

République Algérienne Démocratique et Populaire
Université Abou Bakr Belkaid- Tlemcen
Faculté des sciences
Département d'Informatique

Mémoire de fin d'études pour l'obtention du diplôme de « Master » en informatique

Option:

Modèles d'intelligence et de décision (M.I.D)

Thème

Conception d'un simulateur d'inspection de disques optiques de type DVD lors de la phase de réplication dans un système M.E.S

Réalisé par :

SAKER Sami

Sous la direction de messieurs :

Mr. SMAHI I.

(Encadrant)

Mr. BENDIABDELLAH M^{ed}.H.

(Co-encadrant)

Soutenu devant le jury composé de messieurs :

Mr. HADJILA F.

(Président du jury)

Mr. BELABED A.

(Examineur)

Mr. BENAMAR A.

(Examineur)

Session :

(Octobre 2013)

Année universitaire 2012-2013

Remerciements

Avant toute chose, je remercie le tout puissant de m'avoir accordé la force, la patience et le courage d'entreprendre la réalisation de ce mémoire.

Et c'est avec une immense reconnaissance; que j'adresse mes remerciements les plus sincères à Mr. Roger W. NUTE, sans qui ce travail n'aurait jamais vu le jour, à mon oncle Mr. Hillal FETOUHI qui m'a inculqué des connaissances dans l'optique.

Je remercie grandement mon encadrant Mr. Ismaïl SMAHI qui a fait preuve de patience et m'a orienté dans mon projet, à mon co-encadrant Mr. Hakim M^{ed} BENDIABDELLAH qui a su ménagé mes vœux par rapport au projet et m'a grandement guidé.

Je remercie messieurs les examinateurs et président du jury pour avoir sacrifié de leur temps pour se consacrer à l'évaluation de mon projet de fin d'études.

Je présente à tous ces messieurs tous mes sentiments de gratitude.

Je tiens à témoigner mon enthousiasme et remercie ceux et celles qui m'ont prêtés attention, aide et assistance, de près ou de loin.

Je témoigne entre autre ma reconnaissance à tous nos professeurs du département d'informatique, actuels et antérieurs ; qui ont grandement contribué à ma progression, ainsi qu'à tous les étudiants amis des différentes filières du même département dans lequel j'étudie et que je salue.

Sami

Dédicaces

C'est avec un immense plaisir que je dédie ce mémoire en hommage à feu mon père Mr. Sibawayhi SAKER ainsi que mes grands-parents.

À tous les membres de ma petite famille que sont ma mère à qui je dois tout, à mon frère Zaki et mes sœurs Meriem et Rim et leur belle-famille, à ma tante et seconde mère Zoubida.

À ma grande famille maternelle et paternelle, oncles et tantes, cousins et cousines, je vous dédie ce travail tout autant.

Je dédie ce travail à toutes mes connaissances de la promotion pour leur disponibilité et toute l'aide qu'ils m'ont apportés.

À mes amis de la cité universitaire, à ceux de mon quartier et du centre culturel français. À toute les personnes spéciales que j'ai rencontrées dans ma vie, à mes amis les plus proches, les plus chers et qui se reconnaîtront.

Je préfère ne pas citer de noms afin de négliger personne.

Aami

TABLE DES MATIERES

INTRODUCTION GENERALE	01
CHAPITRE I LE SYSTEME M.E.S., LA MANUFACTURE DES DISQUES OPTIQUES ET LE PRINCIPE D'INSPECTION	03
Introduction	03
I. Le système M.E.S. et les disques optiques	03
I.1. M.E.S (Manufacturing Execution System)	03
I.2. Disques optiques	04
II. La procédure de fabrication d'un disque optique « DVD »	05
II.1. Mastering	06
II.2. Stamper making	06
II.3. Replication	07
II.4. Packaging	08
III. La procédure d'inspection d'un disque optique « DVD »	08
III.1. Inspection de surface	08
III.1.1 Liste des erreurs potentielles sur un disque optique DVD	11
III.2. Inspection de la déviation d'un disque optique	13
Conclusion	14
CHAPITRE II NOTION D'APPRENTISSAGE ET APPROCHES THEORIQUES	15
Introduction	15
I. Définition de l'ensemble des paramètres servant à l'apprentissage	15
I.1. Observations	17
II. Définitions des algorithmes à tester	17
II.1. Réseau de neurones	17
II.2. K plus proche voisin	18
II.3. Naïve Bayes	19
III.3.1 Théorie de la décision bayésienne	20
III. Evaluation des algorithmes d'apprentissage	21

III.1. Mesures de performance d'un algorithme	21
III.1.1 Le rappel d'une classe	21
III.1.2 La précision d'une classe	22
III.1.3 Le taux de classification	22
Conclusion	23
CHAPITRE III CHOIX DE L'ALGORITHME , CONCEPTION DU SIMULATEUR ET DISCUSSION DES RÉSULTATS	24
Introduction	24
I. Expérimentations et discussions des résultats	24
I.1. Expérience sous RapidMiner	24
I.2. Paramétrage des algorithmes	26
I.2.1 Réseau de neurones.....	26
I.2.1.1 Le nombre d'époques	26
I.2.1.2 Le pas d'apprentissage.....	27
I.2.1.3 L'erreur epsilon.....	27
I.2.1.4 Observations.....	28
I.2.2 K plus proche voisin	28
I.2.2.1 Observations.....	29
I.2.3 Naïve Bayes.....	29
I.2.3.1 Corrigé Laplace	29
I.2.3.2 Observations.....	30
I.3. Observations.....	31
I.4. Expérience sous l'application	31
I.4.1 Langage et environnement de développement	31
I.4.2 Descriptif des classes employés sous l'application	32
I.4.3 Présentation de l'application	32
I.5. Observations.....	35
Conclusion	35
CONCLUSIONS GÉNÉRALES ET PERSPECTIVES	36
RÉFÉRENCES BIBLIOGRAPHIQUES, WEB ET MANUELS	36
ANNEXES	38

LISTE DES FIGURES

CHAPITRE I

I.1 Schéma représentant les principes fonctionnels d'un système M.E.S	03
I.2 Vue de surface d'un disque optique DVD avec ses dimensions	04
I.3 Représentation des « pits » et « lands » présents sur un disque optique	05
I.4 Composants d'un disque optique DVD	05
I.5 Vue sur le processus du « Mastering »	06
I.6 Résultats d'un « Stamper making »	07
I.7 Vue sur les étapes de la phase de « Réplication »	07
I.8 Vue sur le concept de l'inspection	08
I.9 Vue sur l'inspection de surface d'un disque optique	09
I.10 Vue sur les étapes de l'inspection de surface d'un disque optique	10
I.11 Capture réelle des défauts détectés lors d'une inspection ainsi que leurs nominations respectives et leurs légendes sous simulateurs	12
I.12 Figure des différentes déformations possibles sur un disque optique	13
I.13 Interprétation géométrique d'une déformation sur un disque optique	13
I.14 Mesure radiale de la déviation	14

CHAPITRE II

II.1 Exemple d'un réseau de neurones	18
II.2 Classification de deux classes avec K-NN, à gauche K=1, à droite K=3	19
II.3 Fonctions discriminantes et décisions	20
II.4 Frontières des décisions dans l'espace des observations	21
II.5 Calcul du taux de classification sous la norme ISO 5725-1	22
II.6 Schéma procédurale de l'apprentissage sous RapidMiner	24
II.7 Résultats de performance de l'algorithme K-NN sous RapidMiner	25

II.6 Taux de classification des RN simulés par différents pas d'apprentissages	27
II.7 Régression du taux de classification de K-NN	11

CHAPITRE III

III.1 Schéma procédurale de l'apprentissage sous RapidMiner	24
III.2 Résultats des performances de l'algorithme « K-NN » sous RapidMiner	25
III.3 Taux de classification des RN simulés par différents pas d'apprentissage	27
III.4 Régression du taux de classification de K-NN	28
III.5 Ensemble des classes servant à l'application	32
III.6 Ecran d'accueil principal de l'application	33
III.7 Choix du type du disque à traiter sous l'unité d'inspection en Spaceline	33
III.8 Interface M.E.S de l'inspection de surface et de la déviation d'un DVD 5	32

ANNEXES

A.1 Réseau de neurones amélioré pour le plus haut taux de classification sous RapidMiner	38
A.2 Algorithme K-NN	40

LISTE DES TABLEAUX

CHAPITRE I

I.1 Extrait d'un fichier log lors d'une inspection de surface d'un disque DVD .10

CHAPITRE II

II.1 L'ensemble des paramètres de l'inspection en surface avec leur descriptif ..16

CHAPITRE III

III.1 Nombre d'erreurs de chaque classe au sein de la base d'apprentissage26

III.2 Comparaison du taux de classifications des différents pas d'apprentissage.28

III.3 Valeurs tirées du graphe de régression des taux de classification K- NN....29

III.4 Valeurs des taux de classification sous Naïve Bayes.....30

III.5 Tableau comparatif des résultats des différents algorithmes simulés.....30

III.6 Vérification des valeurs de déviation sur un même disque optique35

ANNEXES

A.1 Rétropropagation et calcul des fonctions d'erreur pour la couche cachée ..38

A.2 Rétropropagation calcul des fonctions d'activation pour les classes de sort38

INTRODUCTION GÉNÉRALE

La standardisation des secteurs industriels dans le monde économique ainsi que les facteurs qui augmentent son efficacité en production, son raccourcissement en temps d'innovations ainsi que l'intégrité de la qualité de production en question augmentent significativement la pression sur la machine économique, pareillement à plusieurs autres paramètres influents.

Au fil du temps, l'émergence des systèmes informatiques et leur implantation dans les manufactures n'est plus fortuite, tandis qu'ils existaient principalement pour les systèmes orientés en sécurité, ils endossent aujourd'hui un rôle qui offre dans l'industrie de la transparence, de la traçabilité et de l'automatisation ; contribuant au final au bon déroulement du processus lors d'une production industrielle.

Dans un but de rentabilisation, il est nécessaire dans un premier temps de remanier des procédés de productions existants afin d'en augmenter le potentiel, car initialement *la complexité de la manufacture d'un produit affecte significativement la probabilité de l'obtention d'un produit « bon » à la commercialisation*, et étant donné le nombre conséquent planifié lors de la production, il devient dès lors difficile pour ne pas dire impossible d'inspecter avec aisance et précision tous ces produits par des experts, pour remédier à ce fait ; des outils seront requis.

L'informatisation d'un traitement étant un gage de qualité, elle est adoptée par un grand nombre d'entreprises, afin d'atteindre le résultat escompté en production, il fallait auparavant doter ces systèmes de plusieurs extensions et procédures qui ne se limitent pas à des E.R.P (*Planification des ressources d'entreprise*) dont les tâches sont pour la majorité d'ordre administratifs et de comptabilités (aussi complexes soit-elles), il fallait entre autre implanter des tâches de planifications et de contrôles pour les machines ainsi que d'enregistrements pour les produits et cela bien évidemment en temps réel. Pour ces systèmes, le concept du M.E.S (*Système d'exécution de fabrication*) fut conçu.

Partant du principe que le M.E.S est un système modulable ; chaque fabricant dans chaque secteur peut en disposer de manière presque personnelle. En outre, il existe plusieurs fournisseurs de logiciels sur le marché qui offrent des solutions informatiques à base M.E.S, ce qui démontre son potentiel.

Il est à noter que le degré d'automatisation dans le secteur industriel est élevé, de ce fait, recourir à la modification d'un procédé ne va pas pour autant augmenter significativement la production.

Le travail de ce mémoire consistera à reproduire une simulation des fonctions basiques d'une unité d'inspection des disques optique « DVD 5 » dans un système M.E.S, l'inspection en question s'opère d'une phase précise dans une chaîne de fabrication comprenant quatre étapes.

La phase proprement étudiée se nomme « *réplication* », les quatre étapes seront présentées dans le premier chapitre. La réplication est une phase déterminante dans la finalisation d'un produit, par conséquent elle s'avère fondamentale ; le choix d'une telle phase se justifie par le fait que l'inspection emploie un algorithme à notion d'apprentissage ce qui est en soi un concept propre aux modèles d'intelligence et de décision, d'autant que les perspectives proposées par cette discipline offrent des orientations non négligeables aux experts pour une éventuelle aide à la décision.

Le second chapitre expérimente quelques approches théoriques en se basant sur des algorithmes qui seront comparés suivant leur rendement, et par conséquent ; justifieront un éventuel choix pour la conception d'un simulateur.

Le troisième chapitre illustre des captures du simulateur créé dans ce but, tout en expliquant les fonctionnalités propres à ce dernier ainsi que leurs impacts respectifs sur la production en s'appuyant sur des situations probables voire réelles.

Ce mémoire se conclue par une série de perspectives tout en stipulant les points forts et négatifs lors de la conception de ce simulateur pour ladite phase de réplication.

CHAPITRE I

Le système M.E.S.
La manufacture des
disques optiques et le
principe d'inspection.

Introduction

Ce chapitre s'entamera par des définitions qui engloberont le chaîne de fabrication ; en se basant sur le concept d'un système M.E.S, avant d'enchaîner sur la définition du produit à traiter ; à savoir un disque optique « DVD 5», puis ses étapes de fabrication avant de s'axer précisément sur la phase de « *réplication* » tel qu'il a été mentionné dans l'introduction générale.

