64
5. Comparaison des techniques	67
6. Conclusion	67
Annexe :	68
Conclusion :	77
Référence :	78

Liste des Figures :

Figure 1 : Division Cellulaire normale.

Figure 2 : Division Cellulaire Cancéreuse

Figure 3 : Propagation du Cancer

Figure 4 : Cancer du Foie

Figure 5 : Cancer du Pancréas.

Figure 6 : Cancer des Poumons.

Figure 7 : Cancer du Sein.

Figure 8 : Cancer de la peau.

Figure 9 : Coupe de la peau.

Figure 10 : Type de cancer de la peau.

Figure 11 : Schéma sur les étapes du diagnostic au traitement du cancer.

Figure 12 : Domaine d'utilisation du Machine learning.

Figure 13 : SVM linéairement séparable

Figure 14 : les vecteurs support (points encercle) ;

Figure 15 : Noyau

Figure 16 : Arbre de décision pour la base de données pour jouer au golf.

Figure 17 : Exemple K-PPV.

Figure 18 : Transformation d'un neurone biologique en formels

Figure 19 : Neurone simple

Figure 20 : Fonction Heaviside

Figure 21 : Fonction sigmoïde

Figure 22 : Fonction tangent hyperbolique

Figure 23 : Graphe clustering hiérarchique

Figure 24 : Schéma synoptique pour le cancer

Figure 25 : Base de données du cancer de la peau.

Figure 26 : Arbre de décision.

Figure 27 : Graph simplifier de random forest.

Figure 28 : Logo NetBeans.

Figure 29 : Fenêtre principale de l'application.

Figure 30 : Fenêtre pour la classification

Figure 31 : Résultat de la classification avec les arbres de décision.

Figure 32 : Graphe de l'arbre de décision obtenue.

Figure 33 : Graph d'écrivent l'accuracy par rapport au nombre de cross-validation.

Figure 34 : Résultat random forest

Figure 35 : Graph d'écrivent l'accuracy par rapport au nombre de cross-validation.

Liste des Tableaux :

Tableau 1 : Base de données pour jouer au golf.

Tableau 2 : Conversion de la base de données.

Tableau 3 : Résultat de l'arbre de décision.

Tableau 4 : Résultat de Random Forest.

Tableau 5 : Tableau comparative des résultats des deux techniques.

Dans un contexte mondial où les technologies progressent de façon fulgurante, surtout dans le domaine de l'intelligence artificielle (machine learning , deep learning).

Le Machine Learning a émergé dans la seconde moitié du XXème siècle du domaine de l'intelligence artificielle et correspond à l'élaboration d'algorithmes capables d'accumuler de la connaissance et de l'intelligence à partir d'expériences

Le domaine médical est l'un des demandeurs du le machine learning ainsi le personnel médical peuvent être assisté pour diagnostiquer ou pronostiquer des maladies, il permet aussi de gagner du temps et des ressources, pour la détection de maladie ou l'analyse des examens radiologiques.

Notre objectif est de construire un modèle qui pourra aider les oncologues pour le diagnostic du cancer de la peau, grâce à une application et qui sera testée dans la réalité, ainsi le médecin pourra rapidement diagnostiquer son patient.

3. Causes du Cancer :

D'après des études récentes, très peu de cancers ne sont causés que par un seul élément. La plupart semblent être causés par un ensemble complexe de nombreux facteurs de risque, mais il arrive que le cancer apparaisse chez des personnes qui ne présentent aucun facteur de risque. Ce dernier est une substance ou un état qui fait augmenter le risque de cancer. Les facteurs de risque du cancer peuvent jouer différents rôles dans l'apparition du cancer et son développement. Ces facteurs sont:

- **Tabac :**

Le tabac a un lien étroit avec l'apparition de beaucoup de cancer notamment celui du poumon, car les oncologues estiment que le tabac est la première cause du cancer du poumon. Il est à l'origine de 90 % des cas. **Le cancer du poumon est rare chez les non-fumeurs.** Car moins de 1 % des non-fumeurs développent un cancer de poumon. Ou le tabagisme passif augmente de 20 % le risque de cancer du poumon chez le non-fumeur. Le tabagisme est le facteur de risque le plus important du cancer responsable d'environ 22% du total des décès dus au cancer [1].

- **L'exposition aux Rayons :**

Il existe plusieurs rayons (UV, X, ...) qui sont tous dangereux pour la santé et qui peuvent provoquer des cancers de la peau. Le Centre International de Recherche sur le Cancer (CIRC) a estimé qu'au moins **85 % des mélanomes étaient causés par l'exposition au rayonnement solaire.**

- **Prédisposition Génétiques :**

Certaines personnes sont prédisposées au cancer que d'autre dû à une mutation dans des gènes et qui sont transmis de génération en générations

- **Obésité :**

L'obésité est un défi majeur apparu à la fin du XXe siècle apparu suite à la mauvaise alimentation ainsi que la sédentarité qui conduit à un surpoids augmentent ainsi les chances de contracter un cancer (estomac, colon, foie, pancréas, ...).

- **Alimentation Malsaine :**

C'est la principale source d'énergie pour le corps humain mais malheureusement dans nos temps moderne les aliments contiennent beaucoup d'ingrédients industriel mais aussi des résidus de pesticide, selon certaine étude 30% des cancer sont induits par une mauvaise alimenation,

- **Sédentarité :**

L'inactivité physique a un lien très étroit avec le cancer plus la personne est sédentaire plus elle augment la chance d'avoir un cancer mais aussi d'autre maladie comme le diabète, les maladie cardio-vasculaire.

- **L'Alcool :**

L'alcool est un cancérrogène avéré Il serait responsable de 11% des cancers chez les hommes et de 4.5% des cancers chez les femmes (Selon l'OMS).

- **La Pollution :**

La pollution est un vrai problème des temps modernes due à l'industrialisation intensive de plusieurs pays (chine, inde, ...etc.) mais aussi avec les catastrophes climatiques, d'après des recherches ont estimées que la pollution de l'air, de l'eau et des sols seraient impliqués dans le développement d'environ 1 à 4 % de tous les cancers []. La pollution se trouve dans l'amiante, les gaz émis par les échappements des voitures, ainsi que les particules fines, le plastique, ..., etc.

4. Développement d'un Cancer :

Notre corps est composé de milliers de milliards de cellules regroupées pour former les tissus et les organes. Les gènes présents dans le noyau de chaque cellule lui indiquent quand se développer, travailler, se diviser et mourir. En temps normal, nos cellules suivent ces directives (Figure 2) .

Mais lorsque notre **ADN** est modifié ou endommagé, un gène peut subir une mutation. Les gènes mutés ne fonctionnent pas correctement parce que les directives dans leur ADN

sont confuses. Ainsi, les cellules qui devraient se reposer pourraient se diviser et croître de façon désordonnée, ce qui peut mener au cancer (figure 3).

Lorsque les gènes fonctionnent correctement, ils indiquent aux cellules quand le moment est venu de se développer et de se diviser. Quand les cellules se divisent, elles font des copies exactes d'elles-mêmes. Une cellule se divise en 2 cellules identiques puis ces 2 cellules se divisent en 4 et ainsi de suite. Habituellement, chez l'adulte, les cellules se développent et se divisent pour produire plus de cellules uniquement lorsque le corps en a besoin, comme pour remplacer des cellules vieillissantes ou endommagées

- **Division Cellulaire normale :**

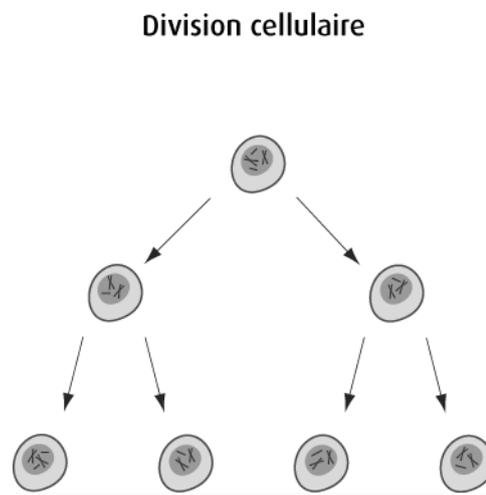


Figure 1 Division Cellulaire Normale[23]

- **Division Cellulaire Cancéreuse :**

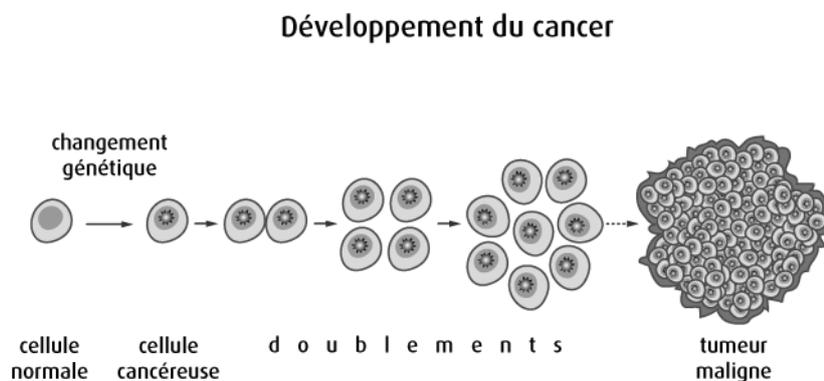


Figure 2 Division Cellulaire Cancéreuse [23]

Mais les cellules cancéreuses sont différentes. Les cellules cancéreuses présentent des mutations génétiques qui ont changé la cellule normale en cellule cancéreuse. Ces mutations génétiques peuvent être héréditaires, se développent avec le temps puisqu'on vieillit et que les gènes s'affaiblissent

La cellule cancéreuse agit différemment de la cellule normale. Elle commence à se développer et à se diviser de façon désordonnée au lieu de mourir quand elle le devrait. Elle ne mûrit pas non plus autant qu'une cellule normale, alors elle reste immature. Bien qu'il existe de nombreux types différents de cancer, ils se forment tous à la suite du développement anormal et incontrôlé de cellules. Le cancer peut prendre naissance dans n'importe quelle cellule du corps

5. Caractéristiques des cellules cancéreuses :

Les cellules cancéreuses ont les caractéristiques suivantes :

- Elles se divisent de façon désordonnée;
- Elles sont immatures et ne deviennent pas des cellules matures qui ont des tâches spécifiques;
- Elles évitent le système immunitaire;
- Elles ignorent les signaux qui leur indiquent de cesser de se diviser ou de mourir quand elles le devraient;
- Elles ne collent pas très bien les unes aux autres et peuvent se propager à d'autres parties du corps par le sang ou le système lymphatique;
- Elles envahissent et endommagent les tissus et les organes.

Comment le cancer se propage

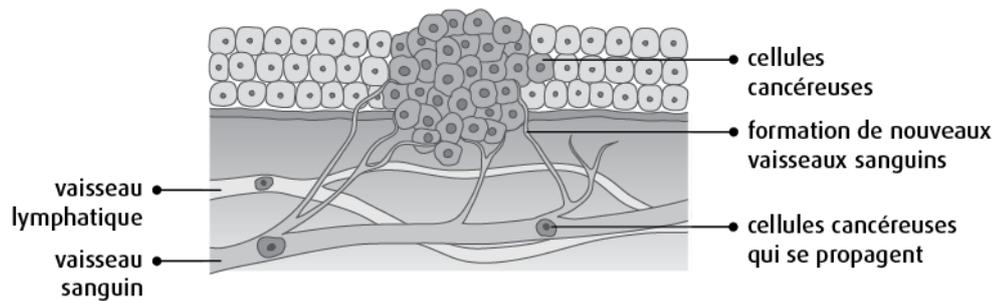


Figure 3. Propagation du Cancer.[23]

6. Stades du Cancer :

- **Stade 1** : Ce Stade correspond à une tumeur qui peut être unique et de petite taille,
- **Stade 2** : Ce Stade correspond à un volume local plus important,
- **Stade 3** : Ce Stade correspond à un envahissement des ganglions lymphatiques ou des tissus avoisinants,
- **Stade 4** : Ce Stade correspond à une extension plus large dans l'organisme sous forme de métastases.

7. Types de Cancers :

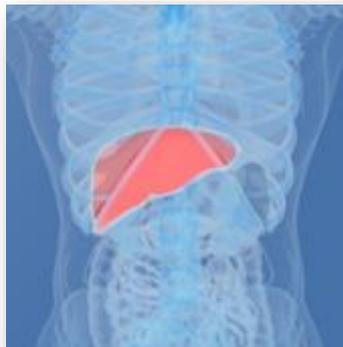


Figure 4. Cancer du Foie.

Le **cancer du foie** (*Figure 4*) se manifeste lorsque des cellules anormales se forment de façon incontrôlée dans ses tissus. Le cancer primitif (aussi appelé **hépatocarcinome**) est un cancer qui prend naissance dans les cellules du foie (appelées hépatocytes). Le cancer secondaire ou **métastatique** provient d'un cancer qui s'est d'abord formé ailleurs dans l'organisme avant de se propager dans le foie par voie sanguine..



Figure 5. Cancer du Pancréas.

Le **cancer du pancréas** (*Figure 5*) prend naissance dans les cellules du pancréas. Les cellules du pancréas subissent parfois des changements qui rendent leur mode de croissance ou leur comportement anormal. Ces changements peuvent engendrer la formation de tumeurs [3].



Figure 6. Cancer des Poumons.

Le cancer du poumon (*Figure 6*) prend naissance dans les cellules du poumon. Dans le monde, le cancer du poumon est le cancer le plus fréquent en termes tant d'incidence que de mortalité (1,35 million de nouveaux cas par an et 1,18 million de décès).

Les cancers du poumon sont divisés en cancer du poumon non à petites cellules et en cancer du poumon à petites cellules selon le type de cellule à partir duquel ils se développent.



Figure 7. Cancer du Sein.

Dans cancer du sein (*Figure 7*), les cellules peuvent rester dans le sein ou se répandre dans le corps par les vaisseaux sanguins ou lymphatiques. La plupart du temps, la progression d'un cancer du sein prend plusieurs mois et même quelques années.

Le **cancer du sein** est le cancer le plus diagnostiqué chez les femmes à travers le monde, autant avant qu'après la ménopause [4].



Figure 8. Cancer de la Peau.

Cancer de la Peau :

La peau (*Figure 8*), est une barrière protectrice entre notre organisme et l'environnement extérieure et que celle ci est composée de plusieurs couche protectrice épiderme, derme, hypoderme (*Figure 9*).

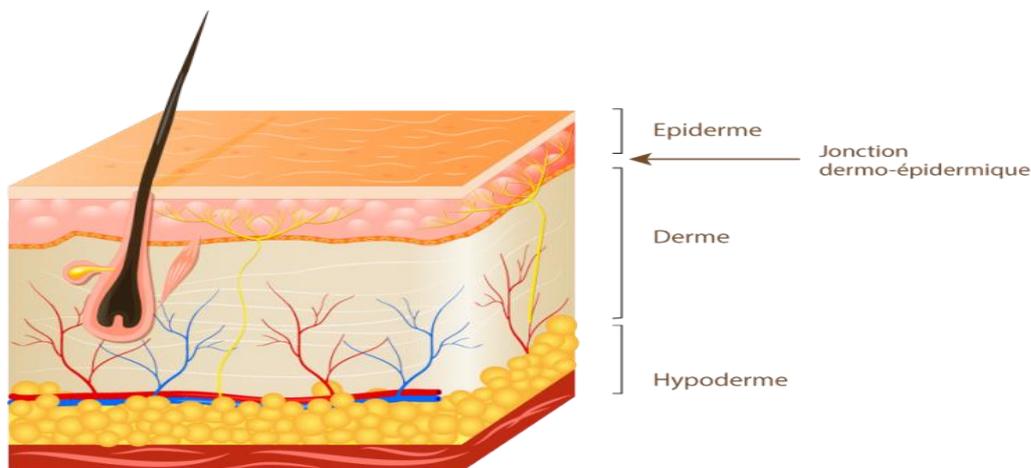


Figure 9. Coupe de la Peau.[24]

Le cancer de la peau se caractérise par des tumeurs malignes qui attaquent les cellules de la peau. Si certains cancers sont qualifiés de bénins ou superficiels, ils sont généralement à surveiller. On compte 3 types de cancer : les épithéliomes, les mélanomes et les sarcomes. L'une des causes du cancer de la peau aurait pour origine une trop grande exposition aux UV, qu'ils soient naturels (le soleil) ou artificiels (cabines de bronzage).

Le cancer de la peau se divise en deux catégories principales : les non-mélanomes et les mélanomes.

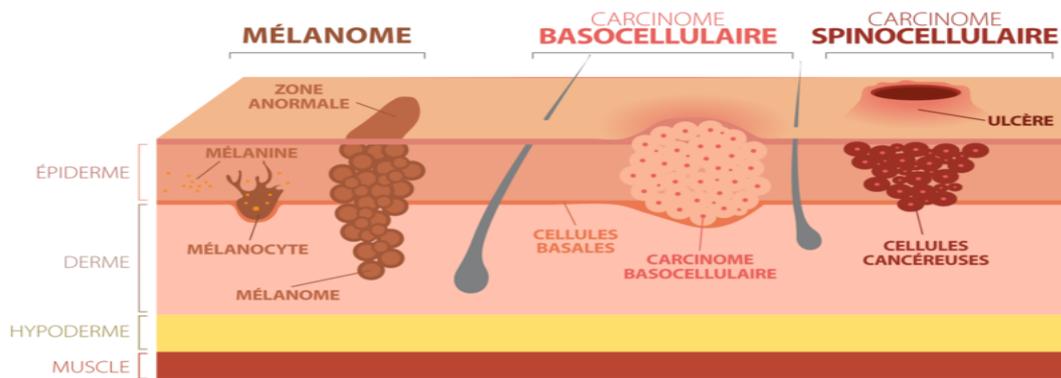


Figure 10. Type de cancer de la peau. [25]

➤ **Non-Mélanomes (Carcinome):**

Les **carcinomes** sont les cancers de la peau les plus fréquents. Ils surviennent généralement après l'âge de 50 ans, sur les zones découvertes du corps (visage, cou, épaules, avant-bras, jambes...). Ils sont le plus souvent dus à une exposition au soleil excessive et chronique.

Les carcinomes sont facilement guérissables dans la majorité des cas. Néanmoins, certains d'entre eux, appelés « carcinomes épidermoïdes », peuvent entraîner des lésions à distance (métastases) s'ils ne sont pas retirés à temps.

➤ **Mélanomes :**

On donne le nom de mélanomes aux **tumeurs malignes** qui se forment dans les mélanocytes, les cellules qui produisent la mélanine (un pigment) et qui se trouvent notamment dans la peau et les yeux. Ils se manifestent habituellement par une **tache noirâtre**.

