

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de La Recherche Scientifique

Université Mohammed Seddik Ben Yahia-Jijel
Faculté des Sciences et de la Technologie
Département d'Electronique



Mémoire de fin d'études présenté en vue de l'obtention du
Diplôme de Master en Télécommunications

Option : Systèmes des Télécommunications

Thème :

Identification des troubles de la voix à l'aide
des techniques d'apprentissage automatique

Présenté par :

M^{elle} Manal BOUDJELLABA

M^{elle} Nadjet BOUDJERIDA

Encadré par :

Dr. Chaâbane BOUBAKIR

Promotion : Juillet 2021.

Remerciements

*Te mémoire n'aurait été possible sans l'aide de Dieu que nous
Remercions de nous avoir donné le courage, la volonté nécessaire et la
Patience durant toutes ces années d'études et que grâce à lui ce travail a
pu être réalisé*

*Nous tenons à remercier très vivement Monsieur Chaïbane
Boubakir, Enseignant à l'université de Tijel qui nous a proposé ce
sujet. Pour avoir dirigé ce travail, pour la confiance et l'intérêt qu'il a
témoigné tout au long de la réalisation de ce travail, pour sa
disponibilité ainsi de nous avoir fait bénéficier de ses compétences.*

*Nous adressons nos vifs remerciements aux membres de jury d'avoir
accepté de juger ce travail.*

*Nous tenons également à remercier tous ceux qui nous ont aidés de près
ou de loin à l'élaboration de ce travail.*

Nous tenons à remercier

Nos parents pour tout ce qui ont fait pour nous.

Manal et Nadjet

Dédicaces

Avec un énorme plaisir,

Un cœur ouvert et une immense joie, que

*Je dédie mon travail à mes très chers, respectueux et
Magnifiques parents qui m'ont soutenu tout au long de
ma vie,*

Que Dieu les garde pour nous.

Ainsi à mes sœurs et mes frères

Je leurs souhaite une belle vie.

A mes nièces que je les adore et apprécie

A mon binôme Manal

A toute ma grande famille

A toutes les personnes qui m'ont encouragé ou aidé.

*A la fin, pour toutes personnes que j'aime et qui
m'aiment.*

Nadjat

Dédicaces

Avec un énorme plaisir,

Un cœur ouvert et une immense joie, que

*Je dédie mon travail à mes très chers, respectueux et
Magnifiques parents qui m'ont soutenu tout au long de
ma vie,*

Que Dieu les garde pour nous.

Ainsi à mes sœurs : Imen, Faten, Rania

Je leurs souhaite une belle vie.

À mon frère Othmane que j'adore et apprécie

À mon binôme Nadjet

À toute ma grande famille

À toutes les personnes qui m'ont encouragé ou aidé.

À la fin, pour toutes personnes que j'aime et qui m'aiment.

Manal

Sommaire

Remerciements.....	i
Dédicaces.....	ii
Sommaire.....	iv
Liste des figures.....	vii
Liste des tableaux	ix
Liste des abréviations	x
Introduction générale.....	1

Chapitre I : Apprentissage automatique et classification

I.1 Introduction	3
I.2 Définition.....	3
I.3 Types d'apprentissage automatique.....	3
I.3.1 Apprentissage supervisé.....	4
I.3.2 Apprentissage non supervisé.....	5
I.3.3 Apprentissage par renforcement.....	5
I.4 Classification.....	5
I.4.1 Classification par SVM.....	6
I.4.1.1 Définition et principe de base des SVM.....	6
I.4.1.2 Formalisme d'un SVM.....	7
I.4.1.3 Cas de SVM non linéaire.....	11
I.4.1.4 SVM pour le cas multi-classes.....	13
I.4.2 Classification par KNN.....	15
I.4.2.1 Principe de fonctionnement.....	15
I.4.2.2 Sélection de la valeur de K dans l'algorithme K-NN	17
I.5 Conclusion.....	18

Chapitre II : Paramètres et pathologies de la voix

II.1 Introduction.....	19
II.2 Mécanisme de production de la parole.....	19
II.3 Les dysphonies.....	20
II.3.1 Dysphonies d'origine morphologique.....	21
II.3.2 Dysphonies d'origine neurologique.....	25
II.4 Paramètres de la voix.....	26
II.4.1 Paramètres acoustiques de la voix.....	26
II.4.2 L'analyse cepstrale.....	27
II.4.3 Le pitch.....	31
II.5 Evaluation des troubles de la voix.....	33
II.5.1 Evaluation subjective.....	32
II.5.2 Evaluation objective	33
II.5.2.1 Indices de perturbation du pitch.....	33
II.5.2.2 Indices de perturbations d'amplitude.....	34
II.5.2.3 HNR (Harmonic to Noise Ratio).....	36
II.5.2.4 NHR (Noise to Harmonic Ratio).....	36
II.6 Conclusion.....	37