I. Le système M.E.S et les disques optiques

I.1 M.E.S (Manufacturing Execution System)

(Ou système d'exécution des fabrications) sont des activités et systèmes permettant de piloter l'ensemble des productions en temps réel et de restituer toutes les conditions dans lesquelles elles se sont déroulées.

Les systèmes M.E.S reposent sur des principes fonctionnels variés, parmi elles :

- La gestion de l'allocation, des états et des règles applicables des ressources requises pour la fabrication.
- Gestion et ordonnancement des ordres de fabrication.
- Suivi de la production en temps réel.
- Analyse et diagnostic de la production.
- Interfaçage avec des systèmes tiers. [1]

L'approche qui sera sollicitée par rapport aux principes fonctionnels d'un système M.E.S concernera « l'analyse et diagnostic d'un produit en cours de fabrication ».

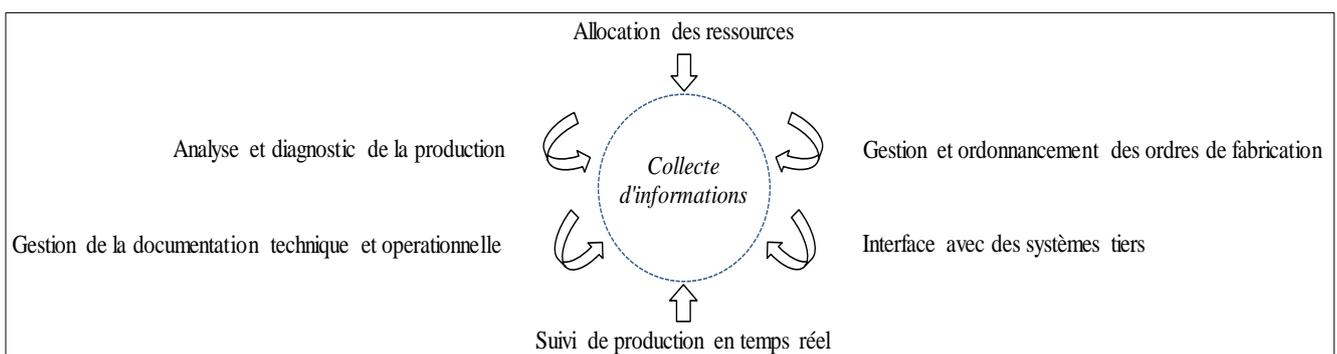


fig. I.1 Schéma représentant les principes fonctionnels d'un système M.E.S

I.2 Disques optiques

Dans le domaine de l'informatique et des technologies optiques, un « *disque optique* » est une masse physique sollicitée dans un but de stockage d'informations ; d'une forme circulaire, un disque optique chiffre des données binaires (bits) sous la forme de « *pits* » (littéralement « creux », qui signifient une absence de réflexion lors d'une lecture, ce qui équivaut à une valeur binaire 0 explicitant l'absence d'une information), et de « *lands* » (signifiant « terre » ; la réflexion étant présente, elle implique une valeur binaire non nulle « 1 »).

La lecture et l'écriture s'opèrent sur un « substrat » (fait en polycarbonate) qui éventuellement contiendra l'information codée en question, alors que la réflexion dépendra d'un autre substrat (généralement en aluminium) qui se nomme « couche réfléchive ». [2]

Il est à noter qu'il existe un grand nombre de types de disques optiques qui se divisent eux-mêmes en plusieurs sous-catégories, ce mémoire s'axera sur le traitement de disques optiques de type « DVD ». [2]

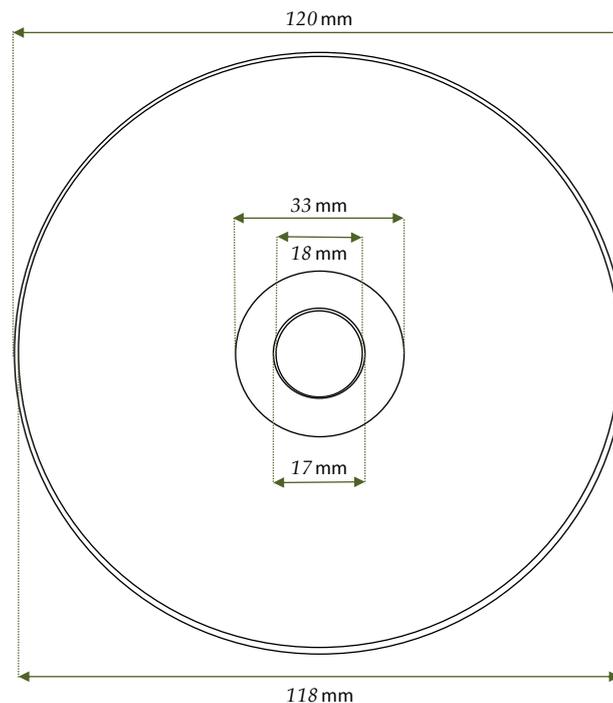


fig. I.2 Vue de surface d'un disque optique DVD avec ses dimensions

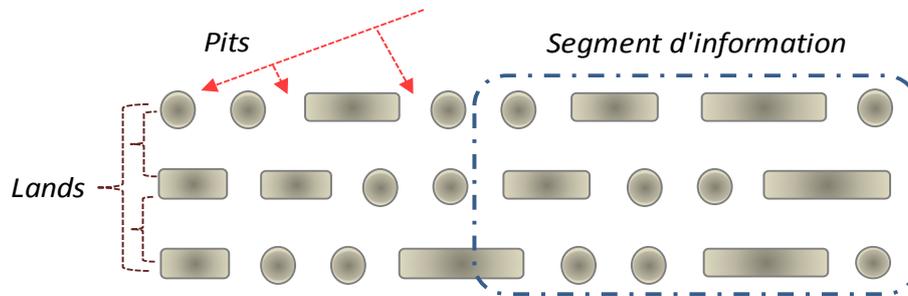
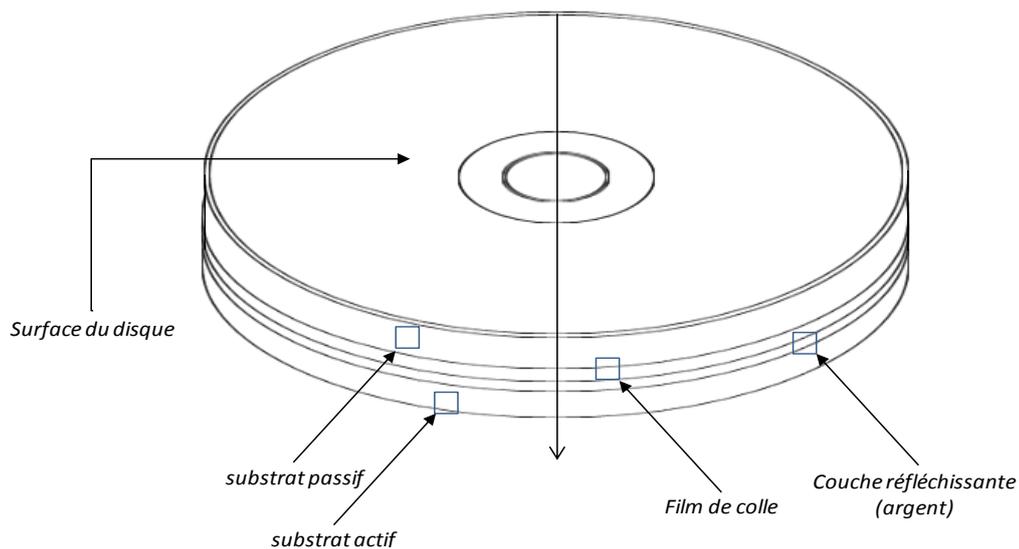


fig. I.3 Représentation des « Pits » et « Lands » présents sur un disque optique DVD

Cette figure détaille les composants d'un disque optique répliqué de type DVD 5 [3]



Couches	Epaisseur	Unité
Substrat passif	0,6	Mm
Film de colle	50 à 55	µm
Couche réfléchissante	± 100	Nm
Substrat actif	0,6	Mm

fig. I.4 Composants d'un disque optique DVD

II. La procédure de fabrication d'un disque optique « DVD »

La conception d'un disque optique passe par plusieurs étapes qui sont laborieuses et par conséquent font appel à plusieurs disciplines, le but de cette section n'étant pas de détailler en profondeur lesdites étapes, le fait de les mentionner tout en vulgarisant le contenu permettra néanmoins de saisir l'enchaînement du procédé et de parvenir de manière claire et rapide à la phase qui nous intéresse ; à savoir la réplique d'un disque optique et plus précisément « l'inspection » au sein de cette même phase.

Les étapes de l'élaboration d'un disque optique sont les suivantes

II.1 Mastering

(Ou matriçage) ; est la phase durant laquelle un programme est enregistré sur la surface de la matrice d'informations qui à partir d'une matière qu'est le « *photo-resist* » sera ancrée sur un disque en verre (*glass disc*). Le mastering comprend cinq procédés individuels au sein du même système. [4]

Le résultat de ce procédé est ainsi appelé « *Photo-resist master* » tel qu'il est montré dans la figure ci-dessous.

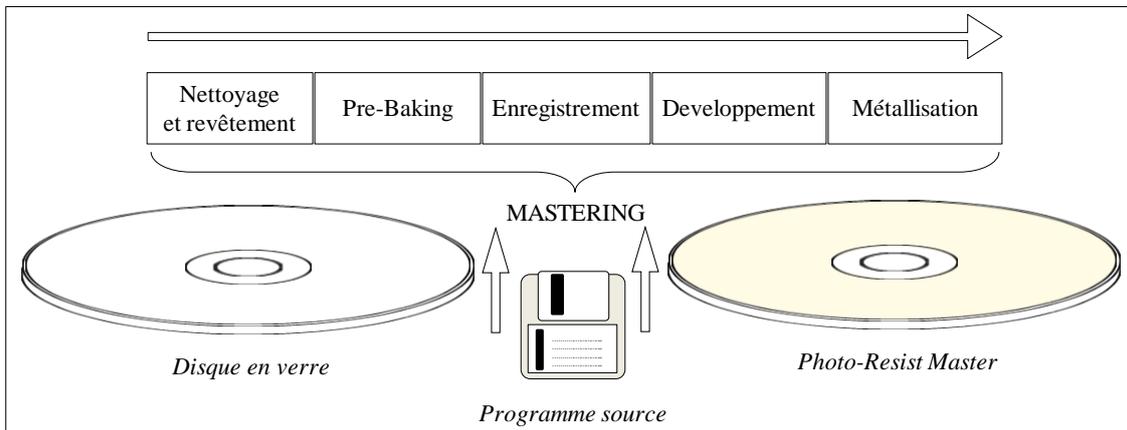


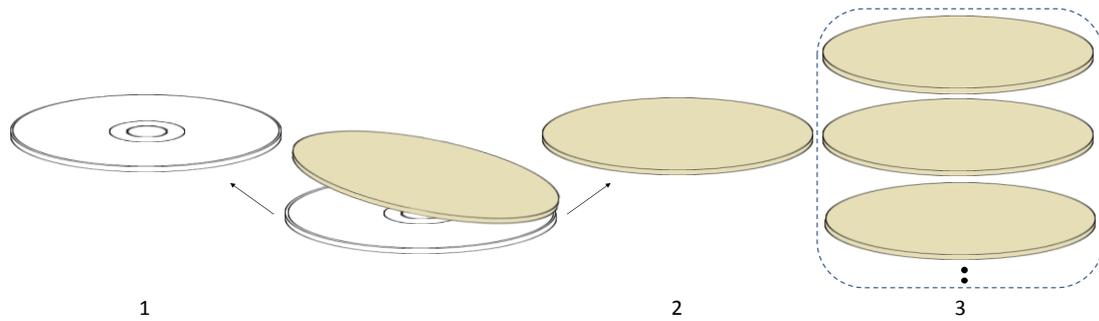
fig. I.5 Vue sur le processus du "Mastering"

II.2 Stamper making

(Fabrication du stamper); est l'étape durant laquelle le « *Photo-resist Master* » obtenu précédemment verra sa métallisation solidifiée par épaissement. Le but étant de séparer en fin compte le « *stamper obtenu* » du « *disque en verre* » utilisé initialement lors du Mastering.

Une fois fait ; il sera dupliqué pour être employé dans la phase de « *réplication* ».