Le **mélanome** est beaucoup plus rare que le carcinome mais c'est le plus grave des cancers de la peau, du fait de son « haut potentiel métastatique », c'est-à-dire de sa capacité à se propager rapidement à d'autres parties du corps.

Le mélanome peut se situer n'importe où sur le corps, y compris sur le cuir chevelu. On le trouve assez fréquemment sur le tronc chez l'homme et sur les jambes chez la femme.

Environ 30% des mélanomes se développent à partir de nævus pigmentés (environ la moitié d'entre eux à partir de nævus atypiques); la quasi-totalité des autres se développent à partir de mélanocytes de la peau saine.

Les nævus atypiques (nævus dysplasiques) peuvent être des précurseurs du mélanome nous allons nous pencher sur la tumeur de Spitz.

Les tumeurs de Spitz (NS) sont des tumeurs fréquentes et le plus souvent bénignes. Il existe néanmoins des formes atypiques à potentiel métastatique pour lesquelles le diagnostic différentiel avec un mélanome malin est délicat.

Le *nævus de Spitz* est un type rare de grain de beauté qui affecte généralement les jeunes et les enfants. Bien qu'elle puisse ressembler à une forme grave de mélanome, le cancer de la peau, une lésion du nerf de Spitz n'est pas considérée comme cancéreuse.

Le *nævus de Spitz* n'est pas très commun. Certaines estimations suggèrent qu'elles touchent environ 7 personnes sur 100 000. Environ 70% des personnes chez qui on diagnostique un nævus de Spitz ont 20 ans ou moins

Les enfants et les adolescents ayant la peau claire sont plus susceptibles de développer un nævus de Spitz.

8. Traitements du Cancer :

Parce qu'il n'existe pas un mais des cancers et parce que chaque malade est unique, différents types de traitements peuvent être utiliser :

▪ La chirurgie :

L'opération chirurgicale est la forme la plus ancienne de traitement des cancers. C'est encore aujourd'hui bien souvent le premier traitement qui est proposé aux personnes chez qui un cancer a été diagnostiqué. Elle est utilisée dans environ 80 % des cas, Une fois la tumeur localisée, on opère pour la retirer. Il existe plusieurs type de chirurgie :

- **Chirurgie diagnostique :**

Elle consiste à prendre une partie ou toute la tumeur pour effectuer un diagnostic à l'aide d'un examen microscopique. [15]

- **Chirurgie à visée curative :**

Elle consiste à enlever la tumeur seule ou l'organe qui le contient pour vérifier l'absence d'extension tumorale en bordure de la tumeur.

- **Chirurgie d'exérèse ganglionnaire :**

Elle fait une ablation des ganglions lymphatiques qui drainent le territoire de la tumeur.

- **Chirurgie de réduction tumorale :**

L'ablation totale de la tumeur n'étant pas possible, elle permet de pratiquer une réduction de la tumeur pour une meilleure action avec un autre traitement comme la radiothérapie ou la chimiothérapie [15].

- **Chirurgie réparatrice ou reconstructrice :**

Elle permet de réparer ou de restaurer la forme des organes ainsi elle améliore la qualité de vie du patient.

- **La radiothérapie :**

La radiothérapie est l'un des traitements les plus fréquemment employés dans la prise en charge des cancers. Plus de la moitié des personnes atteintes d'un cancer reçoivent ce type de traitement.

Il consiste à exposer les cellules cancéreuses à des rayons. Cette exposition provoque une transformation des cellules qui perdent alors leur faculté à se multiplier. Seule la zone où se trouve la tumeur est exposée aux rayons afin d'éviter au maximum que des cellules saines ne soient elles aussi touchées.

Le plus souvent, la radiothérapie est **externe**, les rayons sont émis par un appareil pendant que vous êtes allongé, sans bouger, sur une table. Les séances durent quelques minutes et sont répétées sur plusieurs jours. Au préalable, la zone exposée aura été soigneusement délimitée pour cibler spécifiquement la tumeur. Il existe une autre forme de radiothérapie dite **interne**. Dans ce cas, les rayons sont émis par un implant

placé (au cours d'une intervention chirurgicale) à proximité de la tumeur. Le type et la quantité de rayons prescrits par le médecin ainsi que le nombre de séances sont notamment déterminés en fonction du type de cancer le malade est atteint, de la taille de la tumeur et de sa localisation dans le corps.

- **La chimiothérapie :**

La chimiothérapie est un traitement reposant sur la prise d'une substance chimique, le malade prend un ou plusieurs médicaments actifs contre les cellules cancéreuses, Ces médicaments agissent soit en détruisant les cellules cancéreuses, soit en arrêtant leur multiplication. Certains d'entre eux sont plus actifs contre un type de cancer plutôt que d'autres.

La majorité des médicaments de chimiothérapie se présente sous une forme liquide qui est injectée directement dans une veine. Ils sont ainsi administrés par voie intraveineuse au cours de perfusions.

La chimiothérapie s'accompagne généralement d'effets indésirables, qui sont connus (chute de cheveux, vomissement, ...).

Les médicaments les plus utilisés sont :

- Agent intercalant :**

Obtenu à partir de souches de micro-organisme, ils nécessitent une surveillance du patient pour ces risques de toxicité.

- Alkylants :**

Entraine l'arrêts de la division des cellules.

- Antimétabolites :**

Ils inhibent la synthèse des bases nécessaires à l'ADN.

- **L'hormonothérapie :**

Ce type de traitement repose, comme pour la chimiothérapie, sur la prise de médicaments.

certaines cancers sont sensibles à l'action d'hormones naturellement produites par l'organisme. C'est fréquemment le cas pour les cancers du sein et de la prostate. Les

cellules cancéreuses ont tendance à se multiplier plus vite en présence de ces hormones. L'idée est donc de bloquer la production ou l'activité de ces hormones.

L'hormonothérapie est utile uniquement si le cancer est sensible aux hormones (on dit alors qu'il est hormono-sensible). C'est le cas si les cellules cancéreuses possèdent à leur surface des récepteurs spécifiques à ces hormones. L'étude de la tumeur permet de le déterminer. Si tel est le cas, le traitement repose sur la prise d'un médicament par voie orale. Le type de médicament, la dose prescrite et la durée du traitement sont adaptés à chaque situation.

- **Immunothérapie :**

Le but de ce traitement est de stimuler les défenses du corps contre les cellules cancéreuses, les mécanismes de défense immunitaire sont de mieux en mieux connue et on constate que les tumeurs entourées de nombreuse cellules immunitaire sont d'un meilleur pronostique.

Malheureusement un nombre trop important de cellules cancéreuses bloque les défenses immunitaires.

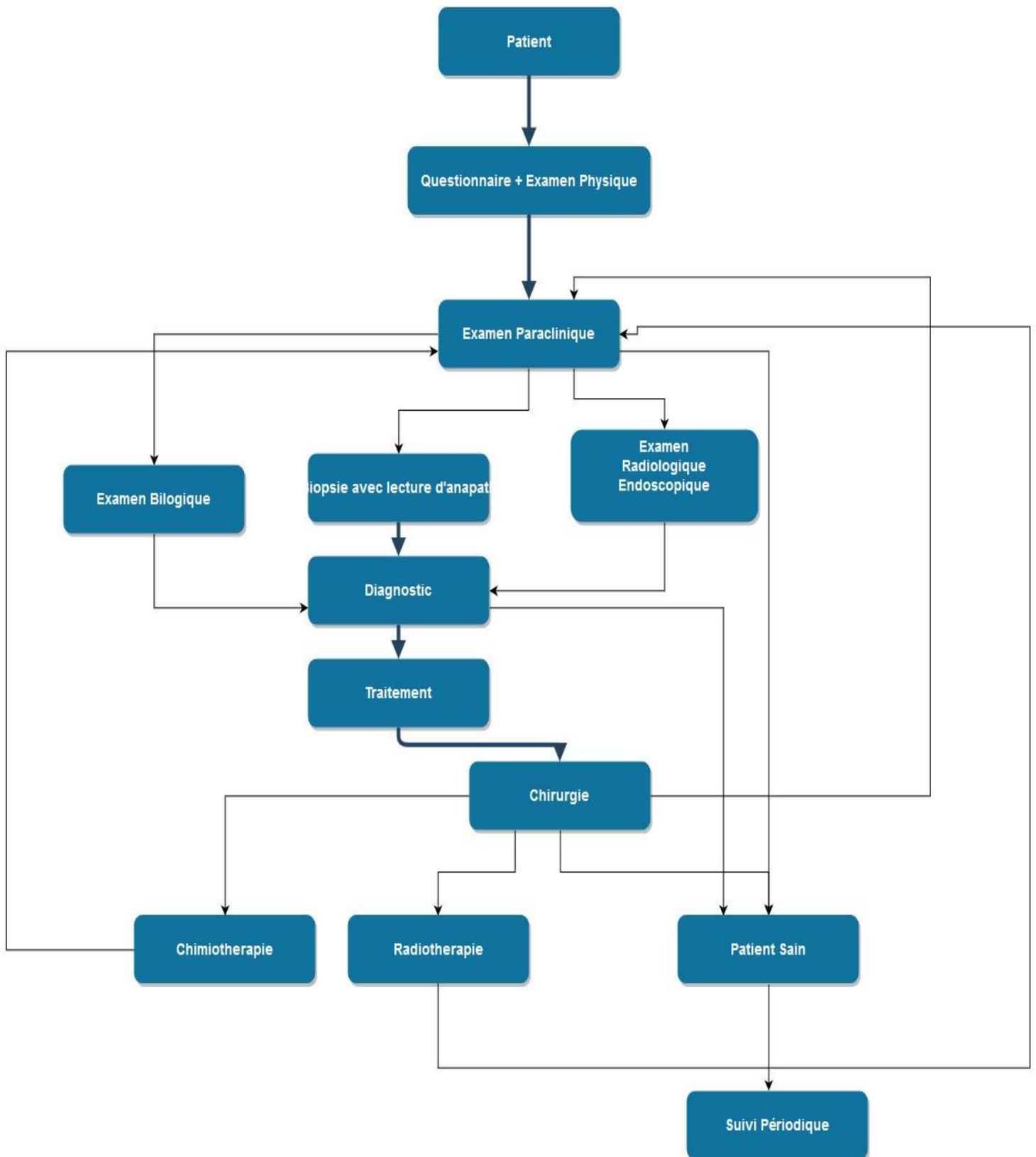


Figure 11. Schéma sur les étapes du diagnostic au traitement du cancer.

9. Conclusion :

Dans ce chapitre nous avons présenté le cancer, quelle sont ces causes, ces types et les traitements pour le vaincre ou du moins le stoppé, plus le diagnostic est pronostiqué plus le patient a de chance de guérir.

Aujourd'hui avec les progrès scientifique et technologique des nouvelles technologies et l'IA beaucoup de travaux et de recherches tentent d'apporter des solutions basées sur les technologies de l'IA.

En particulier celles du Machine Learning qui permettent de présenter des diagnostics aide cancer (DAC) et des pronostics aide cancer (PAC), dans notre prochain chapitre on parlera du Machine Learning

Chapitre 2 : Machine Learning



1. Introduction:

Depuis l'antiquité, le sujet des machines pensantes préoccupe les esprits. Ce concept est la base de pensées pour ce qui deviendra ensuite l'intelligence artificielle, ainsi qu'une de ses sous-branches : l'apprentissage automatique [21].

La concrétisation de cette idée est principalement due à Alain Turing et à son concept de la « machine universelle » en 1936, qui est à la base des ordinateurs d'aujourd'hui. Il continuera à poser les bases de l'apprentissage automatique, avec son article sur « L'ordinateur et l'intelligence » en 1950, dans lequel il développe, entre autres, le test de Turing [21].

Le Machine Learning est sans aucun doute le domaine le plus en vogue ces derniers années il consiste à prédire ou reconnaître automatiquement une forme. Cette branche a conquis tous les domaines de la vie face à l'évolution des et des capteurs.

2. Définition du Machine Learning :

L'apprentissage automatique (Machine Learning) est un type d'intelligence artificielle qui confère aux ordinateurs la capacité d'apprendre sans être explicitement programmés.

Cette technologie s'appuie sur le développement de programmes informatiques capables d'acquérir de nouvelles connaissances afin de s'améliorer et d'évoluer d'eux-mêmes dès qu'ils sont exposés à de nouvelles données.

L'objectif visé est de rendre la machine ou l'ordinateur capable d'apporter des solutions à des problèmes compliqués, par le traitement d'une grande quantité d'informations. Cela offre ainsi une possibilité d'analyser et de mettre en évidence les corrélations qui existent entre deux ou plusieurs situations données, et de prédire leurs différentes implications.

Les algorithmes du Machine Learning sont capables d'évoluer en permanence.

3. Principe du Machine Learning :

Les algorithmes utilisés permettent, dans une certaine mesure, à un ordinateur, d'adapter ses analyses et ses comportements, en se fondant sur l'analyse de données historiques.

La difficulté réside dans le fait que l'ensemble de tous les comportements possibles compte tenu de toutes les entrées possibles devient rapidement trop complexe à décrire (on parle d'explosion combinatoire). On confie donc à des programmes le soin d'ajuster un modèle pour simplifier cette complexité et de l'utiliser de manière opérationnelle. Idéalement, l'apprentissage visera à être non supervisé, c'est-à-dire que la nature des données d'entraînement n'est pas connue.

Ces programmes, selon leur degré de perfectionnement, intègrent éventuellement des capacités de traitement probabiliste des données, d'analyse de données issues de capteurs, de reconnaissance des Formes (image, vidéo, vocale, , d'écriture...).

4. Domaine d'application :

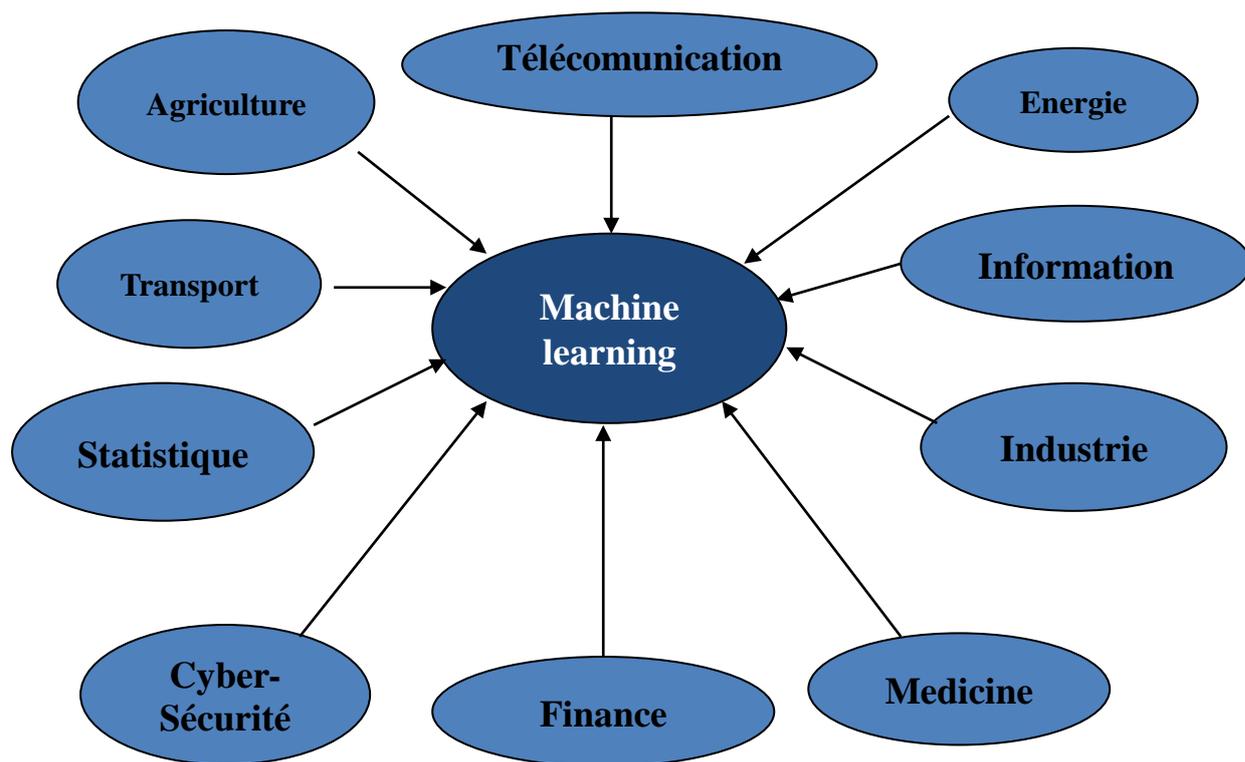


Figure 12. Domaine d'utilisation du Machine Learning.

- **Agriculture :**

Grâce à l'apprentissage automatique, les agriculteurs commencent à personnaliser leurs techniques de culture en fonction de données météo et des caractéristiques singulières d'une portion donnée de leur terrain [20].

- **Energie :**

A l'aide des capteurs posés sur leur infrastructure et des technologies de machine learning qui permettent de traiter les données collectées, les

énergéticiens peuvent prévoir à l'avance quand leurs éoliennes (par exemple) risquent de tomber en panne et la réparé.

○ **Finance :**

Dans la finance, cette technologie permet de débusquer les transactions frauduleuses en analysant les données bancaires des acheteurs, mais également de leur proposer des produits financiers personnalisés[20],

○ **Télécommunication :**

Les télécommunications pourront optimiser les investissements très lourds qu'ils doivent réaliser pour entretenir leur infrastructure en analysant les données collectées sur le terrain par des capteurs. L'apprentissage automatique leur permettra également de créer des assistants vocaux intelligents, capables de prendre en charge la majorité des appels qui sont aujourd'hui gérés par les salariés de leurs call centres.

○ **Médecine :**

L'apprentissage automatique est capables maintenant de poser le diagnostic d'une maladie en analysant les résultats d'un scanner, d'une biopsie ou d'autres types d'analyses médicales [20].

○ **Industrie :**

Dans le monde de l'industrie aussi, le machine learning permet d'ajuster le calendrier de maintenance en fonction des données collectées directement sur les équipements, pour éviter au maximum les pannes et le manque à gagner important qu'elles génèrent pour l'entreprise. L'analyse de données permet également d'optimiser les cycles de production complexes de certaines usines.

○ **Transport :**

Le machine learning permet aux acteurs du monde du transport d'optimiser leur prix grâce à une analyse en temps réel de la demande [20].