Chapitre III : Implémentation, résultats et interprétations de la classification

III.1 Introduction.....	38
III.2 Bases de données utilisées pour la classification des voix saines et pathologiques.....	38
III.2.1 Base de données SVD.....	38
III.2.2 Base de données MEEI.....	38
III.2.3 Base de données utilisée.....	39
III.3 Conditions et Paramètres de simulation.....	39
III.3.1 Conditions d'implémentation.....	39
III.3.2 Sélection des paramètres.....	40
III.3.3 Extraction des paramètres.....	40
III.4 Validation.....	41

III.4.1 Validation simple.....	41
III.4.2 Validation croisée.....	42
III.5 Mesures de performances.....	43
III.6 Implémentations et résultats.....	45
III.6.1 Résultats et interprétations de la détection.....	48
III.6.1.1 Résultats de la détection par la validation simple	48
III.6.1.2 Résultats de la détection par la validation croisée.....	49
III.6.1.3 Résultats en fonction des paramètres choisis.....	51
III.6.2 Résultats et interprétations de la classification des voix pathologiques..	52
III.6.2.1 Résultats de la classification par la validation simple	53
III.6.2.2 Résultats de la classification par la validation croisée.....	54
III.6.3 Résultats de la classification par la méthode SVM Un-contre-Tous.....	56
III.7 Conclusion.....	57

Chapitre IV : Optimisation

IV.1 Introduction.....	58
IV.2 Définitions.....	58
IV.3 Techniques d'optimisation des hyperparamètres.....	58
IV.3.1 Optimisation par recherche de grille.....	59
IV.3.2 Optimisation par recherche aléatoire.....	59
IV.3.3 Optimisation Bayésienne.....	60
IV.4 Résultats d'optimisation des hyperparamètres	60
IV.4.1 K-NN Optimisé.....	60
IV.4.2 SVM Optimisé.....	64
IV.5 Conclusion.....	68
Conclusion générale.....	69
Liste des références.....	71

Liste des figures

Figure I.1 : Fonctionnement de l'apprentissage automatique supervisé.....	4
Figure I.2 : Hyperplan (H), marge maximale et vecteurs de support.....	6
Figure I.3 : Exemple d'un problème de classification multi-classes : 3 classes.....	13
Figure I.4 : Discrimination Un-Contre-Un.....	14
Figure I.5 : Discrimination Un-Contre-Tous.....	14
Figure I.6 : Fonctionnement du K-NN.....	16
Figure II.1 : Illustration de l'appareil phonatoire et de ses différents éléments.....	20
Figure II.2 : Nodule de la corde vocale gauche.....	21
Figure II.3 : Polype de la corde vocale droite.....	22
Figure II.4 : Kyste sur les cordes vocales.....	22
Figure II.5 : Œdème de Reinke.....	23
Figure II.6 : Papillome corné.....	23
Figure II.7 : Granulome inflammatoire.....	24
Figure II.8 : Cancer de la corde vocale droite.....	24
Figure II.9 : La convolution de l'excitation avec le conduit vocal.....	28
Figure II.10 : Phases d'obtention des coefficients MFCC.....	29
Figure II.11 : Approximation de l'échelle de Mel : Banc de filtres triangulaires.....	31
Figure II.12 : Variation de fréquence (Jitter).....	34
Figure II.13 : Variation d'amplitude (Shimmer).....	36
Figure III.1 : Méthode de la validation simple.....	41
Figure III.2 : Etapes de la méthode de validation croisée ($k_v=5$)	42
Figure III.3 : Exemple d'une courbe ROC et de l'AUC d'un classificateur entraîné.....	45
Figure III.4 : Interface graphique de l'application « Classification Learner ».....	45
Figure III.5 : Interface de la nouvelle session.....	46
Figure III.6 : Interface de la nouvelle session après sélection	46