[4]



1	Disque en verre pouvant resservir
2	Obtention du stamper « père » puis sa duplication
3	Obtention des stamper « mère » et « fils », puis l'emploi des « fils » pour la phase de réplcation

fig. I.6 Résultat d'un « Stamper making »

II.3 Replication

Est la phase qui nous fera obtenir des « réplcas » ; soit des disques finis et prêts à l'emploi à partir d'un « stamper », la phase de réplcation comprend quatre procédés individuels que sont :

1. *Injection moulding* : Injection de matière plastique sur un stamper puis son moulage, on en obtient un « *disque moule* ».
2. *Sputtering* : dépôt d'une couche réfléchive (Aluminium) sur le disque moule.
3. *Lacquering* : dépôt d'une couche de laque sur le même disque moule pour protéger la couche réfléchive.
4. *Scanning* : Inspection des défauts présents sur le disque moule obtenu. (L'objet de l'étude). [4]

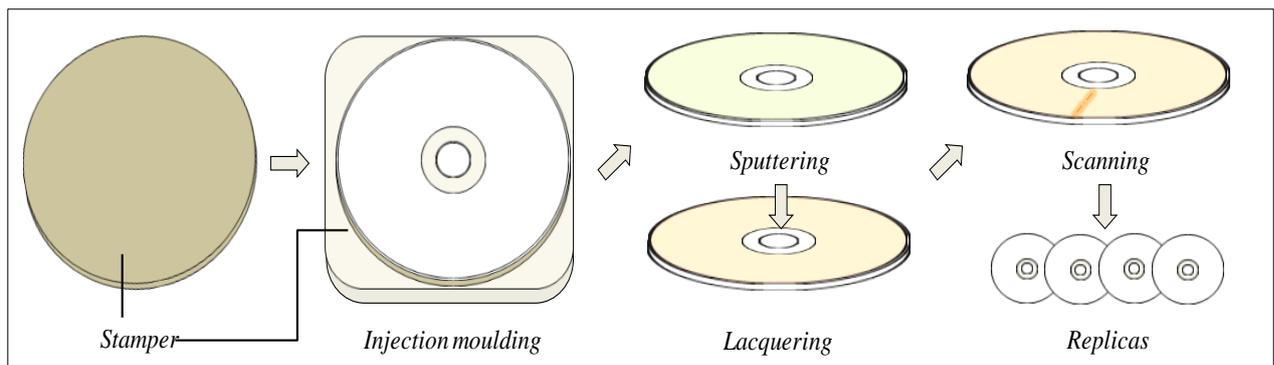


fig. I.7 Vue sur les étapes de la phase « Réplcation »

II.4 Packaging

Les répliques étant des disques finis et prêts à une éventuelle commercialisation, la dernière phase prend en charge l'étiquetage et l'emballage de ces derniers. [4]

III. Le procédé d'inspection d'un disque optique « DVD »

Le procédé d'inspection rentre dans le « contrôle qualité » et est très important car il détermine le rendement d'un processus de production et par conséquent son *succès* ou son *échec* ; le traitement s'appuie sur des détections d'erreurs *physiques* mesurables sur un disque optique et fait appel aux *classifications* pour les déterminer; tout en sollicitant, et cela de manière implicite des notions d'apprentissage.

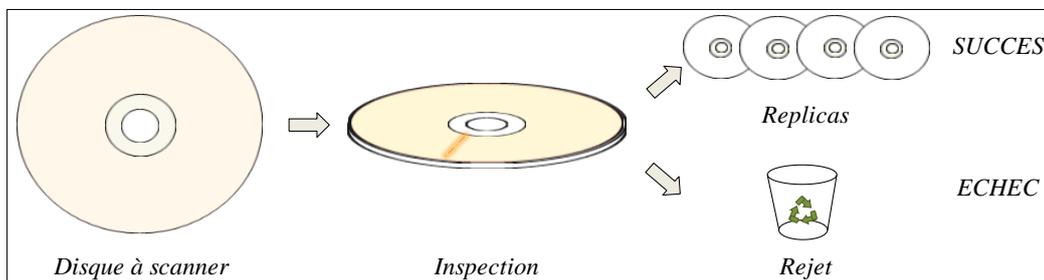


fig. I.8 Vue sur le concept de l'inspection

Le procédé en question s'appuie sur une machinerie dédiée à ce travail (Unité d'inspection) qui est liée via une connectique réseau à un poste PC du technicien sur lequel le logiciel d'inspection est installé.

III.1 Inspection de surface

La première étape du *scanning* consistera à projeter un faisceau lumineux « optique » statique à partir d'un émetteur (Lampe halogène) sur un disque placé sur une plaque tournante. La lumière émise sur le disque optique traité sera interceptée sous réflexion par un récepteur qui est une caméra ; qui dans un premier temps convertira les signaux lumineux en signaux vidéos avant de générer une cartographie du disque traité sur l'interface du technicien. [5]

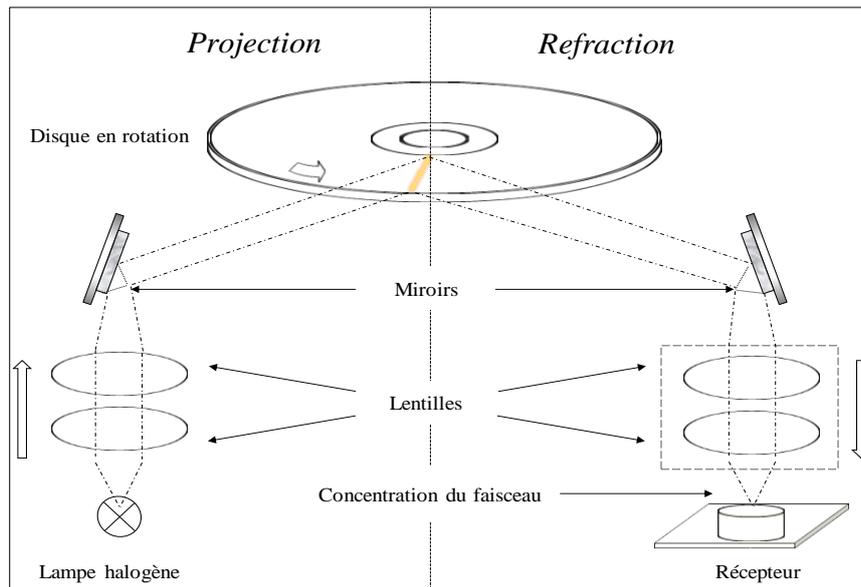


fig. I.9 Vue sur l'inspection de surface d'un disque optique

Il est important de préciser que l'imagerie censée être l'objet d'études et de déduction pour l'inspection de surface se base sur les nuances de gris et que les défauts générés lors de l'inspection sont pour une majorité des différences en contrastes. La figure suivante permettra de savoir comment s'opère l'identification de l'erreur (ou défaut) sur la surface d'un disque optique. [5]

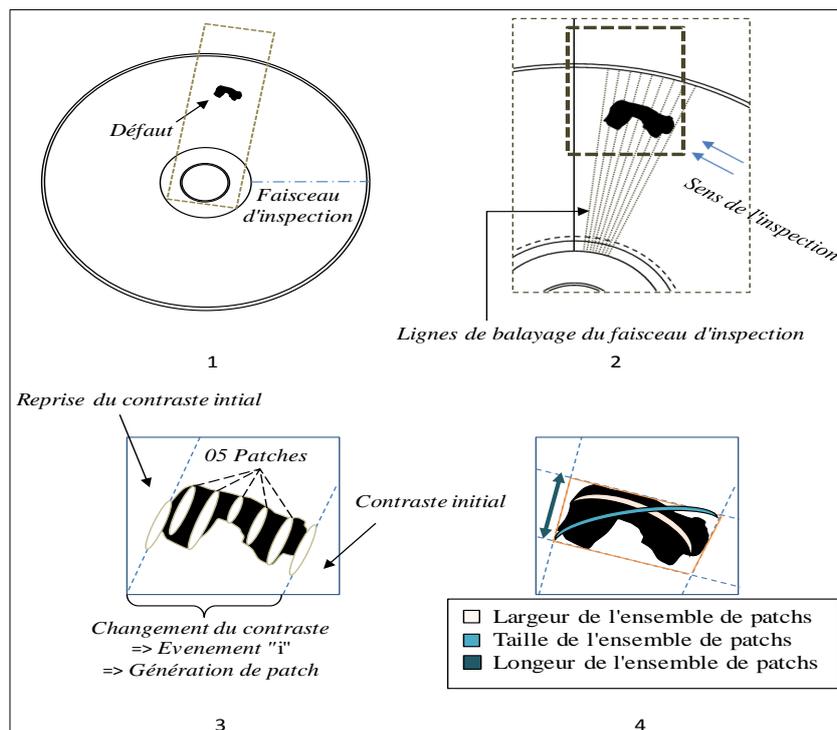


Fig. I.10 Vue sur les étapes de l'inspection de surface d'un disque optique

Le descriptif des étapes de l'inspection du schéma *I.10* sont les suivants :

- 1- Le disque est en rotation et le faisceau d'inspection émis par la machine est sur le point d'identifier un défaut.
- 2- Le balayage du faisceau d'inspection (amas de vecteurs) recouvre une surface partielle du disque jugée anormale.
- 3- Le changement du contraste indique une anomalie et une erreur potentielle, chaque vecteur positionné sur le changement de contraste devient un événement qui enclenche la génération d'un *patch* (une trace) et cela jusqu'à la reprise du contraste initial.
- 4- Une fois tous les patches générés ; le système relie les points les plus éloignés des patches, ce qui finit par matérialiser une surface géométrique qui englobe de près le défaut à étudier.

Un calcul s'opère de suite au sein de l'unité d'inspection afin d'évaluer différentes valeurs de plusieurs paramètres (détaillés à la prochaine section) d'un disque optique, ensuite survient un appariement à une base d'erreur (existante), puis, en fonction de la pertinence des résultats, un défaut en guise d'erreur sera attribué en tant que classe. [5]

Une fois l'évaluation faite, un fichier log « *.txt » sera généré comme rapport ; contenant au final les informations propres aux erreurs évaluées par l'unité d'inspection.

tab. I.1 Extrait d'un fichier log lors d'une inspection en surface d'un disque DVD

Def #	Typ	Pat	Scan (°)	Radius (mm)	Size (mm)	Width (mm)	Length (mm)	Slimf	Attribut
11	NO	M	356	32.520	0.076	0.063	0.043	1	b
51	NO	P	353	24.869	0.045	0.042	0.016	2	t
27	NO	S	353	48.181	0.090	0.063	0.064	1	b,eg
08	NO	S	168	27.398	0.083	0.063	0.054	1	t,eg
11	I	I	182	24.126	0.171	0.170	0.016	7	t,eg
14	NO	H2	191	28.014	0.112	0.063	0.056	3	t,e

III.1.1 Liste des erreurs potentielles sur un disque optique DVD

La base de données contenant toute la liste des erreurs potentielles dans un disque est large car elle prend en charge tous les types de disques optiques, nous nous contenterons dans ce mémoire de présenter les erreurs les plus répandues concernant les disques optiques DVD (5) ; celles obtenues des fichiers log sur lequel ce procédé est effectué.

La liste suivante comprendra un ensemble d'erreurs référencées, de par leurs aspects, leur description ainsi que l'origine potentielle de l'erreur dans un système M.E.S.

Il est à noter que la remédiation d'un défaut en vrai est laborieuse et consiste la plupart du temps à effectuer des vérifications matérielles ainsi qu'un paramétrage poussé qui fait appel à plusieurs notions et disciplines. La présentation des défauts ci-dessous est traduite du Manuel de l'unité d'inspection.

L'erreur « H2 » :

Aspect : Défaut bilatéral, cavité, inclusion

Source : Injection, liaison

Description : Défaut apparent aux mêmes coordonnées sur les deux substrats d'un même disque optique DVD.

- La cavité observée témoigne l'absence d'une partie d'une couche normalement induite.

L'erreur « I » :

Aspect : Injection, mont Fuji.

Source : Moulage.

Description : Traces linéaires, rainures en formes d'aiguilles.

L'erreur « M » :

Aspect : Inclusion, tâches sombres, petites traces noires.

Source : Injection, liaison, moulage.

Description : tâche noire incluse au polycarbonate. [5]

L'erreur « N » :

Aspect : vide de liaison.

Source : Liaison

Description : Sur un disque optique DVD, l'erreur « N » n'est observable qu'à cause de la déformation présente sur la surface du disque.

L'erreur « P » :

Aspect : Trace métallique, trace scintillante

Source : Moulage, liaison

Description : D'apparence brillante, cette erreur est en générale présente sur la couche d'information.

L'erreur « S » :

Aspect : Trace de surface, abrasion

Source : Moulage, Maniement.

Description : Tâche sombre à haut contraste.

L'erreur « T » :

Aspect : Grattement de surface.

Source : Maniement.

Description : Traces linéaires sombres à haut contraste.

- L'orientation des traces peut-être aléatoire.

La figure suivante illustre les erreurs décrites précédemment (captures en nuances de gris) ainsi que la légende propre aux erreurs présentes sur l'interface du simulateur. [5]

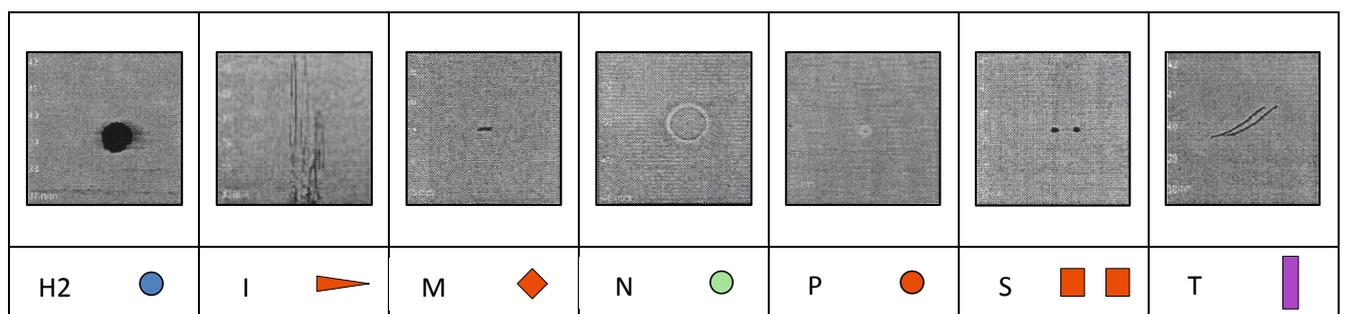


fig. I.11 Capture réelle des défauts détectés lors d'une inspection ainsi que leurs nominations respectives et leur légende sous simulateur.