5. Modèles d'apprentissage :

Les modèles vont permettre au système d'apprendre efficacement selon un biais pour la réalisation de la tâche désirée. Le choix du modèle à employer est donc primordial pour réussir un apprentissage optimal :

-
- **Modèle fonctionnel** : il est basé sur une fonction de décision sur les attributs d'entrées permettant de regrouper les données selon un critère bien définie. La plus célèbre et la plus simple est sans doute la fonction linéaire (SVM support vecteur machine)
 - **Modèle probabiliste** Ce modèle se base sur une distribution de probabilité sur les attributs des entrées. Les lois probabilistes en font une grande partie de ce modèle, les réseaux bayésiens sont sans doute l'exemple typique de ce genre de modèle.
 - **Modèle connexionniste (neuro)**: Ce modèle se base sur un réseau de neurones artificiels. Ce type de modèle s'inspire à la base du fonctionnement du cerveau humain Nous citerons les perceptrons et les réseaux de neurone multicouche (MLP) comme exemples typiques connexionnistes.

6. Types d'Apprentissages :

En résumé, l'apprentissage est un terme très général qui décrit le processus selon lequel la machine peut accroître ses connaissances et améliorer ses performances au cours du temps.

6.1.Apprentissage Supervisé :

L'apprentissage supervisé est simplement une formalisation de l'idée d'apprendre à partir d'exemples. Dans l'apprentissage supervisé, nous devons connaître au préalable la nature des données et la sortie de notre système, ce système reçoit deux ensembles de données, un ensemble de formation et un ensemble de tests. L'apprentissage est constitué de deux phases :

- Phase d'apprentissage (d'entraînement) :

Dans cette phase l'algorithme d'apprentissage reçoit en entrées des exemples d'apprentissage et produit un modèle de prédiction le plus performant possible, c'est-à-dire le modèle qui produit le moins d'erreurs en prédiction.

- Phase de test :

C'est la phase de prédiction des nouveaux exemples en fonction du modèle obtenue lors de la phase d'apprentissage qui va associer à chaque entrée « X »

quelconque une sortie (étiquette) « Y » qui auras normalement été donnée par le superviseur (tester la précision du modèle).

En d'autres termes, le but de l'apprenant est de développer une règle, un programme ou une procédure qui classifie de nouveaux exemples (dans l'ensemble de tests) en analysant des exemples.

Dans l'apprentissage supervisé on distingue deux problèmes :

6.1.1. Régression :

La **régression** est un ensemble de méthodes statistiques très utilisées pour analyser la relation d'une variable par rapport à une ou plusieurs autres. On utilise la régression pour les variables *Quantitative*.

La fonction de prédiction est appelée **Régréseur**

6.1.1.1. Types de Régression :

- **Régression Linéaire :**

Est un modèle de régression qui veut faire une relation linéaire entre une variable dites expliquées et une ou plusieurs variables explicatives.

- **Régression Polynomial :**

C'est l'analyse d'une variable aléatoire à partir d'une fonction polynomiale

On cherche un polynôme :

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

Ainsi la régression polynomiale peut être écrite comme suite :

$$Y_i = a_n X_{i,n} + a_{n-1} X_{i,n-1} + \dots + a_1 X_{i,1} + a_0 + \varepsilon_i$$

- **Régression Logistique :**

C'est un modèle statistique de classification qui a été redécouvert et qui a récemment gagné une grande popularité grâce à sa bonne performance en ce qui a trait à la classification automatique. Ce modèle permet de calculer la probabilité d'appartenance à la catégorie k ,

$\Pr(Y = k|X = x)$, comme suit :

$$P(Y = k|x) = \frac{e^{\beta x}}{1 + e^{\beta x}}$$

Où β est le vecteur de coefficients de régression qui doivent être estimés avec des exemples d'entraînement en utilisant, par exemple, la méthode de moindres carrés.

La catégorie assignée sera celle dont la probabilité est la plus grande, c'est-à-dire :

$$\hat{y} = \arg \max_{k \in \{1, \dots, k\}} P(Y = k|x)$$

La régression logistique a bénéficié de beaucoup de travail de recherches et est devenue un modèle pratique dans de nombreux systèmes commerciaux à grande échelle, en particulier aux grandes sociétés d'Internet comme Google et Yahoo qui l'emploient pour apprendre de grands ensembles de données.

Il existe d'autre type de régression comme la régression Circulaire et la régression non paramétrique.

6.1.2. Classification :

La classification est une discipline qui se base sur plusieurs domaines en particuliers les mathématiques et l'informatique, elle est connue aussi sous

noms variés selon les objets qu'elle traite et les objectifs qu'elle vise à atteindre.

Principalement, elle classe chaque objets (ou individus) dans une classe prédéfinis. La Fonction de prédiction est appelée **Classifieur**. Dans la classification on trouve plusieurs algorithmes :

6.1.2.1. SVM (Support Vector Machine) :

Est une technique d'apprentissage supervisé elle est développée dans les années 1990 par Vladimir Vapnik .

Principe :

Le SVM peut être utilisé pour la classification c'est-à-dire que quelle échantillon appartient à quelle classe. La résolution de ce problèmes passe par la construction d'une fonction h qui a un vecteur d'entrée x fait correspondre une sortie y :

$$y=h(x)$$

Dans le SVM il y a deux cas:

SVM Linéairement séparable :

C'est la recherche d'hyperplans (droite dans le cas de deux dimension) qui sépare le mieux . Elle est Obtenue par la combinaison linéaire de vecteur d'entrée $x = (x_1, \dots, x_n)^T$ est un vecteur de poids $w = (w_1, \dots, w_n)^T$

$$h(x) = w^T x + b$$

Où w est un vecteur de m dimensions et b est un terme. La fonction de décision, pour un

Exemple x est exprimé comme suit :

$$\begin{cases} \text{classe} = 1 & \text{si } h(x) > 0 \\ \text{classe} = -1 & \text{si } h(x) < 0 \end{cases}$$

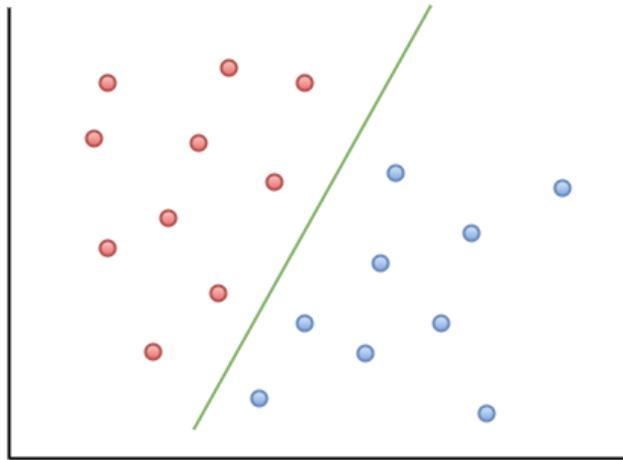


Figure 13. SVM linéairement séparable.

Le choix de l'hyperplan séparateur est très difficile par ce qu'il y a un nombre infini d'hyperplans séparateur pour ce fait le meilleur hyperplan est celui qui maximise la marge

L'hyperplans séparateur peut être trouver en résolvent l'équation suivante :

$$\left\{ \begin{array}{l} \text{Maximiser} \\ \text{sous constraint :} \\ y_i(w^T x_i + b) \geq 1 \quad \forall i = 1 \dots n \end{array} \right. \quad \frac{2}{\|w\|^2}$$

Mais aussi il y a les vecteur support qui sont les vecteurs vérifient l'équation suivante:

$$y_i(w^T x_i + b) = 1$$

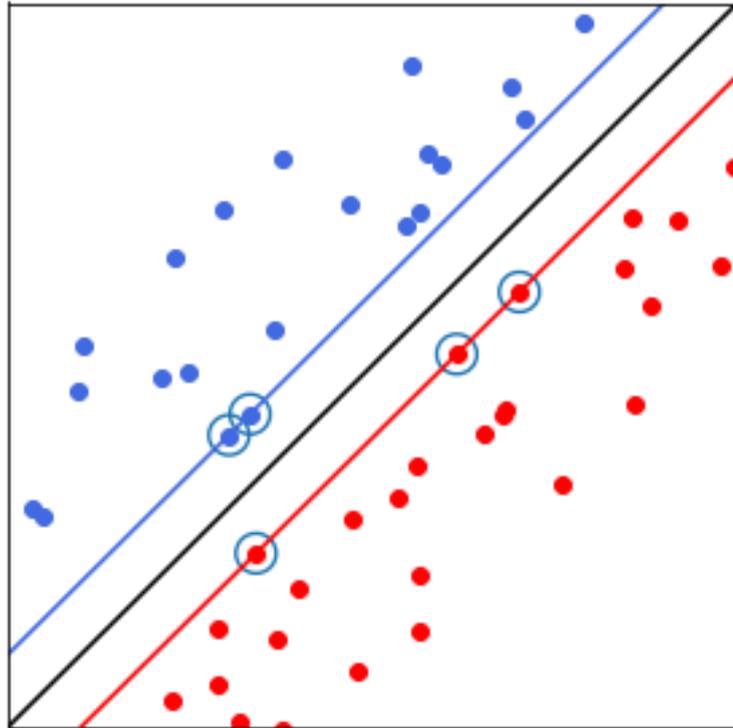


Figure 14. Les Vecteur support (points encerclé)

SVM linéairement non séparable :

Parfois un hyperplan n'existe pas toujours donc il est linéairement non séparable alors on utilise l'astuce du noyau (kernel trick) mais la détermination d'une telle fonction reste difficile voire impossible du coup on transforme l'espace de représentation des données d'entrées en un espace de plus grande dimension, où un classifieur linéaire peut être utilisé.

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \phi(x_i), \phi(x_j) \rangle$$

Où $\langle \phi(x_i), \phi(x_j) \rangle$ est le produit scalaire des deux images des vecteurs x_i et x_j dans le nouvel espace et dont le résultat est un scalaire. au lieu de calculer $\phi(x_i), \phi(x_j)$ et leur produit scalaire, on calcule plutôt une fonction

$K(x_i, x_j)$ Qui représente à la fois les deux transformations (qui peuvent être inconnues) alors la fonction objective peut-être calculer comme suite :

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

Et la fonction de décision est :

$$h(x) = \sum_{i \in S} \alpha_i y_i K(x_i, x) + b$$

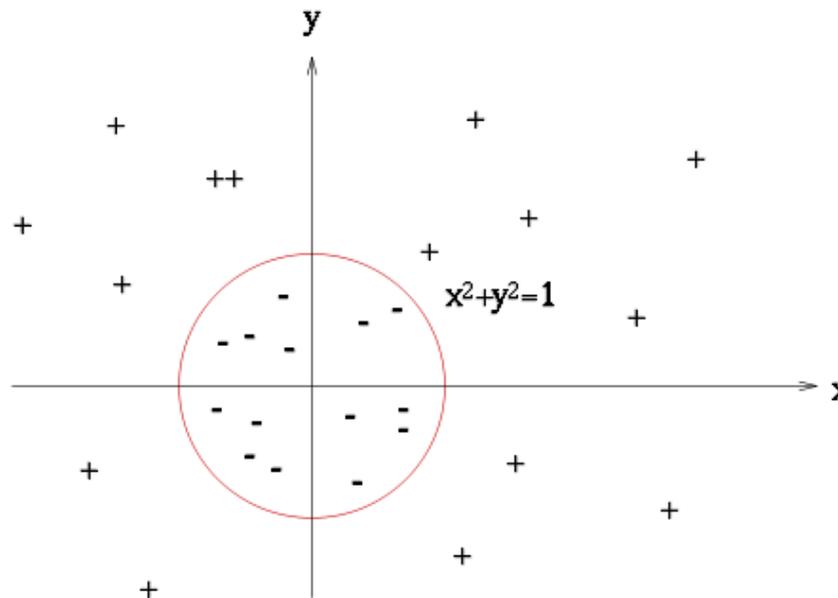


Figure 15. Noyau

Type de Noyau :

Noyau linéaire :

$$K(x_i, x_j) = x_i^T x_j$$

Noyau polynomial : élève la puissance a d :

$$K(x_i, x_j) = (x_i^T x_j)^d$$

6.1.2.2. Classification naïve bayésiennes :

La Classification naïve bayésiennes est un type de classification probabiliste simple basé sur le théorème de Bayes avec une forte indépendance des hypothèses, chaque hypothèse on associe une probabilité (qu'elle soit la solution)

Théorème de Bayes :

$$\left\{ \begin{array}{l} P(A|B) = P(A \cap B)/P(B) \\ P(A \cap B) = P(B|A) \cdot P(A) \\ \text{Donc } P(A|B) = P(B|A) \cdot P(A)/P(B) \end{array} \right.$$

Dans la classification on cherche la probabilité qu'un objet appartient à une classe i sachant qu'il y a de propriété x et y [19]:

$$P_{x,y}(C_i)$$

6.1.2.3. Arbre de Décision :

Un arbre de décision est un algorithme utilisé dans la classification qui représente un ensemble de décisions sous la forme d'un arbre où chaque nœud final représente une décision et chaque nœud interne représente un test les branches représentent le résultat des tests.

N°	Ensoleillement	Température	Humidité	Vent	Jouer
1	Soleil	75	70	Oui	Oui
2	Soleil	80	90	Oui	Non
3	Soleil	85	85	Non	Non
4	Soleil	72	95	Non	Non
5	Soleil	69	70	Non	Oui
6	Couvert	72	90	Oui	Oui
7	Couvert	83	78	Non	Oui
8	Couvert	64	65	Oui	Oui
9	Couvert	81	75	Non	Oui
10	Pluie	71	80	Oui	Non
11	Pluie	65	70	Oui	Non
12	Pluie	75	80	Non	Oui
13	Pluie	68	80	Non	Oui
14	Pluie	70	96	Non	Oui

Tableau 1. BD pour jouer au golf.

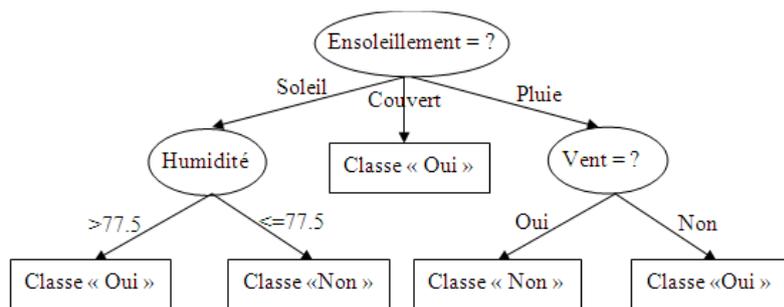


Figure 16. Arbre de Décision pour la base de données pour jouer au golfe

Avantages:

- Lisibilité.
- Capacité à sélectionner automatiquement les variables.
- Robuste au bruit et aux valeurs manquantes.
- Classification rapide (parcours d'un chemin dans un arbre)

Inconvénients :

- Sensibles au nombre de classes.
- Evolutivité dans le temps (le cas où les données évoluent dans le temps il faut refaire la phase d'apprentissage)

6.1.2.4. K-plus proche voisin (K-PPV):

C'est une méthode qui classe les exemples non étiquetés sur la base de leur similarité avec les exemples de la base d'apprentissage

La distance est une fonction respectant les 3 propriétés suivantes[18] :

7. Symétrie : $\forall a, b \in \mathbb{R}^d, dist(a, b) = dist(b, a)$
8. Séparabilité : $\forall a, b \in \mathbb{R}^d, dist(a, b) = 0 \leftrightarrow a = b$
9. Inégalité triangulaire : $\forall a, b, c \in \mathbb{R}^d, dist(a, c) = dist(a, b) + dist(b, c)$

Distance :

$$\sqrt{\sum_{i=1}^d (a_i - b_i)^2}$$

Avec : d vecteur de taille réel

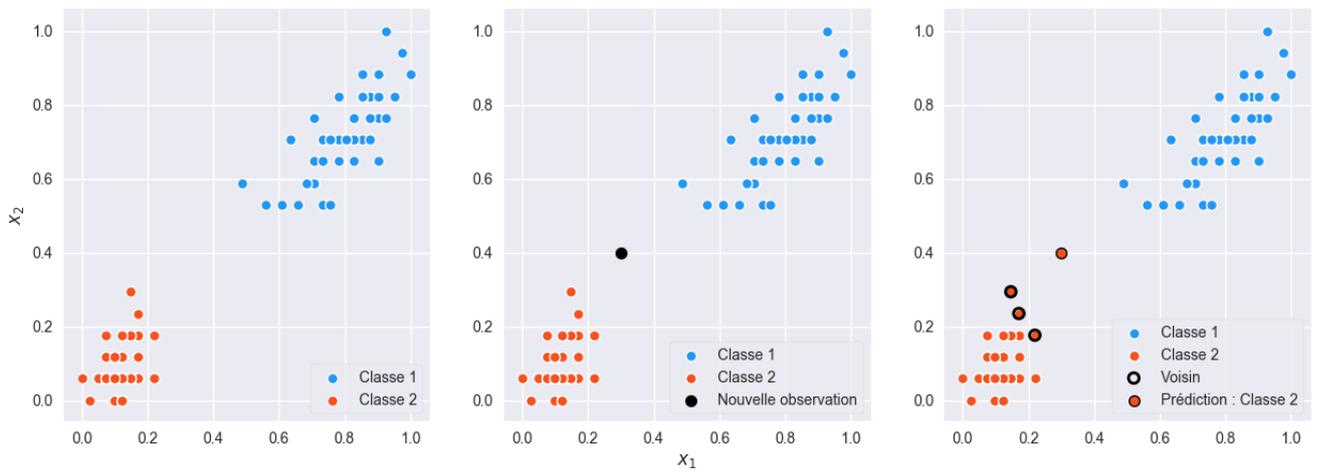


Figure 17. Exemple K-PPV

6.1.2.5. Réseaux de neurones artificiels (RNA) :

Un réseau de neurone artificiel est un modèle qui s'inspire du fonctionnement des neurones du cerveau humain.

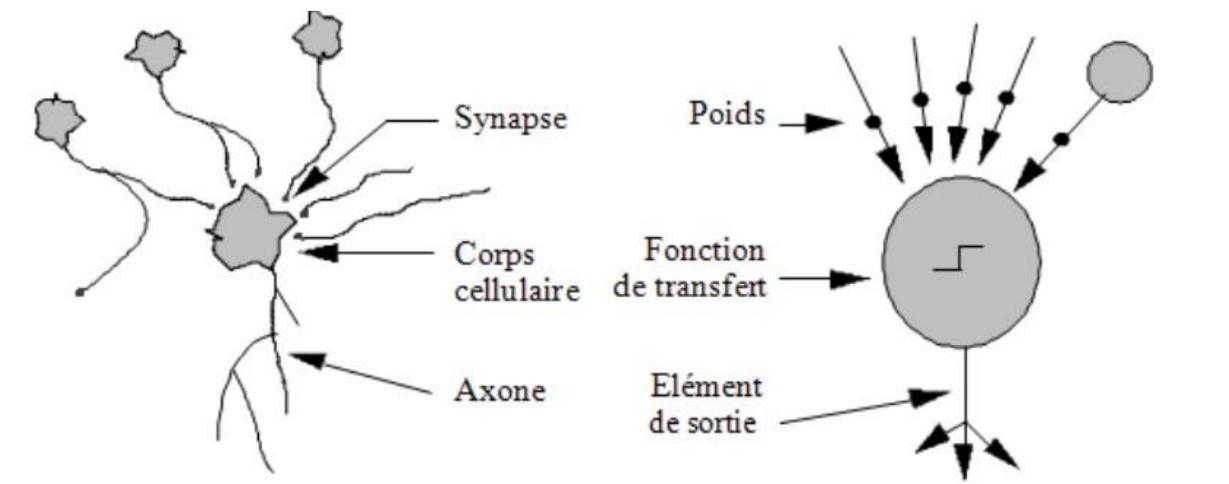


Figure 18. Transformation d'un neurone biologique en formels [17].