Figure III.7 : Interface de l'application avant apprentissage.....	47
Figure III.8 : Courbe ROC pour un classificateur SVM cubique en sélectionnant les MFCC uniquement.....	52
Figure III.9 : Courbe ROC pour un classificateur SVM cubique en sélectionnant les paramètres Shimmer, Jitter et HNR uniquement.....	52
Figure III.10 : Système proposé pour l'identification des troubles de la voix.....	53
Figure IV.1 : Exemple de recherche de grille.....	59
Figure IV.2 : Exemple de recherche aléatoire.....	59
Figure IV.3 : Modèle K-NN optimisé par recherche aléatoire.....	61
Figure IV.4 : Modèle K-NN optimisé par recherche de grille.....	62
Figure IV.5 : Modèle K-NN optimisé par optimisation Bayésienne.....	63
Figure IV.6 : Modèle SVM optimisé par recherche aléatoire.....	64
Figure IV.7 : Modèle SVM optimisé par recherche de grille.....	65
Figure IV.8 : Modèle SVM optimisé par optimisation Bayésienne.....	66
Figure IV.9 : Interface indiquant l'option du réglage des paramètres.....	67
Figure IV.10 : Réglage des hyperparamètres pour le classificateur K-NN grossier.....	67
Figure IV.11 : Réglage des hyperparamètres pour le classificateur SVM gaussien moyen....	68

Liste des tableaux

Tableau III.1 : Distribution des fichiers normaux et pathologiques utilisés.....	39
Tableau III.2 : Les seuils normal / pathologique du Jitter selon le logiciel PRAAT.....	40
Tableau III.3 : Les seuils normal / pathologique du Shimmer selon le logiciel PRAAT.....	40
Tableau III.4 : Matrice de confusion.....	43
Tableau III.5 : Propriétés par défaut des modèles K-NN.....	47
Tableau III.6 : Propriétés par défaut des modèles SVM.....	47
Tableau III.7 : Mesures des performances pour la détection voix normale/ voix pathologique par validation simple 20% avec K-NN et SVM.....	48
Tableau III.8 : Mesures des performances pour la détection voix normale/ voix pathologique par validation croisée kv=5 (K-NN et SVM).....	49
Tableau III.9 : Mesures des performances pour la détection voix normale/ voix pathologique par validation croisée kv=10 (K-NN et SVM).....	50
Tableau III.10 : Mesures des performances pour la détection voix normale/voix pathologique avec sélection de paramètres.....	51
Tableau III.11 : Mesures des performances pour la classification des pathologies par validation simple 20% (K-NN et SVM).....	53
Tableau III.12 : Mesures des performances pour la classification des pathologies par validation croisée kv=5 (K-NN et SVM).....	54
Tableau III.13 : Mesures des performances pour la classification des pathologies par SVM et K-NN avec une validation croisée kv=10.....	55
Tableau III.14 : Mesures des performances pour la classification des pathologies par la méthode multi-classes Un-contre-Tous.....	56
Tableau IV.1 : Comparaison entre les performances du K-NN grossier et du SVM gaussien moyen avant et après le réglage de leurs hyperparamètres.....	68

Liste des abréviations

APQ	Amplitude Perturbation Quotient
AUC	Area Under Curve
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
F0	Fréquence fondamentale
FFT	Fast Fourier Transform
FN	Faux Négatif
FP	Faux Positif
GRBASI	Grade, Roughness, Breathiness, Asthenia, Strain, Instability
HNR	Harmonic to Noise Ratio
Jitt	Jitter local relatif
Jitta	Jitter local absolu
KKT	Karush Kuhn et Tucker
KNN	K-Nearest Neighbors
K-PPV	K-Plus Proches Voisins
Kv	K-fold cross validation
MEEI	Massachusetts Eye and Ear Infirmary

MFCC	Mel-scaled Frequency Cepstral Coefficients
NHR	Noise to Harmonic Ratio
PPQ	Period Perturbation Quotient
RAP	Relative Average Perturbation
RIF	Réponse Impulsionnelle Finie
ROC	Receiver Operating Characteristic
ShdB	Shimmer local en (dB)
Shim	Shimmer local
SVD	Saarbrücken Voice Database
SVM	Support Vector Machines
TCD	Transformée en Cosinus Discrète
TFN	Taux de Faux Négatifs
TFP	Taux de Faux Positifs
VHI	Voice Handicap Index
VN	Vrai Négatif
VP	Vrai Positif

Introduction générale

Introduction générale

La communication fait partie de notre vie et surtout la communication orale en utilisant notre voix qui est un outil puissant pour transmettre le message.

Ces dernières années et avec la technologie et ses évolutions, cette voix a subi des dégradations ou ce qu'on appelle la dysphonie qui indique que certains changements négatifs se sont produits dans les organes de phonation.

C'est pourquoi les gens n'ont cessé de perfectionner cette technologie et développer des techniques pour faciliter l'identification de ces dégradations et pour que chaque individu puisse faire un diagnostic chez soi. Mais rien de cela ne serait possible sans l'apprentissage automatique qui est une branche de l'intelligence artificielle.