III.2 Inspection de la déviation d'un disque optique

Lors de la manufacture d'un disque optique, rien ne garantit qu'un disque soit plat, l'étude d'une déviation sera fondamentale car elle permettra de vérifier si le disque est déformé ou non, en supposant que ce soit le cas ; la lecture du disque, suivant un *seuil de tolérance* sera (peut-être) compromise et il y'aura dès lors une perte d'information. [6]

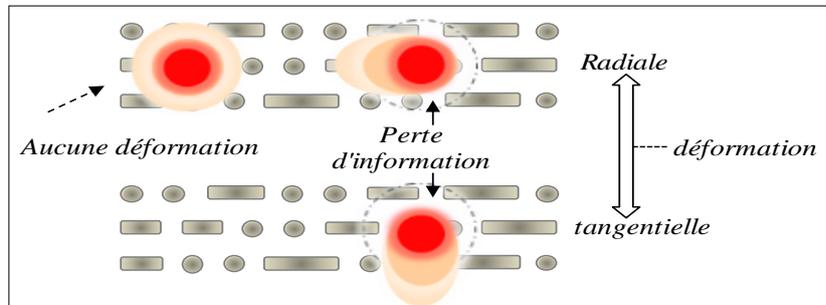


fig. I.12 Figure des différentes déformations possibles sur un disque optique [6]

La figure précédente stipule deux types de déformations : radiale et tangentielle, qui, au sein de l'inspection sont mesurables. Lors du test de la déviation qui implique une projection d'un faisceau d'évaluation ; la déformation peut être interprétée par une géométrie basique. [6]

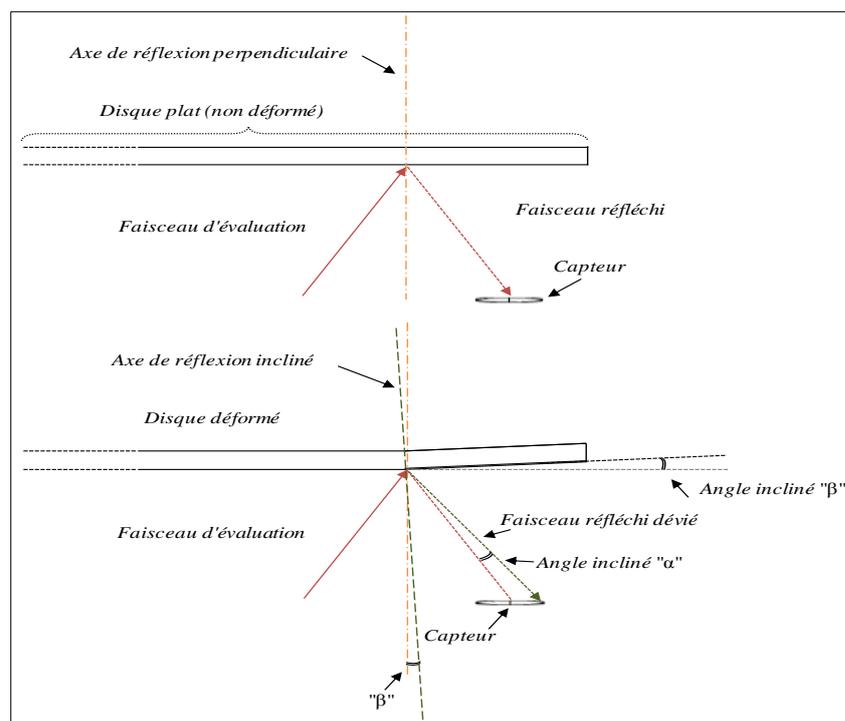


fig. I.13 Interprétation géométrique d'une déformation sur un disque optique

L'inspection en question évaluera la projection du faisceau sur trois à quatre longueurs (rayons) différentes : 23 mm, 37 mm, 47mm et 57 mm. [6]

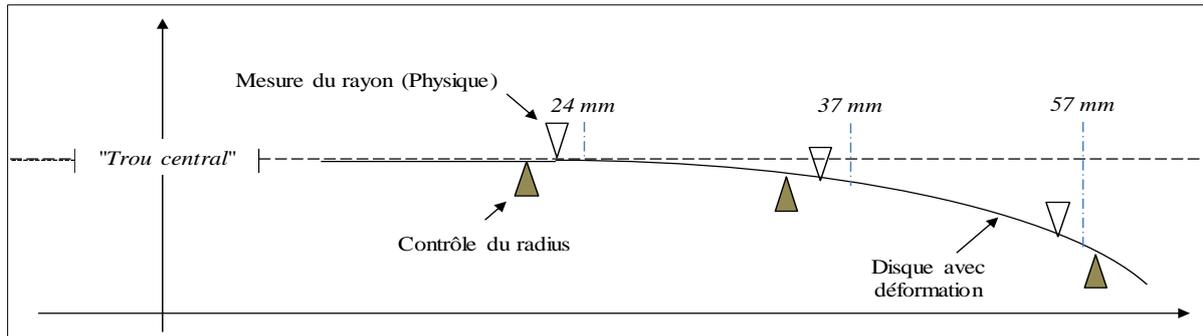


fig. I.14 Mesure radiale de la déviation

Les calculs s'effectuent comme il suit :

$$\text{Déviation radiale} = (0.03^\circ \pm 2\% \text{ de la mesure réelle})$$

$$\text{Déviation tangentielle} = (0.02^\circ \pm 2\% \text{ de la mesure réelle})$$

La mesure réelle est une métrique qui détermine la profondeur de la déviation.

Selon les normes de fabrication d'un disque optique DVD, le seuil toléré maximal lors d'une déviation radiale est de $\pm 0.08^\circ$, quant au tangentiel, sa valeur est nettement inférieure ; elle est estimée à $\pm 0.03^\circ$. Les valeurs de l'inspection en déviation sont elles aussi disponibles sous fichiers log. [6]

Conclusion

Ce chapitre a recensé toutes les informations explicatives du processus général de la conception d'un disque optique. Le prochain chapitre mettra en avant les ressources qui permettront d'introduire le concept de « l'intelligence artificielle » ; à savoir les paramètres qui serviront à un éventuel apprentissage ainsi que les différentes approches théoriques qui permettront de justifier la conception ultérieure d'un simulateur.

CHAPITRE II

Notion d'apprentissage et approches théoriques

Introduction

L'unité d'inspection des disques optiques « *Dr. Schenk* » opère sa vérification suivant un algorithme qui lui est propre. Cependant, les résultats fournis à partir des fichiers *d'erreurs* pointés précédemment, permettront de cerner approximativement une approche (une méthode apprentissage) ; *qui aurait pu* implicitement servir à élaborer la solution sollicitée par l'unité d'inspection, cela permettra de converger ou mieux, d'optimiser le rendement d'une classification par rapport aux résultats présents sur les rapports par le biais des prédictions obtenus lors d'une phase d'apprentissage.

Le choix de la méthode ne pouvant être aléatoire compte tenu du nombre conséquent de classificateurs existants tout comme l'adéquation de la méthode, et cela en fonction des données de ce projet. Ce chapitre est un préquelle à la sélection de l'algorithme du simulateur, et s'intéressera en premier lieu à mettre en lumière quelques algorithmes d'apprentissages assimilés, de par leur définition globale et leur mode opératoire.

I. Définition de l'ensemble des paramètres servant à l'apprentissage

Avant de s'intéresser aux algorithmes d'apprentissages, il est préférable dans un premier temps de définir ce qui les caractérise, à savoir les paramètres.

Le contenu du tableau *I.1* figurant au chapitre précédent (Page 10) dispose du contenu suivant : des champs dont les colonnes sont en fait les paramètres servant à une inspection de surface d'un disque optique.

Chaque paramètre disposant d'un rôle précis ; en définissant la fonction de chacun d'eux, il nous sera dès lors possible d'assigner ce qui doit être classifié (sortie) et ce qui aidera pour ce faire (entrées).

La table suivante est traduite du manuel de l'unité d'inspection et renvoie le rôle de chaque paramètre.

tab. II.1 L'ensemble des paramètres de l'inspection en surface avec leur descriptif.

Entête	Signification
<i>Def #</i>	Indique l'indice du défaut au sein d'un disque traité.
<i>Typ</i>	Affiche le défaut local sur le disque traité, initialement en valeur « NO », la valeur de <i>typ</i> ne change que quand le seuil géométrique du défaut affiché en <i>Pat.</i> est dépassé.
<i>Pat.</i>	Affiche le défaut évalué par l'ensemble des patches générés lors de l'inspection.
<i>Scan</i> °	(En degrés) : Evalue l'angle entre la position initiale du faisceau avant l'inspection et la détection du <i>Pat.</i> traité.
<i>Radius</i>	(En mm) : Est une métrique évaluant la longueur du rayon du <i>Pat.</i> traité.
<i>Size</i>	(En mm) : Est une métrique évaluant la « taille » du <i>Pat.</i> traité.
<i>Width</i>	(En mm) : Est une métrique évaluant la « largeur » du <i>Pat.</i> traité.
<i>Length</i>	(En mm) : Est une métrique évaluant la « longueur » du <i>Pat.</i> traité.
<i>Slimf</i>	Est un calcul ratio ($Size / Length$), le résultat affiché est l'arrondissement de la valeur obtenue.
<i>Attribute</i>	Réfère à une notion en optique qu'est « l'écho » (<i>duplication d'un défaut à proximité du même défaut lors d'une inspection</i>). Elle comprend quatre types d'erreurs, deux d'entre elles pointent « l'écho » en question (e, eg), les deux autre indiquent le substrat (Actif ou passif) en lequel se situe ledit défaut dupliqué. (b, t).

I.1 Observations

- Dans le précédent chapitre, il est stipulé que l'inspection de la surface d'un disque optique se focalise en priorité sur la délimitation et l'identification d'un défaut en ayant recours aux ensembles de patches. Suivant les données recensés dans le tableau ; le paramètre *Pat.* est celui qui correspond à cet effet, en conséquence ; ce paramètre fera office de « sortie ».

- Les entêtes *Def#* et *Scan°* recensent des informations *abstraites* par rapport à la classification ; la première renvoie un classement de l'ordre de rencontre d'une erreur sur un disque et la seconde une coordonnée angulaire, les utiliser lors d'une classification pourrait fausser l'opération, dès lors; il est préférable de ne pas les prendre en compte.

- Le contenu du paramètre en sortie *Pat.* contient la liste d'erreurs présentes sur un disque optique (Chapitre I, fig.I.II, page 12), le nombre de ces derniers n'étant pas binaire, en conséquence, les algorithmes en « arbres » ne seront pas employés pour ce projet.

II. Définitions des algorithmes à tester

- Les algorithmes désignés dans cette section ont été choisis en fonction de leur simplicité que ce soit en utilisation ou en interprétation par rapport aux résultats obtenus lors d'une classification. Les méthodes choisies et qui seront expérimentées sont les suivantes

II.1 Réseau de neurones

Un réseau de neurones artificiels (ANN), généralement appelé réseau de neurones (NN), est un modèle mathématique ou un modèle de calcul qui est inspiré par la structure et les aspects fonctionnels de réseaux de neurones biologiques. Un réseau de neurones est constitué d'un groupe interconnecté de neurones artificiels, et traite l'information en utilisant une approche connexionniste de calcul.

Dans la plupart des cas, un ANN est un système adaptatif qui modifie sa structure basée sur des informations externes ou internes qui s'acheminent à travers le réseau au cours de la phase d'apprentissage.