Un neurone artificiel (figure 17). reçoit en entrée des entrées numériques x_i est le nombre de données que va recevoir le neurone valorisée chacune par un coefficient w_i .

Le neurone artificiel effectue alors une somme pondérée de ses entrées et lui ajoute un coefficient w_0 (Figure 18).

$$S = \sum_{i=1}^n w_i \cdot x_i - w_0$$

La valeur de la fonction d'activation est la sortie y du neurone de sortie :

$$y = f\left(\sum_{i=1}^n w_i \cdot x_i - w_0\right)$$

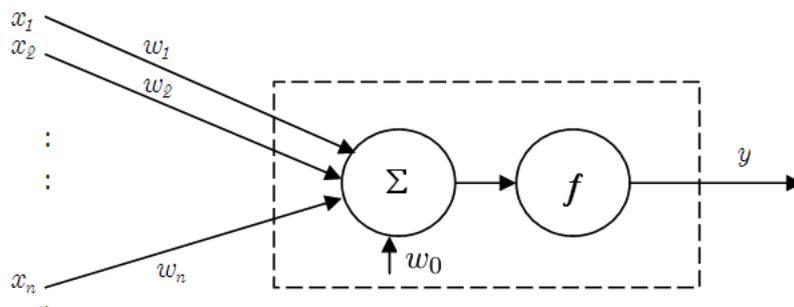


Figure 19. Neurone Simple.

Fonction d'activation :

Fonction Heaviside :

$$\begin{cases} f(x) = 1 \text{ si } x \geq 0 \\ f(x) = 0 \text{ si } x < 0 \end{cases}$$

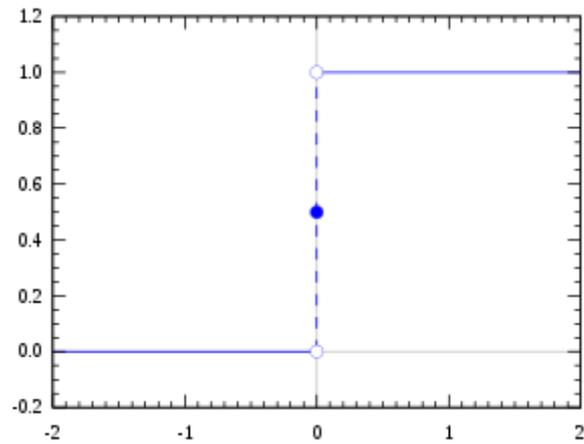


Figure 20. Fonction Heaviside.

Fonction Sigmoide :

$$f(x) = \frac{1}{1 + e^{-x}}$$

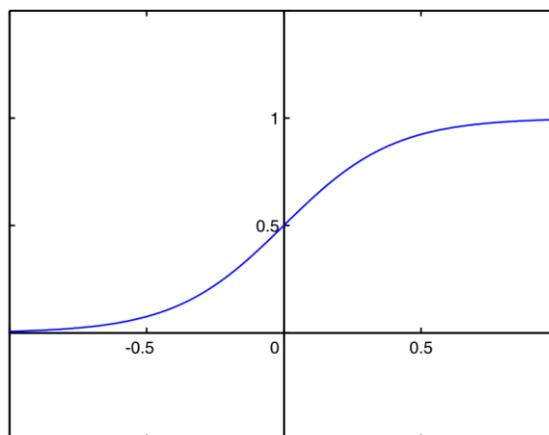


Figure 21. Fonction Sigmoide.

Fonction tangente hyperbolique :

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = 1 - \frac{2}{e^{2x} + 1}$$

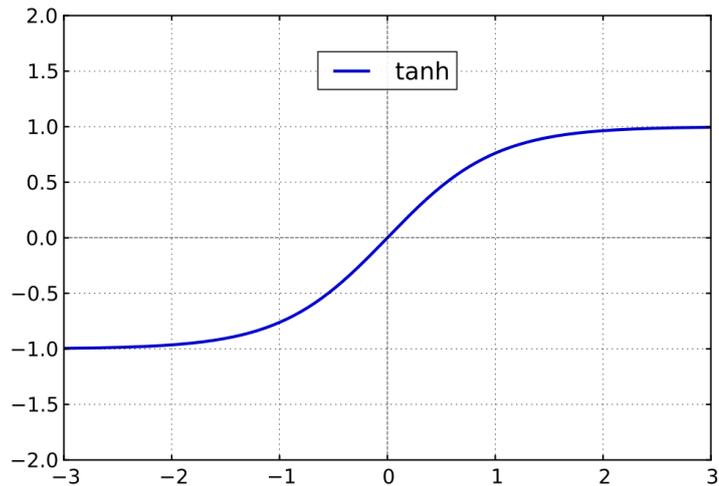


Figure 22. Fonction Tangent Hyperbolique.

6.2. Apprentissage Non Supervisé :

Dans l'apprentissage non supervisé, les données fournies au système ne sont ni étiquetées, ni classées, et les algorithmes du système traitent les données sans aucun entraînement préalable. La sortie dépend des algorithmes

7.2.1. Clustering :

Elle vise à diviser un ensemble de données en différents groupes homogènes

De manière générale, l'algorithme cherche à maximiser d'une part l'homogénéité des données au sein des groupes de données et à former des groupes aussi distincts que possible : selon le contexte, on choisit d'utiliser tel ou tel algorithme pour classer les données par exemple selon leur densité ou leur gradient de densité.

Le clustering prend en entrant un ensemble de données D et une mesure de similarité a et a comme sortie des clusters (groupe) qui ont une similarité communes. la similarité se calcule par la distance entre les données.

Type de Distance :

Distance euclidienne :

$$D_n(x_i, x_j) = \sqrt{\sum_{k=1}^{n_n} (x_{ik} - x_{jk})^2}$$

Distance City blocs :

$$D_n(x_i, x_j) = \sum_{k=1}^{n_n} |x_{ik} - x_{jk}|$$

Distance Minkowski :

$$D_n(x_i, x_j) = \left(\sum_{k=1}^{n_n} (x_{ik} - x_{jk})^2 \right)^2$$

Algorithme utiliser dans le Clustering :

- **K-Means :**

Est une méthode de clustering qui fait diviser les données en k clusters la distance d'un point à la moyenne des points de son cluster.

On a un ensemble de points (x_1, x_2, \dots, x_k) est on partitionner les n points en k ensemble $S = \{S_1, S_2, \dots, S_k\}$ en minimisant les distance entre les points :

$$\operatorname{argmin}_S \sum_{i=1}^k \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

μ_i est le centre de gravité du clustre

- **Regroupement Hiérarchiques :**

Les clusters prennent un ensemble de données comme entrées et donnent en retour une arborescence de clusters, et on peut trouver deux types de méthodes :

Les algorithmes divisibles : commencent à partir d'un ensemble de données et le subdivisent en sous-ensembles puis subdivisent chaque sous-ensemble en d'autres plus petits, et ainsi de suite, pour générer une séquence de clusters ordonnée du plus général au plus fin.

Les algorithmes agglomératifs qui considèrent chaque enregistrement comme étant un cluster indépendant puis rassemblent les plus proches en des clusters plus importants, et ainsi de suite jusqu'à atteindre un seul cluster contenant toutes les données.

Exemple : soit les données suivantes :

$$X_1(0,2), X_2(0,0), X_3(1.5,0), X_4(5,0), X_5(5,2)$$

On utilise la distance euclidienne :

$$D(X_1, X_2) = \sqrt{(0-0)^2 + (2-0)^2} = 2$$

$$D(X_1, X_3) = 2,5$$

....

$$D(X_2, X_3) = 1,5$$

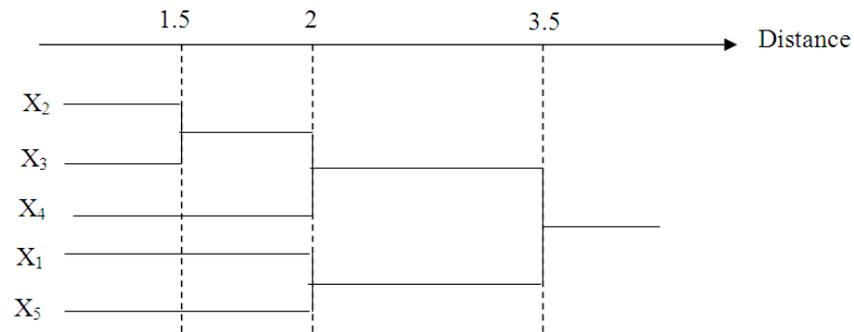


Figure 23. Graphe Clustering Hiérarchique.

- **Maximisation de l'Espérance (EM) :**

Est un algorithme itératif qui permet de trouver les paramètres du maximum de vraisemblance.

Cette algorithme calcule l'espérance d'un point ensuite maximise cette espérance de vraisemblance.

O a un ensemble d'individus $x = (x_1, x_2, \dots, x_n)$ suivant un phénomène f

$f(x_i, \theta)$ paramètre par θ on chercha trouver θ toute en maximisant la log variance :

$$L(x; \theta) = \sum_{i=1}^n \log f(x_i, \theta)$$

8. Conclusion :

Le Machine Learning est un domaine qui nous présente des techniques variés et efficaces pour la classification, le clustering et la prédiction.

Dans notre travail, on s'intéresse à l'application des techniques du machine learning dans le domaine médical et spécialement le diagnostic pour le cancer de la peau.

Cette solution va aider les oncologues et les pathologistes à prendre une décision fiable concernant le type du cancer. Cette étape est très importante pour la thérapie et le traitement.

Chapitre 3 :_Conception

1. Introduction :

Notre objectif est de présenter une solution adaptée basée sur l'Intelligence Artificielle pour l'aide au diagnostic du cancer de la peau. Cette opération est nécessaire pour pouvoir prescrire le bon traitement.

Pour ce faire nous allons utiliser une base de données réelle pour des malades atteints du cancer de la peau.

2. Schémas Synoptique :

La figure 24 représente un schéma synoptique du processus d'aide au diagnostic du cancer de la peau.

Ce processus est constitué de 3 phase : diagnostic puis traitement et la convalescence(suivi).

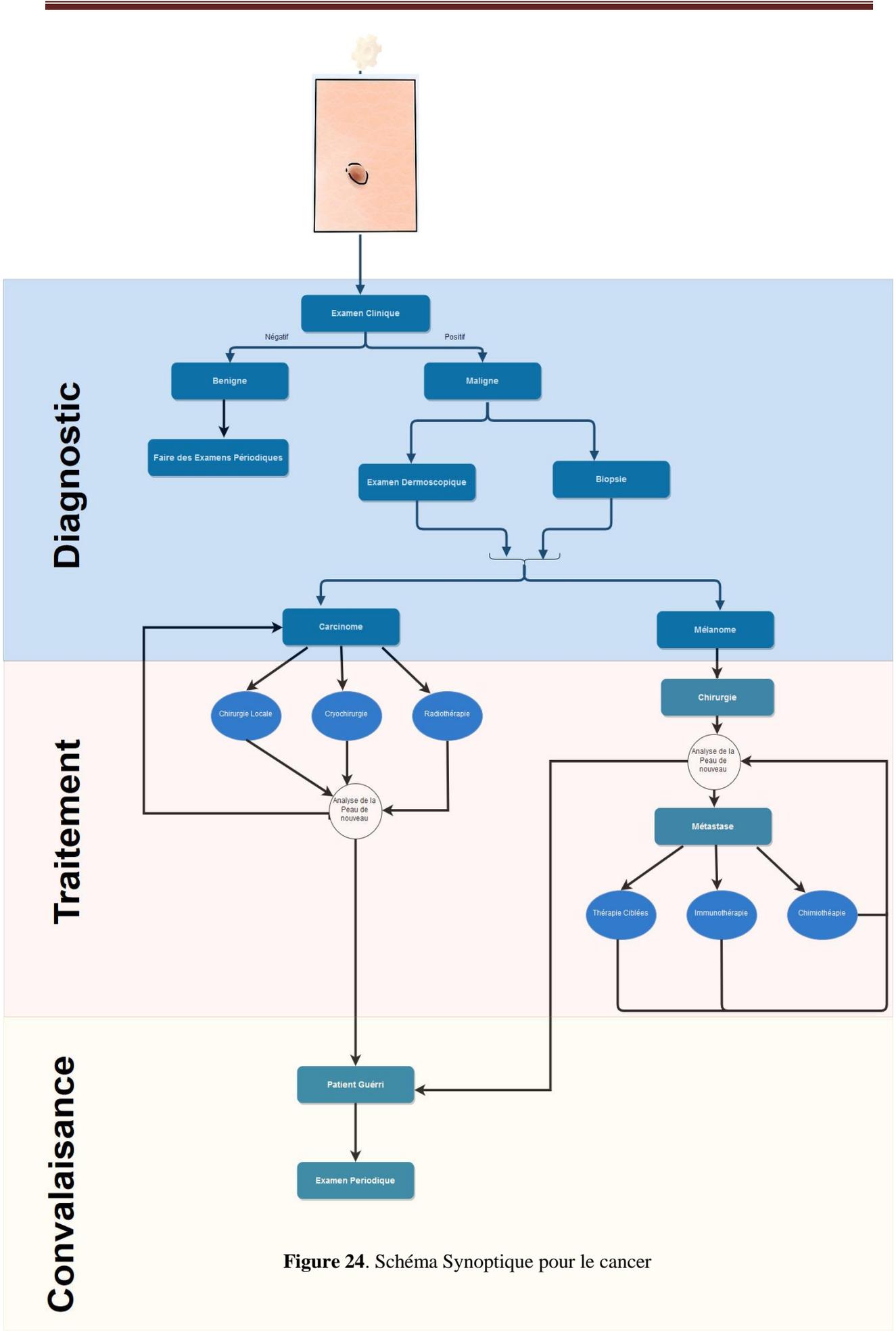


Figure 24. Schéma Synoptique pour le cancer

3. Objectif :

Réaliser une solution efficace pour le diagnostic automatique du cancer en utilisant les techniques du Machine Learning beaucoup de travaux ont traité le problème du cancer de la peau mais notre spécificité pour cette étude est basée sur des données réelles en provenance du département d'oncologie de l'hôpital de Franche-Comté.

Cette application sera testée dans ce service pour avoir un retour réel et concret sur notre travail.

Notre application a été précédé par une autre application pour la gestion du fichier patient.

4. Description de la Base de Données :

Nous possédons une base de données sur le nævus de Spitz sous le format Excel qui se compose de 54 patients (enregistrements) et qui est composé en 37 caractéristiques.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
	Numéro c	Numéro li	NOM	Sexe	date du diag	Age du di	Colonne1	diamètre	localisatic	forme	épaisseur	BRAF	ALK IH	ALK FISH	p16	Ly	pig mél	Ki-67	Ki67>1%	desmo	ind
1	1 00.108417			M	07/09/2001	33		0,5 tronc	jonctionn		0,1 muté	negatif			pas de pei2+	3+			2 oui	non	
2	2 07B08580			M	28/08/2007	33		1,2 tronc	dermique		2 non muté	negatif			pas de pei non		1+		0 non	oui	
3	3 06M11184			F	24/10/2006	39		0,6 MI	jonctionn		0,1 non muté	negatif			pas de pei2+	3+			2 oui	non	
4	4 09M03245			F	20/03/2009	41		0,6 MI	composé		0,4 non muté	negatif			pas de pei1+	3+			5 oui	non	
5	5 10M01320			F	02/02/2010	24		0,3 MS	composé		2,2 non muté	positif			pas de pei	0 1+			10 oui	non	
6	6 00.11480			F	14/11/2000	32		0,5 MI	composé		1 non muté	negatif			pas de pei1+	1+			2 oui	non	
7	7 09M13979			F	04/01/2010	46		0,3 MI	jonctionn		0,1 non muté	negatif			pas de pei1+	3+			2 oui	non	
8	8 09M08280			F	22/07/2009	18		0,6 abdo	composé		0,4 muté	negatif			pas de pei2+	2+			2 oui	non	
9	9 02M09705			F	08/10/2002	39		0,8 tronc	composé		0,3 muté	negatif			pas de pei3+	2+			3 oui	non	
10	10 15H00915			F	21/01/2015	33		0,6 MI	composé		0,9 non muté	negatif			pas de pei1+	2+			7 oui	non	
11	11 08B11870			F	04/12/2008	23		0,7 TC	dermique		2 non muté	negatif			pas de pei	0	0		0 non	non	
12	12 000.6892			F	30/06/2000	37		0,4 MI	composé		0,5 non muté	negatif			pas de pei2+	3+			5 oui	non	
13	13 15H11061			F	21/09/2015	20		0,6 MS	jonctionn		0,5 non muté	negatif			pas de pei2+	3+			5 oui	non	
14	14 0.102703			F	16/03/2001	20		0,8 MS	composé		0,6 non muté	negatif			pas de pei3+	2+			6 oui	non	
15	15 02B05162			F	04/06/2002	25		0,6 MS	dermique		2 non muté	negatif			pas de pei2+	1+			1 non	non	
16	16 00.12909			F	22/12/2000	52		0,4 MS	composé		0,8 non muté	negatif			pas de pei1+	1+			0 non	oui	
17	17 03M04763			M	16/05/2003	54		0,7 abdo	composé		0,8 non muté	negatif			pas de pei3+	3+			6 oui	non	
18	18 07B08752			F	31/08/2007	14		0,8 TC	dermique		1,3 non muté	positif			pas de pei	0	0		1 non	non	
19	19 03B00994			F	29/01/2003	7		0,3 tronc	composé		0,4 non muté	negatif			pas de pei	0 3+			2 oui	non	
20	20 05B01352			M	10/02/2005	17		0,8 TC	composé		0,8 muté	negatif			pas de pei3+	2+			4 oui	non	
21	21 09M12088			F	02/11/2009	16		0,5 tronc	jonctionn		0,5 muté	negatif			pas de pei2+	3+			0 non	non	
22	22 11B12456			M	01/12/2011	3		0,4 TC	composé		1 muté	negatif			pas de pei1+	1+			1 non	non	

Figure 25. Base de données du cancer de la peau

Numéro : Numéro d'ordre du patient.