L'intelligence artificielle est un domaine qui est apparu récemment il y'a une soixantaine d'années environ. Ce domaine complexe qui mélange sciences, théories et techniques a pour ambition d'imiter les capacités cognitives et intelligentes d'un être humain [1].

L'un des sous-ensembles de ce domaine complexe est l'apprentissage automatique ou artificiel (Machine Learning en anglais), qui permet à un système d'apprendre à partir des données sans être explicitement programmé. Il se divise à leur tour en différentes approches qui varient selon le type et le volume des données : apprentissage supervisé, apprentissage non supervisé et apprentissage par renforcement.

La détection automatique et la classification des pathologies est un domaine d'actualité et toujours exploré par la communauté des chercheurs [1-3].

Nous nous intéressons ici à l'apprentissage supervisé qui est le plus couramment utilisé, dont le spécialiste de données agit en tant que guide et apprend au système d'apprentissage les conclusions qu'il devrait tirer.

Il peut être regroupé en deux catégories d'algorithmes : Classification et Régression.

Parmi les algorithmes de la classification, nous nous intéresserons au classificateur des K plus proches voisins (K-NN pour K-Nearest Neighbors en anglais) et les machines à vecteurs de support (SVM pour Support Vector Machines).

L'apprentissage automatique est tout autour de nous, il est utilisé dans de très nombreux domaines, il peut s'agir :

De l'industrie, pour éviter par exemple les pannes et optimiser les cycles de production complexes dans les usines, et aussi l'industrie des transports pour le développement d'un système de navigation sans conducteur, etc.

De l'agriculture, où les agriculteurs customisent leurs techniques en fonction de données météo et les caractéristiques de leur terrain.

Des Télécommunications, en créant des assistants vocaux intelligents pour gérer les appels, etc.

Du secteur médical, où les machines aident à diagnostiquer des différentes maladies.

Et aussi pour la reconnaissance d'images, la détection des activités frauduleuses, ...

L'objectif de notre mémoire est l'utilisation des techniques d'apprentissage automatique pour la détection et la classification des pathologies de la voix et le choix des meilleures mesures de performances.

C'est pourquoi le premier chapitre va introduire quelques concepts entourant l'apprentissage automatique, les classificateurs qui seront utilisés ainsi que leurs fonctionnements.

Le deuxième chapitre sera consacré à la présentation du mécanisme de production de la parole, les dysphonies morphologiques et neurologiques, les paramètres de la voix qui seront utilisés, et aussi l'évaluation subjective et objective de la voix.

Dans le troisième chapitre, nous présenterons la base de données des fichiers de parole pathologique utilisée, les conditions d'implémentation, les mesures de performances à utiliser, les résultats de la détection et la classification par apprentissage automatique obtenus et des discussions.

Le quatrième chapitre sera consacré aux différentes méthodes d'optimisation des hyperparamètres des modèles d'apprentissage automatique, suivi par un exemple d'optimisation des modèles d'apprentissage automatique présentés dans le premier chapitre, et des interprétations des résultats obtenus.

Enfin, nous terminerons ce document par une conclusion générale.

Chapitre I

Apprentissage automatique et classification

I.1 Introduction

L'apprentissage automatique et les méthodes de classification ont pu provoquer une révolution importante dans le développement et l'amélioration des techniques de diagnostic des troubles vocaux.

Parmi les différentes techniques d'apprentissage automatique qui existent dans la littérature, on va s'intéresser principalement dans notre travail à deux méthodes qui sont : la classification par SVM (Support Vector Machines) et la classification par les K plus proches voisins K-NN (K-Nearest Neighbors).

Avant de détailler le principe et les fondements mathématiques de chaque méthode, le principe de l'apprentissage automatique d'ordre général et ses notions de base seront présentés au début.

I.2 Définition

L'apprentissage automatique (ou artificiel), ou Machine Learning en anglais est une branche de l'intelligence artificielle permettant aux ordinateurs d'apprendre sans être explicitement programmés selon le mathématicien américain Arthur Samuel.

L'apprentissage automatique englobe toute méthode permettant de construire un modèle de la réalité à partir de données, soit en améliorant un modèle partiel ou moins général, soit en créant complètement le modèle [2].

Au contraire d'un programme informatique traditionnel, l'apprentissage automatique ne suit pas des instructions mais plutôt s'appuie sur des motifs et de l'inférence à partir de l'expérience.