Les réseaux de neurones modernes sont généralement utilisés pour modéliser les relations complexes entre les entrées et les sorties ou pour élaborer des modèles au sein des données. [7]

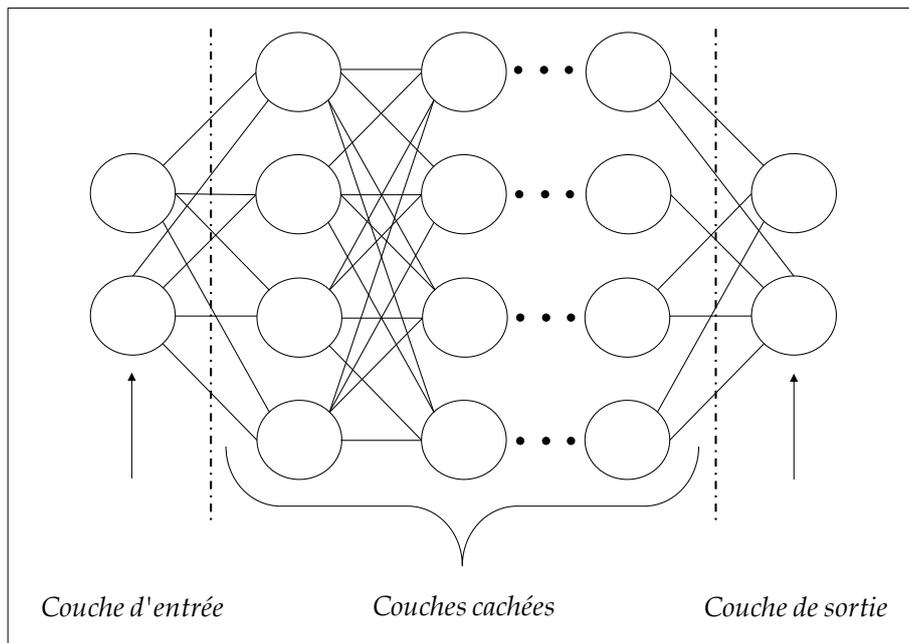


fig. II.1 Exemple d'un réseau de neurones

II.2 K plus proche voisin

L'algorithme « le plus proche voisin k » (*k-Nearest Neighbour*) est une méthode de classification des éléments à partir des exemples d'apprentissage. k-NN est un type d'apprentissage paresseux où la fonction est seulement approchée localement et tout le calcul est reporté à la classification. L'algorithme du k plus proche voisin est parmi les plus simples de tous les algorithmes d'apprentissage machine: un objet est classé par un vote à la majorité de ses voisins, dans le but d'être affecté à la classe la plus commune parmi ses k plus proches voisins (k est un entier positif, généralement de petite taille). Si $k = 1$, alors l'objet est simplement attribué à la classe de son plus proche voisin. Les voisins sont pris à partir d'un ensemble d'éléments pour lesquels la

classification correcte est connue. Cela est considéré comme l'ensemble d'apprentissage de l'algorithme. [8]

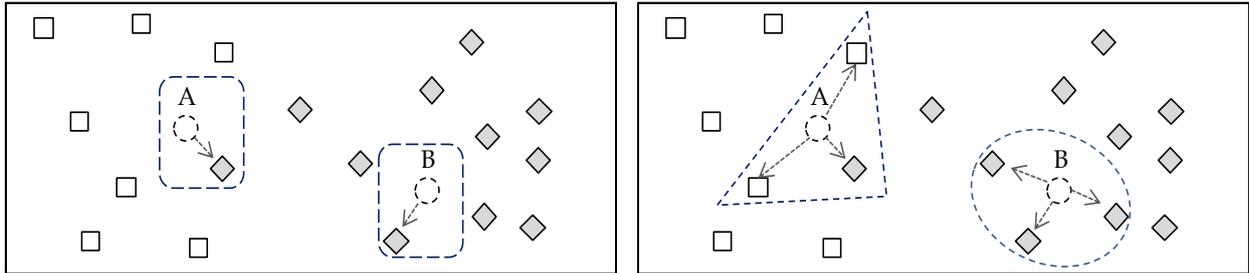


fig. II.2 Classification de deux classes avec K-NN, à gauche K=1, à droite K=3

II.3 Naïve Bayes

Est une classification Bayésienne probabiliste simple basée sur le théorème de Bayes avec une forte indépendance (dite naïve) des hypothèses. Elle met en œuvre un classifieur bayésien naïf, ou classifieur naïf de Bayes, appartenant à la famille des classifieurs Linéaires.

En termes simples, un classifieur bayésien naïf suppose que l'existence d'une caractéristique pour une classe, est indépendante de l'existence d'autres caractéristiques.

Malgré leur modèle de conception « naïf » et ses hypothèses de base extrêmement simplistes, les classifieurs bayésiens naïfs ont fait preuve d'une efficacité plus que suffisante dans beaucoup de situations réelles complexes [9]. En 2004, un article a montré qu'il existe des raisons théoriques derrière cette efficacité inattendue [10]. Toutefois, une autre étude de 2006 montre que des approches plus récentes (arbres renforcés, forêts aléatoires) permettent d'obtenir de meilleurs résultats [11].

L'avantage du classifieur bayésien naïf est qu'il requiert relativement peu de données d'entraînement pour estimer les paramètres nécessaires à la classification, à savoir moyennes et variances des différentes variables. En effet, l'hypothèse d'indépendance des variables permet de se contenter de la variance de chacune d'entre elle pour chaque classe, sans avoir à calculer de matrice de covariance. [9]

II.3.1 Théorie de la décision bayésienne

Soit les définitions suivantes :

- $W = \{w_1, w_2, \dots, w_s\} \rightarrow$ L'ensemble des s classes d'appartenance possibles.
 - $X = \{x_1, x_2, \dots, x_d\} \rightarrow$ Vecteur de paramètres décrivant une forme.
 - $P(x/w_j) \rightarrow$ Fonction de densité de probabilité de la classe w_j .
 - $P(w_j) \rightarrow$ Probabilité à priori de la classe w_j .
- Pour disposer d'une règle de décision chaque classe recherchée est caractérisée par une fonction discriminante : $g_i(x)$ pour $i=1 \dots s$.
 - Et une forme x est classée dans la classe w_i si $(g_i(x) > g_j(x))$ pour tout $j \neq i$.

Le classifieur obtenu calcule donc pour un vecteur x donné les fonctions discriminantes puis recherche le maximum qui permet de choisir la classe. Ce classifieur peut être schématisé de la façon suivante :

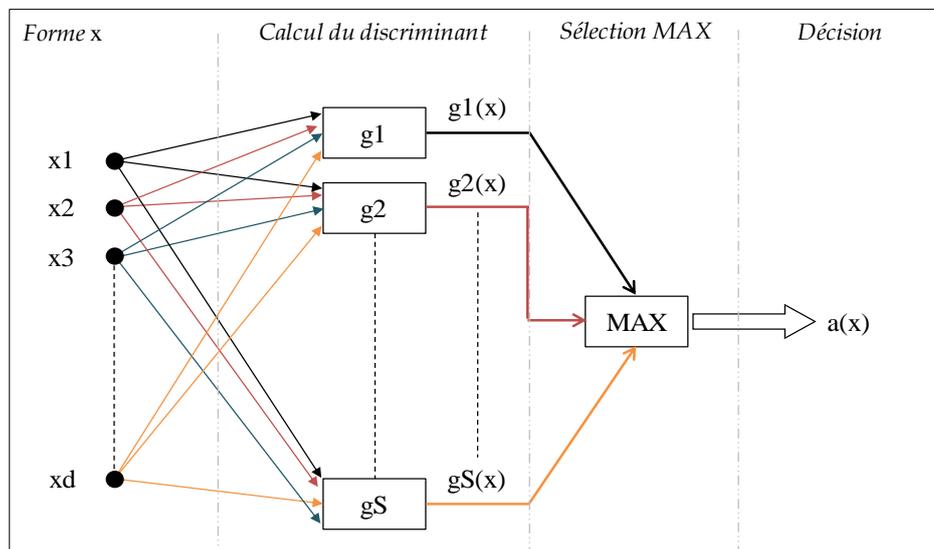


fig. II.3 Fonctions discriminantes et décisions

La fonction discriminante est définie comme suit :

- $G_i(x) = (x/w_i) P(w_i)$
- L'utilisation des s fonctions discriminantes pour réaliser la classification de forme amène à découper l'espace des observations (espace des paramètres x) en s régions $\{R_1, R_2, \dots, R_s\}$.

- Chaque région est définie comme suit : Si $g_i(x) > g_j(x)$ pour tout $(j \neq i)$ alors x appartient à R_i , donc assigner x à w_i .

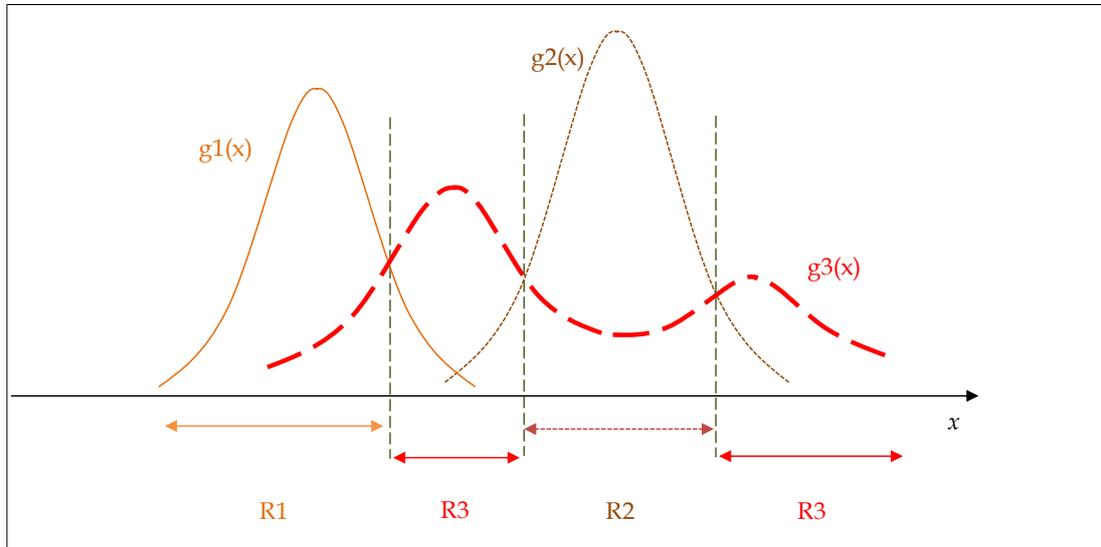


fig. II.4 Frontières des décisions dans l'espace des observations

III. Evaluations des algorithmes d'apprentissages

Dans l'ensemble un algorithme d'apprentissage est qualifié d'*efficace* que si son rendement s'avère *bon*, ce rendement se traduit par des performances ; qui par des calculs matriciels sont mesurables.

III.1 Mesures de performances d'un algorithme d'apprentissage

Il est fondamental de préciser que la mesure de performance d'un algorithme d'apprentissage fait appel à trois critères : *le rappel de classe*, *la précision de classe* et *le taux de classification*.

III.1.1 Le rappel d'une classe

(*Class recall* en anglais) Le rappel est un taux défini par le nombre de résultats bien classés par rapport au nombre de résultats pertinents que possède la base de données. [12]

$$\text{Rappel } i = \frac{\text{Résultats correctement attribués à une classe } i}{\text{Nombre de résultats attribués à une classe } i}$$

III.1.2 La précision d'une classe

(*Class precision* en anglais) La précision est un taux défini par le nombre de résultats bien classés par rapport nombre total des résultats que possède la base de données. [12]

$$\text{Précision } i = \frac{\text{Résultats correctement attribués à une classe } i}{\text{Nombre de résultats appartenant à une classe } i}$$

III.1.3 Le taux de classification

(*Accuracy* en anglais) est un taux global qui s'interprète comme étant le degré de convergence des prédictions d'un apprentissage par rapport aux données réelles. Elle se calcule en générale comme incertitude, la marge qui la compose est le degré auquel les mesures répétées dans des conditions inchangées affichent les mêmes résultats. [12]

Sous *RapidMiner*, le calcul du taux de classification suit la norme *ISO 5725-1*.

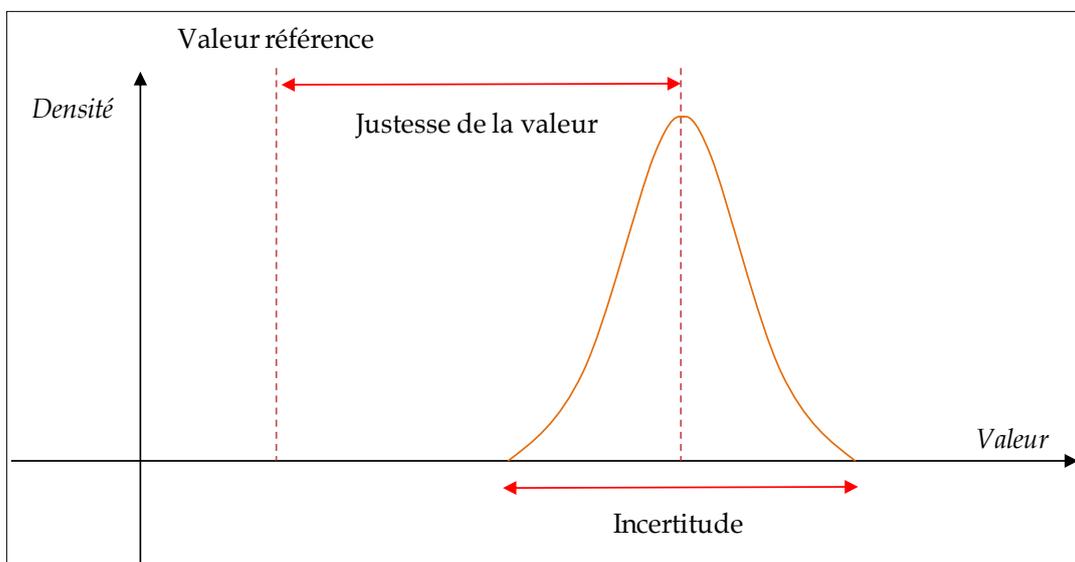


fig. II.3 Calcul du taux de classification sous la norme *ISO 5725-1*

Conclusion

Les algorithmes étant définis tous comme leurs critères de performances en général, il ne reste plus qu'à les tester, et choisir la méthode qui sied le mieux au besoin du projet.