Numéro lame : Numéro de la lame qui vas être analysée.

Nom : Nom du patient non communiqué pour des raisons de confidentialité.

Sexe : Sexe du patient (Masculin ou Féminin).

Date du Diagnostic : La date du diagnostic de la maladie

Age du Diagnostic : Age au diagnostic

Diamètre : Le diamètre de la tumeur.

Localisation : Localisation de la tumeur dans le corps.

Forme : La forme de la tumeur (Jonctionnel,dermique,composé).

Epaisseur :

BRAF : Est un gène humain situé sur le chromosome 7 ,ça mutation entraine des cancer qui peuvent toucher l'homme.

ALK IH : Le gène de la lymphome kinase anaplasique (ALK) joue un rôle physiologique important dans le développement du cerveau et peut être altéré de manière oncogène lors de plusieurs affections malignes.

P16 : Il est une protéine supresseurs de tumeur.

Ly : Le lymphome est un cancer qui infectent les cellules du système immunitaire qui combattent les infections, appelées lymphocytes.

Pig mel : Pigmentation du mélanome.

Ki-67 : est un antigènes qui marque la prolifération

Ki-67>1% : si Ki-67 est supérieure a 1%.

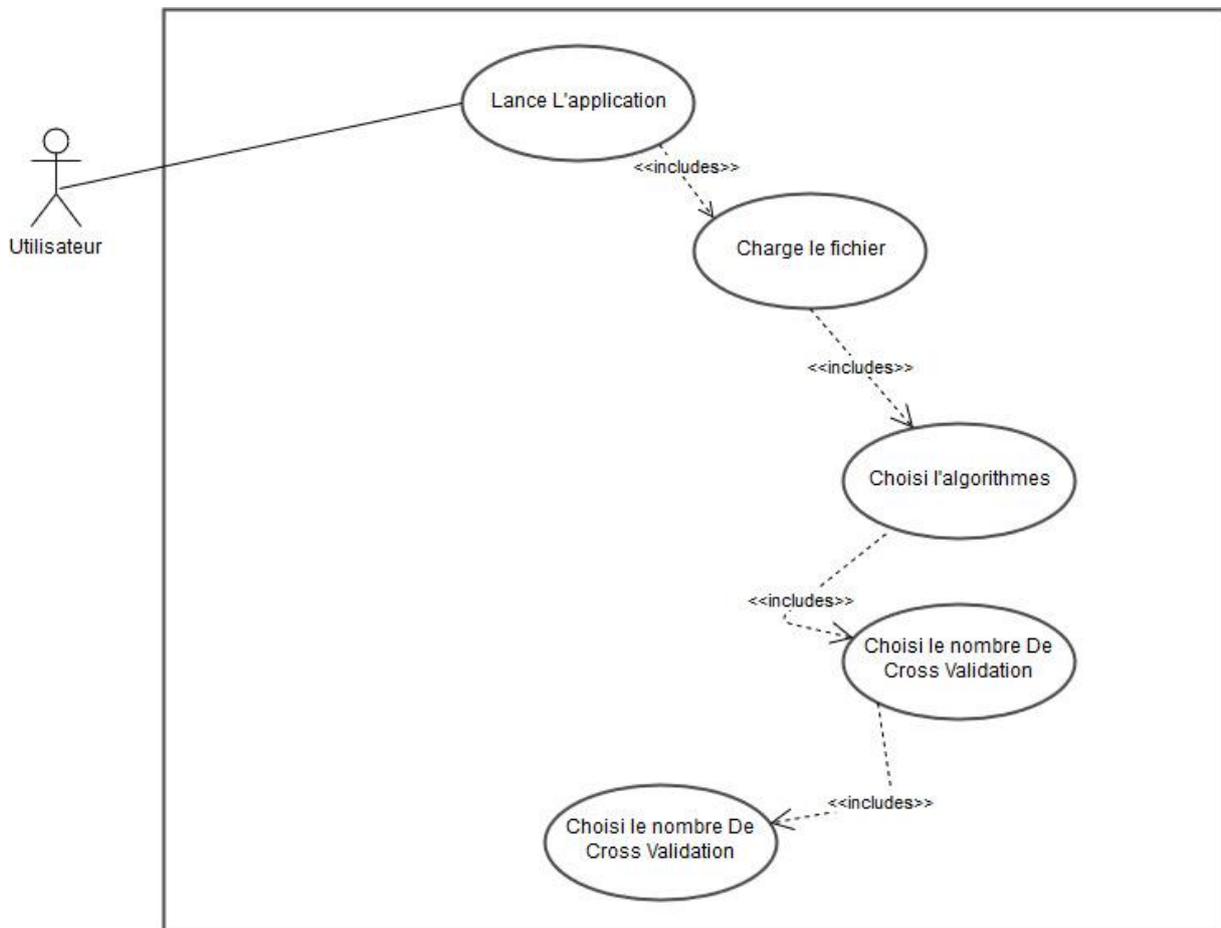
Desmo : désmoplastique.

Mais aussi des données cliniques :

*Index mitotique,Mitoses profondes,Mitoses atypiques,Atypies
CN ,hyperCELL ,kamino ,hypoderme ,Asymetries ,Lim Floues ,Migration ,
Diametre>1 ,Th irreg ,Pas de grad, Grenz, Ulceration, Modif epid ,Diagnostic :*

5. Diagramme de Cas d'utilisation :

Le diagramme suivant présente le diagramme de cas d'utilisation de notre application.



6. Prétraitement :

Dans cette partie nous expliquons comment avoir un fichier, qui pourras plus tard être utilisé dans la partie traitement. Les étapes sont :

Normalisation :

La normalisation permet de faire entrer les valeurs de donnée dans un intervalle il existe plusieurs méthodes [16] :

Min-Max :

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Z-Score :

$$X' = \frac{x - avg_A}{\sigma_A}$$

Données Non Pertinentes :

Enlever les caractéristiques qui ne s'ont pas intéressante ou qui n'ont pas d'influence sur la prédiction, ces données peuvent avoir tous la même valeur ou : date du diagnostic, P16(tous 'pas de perte'), mitoses atypique (tous 'non')

Valeurs Manquante :

Il existe plusieurs technique pour remédier aux valeurs manquante :

Imputation Simple :

Produit une valeur artificielle pour remplacer la valeur Non Attribué (NA).

Hot-Deck :

La valeur manquante est remplacée par une valeur observée chez un individu ayant les mêmes caractéristiques

Cold-Deck :

La valeur manquante est remplacée par une valeur observée chez un individu ayant les mêmes caractéristiques, mais provenant d'une autre source d'information.

Moyennes :

Calcule simplement la moyenne des valeurs observées pour cette variable pour tous les individus non manquants.

Régression :

La valeur prédite obtenue en faisant régresser la variable manquante sur d'autres variables.

Conversion de données :

C'est de convertir les données catégorielle en données numérique comme , nous l'avons fait dans notre base de données du cancer de la peau :

Features	Avant	Aprés
Sexe	Male	0
	Feminin	1
Localisation	Trone	1
	MI	2
	MS	3

	Abdo	4
	TC	5
Forme	Jonctionnel	1
	Dermique	2
	Composée	3
BRAF	Mute	1
	Non Muté	2
ALK IH	Negative	1
	Positive	2
Ly	2+	2
	1+	1
	0	0
	3+	3
Pig Mel	3+	3
	2+	2
	1+	1
	0	0
KI67>1%	Oui	1
	Non	0
Desome	Oui	1
	Non	0
Diagnostic	NS	1
	TSA	0

Tableau2. Conversion de la base de donnée

7. Traitement :

Dans notre travaille nous avons choisi deux méthode de classification qui sont les arbres de décision et les forets aléatoire (Random Forest) :

7.1.Arbre de Décision :

Est une méthode de classification simple et compréhensible pour l'homme

Elle est utilisée dans la classification tous comme dans la régression.

Pour construire un arbre ils existent plusieurs algorithmes :ID3, CART, C4.5, CHAID...,etc. On commence généralement par le choix d'un attribut puis le choix d'un nombre de critères pour son nœud. On crée pour chaque critère un nœud concernant les données vérifiant ce critère. L'algorithme continue d'une façon récursive jusqu'à obtenir des nœuds concernant les données de chaque même classe

Les Arbres de Décision ont prouvé leur qualification au domaine d'apprentissage automatique. Leur

Succès est notamment dû à leur aptitude à traiter des problèmes complexes de classification.

En effet, ils offrent une représentation facile à comprendre et à interpréter ainsi qu'ils possèdent une capacité à produire des règles logiques de classification.

Leur rapidité en classement dépend tout simplement du nombre majeur des nœuds en partant de la racine aux feuilles

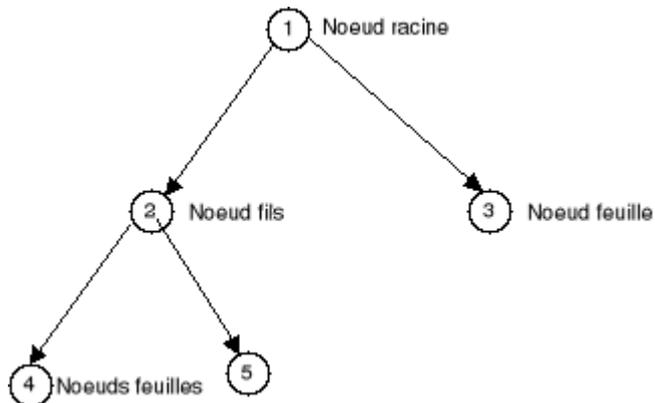


Figure 26. Arbre de Décision.

Algorithmes de Construction d'arbre :[14]

Algorithme 3 CONSTRUIRE-ARBRE(D : ensemble de données)

- 1: Créer nœud N
 - 2: **Si** tous les exemples de D sont de la même classe C alors
 - 3: Retourner N comme une feuille étiquetée par C ;
 - 4: **Si** la liste des attributs est vide alors
 - 5: Retourner N Comme une feuille étiquetée de la classe de la majorité dans D ;
 - 6: Sélectionner l'attribut A du meilleur Gain dans D ;
 - 7: Etiqueter N par l'attribut sélectionné;
 - 8: Liste d'attributs \leftarrow Liste d'attributs - A ;
 - 9: **Pour** chaque valeur V_i de A
 - 10: Soit D_i l'ensemble d'exemples de D ayant la valeur de $A = V_i$;
 - 11: Attacher à N le sous arbre généré par l'ensemble D_i et la liste d'attributs
 - 12: **FinPour**;
 - 13: **Fin**;
-

Choix de la variable de segmentation :

Gain d'information :

Appeler aussi l'entropie qui est utiliser par l'algorithme C4.5,ID3 , on sélectionne l'attribut qui a le gain le plus élevé.

L'entropie : soit un ensemble X d'exemple dont une partie y et une partie z :

$$H(X) = y \log_2(y) - z \log_2(z)$$

$$0 \leq H(X) \leq 1$$

$$\text{Gain}(X, a) = H(X) - \sum \frac{a_y}{|X|} H(X_{a=y})$$

Gini Index :

Soit une base de X exemple contient y classe :

$$Gini(X) = 1 - \sum_{i=1}^y p_i^2$$

p_i La fréquence de la classe i dans la base X .

Pour diviser un nœud on choisit le minimum :

$$Gini_{split}(X) = \frac{n_1}{N} Gini(X_1) + \frac{n_2}{N} Gini(X_2)$$

J48 :

Est une implémentation open source pour Java basé sur l'algorithme C4.5

7.2.Forets Aléatoire (Random Forest) :

Random forest (Figure25) est un algorithme appartient à la famille des agrégations de modèles, c'est en fait un cas particulier de bagging (bootstrap aggregating) appliqué aux arbres de décision de type CART.

Le principe des méthodes de Bagging, et donc en particulier des forêts aléatoires c'est de faire la moyenne des prévisions de plusieurs modèles indépendants pour réduire la variance et donc l'erreur de prévision. Pour construire ces différents modèles, on sélectionne plusieurs échantillons bootstrap, c'est à dire des tirages avec remises.

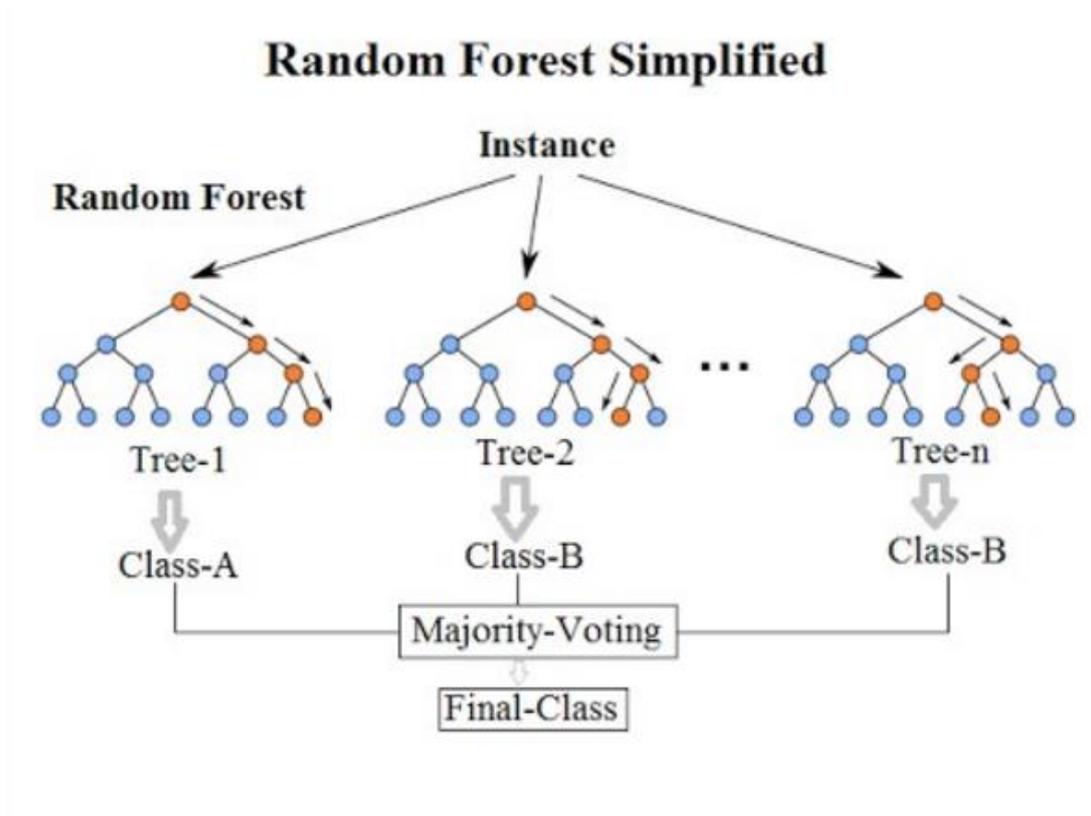


Figure 27. Graph simplifier de Random Forest

7.3.Evaluation des Performances :

1) Matrice de Confusion :

Une Confusion Matrix est un **résumé des résultats de prédictions sur un problème de classification**. Les prédictions correctes et incorrectes sont mises en lumière et réparties par classe. Les résultats sont ainsi comparés avec les valeurs réelles.

	Oui	non
Oui	Vrai Positif (TP)	Faux Négatif (FN)
NON	Faux Positif (FP)	Vrai Négatif (TN)

(True Positives) : les cas où la prédiction est positive, et où la valeur réelle est effectivement positive.

- **TN (True Negatives)** : les cas où la prédiction est négative, et où la valeur réelle est effectivement négative.
- **FP (False Positive)** : les cas où la prédiction est positive, mais où la valeur réelle est négative
- **FN (False Negative)** : les cas où la prédiction est négative, mais où la valeur réelle est positive.

Accuracy :

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

Précision :

$$Précision = \frac{TP}{TP + FP}$$

Rappel (Recall) :

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

F-mesure :

$$f - mesure = \frac{2 \times Précision \times Rappel}{Précision + Rappel}$$

Correctly Classified Instances :

Les instances qui sont correctement classifiées.

Incorrectly Classified Instances :

Les instances qui ne sont pas bien classifiées.

Kappa statistic :

Kappa statistique est :

$$\frac{\text{le taux de réussite du prédicteur réel} - \text{taux de réussite du prédicteur aléatoire}}{1 - \text{taux de réussite du prédicteur aléatoire}}$$

Mean absolute error :

$$\frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n}$$

Root mean squared error :

Valeurs cibles réelles:: $a_1 a_2 \dots a_n$

Valeurs cibles prédites : $p_1 p_2 \dots p_n$

$$\sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}}$$

Relative absolute error :

$$\frac{|p_1 - a_1| + \dots + |p_n - a_n|}{|a_1 - \bar{a}| + \dots + |a_n - \bar{a}|}$$

Root relative squared error :

\bar{a} la valeur moyenne sur les données d'entraînement

$$\sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{(a_1 - \bar{a})^2 + \dots + (a_n - \bar{a})^2}}$$

Total Number of Instances :

Le nombres totale d'instance utiliser dans l'apprentissage.

TP Rate :

C'est le pourcentage des Vrai Positif.

FP Rate :

C'est le pourcentage des Faux Positif.

MCC :

Est utilisé dans l'apprentissage automatique en tant que mesure de la qualité des classifications binaires (à deux classes). Il prend en compte les vrais et les faux positifs et négatifs et est généralement considéré comme une mesure équilibrée pouvant être utilisée même si les classes sont de tailles très différentes.

ROC Area (Receiver Operating Characteristics):

L'une des valeurs les plus importantes de Weka. Ils nous donnent une idée de la performance des classificateurs en général.

8. Conclusion :

Dans ce chapitre nous avons présenté le concept de notre solution et qui consiste à développer une application d'aide au diagnostic du cancer de la peau.

Dans ce qui suit, nous allons détailler la partie de l'implémentation et discuté les résultats obtenus.

Chapitre 4 : Implémentation et discussion

1. Introduction :

Dans ce chapitre nous présentons notre application et les outils de programmation utilisé pour son implémentation.

2. Outils de Programmation :

2.1.Langage de programmation :

Dans notre application nous avons utilisé le langage de programmation Java, ce choix est basé sur le fait que java est un bon langage pour la partie graphique mais aussi pour le machine learning avec Weka est que celle-ci est elle-même crée avec java.

2.2.Outils de Développement :

NetBeans est un environnement de développement intégré (EDI ou IDE - IntegratedDevelopment Environment) et il est pleinement fonctionnel pour les développeurs d'application Java.