Un système d'apprentissage automatique apprend à partir des données historiques, construit les modèles prédictifs et prédit la sortie pour chaque fois qu'il reçoit de nouvelles données. La précision de la sortie attendue dépend de la quantité de données, car la grande quantité de données permet de créer un meilleur modèle qui prédit la sortie avec plus de précision.

Supposons que nous ayons un problème complexe sur lequel nous devons faire des prédictions. Au lieu d'écrire du code pour cela, nous avons juste besoin de nourrir les données dans des algorithmes génériques et, à l'aide de ces algorithmes, la machine construit la logique en fonction des données et prédit la sortie.

I.3 Types d'apprentissage automatique

Il existe différents types d'apprentissage automatique. Généralement, ils peuvent être divisés, selon le mode d'apprentissage qu'ils emploient, en trois catégories :

- L'apprentissage supervisé.
- L'apprentissage non supervisé.
- L'apprentissage par renforcement.

I.3.1 Apprentissage supervisé

L'apprentissage supervisé est un type de méthode d'apprentissage automatique dans lequel nous fournissons des échantillons de données étiquetées au système d'apprentissage automatique afin de le former, et sur cette base, il prédit le résultat. Le système crée un modèle en utilisant des données étiquetées pour comprendre les ensembles de données et en apprend davantage sur chaque donnée, une fois que la formation et le traitement sont terminés, nous testons le modèle en fournissant un échantillon de données pour vérifier s'il prédit ou non la sortie exacte.

Le diagramme suivant (figure I.1) explique le fonctionnement de l'algorithme d'apprentissage automatique supervisé.

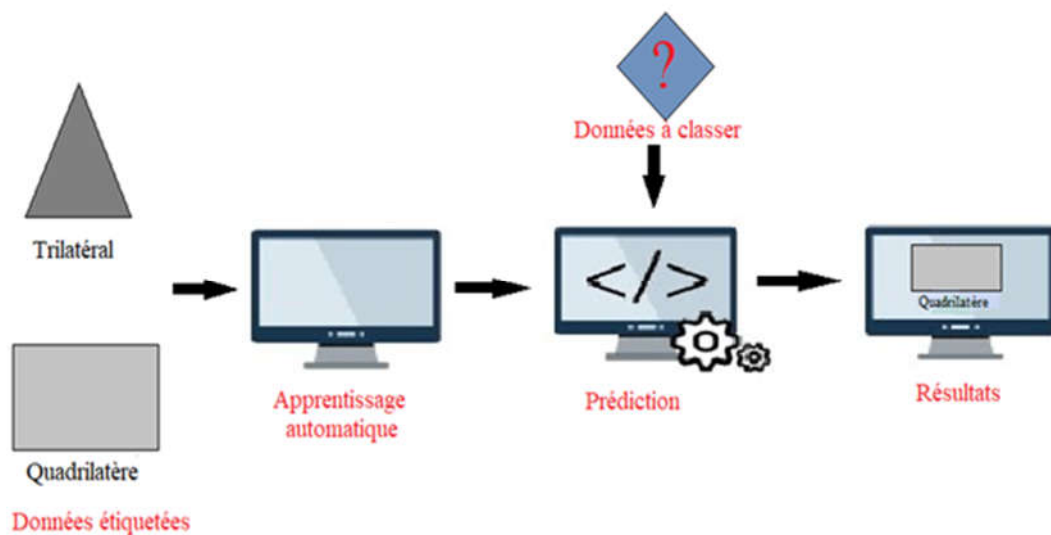


Figure I.1 : Fonctionnement de l'apprentissage automatique supervisé.

L'objectif de l'apprentissage supervisé est de faire correspondre les données d'entrée avec les données de sortie.

L'apprentissage supervisé peut être regroupé en deux catégories d'algorithmes :

- Classification.
- Régression.

Un problème de classification consiste à prédire la valeur d'une propriété d'un individu à partir des valeurs de ses autres propriétés. Alors que la régression, consiste à prédire une valeur continue (prix d'un produit), en classification on cherche à prédire une valeur catégorielle, comme l'état de santé d'un patient.

I.3.2 Apprentissage non supervisé

L'apprentissage non supervisé est une méthode d'apprentissage dans laquelle une machine apprend sans aucune supervision. La formation est fournie à la machine avec l'ensemble de données qui n'a pas été étiqueté, classé ou catégorisé, et l'algorithme doit agir sur ces données brutes sans aucune supervision. Le but de l'apprentissage non supervisé est de restructurer les données d'entrée en de nouvelles fonctionnalités ou un groupe d'objets avec des modèles similaires. Dans l'apprentissage non supervisé, nous n'avons pas de résultat prédéterminé. La machine essaie de trouver des informations utiles à partir de l'énorme quantité de données.