Le prochain chapitre s'approfondira en détail sur le procédé d'apprentissage et les performances individuelles de chaque algorithme avant d'entamer une vue sur l'application.

CHAPITRE III

Choix de l'algorithme,
conception du
simulateur et
discussions des
résultats

Introduction

Avant d'entamer une quelconque implémentation, nous devons fixer un choix sur une méthode parmi les trois cités précédemment, la meilleure stratégie à adopter est d'avoir recours à une étude comparative par le biais d'un environnement d'apprentissage, d'analyse de données et de prédictions. L'application choisie à cet effet est *RapidMiner*. S'en suivra une présentation détaillée des fonctionnalités de l'application et une discussion des résultats retournés par le simulateur.

I. Expérimentations et discussion des résultats

I. 1 Expérience sous RapidMiner

L'expérience prendra en compte une base de données comme entrée qui servira de base d'apprentissage. La base contient dans cette expérience 569 entrées, nous ne sélectionneront pas manuellement une base de test, car chaque base de test choisie dispose de sa propre pertinence et favorisera à coup sûr un algorithme d'apprentissage.

Dans un but d'estimation de fiabilité d'une méthode de classification, la meilleure approche à adopter est *un échantillonnage multiple* ; ce qui correspond à une « *validation croisée* ».

L'opération se déroulera comme suit :

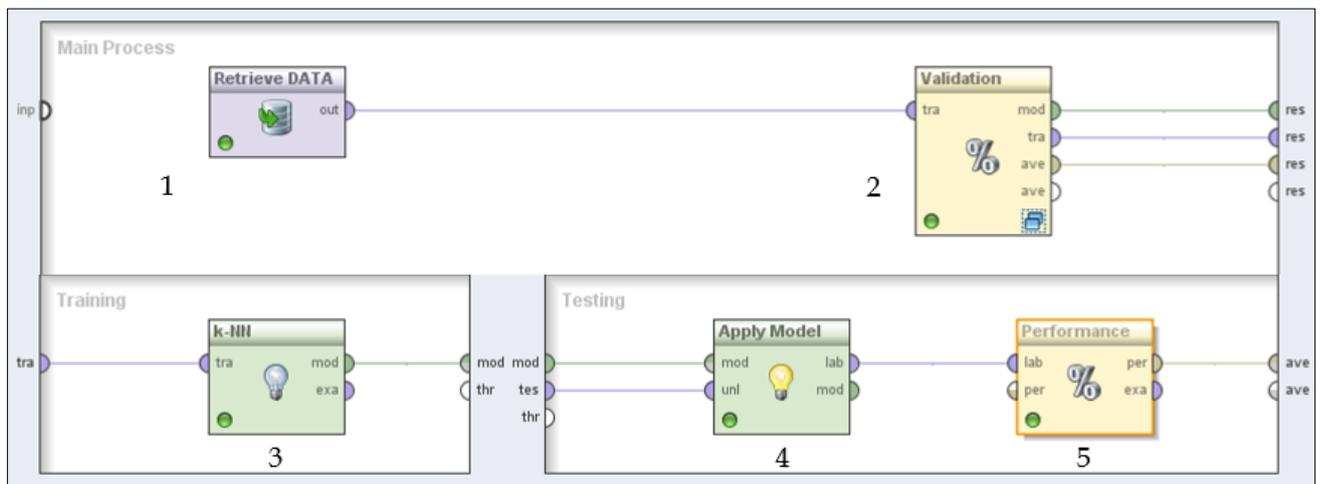


fig. III.1 Schéma procédurale de l'apprentissage sous RapidMiner

La numérotation présente sur le schéma ci-dessus se définit comme suit :

Procédé principal (Main process) :

- 1- Représente la base de données d'apprentissage choisie.
- 2- Cet opérateur effectue une validation croisée (10-fold) pour évaluer les performances sur toute la base de données.
 - L'opérateur de la validation croisée sur *RapidMiner* est une procédure en soi ; elle se divise en deux parties et comprend ce qui suit:

Partie apprentissage (Training) :

- 3- L'opérateur en question est l'algorithme de classification sur lequel va se reposer l'apprentissage (k-NN dans cet exemple).

Partie Test (Testing) :

- 4- Cet opérateur modélise l'apprentissage reçu et le sauvegarde dans un premier temps, ce dernier sera ensuite sollicité dans la prédiction d'une base de test.
- 5- Cet opérateur est utilisé pour estimer la performance d'une méthode d'apprentissage, il fournit une liste de valeurs des critères de performance.

Exemple :

La figure du dessous est une capture dudit résultat (*K-NN dans cet exemple*) :

accuracy: 93.33% +/- 2.57% (mikro: 93.32%)								
	true S	true M	true H2	true P	true N	true T	true I	class precision
pred. S	118	13	0	0	0	0	1	89.39%
pred. M	15	284	0	1	0	0	0	94.67%
pred. H2	0	0	59	0	0	0	1	98.33%
pred. P	0	3	0	12	0	0	0	80.00%
pred. N	0	0	0	0	50	0	0	100.00%
pred. T	0	0	0	0	0	6	2	75.00%
pred. I	1	0	0	0	0	1	2	50.00%
class recall	88.06%	94.67%	100.00%	92.31%	100.00%	85.71%	33.33%	

fig. III.2 Résultats des performances de l'algorithme « K-NN » sous RapidMiner

Il est à noter que le nombre d'erreurs appartenant à une classe au sein de la même base d'apprentissage correspond à la somme des valeurs d'une même ligne

$$\text{Nb d'erreurs appartenant à une classe } C = \sum_{i=1}^{\text{nombre de classe}} \text{Valeurs ligne Classe } C$$

On dénombre sept types d'erreurs qui font office de classes, le nombre d'entrées des classes est dans la table suivante :

tab. III.1 Nombre d'erreurs de chaque classe au sein de la base d'apprentissage

Types d'erreurs	S	M	H2	P	N	T	I
Nombre d'erreurs	132	300	60	15	50	8	4

I. 2 Paramétrage des algorithmes

Avant d'amorcer un quelconque apprentissage, les mesures de performances d'un algorithme affichées sont intrinsèquement liées aux valeurs qui leur ont été attribuées, il est de ce pas fondamental de paramétrer chaque spécificité d'un algorithme afin d'en maximiser le rendement pour une même base d'apprentissage.

I.2.1 Réseau de neurones

Le paramétrage prendra en compte trois critères : *Le nombre d'époque*, *le pas d'apprentissage* et *le seuil epsilon*. La définition de ces trois critères ainsi que leur domaine de définition dans cette expérience sont les suivants :

I.2.1.1 Le nombre d'époques

Ce critère spécifie le nombre de cycle d'apprentissage pour un réseau de neurones. Lors d'une rétro-propagation, les valeurs en sorties sont comparées avec les réponses justes afin de calculer la valeur des fonctions d'erreurs. L'erreur est renvoyée à travers le réseau de neurones.

En utilisant cette information, l'algorithme ajuste le poids de chaque neurone afin d'en minimiser la valeur de l'erreur par de petits ajustements.

Le procédé se répète n fois, n peut être spécifié comme étant la valeur du critère [7].

Domaine de définition choisi : {100, 200, 300, 400, 500}.

I.2.1.2 Le pas d'apprentissage

Ce critère spécifie la valeur de l'ajustement à chaque étape. [7]

Domaine de définition choisi : {0.1, 0.2, 0.3}.

I.2.1.3 L'erreur epsilon

L'optimisation s'arrête lorsque la valeur de l'erreur dépasse la valeur d'epsilon. [7]

Domaine de définition choisi : $\{10^{-3}, 10^{-4}, 10^{-5}\}$.

Le graphe (sous *MATLAB*) obtenu est le suivant :

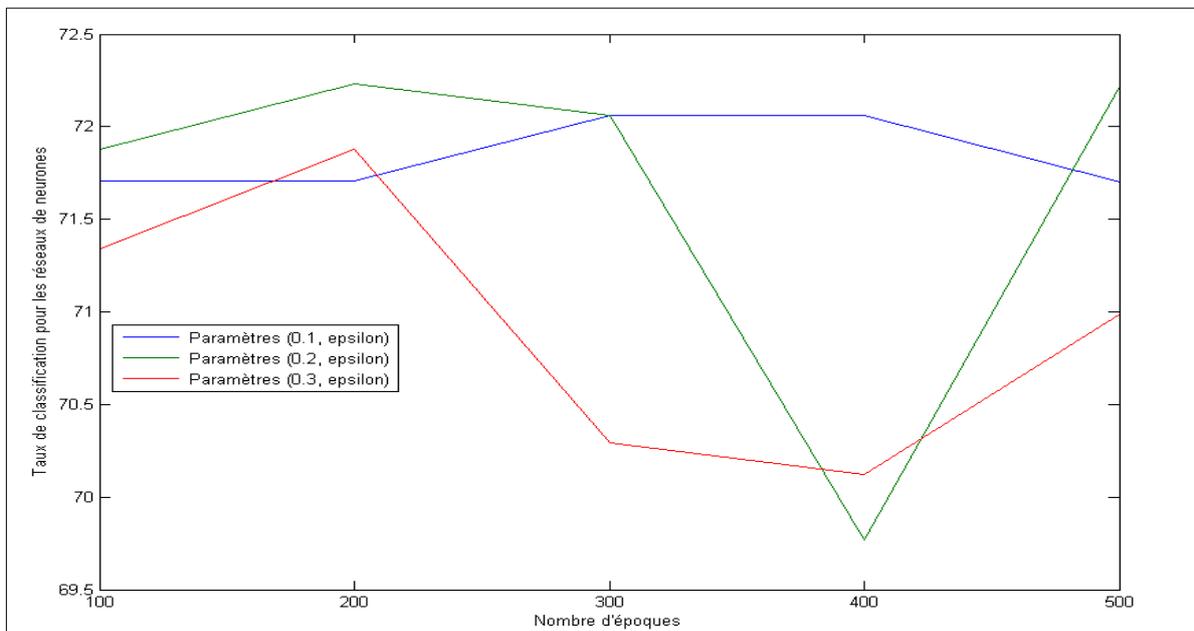


fig. III.3 Taux de classification des RN simulés par différents pas d'apprentissages

I.2.1.4 Observations

- Les plus grandes valeurs du taux de classification des trois graphes sont recensées dans la table suivante :

tab. III.2 Comparaison du taux de classifications des différents pas d'apprentissage.

	Graphe 1	Graphe 2	Graphe 3
Pas d'apprentissage	0,1	0,2	0,3
Nombre d'époques	[300,400]	200	200
Taux de classification (%)	72,06	72,23	71,88

- Toutes les valeurs d'épsilon n'ont nullement influencés l'apprentissage.
- Les critères choisis pour cet algorithme correspondront à la valeur du taux de classification la plus grande : $\{(0.2, 200, \epsilon)\}$.

I.2.2 K plus proche voisin

Le paramétrage concernera l'indice « K », le graphe suivant (Sous *MATLAB*) est une évaluation du taux de classification pour différentes valeurs attribuées à l'indice en question.

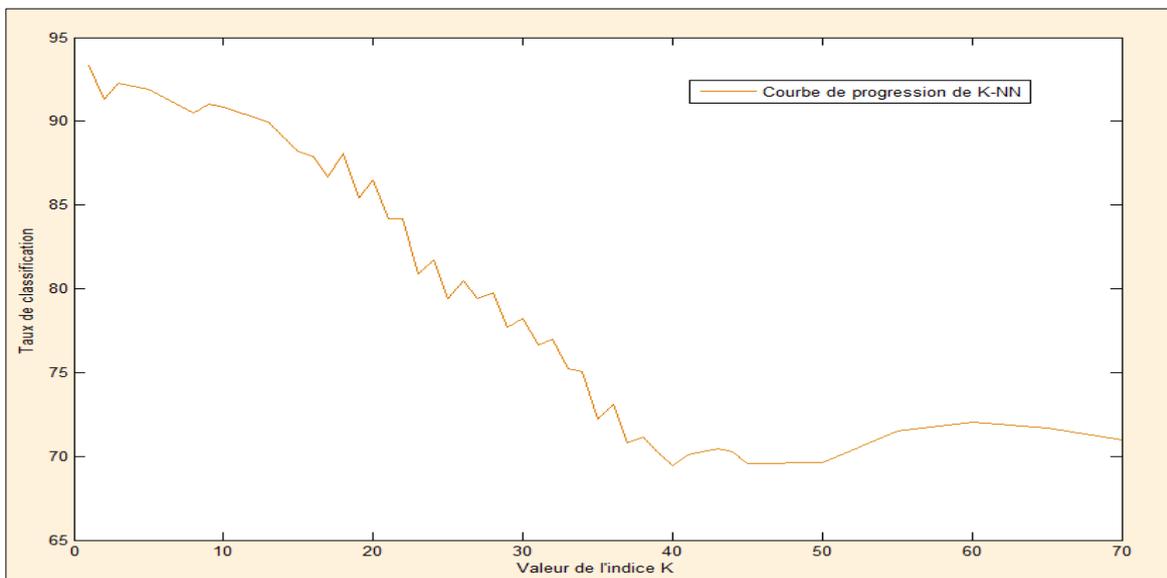


fig. III.4 Régression du taux de classification de K-NN

I.2.2.1 Observations

- L'indice K dispose d'une valeur initiale de 1, le graphe ci-dessus prend en compte les 70 premières valeurs de K.
- La table suivante recense les résultats les plus explicites de *la figure II.5*

tab. III.3 Valeurs tirées du graphe de régression des taux de classification K- NN

Valeur de l'indice K	1	10	20	30	40	50	60	70
Taux de classification (%)	93,33	90,87	86,47	78,21	69,94	69,61	72,06	71,01

- En modifiant l'indice K, la tolérance (valeur de la marge) peut augmenter mais au détriment d'un taux de classification diminué et d'une précision de classe dont la valeur peut s'avérer catastrophique ($\leq 1\%$).
- La valeur choisie pour la méthode *K-NN* correspond au plus grand taux de classification ($K=1$)

I. 2.3 Naïve Bayes

Comme il a été cité dans sa définition, *Naïve Bayes* adopte un calcul probabiliste (moyennes et variances), néanmoins, il n'existe aucun critère numérique à assigner sous *RapidMiner* pour cet algorithme, cependant, il y'a un critère qui permet d'occasionner selon la base d'apprentissage choisie une différence en taux de classification nommé *corrigé Laplace*.