Figure 28. Logo NetBeans.

2.3.Les Packages :

Les Package utiliser sont :

-Java Swing : package pour les interfaces graphiques de l'application avec une multitude de composant graphique.

-**Java Weka.jar** : c'est un fichier qui contient plusieurs classes prédéfinis, spécialement pour le machine learning, il permet de faire tous les type de méthodes d'apprentissage tel que :la classification, la régression, le clustering,et contient un multitude d'algorithmes comme les arbres de décision ,SVM, Réseaux de Neurons Artificiels,...,etc.

3. Implémentation :

Cette application est destinée pour un usage médicale. Ces utilisateurs peuvent être des médecins ou du personnel médical.

3.1. Les Interfaces :

Notre application se compose de 2 fenêtre principale :

Fenêtre d'accueil :

C'est fenêtre d'accueil de notre application

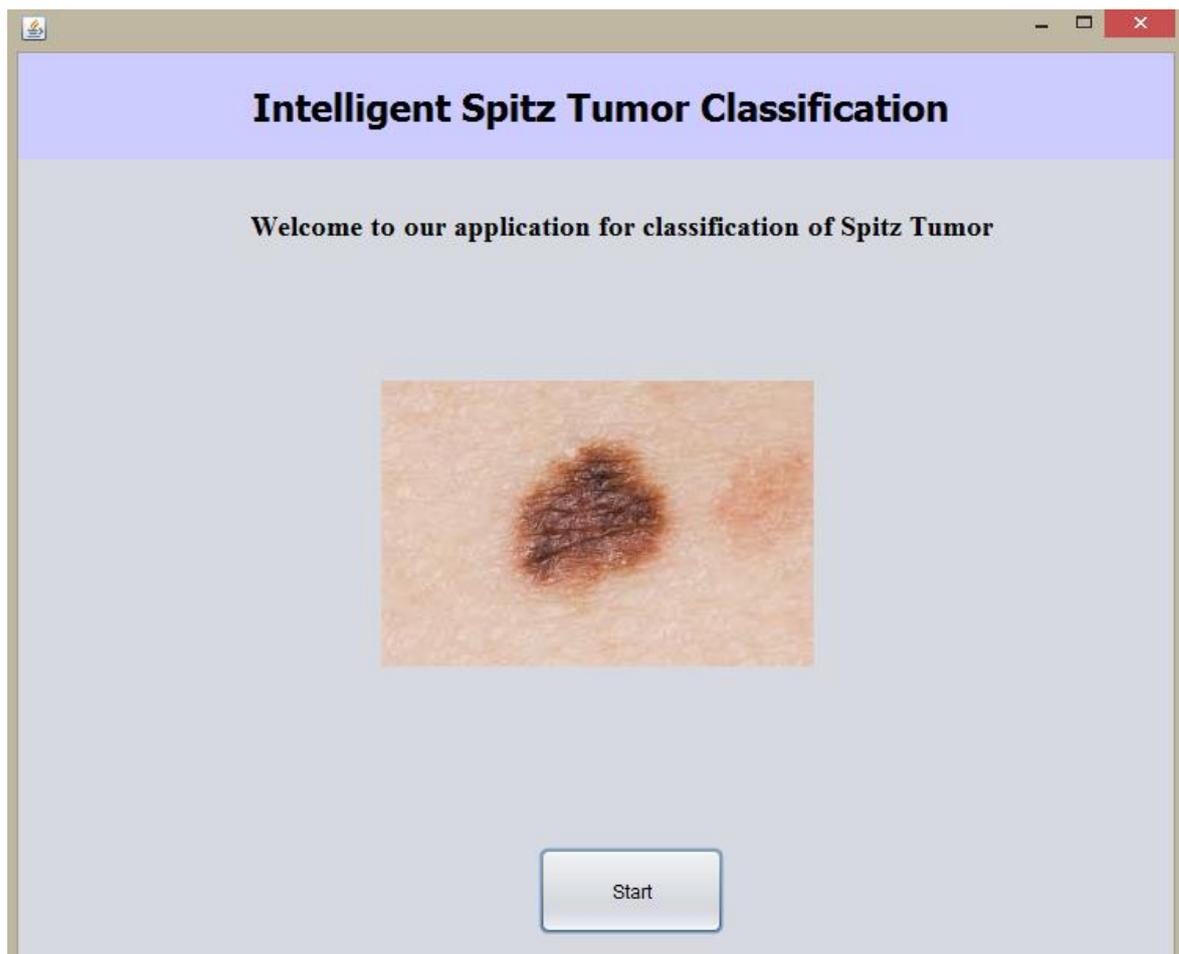


Figure 29. Fenêtre principale de l'application.

Fenêtre principale :

C'est la fenêtre principale de notre application là où on utilise l'une des méthodes de classification suivantes arbre de décision ou Random Forest.

On commence par charger le fichier Arff puis on choisit l'algorithmes voulus avec le nombre de cross-validation et on lance l'application.

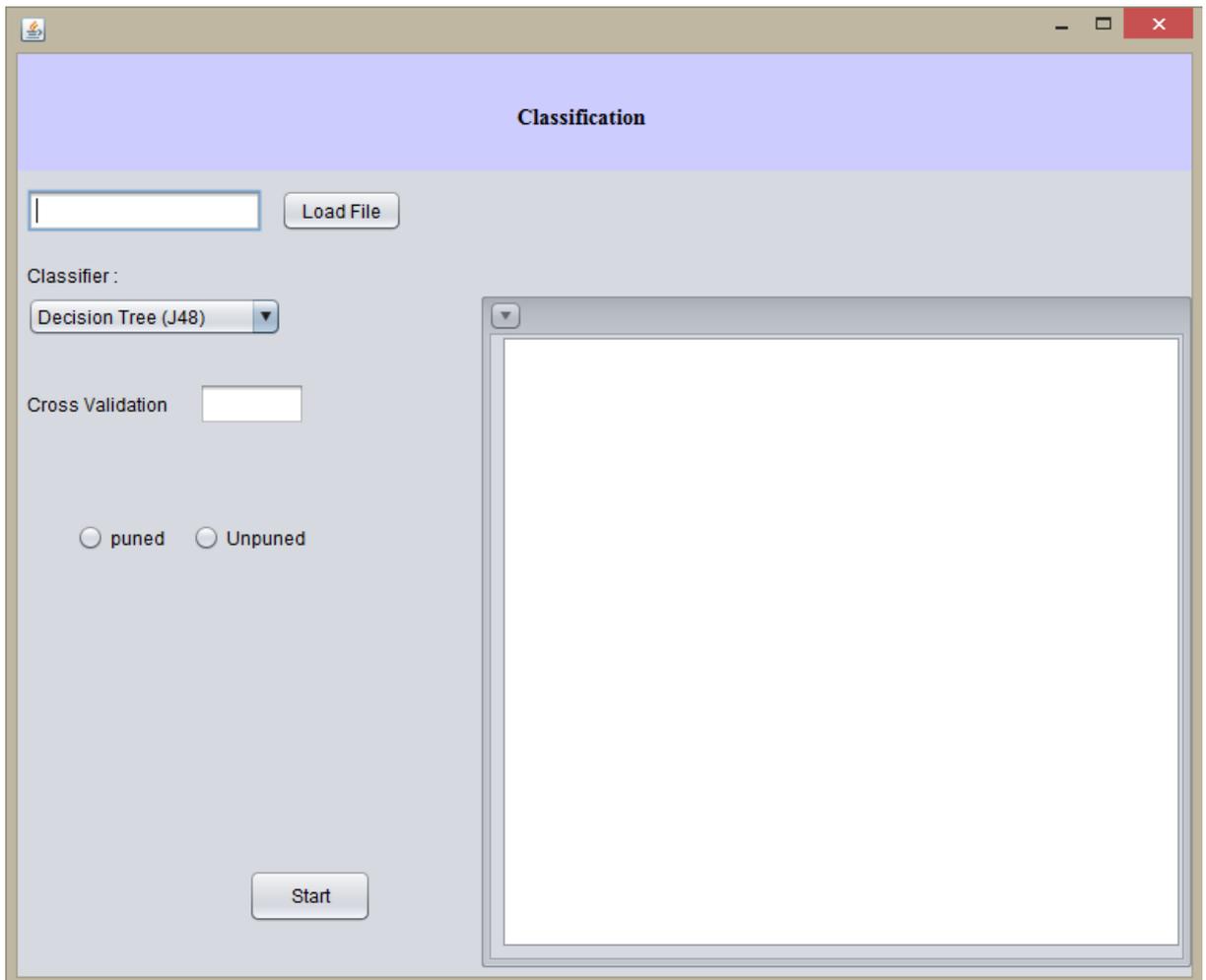


Figure 30. Fenêtre pour la classification.

4. Résultats de l'application :

4.1. Arbre de décision :

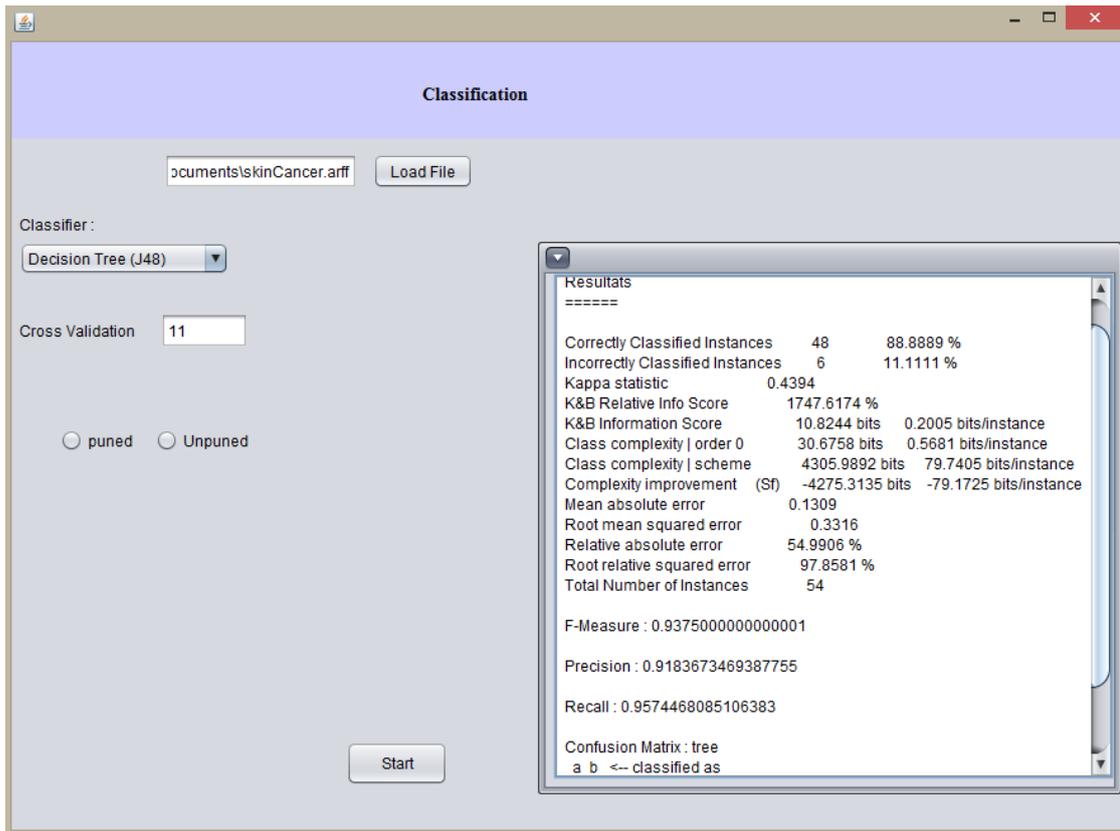


Figure 31. Résultat de la classification avec les arbres de décision.

Graphe de l'arbre de décision :

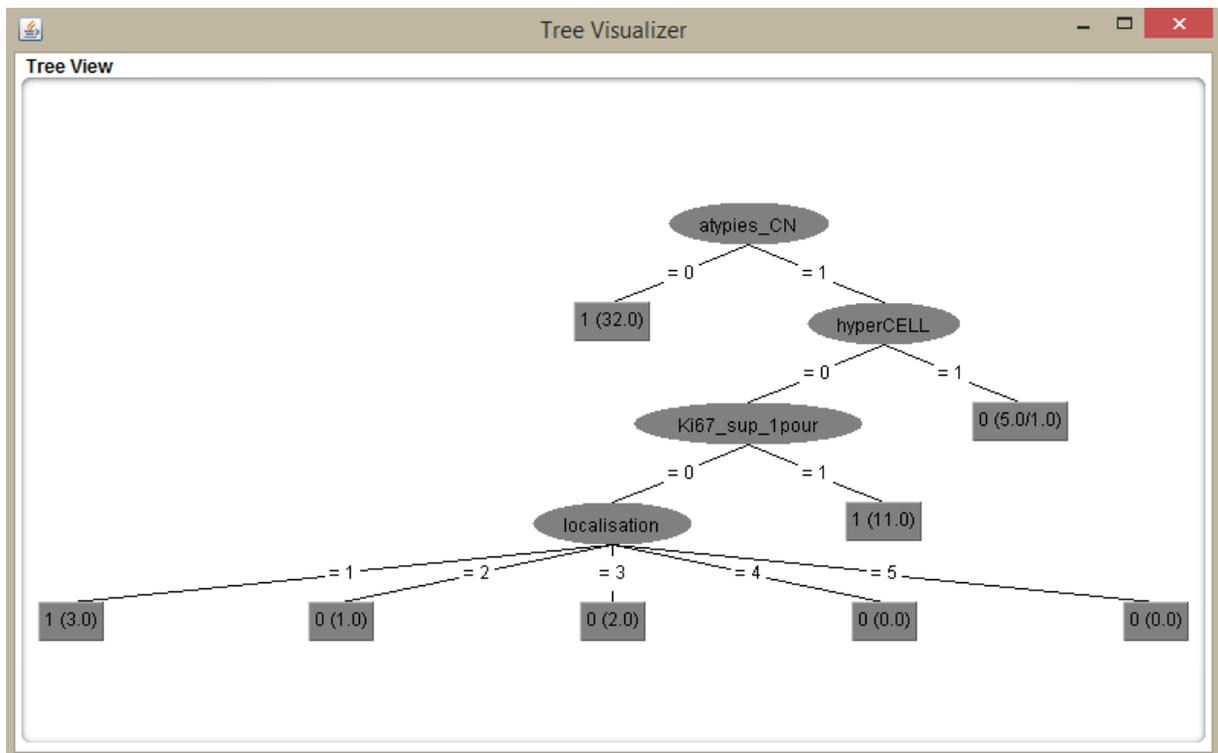


Figure 32. Graphe de l'arbre de décision obtenue.

Description des Résultats :

Correctly classified instances	Incorrectly classified instances	Accuracy	Precision	Recall	F-Measure	Kappa Statistic	Means Absolute error
48	6	0,87	0,92	0,95	0,93	0,44	0,13

Tableau 3. Résultat de l'arbre de décision.

Précision de l'arbre de décision :

Ce graphe représente la précision du modèle avec un changement du nombre de partitionnement (Cross-Validation) la précision maximale est de 88,88 avec 11 partitionnement.

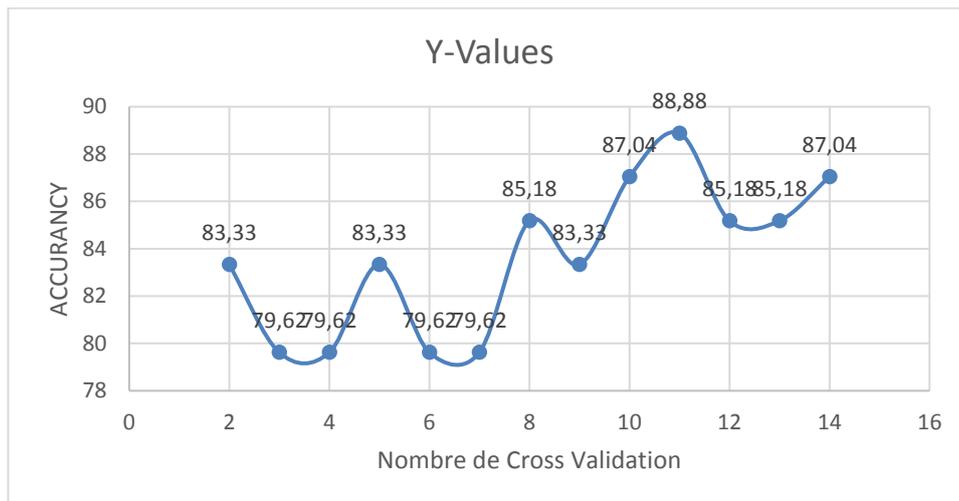


Figure 33. Graph décrivent l'accuracy par rapport au nombre de cross-validation

4.2. Random Forest :

Avantages :

- Bonnes performances en prédiction
- Paramétrage simple
- Pas de problème d'overfitting

- Mesure de l'importance des variables
- Possibilité de programmation parallèle

Inconvénients :

- Problème si nombre de variables pertinentes très faibles.
- Déploiement d'un tel modèle reste compliqué.