Il peut être classé en deux catégories d'algorithmes :

- Regroupement (Clustering).
- Association.

I.3.3 Apprentissage par renforcement

L'apprentissage par renforcement est une méthode d'apprentissage basée sur la rétroaction, dans laquelle un agent d'apprentissage obtient une récompense pour chaque bonne action et une pénalité pour chaque mauvaise action.

L'agent apprend automatiquement avec ces retours et améliore ses performances. Dans l'apprentissage par renforcement, l'agent interagit avec l'environnement et l'explore.

Le but d'un agent est d'obtenir le plus de points de récompense, et par conséquent, il améliore ses performances.

I.4 Classification

Les méthodes de classification font partie des méthodes d'apprentissage supervisé, car elles permettent de grouper des objets (observations ou individus) dans des classes de manière à ce que les objets appartenant à la même classe soient plus similaires entre eux qu'aux objets appartenant aux autres classes.

Dans le contexte supervisé on dispose déjà d'exemples dont la classe est connue et étiquetée. Les données sont donc associées à des labels des classes.

L'objectif est alors d'apprendre à l'aide d'un modèle d'apprentissage des règles qui permettent de prédire la classe des nouvelles observations ce qui revient à déterminer une fonction qui à partir des descripteurs de l'objet associe une classe et de pouvoir aussi affecter toute nouvelle observation à une classe parmi les classes disponibles.

Parmi les méthodes de classification les plus utilisées, nous nous intéressons aux méthodes des k-plus proches voisins (K-NN) et machines à vecteurs de support (SVM) [3].

I.4.1 Classification par SVM

Les machines à vecteurs de support ou séparateurs à vaste marge sont un ensemble de techniques d'apprentissage supervisé conçues pour résoudre des problèmes de régression ou de classification.

Ils ont été introduits par V.Vapnik en 1995 dans son livre « The nature of statistical learning theory », mais leur première apparition était en 1992 après être publiées par Boser, Guyon et Vapnik dans un article [4].

Sa robustesse vis-à-vis de la dimensionnalité des données et son pouvoir accru de généralisation, font que le SVM est nettement plus avantageux [5].

I.4.1.1 Définition et principe de base des SVM

Le SVM est un classificateur binaire largement utilisé pour la classification des données dans la plupart des domaines. Son objectif est de trouver la frontière optimale qui sépare deux classes avec la plus grande marge entre la frontière de séparation et les vecteurs de support [6]. La figure (I.2) illustre le principe du SVM et montre les vecteurs de support, l'hyperplan optimal et la marge maximale.

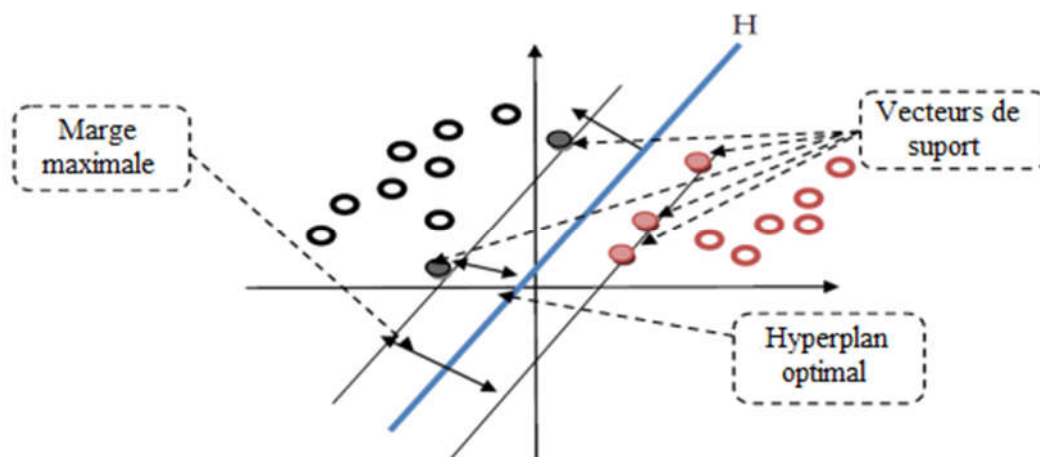


Figure I.2 : Hyperplan (H), marge maximale et vecteurs de support [7].

Ainsi, l'algorithme de base de SVM est la projection de données d'apprentissage de faible dimension dans un espace de caractéristiques de dimension supérieure ou infinie, afin que les données non linéairement séparables dans l'espace d'entrée deviennent linéairement séparables dans l'espace de caractéristiques [6].