I.2.3.1 Corrigé Laplace

Ce critère est sollicité pour éviter les influences des probabilités nulles, il peut à titre d'exemple ajouter une valeur infinitésimale aux valeurs concernées afin de conserver l'intégrité des calculs probabilistes tout en

éliminant l'existence des valeurs nulles. Il est à noter que son emploi est facultatif, même sélectionné, ce critère n'agit que quand il est requis [13].

I.2.3.2 Observations

- La table suivante recense les résultats obtenus :

tab. III.4 Valeurs des taux de classification sous Naïve Bayes

	Avec corrigé Laplace	Sans corrigé Laplace
Taux de classification (%)	78,03	78,03

- Le taux de classification demeure identique pour les deux cas.

Les critères des algorithmes choisis et l'apprentissage amorcé, les résultats d'apprentissage de tous les algorithmes en tenant compte des critères de performances figurent dans la table ci-dessous

tab. III.6 Tableau comparatif des résultats des différents algorithmes simulés

		Algorithme d'apprentissage choisi						nb entrées
		Naïve Bayes		k-NN		Réseau de neurones		
		Rappel / Précision	Rappel / Précision	Rappel / Précision	Rappel / Précision	Rappel / Précision	Rappel / Précision	
Résultats d'apprentissage	H2	100,0 %	100,0 %	100,0 %	98,33 %	100,0 %	98,33 %	60
	I	50,00 %	33,33 %	33,33 %	89,39 %	0,00 %	0,00 %	04
	M	75,67 %	83,46 %	94,67 %	94,67 %	99,00 %	66,00 %	300
	N	98,00 %	100,0 %	100,0 %	100,0 %	100,0 %	98,04 %	50
	P	100,0 %	31,71 %	92,31 %	80,00 %	00,00 %	0,00 %	15
	S	64,93 %	68,50 %	88,06 %	89,39 %	2,99 %	57,14 %	132
	T	85,71 %	50,00 %	85,71 %	75,00 %	14,29 %	100,0 %	08
Taux de classification		78,03 %	93,33 %	72,23 %				

I. 3 Observations

- Initialement, les trois algorithmes donnent de bons *taux de classification* (> 70 %).
- Le taux de classification le plus grand oriente vers « K-NN », néanmoins, il est préférable de ne pas fixer son choix directement; on doit au préalable vérifier les rappels de classes et les précisions de classes pour ces algorithmes et cela afin de choisir le meilleur compromis, le choix doit se porter sur l'algorithme le plus précis.
- Les résultats I et P pour l'algorithme des *réseaux de neurones* retournent des valeurs nulles, par défaut, cet algorithme n'est pas retenu.
- Majoritairement, la précision de classe favorise l'algorithme K-NN à 66,66 % (I, M, P, S). Idem pour le rappel de classe qui favorise le même algorithme à 60 % (M, N, S).

L'apprentissage étant un succès, il a été possible d'évaluer concrètement les critères de performance de chaque algorithme initialement choisi et d'assigner celui qui présente le plus d'affinité en termes de traitement par rapport à celui employé par l'unité d'inspection. La méthode K-NN sera directement implémentée et testée en y incorporant des fonctionnalités similaires à l'unité d'inspection dans un système M.E.S

I. 4 Expérience sous l'application

Avant d'expérimenter le simulateur, il est intéressant de s'intéresser à la structure de ce dernier ; de par son contenant comme son contenu.

I.4.1 Langage et environnement de développement

Pour le développement de cette application, le langage sollicité est JAVA avec l'IDE *NetBeans*. Ceci permet de développer un programme pouvant fonctionner sur différents environnements (*émulation possible*).

I.4.2 Descriptif des classes employées pour le simulateur

Dans cette partie nous allons citer les classes principales constituant l'application :

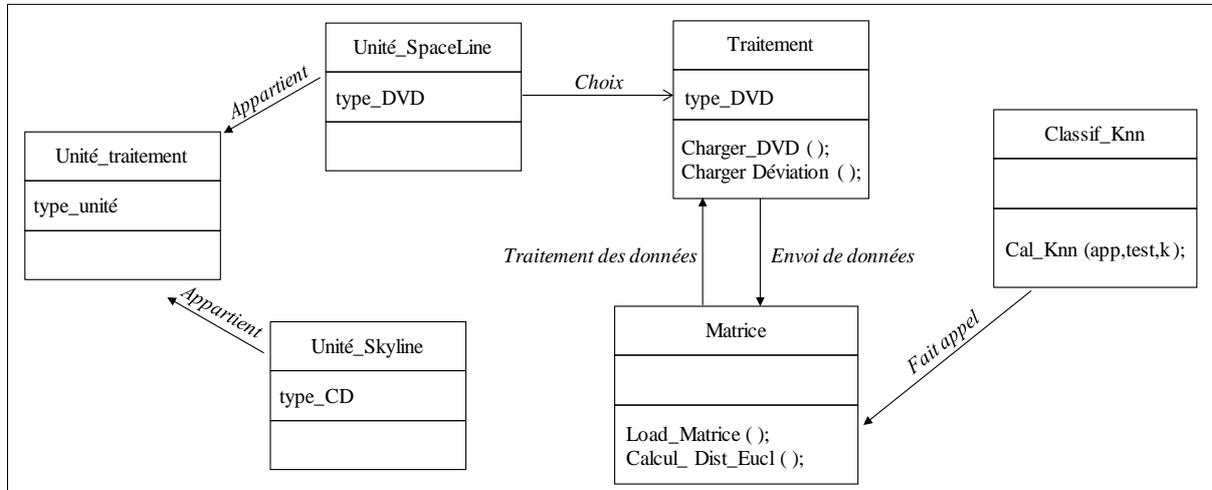


fig. III.5 Ensemble des classes servant à l'application

- *Classif_Knn* : Cette classe est utilisée pour la classification avec l'approche du K plus proche voisin.
- *Matrice* : Le classifieur précédent fait appel à cette classe pour faire les différentes opérations (avoir la classe ou les caractéristiques d'un cas donné, calcul de la pertinence d'une classe, et les différentes fonctions nécessaires pour les classifieurs).
- *Traitement* : C'est au sein de cette classe que s'effectue l'importation des données d'un disque pour un éventuel traitement.

I.4.2 Présentation de l'application

Il est nécessaire dans un premier temps de se familiariser avec les fonctions de l'application qui se veut similaire à l'interface de l'unité d'inspection « Dr. Schenk ». Les figures ci-dessous illustrent l'interface dans chaque étape, les numéros présents sur les figures sont expliqués directement après.

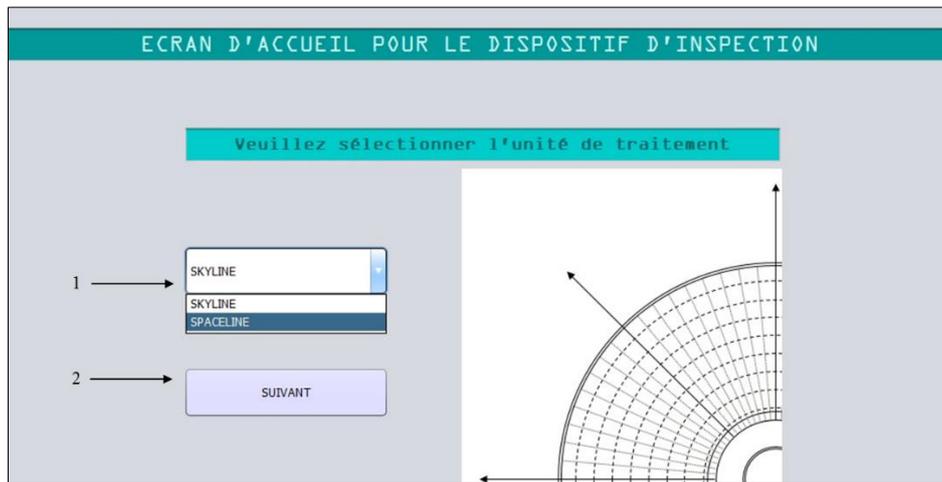


fig. III.6 Ecran d'accueil principal de l'application

1- Menu déroulant qui propose l'unité d'inspection à solliciter.

- *Skyline* : Renvoie l'unité d'inspection qui prend en charge les disques optiques de type CD.
- *Spaceline* : Renvoie l'unité d'inspection en charge des disques optiques de type DVD. (*Option choisie*)

2- Bouton qui pointe sur la prochaine section.



fig. III.7 Choix du type du disque à traiter sous l'unité d'inspection en Spaceline

1- Malgré une similitude des disques optiques DVD, chaque modèle dispose de sa propre architecture ; ce qui implique que la liste d'erreurs potentielles sur les différents disques n'est pas identique. Ce bouton fait sélectionner un disque et charge le procédé d'évaluation adéquat.

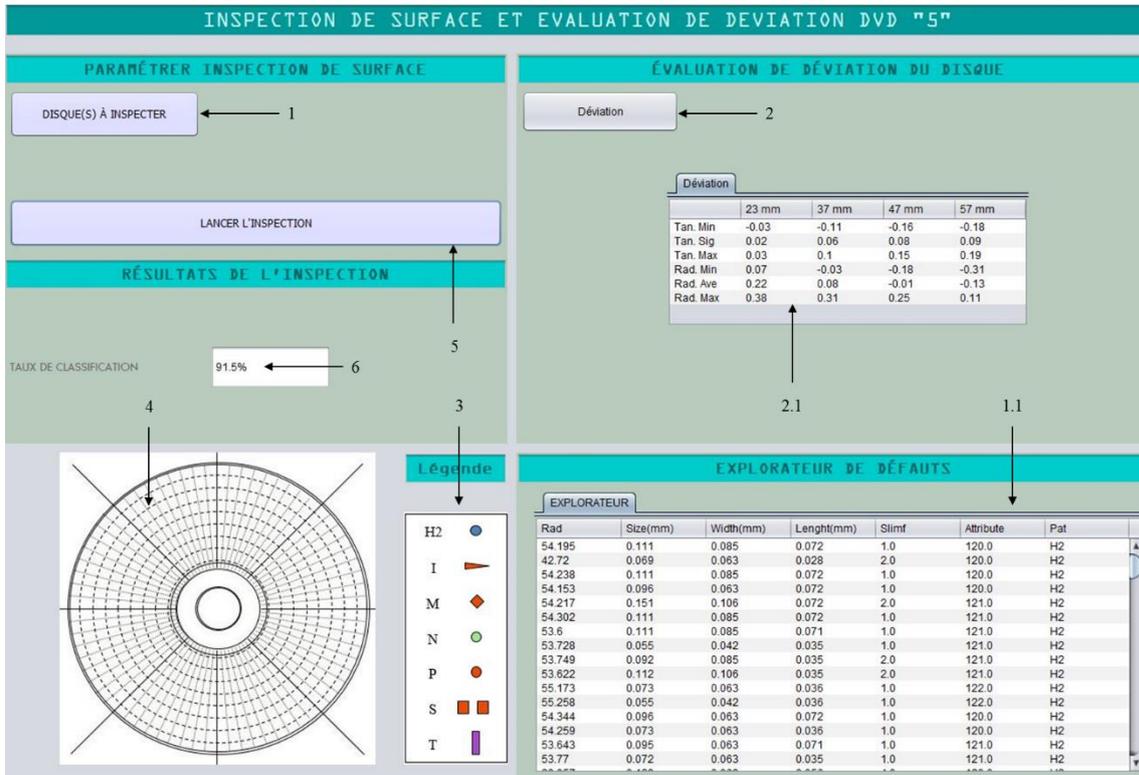


fig. III.8 Interface M.E.S de l'inspection de surface et de la déviation d'un DVD 5

- 1- Bouton servant à charger le disque à inspecter, en réalité, l'unité d'inspection charge une population de disques.
 - 1.1 Génère un ensemble d'informations (paramètres) sur l'explorateur de défauts pour un disque chargé.
- 2- Bouton qui charge les données de déviations à calculer.
 - 2.2 Renvoie les résultats de calculs de déviation pour un disque chargé.
- 3- Légende des erreurs potentielles présentes sur un disque optique « DVD-5 ».
- 4- Surface du disque sur lequel est censé apparaître l'ensemble des erreurs de l'explorateur de défauts en tant que de différences de contraste.
- 5- Amorces la classification de l'application.
- 6- Renvoie le taux de classification sous l'algorithme K-NN pour le disque traité.