Résultat de Random Forest :

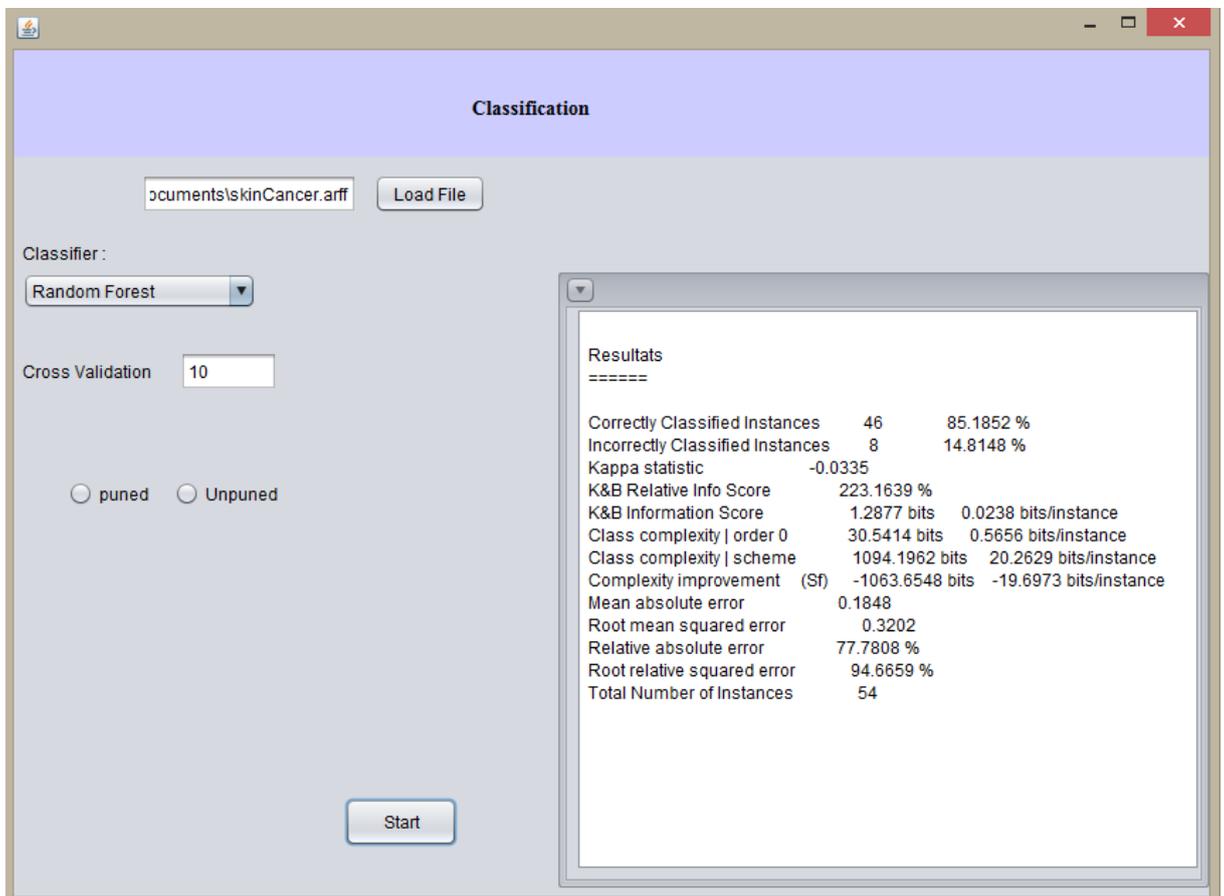


Figure 34. Résultat de Random Forest

Description des Résultats :

Correctly classified instances	Incorrectly classified instances	Accuracy	Precision	Recall	F-Measure	Kappa Statistic	Means Absolute error
47	7	0,87	0,87	1	0,93	0	0,19

Tableau 4. Résultat de Random Forest.

Précision de Random Forest :

Ce graphe représente la précision du modèle avec le changement du nombre de partitionnement (Cross-Validation) la précision maximale est de 88,88 avec 4 partitionnement

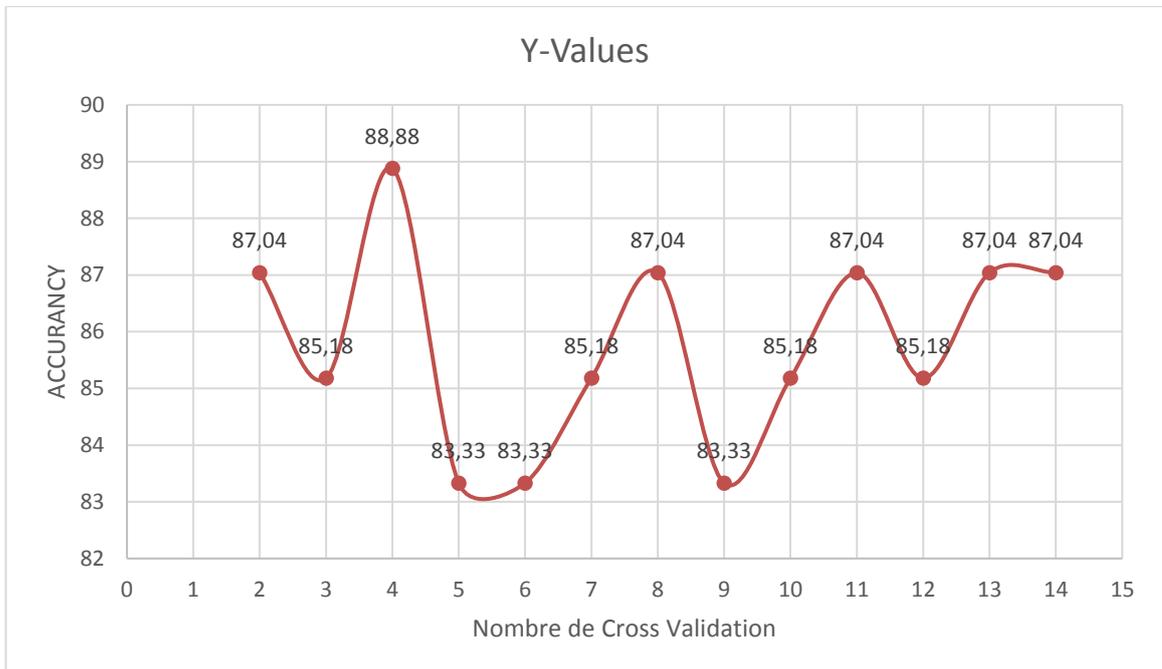


Figure 35. Graph décrivent l'accuracy par rapport au nombre de cross-validation

5. Comparaisons des Techniques :

	Accuracy	Précision	Recall	F- mesure	Instance correctly classified	Instances incorrectly classified	Error rate
Arbre de décision	0,87	0,92	0,95	0,93	48	6	0,11
Random Forest	0,87	0,87	1	0,93	47	7	0,13

Tableau 5. Tableau comparative des résultats des deux techniques.

6. Conclusion :

L'implémentation de notre solution nous a permis d'obtenir une application conviviale pour l'aide au diagnostic du cancer de la peau.

Nous avons utilisé deux techniques à savoir les arbres de décision et Random Forest.

Les résultats ont montré que les arbres de décision donnent le meilleur résultat car on a obtenu un meilleur taux de précision moins d'erreur.

Weka

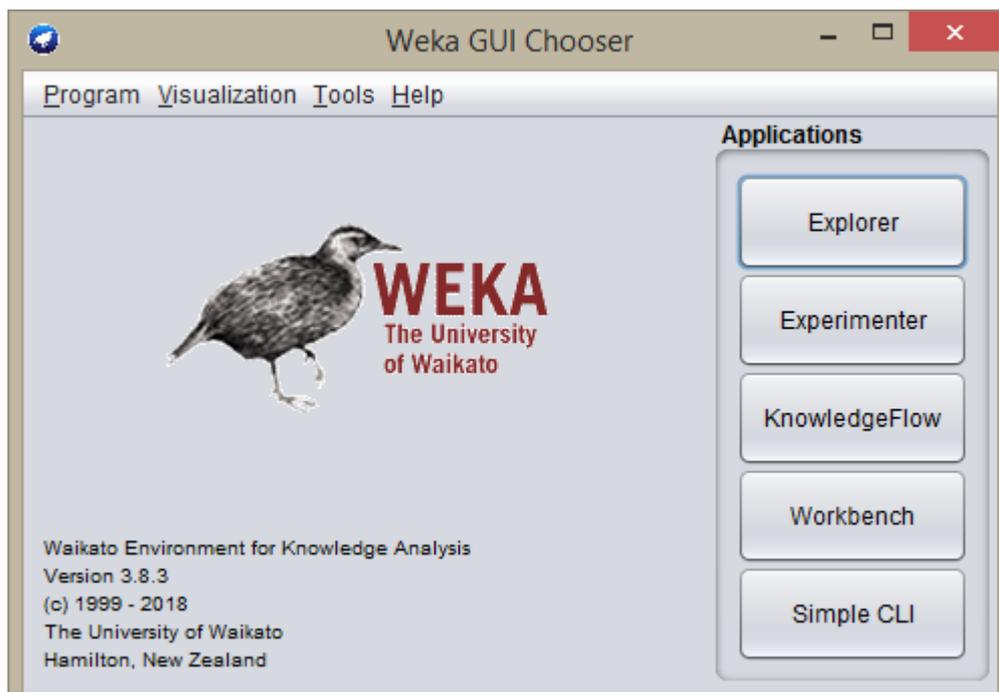
Est un logiciel (environnement) utiliser pour le data mining (analyse des connaissance) et le Machine Learning.

Installation de l'application :

Pour installer weka il suffit de d'entré dans son site web

<https://www.cs.waikato.ac.nz/ml/weka/downloading.html>

est de choisir le système d'exploitation(windows,mac et linux) et après la fin du téléchargement on clique sur le fichier à installer après que l'installation soit faite il nous reste juste a lancer l'application.



WEKA possèdent des Outils de prétraitement des données (filtering) : Sélection, transformation, combinaison d'attributs, normalisation, ré-échantillonnage, ..., etc. Ainsi plusieurs algorithmes pour le Machine Learning: Clustering, classification, régression, etc.

WEKA permet aussi l'analyse de résultats avec une évaluation des performances, comparaison d'algorithmes, ...etc. Possèdent plusieurs interfaces : Graphiques (Explorer, Expérimenter et Knowledge Flow) et en ligne de commande (CLI)

Traitements des données :

Le format des fichiers de Weka est ARFF (Attribute Relation File Format)

Il est toutefois possible d'utiliser d'autres formats comme les fichiers CSV, binaire, BDD SQL (avec JDBC), à partir d'une URL, etc.

Les Caractéristiques des fichiers arff :

1) Le nom de l'ensemble des données on l'écrit :

@relation nom pour ensemble_des_données

2) On définit les *features* par **@attribute**

Attribut nominative :

@attribute outlook { sunny, overcast, rainy }

Attribut réel:

@attribute prix *real*

Attribut chaîne avec *string* et date avec *date*.

3) On met **@data** pour commencer les instances

```

C:\Program Files\Weka-3-8\data\weather.nominal.arff - Notepad++
Fichier Édition Recherche Affichage Encodage Langage Paramètres Outils Macro Exécution Compléments Documents ?
weather.nominal.arff
1 @relation weather.symbolic
2
3 @attribute outlook {sunny, overcast, rainy}
4 @attribute temperature {hot, mild, cool}
5 @attribute humidity {high, normal}
6 @attribute windy {TRUE, FALSE}
7 @attribute play {yes, no}
8
9 $data
10 sunny,hot,high,FALSE,no
11 sunny,hot,high,TRUE,no
12 overcast,hot,high,FALSE,yes
13 rainy,mild,high,FALSE,yes
14 rainy,cool,normal,FALSE,yes
15 rainy,cool,normal,TRUE,no
16 overcast,cool,normal,TRUE,yes
17 sunny,mild,high,FALSE,no
18 sunny,cool,normal,FALSE,yes
19 rainy,mild,normal,FALSE,yes
20 sunny,mild,normal,TRUE,yes
21 overcast,mild,high,TRUE,yes
22 overcast,hot,normal,FALSE,yes
23 rainy,mild,high,TRUE,no
24
Normal text file length: 587 lines: 24 Ln: 1 Col: 27 Sel: 0|0 Unix (LF) UTF-8 INS

```

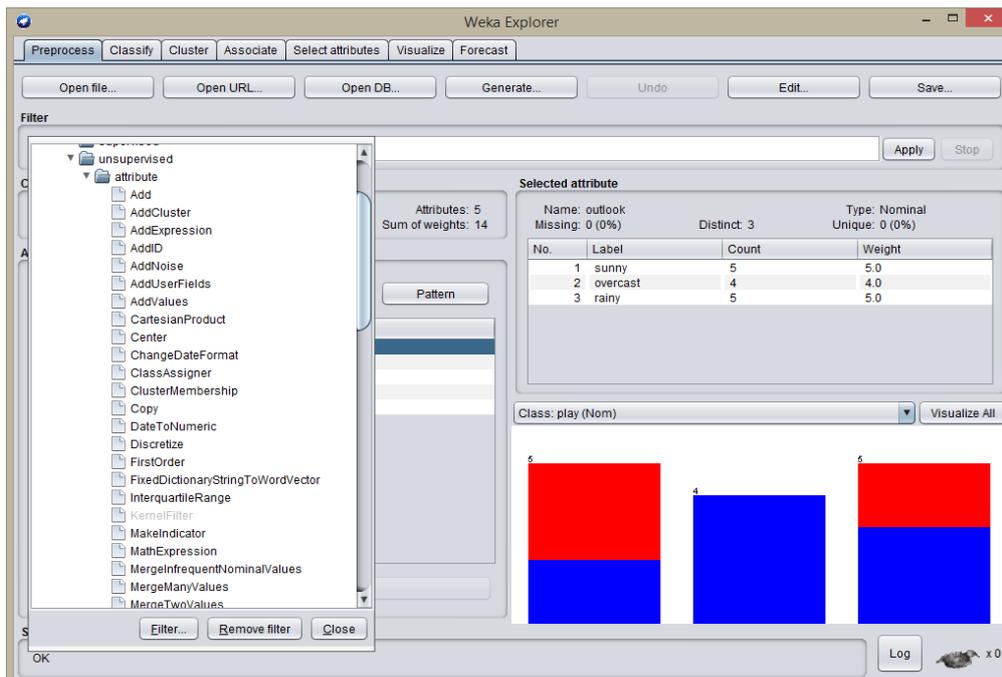
Pré-traitement de données :

Pour faire des pré-traitement sur les données on utilise des *Filtres*

Les filtres permettent de modifier les ensemble de données : supprimer ou ajouter des attributs, ré-échantillonner, supprimer des exemples, etc.

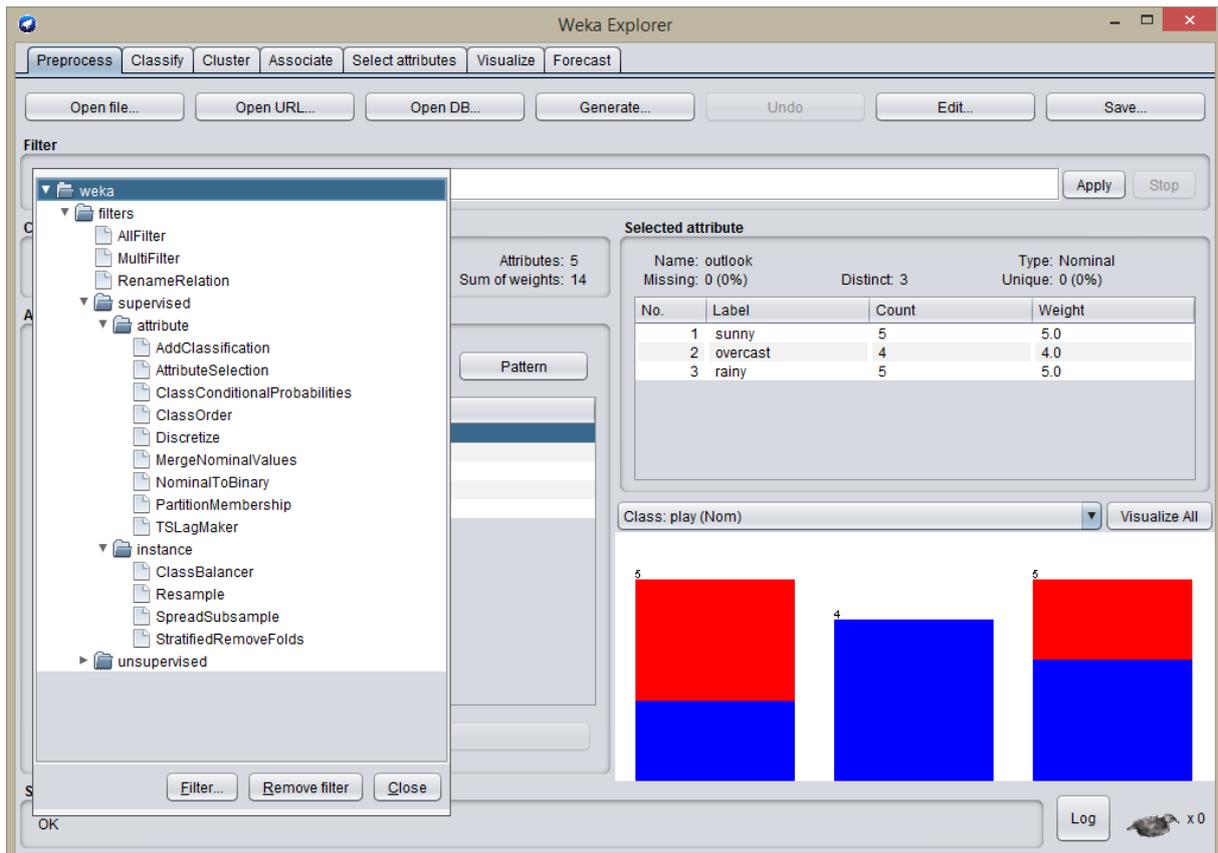
weka dispose de deux types de filtres

Filtres non supervisés.



Filtres supervisés.

Il existe des filtres pour les attributs et pour les exemples



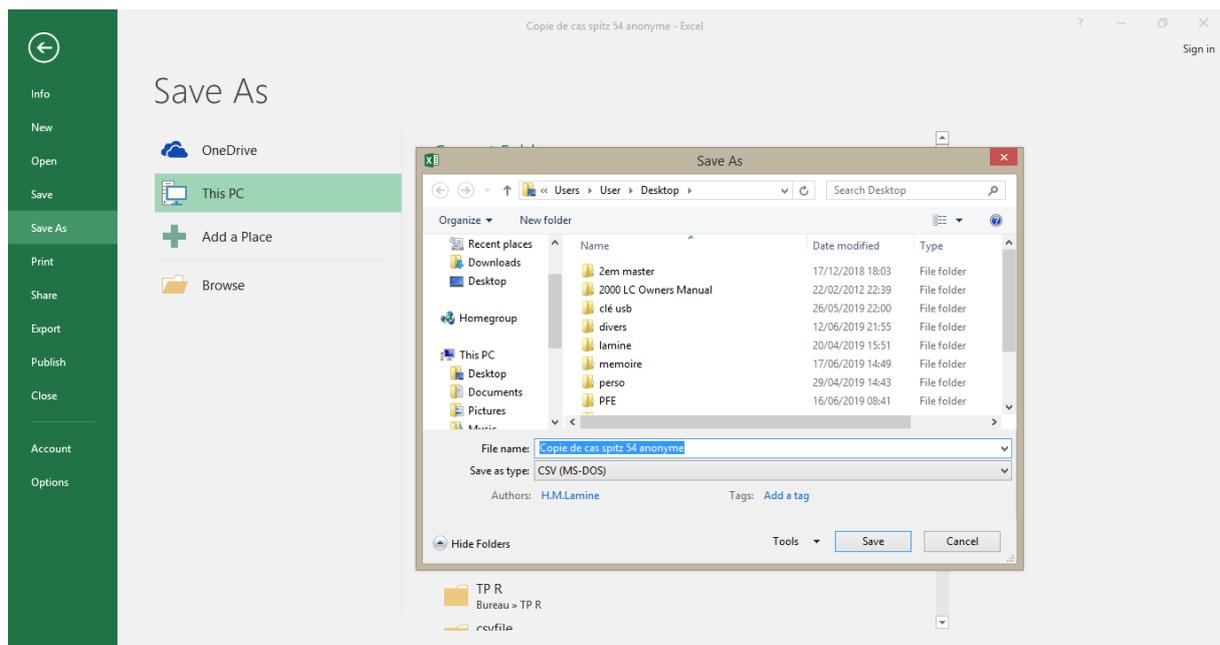
Exemples de Filtres :

- ❖ Ajout, suppression et copie d'attributs (*Add, Remove et Copy*).
- ❖ Ajout de bruit (*AddNoise*).
- ❖ Fusion d'attributs (*Merge*).
- ❖ Conversion (*NominalToBinary, StringToBinary, etc.*).
- ❖ Méthode de discrétisation (*Discretize*).