Pour deux classes d'exemples donnés, le but de SVM est de trouver un classificateur qui va séparer les données et maximiser la distance entre ces deux classes. Avec SVM, ce classificateur est un classificateur linéaire appelé hyperplan.

- **Hyperplan**

Il peut y avoir plusieurs lignes/frontières de décision pour séparer les classes dans un espace à n dimensions, mais nous devons trouver la meilleure frontière de décision qui aide à classer les points de données. Cette meilleure frontière est connue sous le nom d'hyperplan de SVM [3].

Dans un espace de dimensions 2, l'hyperplan sera une droite, dans un espace de dimensions 3, l'hyperplan sera un plan, etc. L'hyperplan est indiqué sur la figure (I.2) [4].

- **Vecteurs de support**

Les vecteurs de support sont les points les plus proches de l'hyperplan et qui affectent sa position parmi l'ensemble total d'apprentissage. Ce sont les points de données les plus difficiles à classer. Ils ont une incidence directe sur l'emplacement optimal de la surface de décision. Les vecteurs de support sont indiqués sur la figure (I.2).

- **Marge :**

Il existe une infinité d'hyperplans capables de séparer parfaitement les deux classes d'exemples. Le principe des SVM est de choisir celui qui va maximiser la distance minimale entre l'hyperplan et les vecteurs de support, cette distance est appelée la marge. La marge maximale est indiquée dans la figure (I.2).

I.4.1.2 Formalisme d'un SVM

Dans un espace vectoriel de dimension finie N , un hyperplan est un sous-espace vectoriel de dimension $N-1$. Ainsi, dans un espace de dimension 2 un hyperplan sera une droite, dans un espace de dimension 3 un hyperplan sera un plan, etc.

Soit un espace vectoriel E de dimension N . L'équation caractéristique d'un hyperplan est de la forme :

$$w_1x_1 + w_2x_2 + \dots + w_Nx_N = 0 \quad (\text{I.1})$$

Où w_1, \dots, w_N sont des scalaires.

Par définition, tout vecteur $x = (x_1, \dots, x_N) \in E$, vérifiant l'équation (I.1) appartient à l'hyperplan. De plus, un hyperplan sépare complètement l'espace vectoriel en deux parties distinctes.

Ainsi, une ligne droite sépare le plan en deux régions, il est donc possible de diviser notre espace vectoriel en deux catégories distinctes. Comme nous pouvons le constater, un hyperplan vectoriel passe toujours par 0. C'est pour cette raison qu'on utilisera un hyperplan affine, qui n'a pas quant à lui l'obligation de passer par l'origine.

Ainsi, si l'on se place dans \mathbb{R}^N , pendant son entraînement le SVM calculera un hyperplan vectoriel d'équation (I.1) et un scalaire (un nombre réel) b . C'est ce scalaire b qui va nous permettre de travailler avec un hyperplan affine, comme nous allons le voir.

Le vecteur $w = (w_1, \dots, w_N)$ est appelé vecteur de poids, le scalaire b est appelé biais. Une fois l'entraînement terminé, pour classer une nouvelle entrée $x = (a_1, \dots, a_N)$, le SVM regardera le signe de :

$$h(x) = w_1 a_1 + \dots + w_N a_N + b = \sum_{i=1}^N w_i \cdot a_i + b = w^T \cdot x + b \quad (\text{I.2})$$

Si $h(x)$ est positif ou nul, alors x est d'un côté de l'hyperplan affine et appartient à la première catégorie, sinon x est de l'autre côté de l'hyperplan, et donc appartient à la seconde catégorie.

En résumé, on souhaite savoir, pour un point x , s'il se trouve d'un côté ou de l'autre de l'hyperplan. La fonction h nous permet de répondre à cette question, grâce à la classification suivante :

$$h(x) \geq 0 \Rightarrow x \in \text{catégorie 1} \quad (\text{I.3})$$

$$h(x) < 0 \Rightarrow x \in \text{catégorie 2} \quad (\text{I.4})$$

Ainsi, étant donné un hyperplan de vecteur de poids w , et de biais b , nous pouvons calculer si un point x_k appartient à telle ou telle catégorie, grâce au signe de $h(x_k)$.