I. 5 Observations

- Le taux de classification pour un disque choisi sous l'application est à 91,5 %.
- La base de test est un exemple à part entière (*un fichier log*), l'application n'est pas en *validation croisée*. L'apprentissage opéré converge aux résultats obtenus sous *RapidMiner*, ce qui dénote un certain succès quant à l'opération, la différence étant inférieure à 2%.
- Le calcul de la déviation d'un disque dispose de ses propres contraintes, indépendantes de l'inspection de surface mais reste sa suite logique durant le procédé de vérification. Si on suit les normes pour lesquels un disque optique est considéré comme déformé, (Chapitre I, page 14), les résultats seront comme suit :

tab. III.5 Vérification des valeurs de déviation sur un même disque optique

	23 mm	37 mm	47 mm	57 mm
Tan. Min	- 0.03	-0.11	-0.16	-0.18
Tan. Sig	0.02	0.06	0.08	0.09
Tan. Max	0.03	0.10	0.15	0.19
Rad. Min	0.07	- 0.03	-0.18	-0.31
Rad. Ave	0.22	0.08	-0.01	-0.13
Rad. Max	0.38	0.31	0.25	0.11

- Les cases en surbrillances contiennent des valeurs qui excèdent les normes, compte tenu de l'importante présence de ces dépassements, le disque est considéré comme déformé et inapte à la commercialisation.

Conclusion

Ce chapitre se conclue sur le succès de l'opération qui reposait initialement sur la reconnaissance des erreurs présentes sur un disque via un apprentissage en s'appuyant sur des approches pondérées de par leur traitement, puis un choix parmi ses approches et enfin sa démonstration sous l'application qui converge aux résultats de l'environnement d'apprentissage.

CONCLUSIONS
GÉNÉRALES ET
PERSPECTIVES

Le projet touche à sa fin et à sa clé une liste de perspectives et de conclusions qui pourraient élargir le champ de ce travail et aboutir en fin de compte à des résultats intéressants.

Une des améliorations à envisager serait d'étendre le champ expérimental des algorithmes testés jusqu'à maintenant, que ce soit dans un souci d'optimisation ou de recherche, c'est-à-dire ajouter d'autres algorithmes au tableau tel que les SVM (*Support Vector Machine*).

Les solutions qui n'ont pas été retenus pour l'application ne sont pas obsolètes, aucun de ces mêmes algorithmes n'a donné de mauvais résultats, en premier lieu, le rendement d'une classification dépend de l'ensemble de données à traiter (base d'apprentissage) que ce soit en nombre ou en structure avant même de les paramétrer, le travail qui a été fait ne reflète qu'une expérience parmi tant d'autres.

Le travail qui s'est fait jusqu'à maintenant avait un but fondamental, celui de démontrer le rôle des modèles d'intelligence et de décision au sein de l'industrie commerciale, de par leur existence, aussi subtile soit-elle, tout comme le fait est qu'elles soient sujettes aux améliorations, ce qui occasionnerait des gains non négligeables.

L'interface implémentée pour l'application disposait d'autres options, telles que la possibilité de recourir à une unité d'inspection en *Skyline*, et d'autres types de disques optiques DVD non prises en considération dans ce projet, il serait bon de combler ces fonctionnalités en étant conforme aux normes de leurs inspections.

Une autre amélioration et des plus grandes consisterait à agencer un système communicatif (*S.M.A*) entre plusieurs systèmes à base M.E.S, de collecter en temps réel l'ensemble des informations ainsi que restituer les performances des processus au sein d'une seule interface.

Ce projet n'a été rendu expérimentable que par le fait d'avoir eu recours à un environnement d'apprentissage, leur rapidité d'analyse et leur précision permet de saisir l'orientation et l'évolution d'une méthode de classification par rapport à une même base d'apprentissage, et au final; se fixer un choix.

RÉFÉRENCES

BIBLIOGRAPHIQUES,

WEB ET MANUELS

Références bibliographiques :

- [1] H. Meyer, F. Fuchs, K. Thiel, “*Manufacturing Execution System*”, McGraw-Hill Professional, 2009.
- [2] T. O’Kelly, “*Reference guide for Optical Media*” Memorex, 2008.
- [3] J.J. Wegener, “*L’enregistrement optique*”, GIS-DON, 2011.
- [8] Bermejo, S. Cabestany, “*J. Adaptive soft k-nearest-neighbour classifiers*”, Vol. 33, 2000.
- [10] H. Zhang, “*The Optimality of Naive Bayes*”. University of New Brunswick, FLAIRS, 2004.
- [11] R. Caruana, A. Niculescu-Mizil : “*An empirical comparison of supervised learning algorithms*”. Proceedings of the 23rd international conference on Machine learning, 2006.
- [12] J.R. Taylor. “*An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*”, University Science Books, 2000.

Références web :

- [4] Nom de l’article : “*Qu’est-ce qu’un disque optique ?*”
Lien URL : http://sandsgroup-dz.com/docs/WHAT_IS_A_CD_french_version.pdf
Année de publication : 2013.
- [9] Nom de l’article : “*Classification naïve bayésienne*”
Lien URL : http://fr.wikipedia.org/wiki/Classification_naïve_bayésienne
Date de publication : 09 Juin 2008.
Date de dernière révision : 10 Juillet 2013.

Manuels :

- [5] *Dr.Schenk* Industriemesstechnik, “*Operator Manual, Training Manual*”, 2005.
- [6] *Dr.Schenk* Industriemesstechnik, “*Operator Manual, ISM.xxx for VPC-Visualization*”, 2005.
- [7] *RapidMiner* (ver. 5.3.012) documentation, “*Synopsis of Neural network*”, 2012.
- [13] *RapidMiner* (ver. 5.3.012) documentation, “*Synopsis of Naïve Bayes*”, 2012.

ANNEXES

Réseaux de Neurones

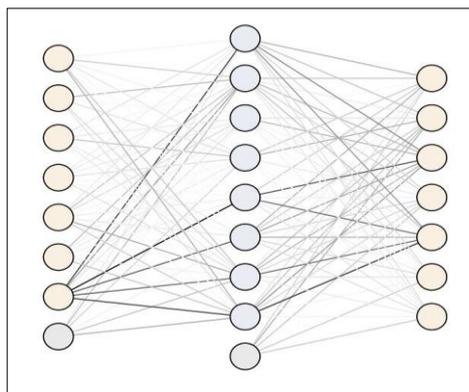


fig. A.1 Réseau de neurones amélioré pour le plus haut taux de classification sous RapidMiner

tab. A.1 Rétropropagation et calcul des fonctions d'erreur pour la couche cachée

	Nœud 1	Nœud 2	Nœud 3	Nœud 4	Nœud 5	Nœud 6	Nœud 7	Nœud 8
<i>Typ</i>	-0.475	2.266	0.582	1.051	-0.295	-0.184	2.363	3.262
<i>Radius</i>	-0.050	3.053	0.252	-1.304	0.057	-0.038	1.537	-1.247
<i>Size</i>	1.201	0.386	0.954	2.659	-0.197	0.781	-0.884	-0.126
<i>Width</i>	0.681	0.962	0.896	1.376	0.327	0.313	-1.358	0.339
<i>Length</i>	-0.995	-2.471	1.479	-0.352	0.744	-0.961	3.816	-2.225
<i>Slimf</i>	0.152	2.760	0.543	0.782	-0.663	0.180	-0.648	2.479
<i>Attribute</i>	8.537	-4.217	-1.087	-0.173	-11.553	7.561	-7.290	-8.954
<i>Bias</i>	0.852	-2.772	-0.404	1.868	0.113	0.671	-2.744	-3.782

tab. A.2 Rétropropagation, calcul des fonctions d'activation pour les classes de sortie

	Nœud 1	Nœud 2	Nœud 3	Nœud 4	Nœud 5	Nœud 6	Nœud 7	Nœud 8	Seuil
<i>S</i>	-3.561	-2.515	-0.254	-2.375	2.154	-3.478	0.588	-1.238	-1.877
<i>M</i>	-2.688	0.837	-1.131	-3.648	0.099	-2.168	2.830	2.516	-3.501
<i>H2</i>	5.213	-3.058	-3.336	-0.158	-8.080	4.302	-3.194	-4.428	0.047
<i>P</i>	-1.808	-1.869	-0.405	-0.948	-0.653	-1.855	-1.776	1.732	-2.406
<i>N</i>	-4.518	-2.576	-0.727	-0.147	6.845	-3.796	-6.752	-9.465	0.127
<i>T</i>	-1.568	-1.077	-0.255	-0.278	-0.745	-1.618	-1.289	0.390	-2.278
<i>I</i>	-1.239	-0.854	-0.602	-0.431	-1.191	-1.346	-0.865	-0.490	-2.377

Informations supplémentaires

RapidMiner ajuste initialement et sans paramétrage au préalable deux critères lors d'un apprentissage d'un réseau de neurones, *le poids d'un neurone* et *la valeur de la fonction d'activation* d'un neurone, leur ajustement étant simultané ; le moindre changement peut fausser les résultats d'une couche précédente, résultat qui aurait pu être le meilleur ajustement pour cette même couche.

Pour pallier cette incommodité, un neurone supplémentaire caché est ajouté (*Bias*) qui se trouve au sein d'une couche précise et est liée à l'ensemble des neurones de la couche suivante. Ce même neurone n'est pas lié à la couche précédente et n'est non plus lié à un autre neurone *Bias*.

Le neurone *bias* contient la valeur 1 qui sera induit dans la formule de la somme des poids des neurones lors de la propagation :

$$y = g(ak) = g\left(\sum_{j=0,d} w_{jk} \theta_j\right)$$

- w_{jk} : poids de la connexion de la cellule j à la cellule k
- a_k : activation de la cellule k
- g : fonction d'activation

Les neurones *bias* dans la figure A.1 sont les deux neurones inférieurs dans la couche d'entrée et la couche cachée.

K plus proche voisin

Les exemples d'apprentissage sont des vecteurs en espace de caractéristiques multidimensionnelles, chacun avec une classe. La phase d'apprentissage de l'algorithme se limite à stocker les vecteurs de caractéristiques et les classes des échantillons d'apprentissage.

Dans la phase de classification, k est une constante définie par l'utilisateur, et un vecteur non marqué (une requête ou d'un point de test) est classé en attribuant la classe qui est plus fréquente parmi les k échantillons d'apprentissage plus proche de ce point de requête.

- Soit $A = \{(x', c) | x' \in \mathbb{R}^d, c \in C\}$ l'ensemble d'apprentissage.
- Soit x l'exemple dont on souhaite déterminer la classe.
- $D(x, x')$ la distance entre le vecteur de test et d'apprentissage, généralement on utilise la distance Euclidienne.

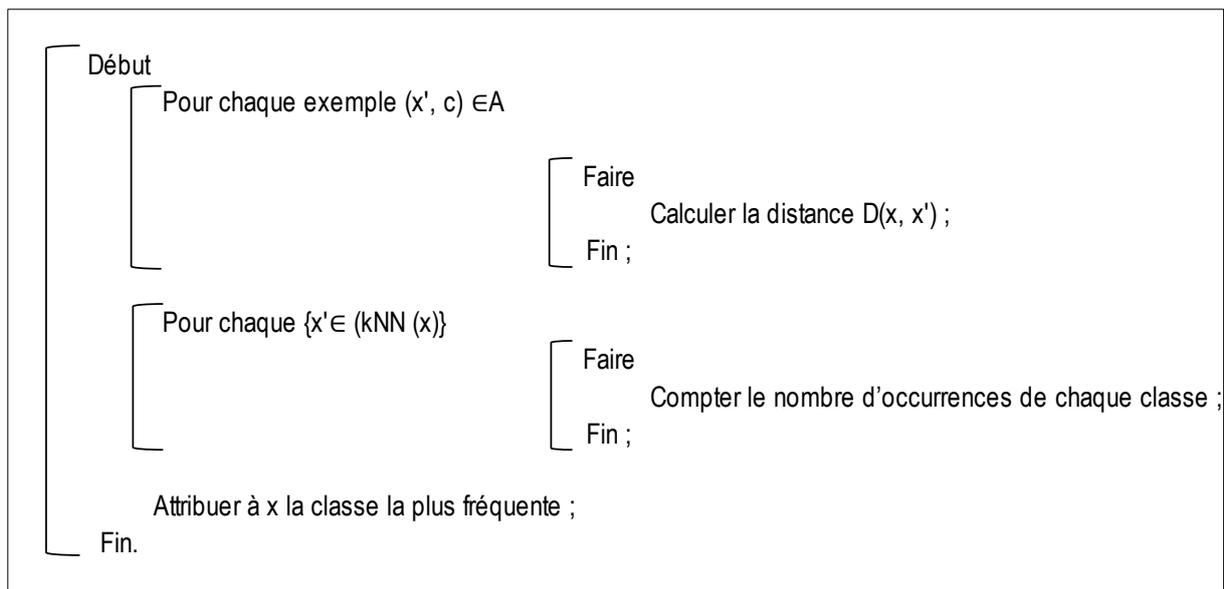


fig. A.2 Algorithme K-NN