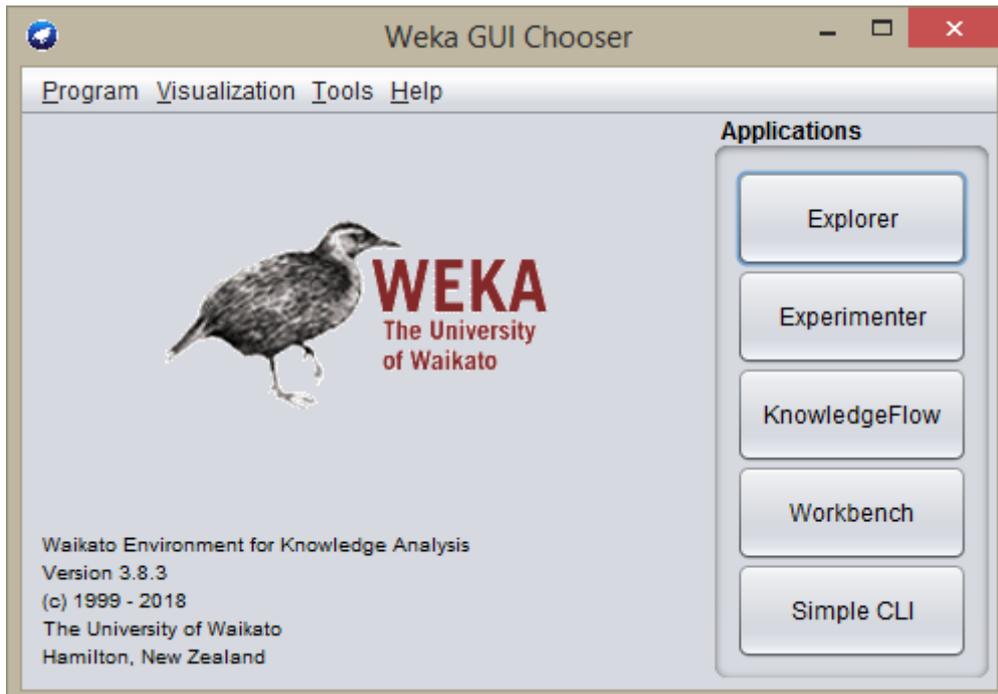
Les Algorithmes :

Il existe des algorithmes pour la Classification ,la régression et le clustring parmi ces algorithmes :

- ✚ Arbres de Décision.
- ✚ Support Vector Machine (SVM).
- ✚ Classification Bayésienne naïve.
- ✚ Réseaux Bayésien.
- ✚ Perceptron Multi-Couches.
- ✚ Règles d'associations.



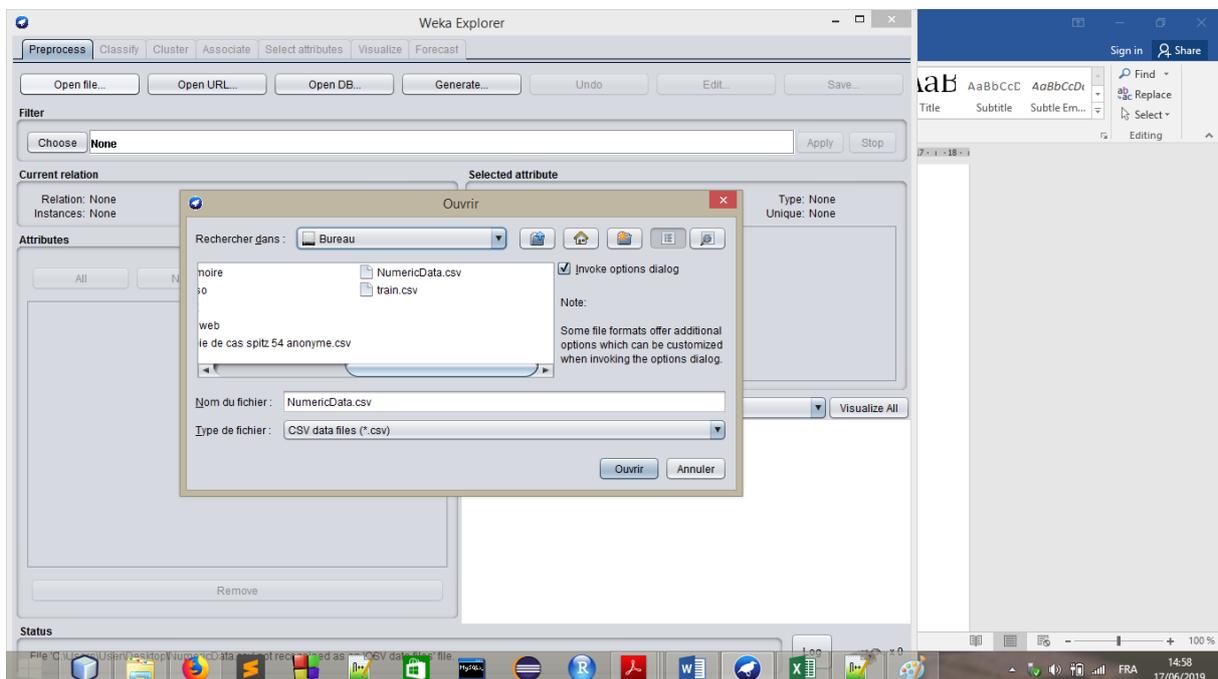
Ouvrir Weka -> ouvrir Explorer

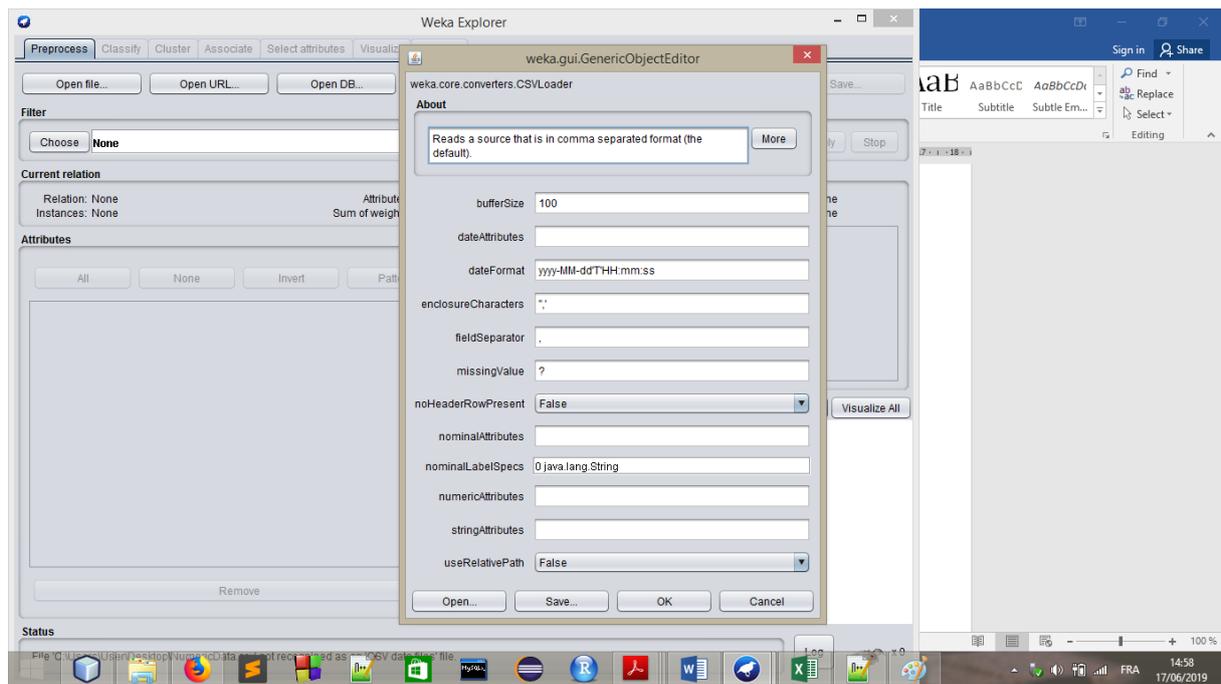


Puis cliquer sur open file :

Choisir le fichier voulu

Cocher "Invoke options Dialog"

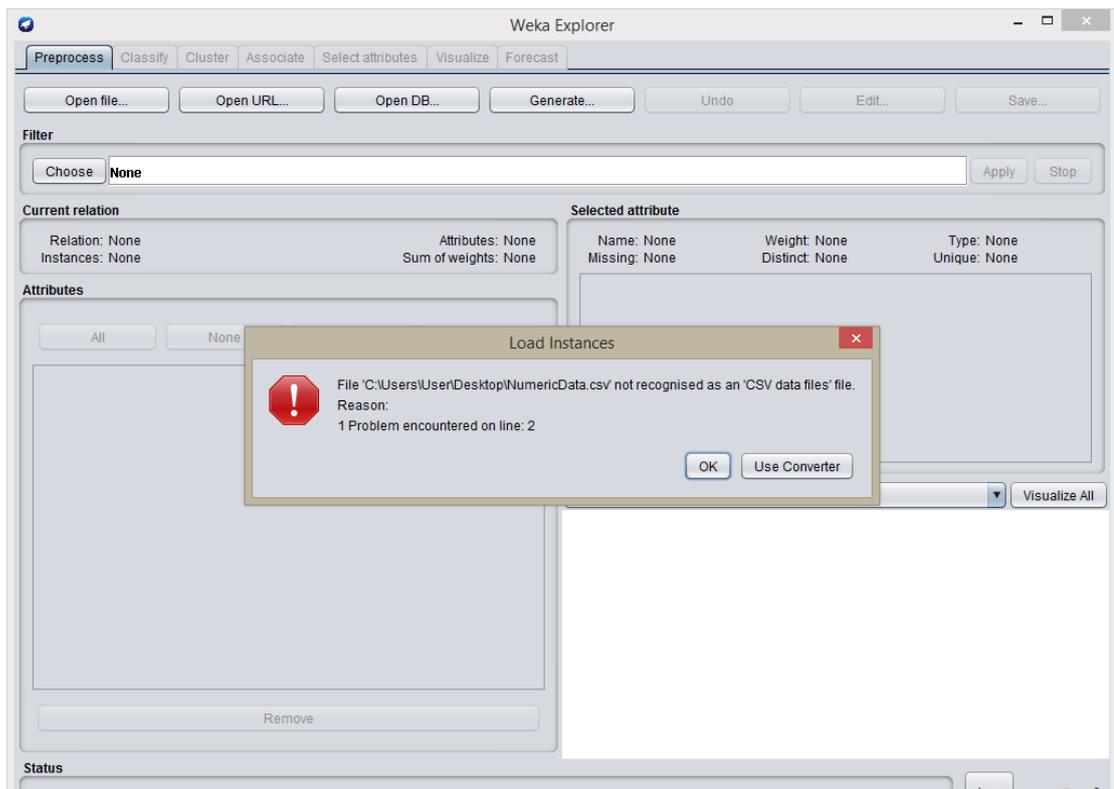




On règle

- Les séparateur , ou ;
- Valeur manquante par ? ou ‘ ‘
- On clique sur ok après normalement il s'il n'a pas d'erreur le fichier est ouvert dans le cas contraire une fenêtre d'erreur s'affiche avec la ligne ou se trouve l'erreur comme

ceci :



Dans mon cas j'ai Créé un fichier ARFF a l'aide d'un éditeur de texte.

```
C:\Users\User\Desktop\skinCancer.arff - Notepad++
Fichier Édition Recherche Affichage Encodage Langage Paramètres Outils Macro Exécution Compléments Documents ?
skinCancer.arff X
1 @relation skinCancer
2
3
4
5 @attribute id numeric
6 @attribute Sexe {0,1}
7 @attribute Age_du_diagnostic numeric
8 @attribute diametre numeric
9 @attribute localisation {1, 2, 3, 4, 5}
10 @attribute forme {1, 2, 3}
11 @attribute epaisseur numeric
12 @attribute BRAF {1, 2}
13 @attribute ALK_IH {1, 2}
14 @attribute Ly {0, 1, 2, 3}
15 @attribute pig_mel {0, 1, 2, 3}
16 @attribute Ki-67 numeric
17 @attribute Ki67_sup_lpour {0,1}
18 @attribute desm {0,1}
19 @attribute index_mitotique numeric
20 @attribute mitoses_profondes {0,1}
21 @attribute atypies_CN {0,1}
22 @attribute hyperCELL {0,1}
23 @attribute Kamin0 {0,1}
24 @attribute hypoderme {0,1}
25 @attribute asymetrie {0,1}
26 @attribute lim_floues {0,1}
27 @attribute migration {0,1}
28 @attribute diametre_suppl {0,1}
29 @attribute Th_irreg {0,1}
30 @attribute pas_de_grad{1,2}
31 @attribute grenz{1,2}
32 @attribute modif_epid {0,1}
33 @attribute diagnostic {0,1}
34
35
36
37 @data
38 0,0,33,0,5,1,1,0,1,1,1,2,3,2,1,0,0,0,1,0,0,0,0,0,1,0,0,1,2,0,1
Normal text file length: 4387 lines: 92 Ln: 11 Col: 29 Sel: 0|0 Windows (CR LF) UTF-8 INS
```

On fait comme on fait pour un fichier csv sauf que type de fichier est maintenant ARFF

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | Forecast

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** [Apply] [Stop]

Current relation
 Relation: skinCancer
 Instances: 54
 Attributes: 29
 Sum of weights: 54

Attributes
 All | None | Invert | Pattern

No.	Name
1	<input checked="" type="checkbox"/> id
2	<input type="checkbox"/> Sexe
3	<input type="checkbox"/> Age_du_diagnostic
4	<input type="checkbox"/> diametre
5	<input type="checkbox"/> localisation
6	<input type="checkbox"/> forme
7	<input type="checkbox"/> epaisseur
8	<input type="checkbox"/> BRAF
9	<input type="checkbox"/> ALK_IH
10	<input type="checkbox"/> Ly
11	<input type="checkbox"/> pig_mel
12	<input type="checkbox"/> Ki-67
13	<input type="checkbox"/> Ki67_sup_1pour
14	<input type="checkbox"/> desmo
15	<input type="checkbox"/> index_mitotique
16	<input type="checkbox"/> mitoses_profondes

Remove

Selected attribute
 Name: id
 Missing: 0 (0%)
 Distinct: 54
 Type: Numeric
 Unique: 54 (100%)

Statistic	Value
Minimum	0
Maximum	53
Mean	26.5
StdDev	15.732

Class: diagnostic (Nom) [Visualize All]

Status: OK [Log] x 0

Windows taskbar: 15:08 17/06/2019

Ce mémoire avait pour objectif de proposer une approche intelligente pour l'aide au diagnostic du cancer de la peau basée sur le Machine learning. Dont le but est de satisfaire les besoins des oncologues, Tout en proposant une application ergonomique qui aidera les médecins au diagnostic.

La réalisation de cette étude nous a conduites à suivre les étapes suivantes :

- En premier lieu, nous avons défini le cancer, nous avons détaillés ces causes, son fonctionnement, ces traitements possibles.
- En suite nous avons présenté le machine learning avec son principe, ces algorithmes.
- Apres nous vous avons montré la partie conception est ces différentes étapes.
- en dernier nous avons présenté l'implémentation de notre approche.

Ce mémoire constitue une base de travail à partir du quelle, de nouvelles activités de recherche peuvent être lancées afin d'améliorer le travail présenté.

Références :

- 1) <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/lung/lung-cancer/?region=on>
- 2) <https://www.who.int/fr/news-room/fact-sheets/detail/cancer>
- 3) <https://www.cancer.ca/fr-ca/cancer-information/cancer-type/pancreatic/pancreatic-cancer/?region=on>
- 4) https://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer_sein_p_m
- 5) <https://www.e-cancer.fr/Comprendre-prevenir-depister/Qu-est-ce-qu-un-cancer/Etapes-et-stades-du-cancer>
- 6) <http://www. Roche.fr/patients/info-patients-cancer/traitement-cancer/traitements-cancer.html>
- 7) <https://www.radioalgerie.dz:9090/news/fr/article/20171203/127573.htm>
- 8) Introduction to machine learning : Alex Smola and S.V.N. Vishwanathan
- 9) Machine learning by Margaret Rouse
<https://whatis.techtarget.com/fr/definition/Machine-Learning>
- 10) <https://dataanalyticspost.com/Lexique/apprentissage-non-supervise/>
- 11) <https://www.fondation-arc.org/cancer/cancer-peau/traitement-cancer>
- 12) https://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer_peau_pm
- 13) <https://www.msdmanuals.com/fr/professional/troubles-dermatologiques/cancers-cutan%C3%A9s/m%C3%A9lanome>
- 14) Cours Dr Djeflal sur la Classification 2018/2019
- 15) Traitement des Cancer par institut national contre le cancer.
- 16) Yoann Pitarch : Exploration d'information et fouilles de données.
- 17) <http://penseeartificielle.fr/focus-reseau-neurones-artificiels-perceptron-multicouche/>
- 18) http://mediamining.univ-lyon2.fr/people/guille/m4101/k_plus_proches_voisins.html
- 19) <https://maximilienandile.github.io/2016/09/29/Machine-Learning-comment-fonctionne-la-classification-naive-Bayesienne/>
- 20) <https://www.journaldunet.com/solutions/reseau-social-d-entreprise/1191979-machine-learning-12-secteurs/>
- 21) https://fr.wikipedia.org/wiki/Apprentissage_automatique
- 22) <https://www.e-cancer.fr/Dictionnaire/C/cancer>
- 23) <https://www.cancer.ca/fr-ca/cancer-information/cancer-101/what-is-cancer/how-cancer-starts-grows-and-spreads/?region=qc>
- 24) <https://flamigel.fr/anatomie-de-la-peau/>
- 25) <https://www.ufcnouvellecaledonie.nc/cancers-de-peau>