En particulier, supposons que l'on assigne à tout point x_k un label l_k qui vaut (1) si x_k appartient à la première catégorie, et (-1) si x_k appartient à la seconde catégorie. Alors, si le SVM est correctement entraîné, on a toujours :

$$l_k \cdot h(x_k) \geq 0 \quad (\text{I.5})$$

C'est-à-dire :

$$l_k (w^T \cdot x_k + b) \geq 0 \quad (\text{I.6})$$

Le but d'un SVM, lors de l'entraînement, est donc de trouver un vecteur de poids w et un biais b tels que, pour tout x_k de label l_k appartenant aux données d'entraînement, l'équation (I.6) est satisfaite. Autrement dit, de trouver un hyperplan séparateur entre les deux catégories

- **Calcul de la marge**

Si l'on prend un point $x_k \in \mathbb{R}^N$, on peut prouver que sa distance à l'hyperplan de vecteur support w et de biais b est donnée par :

$$\frac{l_k (w^T \cdot x_k + b)}{\|w\|} \quad (\text{I.7})$$

Où $\|w\|$ désigne la norme euclidienne de w . La marge d'un hyperplan de paramètres (w, b) par rapport à un ensemble de points (x_k) est donc :

$$\min \frac{l_k(w^T \cdot x_k + b)}{\|w\|} \quad (\text{I.8})$$

On veut trouver l'hyperplan de support w et de biais b qui permettent de maximiser cette marge, donc on cherche un unique hyperplan dont les paramètres (w, b) sont donnés par la formule suivante :

$$\arg \max_{w,b} \min_k \left\{ \frac{l_k(w^T \cdot x_k + b)}{\|w\|} \right\} \quad (\text{I.9})$$

Même si on peut montrer que l'hyperplan optimal est unique, il existe plusieurs couples (w, b) qui décrivent ce même hyperplan. On décide de ne considérer que l'unique paramétrage (w, b) tel que les vecteurs supports x_s vérifient

$$(w^T \cdot x_s + b) = 1 \quad (\text{I.10})$$

Par conséquent $\forall k, l_k(w^T \cdot x_k + b) \geq 1$, et l'égalité est atteinte si x_k est un vecteur support. Autrement dit, cette normalisation sur w et b permet de garantir que la marge :

$$\min \frac{l_k(w^T \cdot x_k + b)}{\|w\|} = \frac{1}{\|w\|} \quad (\text{I.11})$$

On se retrouve donc avec le problème suivant :

$$\left\{ \begin{array}{l} \text{maximiser } \frac{1}{\|w\|} \\ \text{sous les contraintes } \forall k, l_k(w^T \cdot x_k + b) \geq 1 \end{array} \right. \quad (\text{I.12})$$

Que l'on peut reformuler de la façon suivante :

$$\left\{ \begin{array}{l} \text{minimiser } \|w\| \\ \text{sous les contraintes } \forall k, l_k(w^T \cdot x_k + b) \geq 1 \end{array} \right. \quad (\text{I.13})$$

Que, pour des raisons pratiques, on reformule à nouveau :

$$\left\{ \begin{array}{l} \text{minimiser } \frac{(\|w\|)^2}{2} \\ \text{sous les contraintes } \forall k, l_k(w^T \cdot x_k + b) \geq 1 \end{array} \right. \quad (\text{I.14})$$

Ce genre de problème est appelé problème d'optimisation quadratique, et il existe de nombreuses méthodes pour le résoudre. Dans le cas présent, on utilise la méthode des multiplicateurs de Lagrange.

- **Méthode des multiplicateurs de Lagrange**

La méthode de Lagrange est une méthode mathématique utilisée pour résoudre des problèmes d'optimisation quadratique sous contraintes, ça signifie que lorsque nous voulons optimiser (trouver les min ou max extrêmes) une fonction de base $f(x_1, \dots, x_N) : \mathbb{R}^N \rightarrow \mathbb{R}$, vous pouvez simplement utiliser le test de la dérivée seconde, en plus de cette fonction, vous avez également une contrainte $g(x_1, \dots, x_N) = 0$. Donc, nous essayons d'optimiser f tout en contraignant f avec g . Vous pouvez considérer une contrainte comme une frontière.

Par exemple, disant qu'on est dans une pièce et nous voulons connaître la distance la plus élevée pour laquelle nous pouvons lancer une balle, alors nous sommes conditionné par le plafond, nous ne pouvons pas lancer la balle plus haut que le plafond. Alors, la contrainte est le plafond. Le cœur de Lagrange est l'équation suivante :

$$\nabla f(x) = \lambda \nabla g(x) \quad (\text{I.15})$$

Cela dit que le gradient de f est égal à un scalaire multiplier par le gradient de g , rappeler que :

$$g(x) = 0 \quad (\text{I.16})$$

Souvent et surtout dans le contexte SVM, les deux équations (I.14) et (I.15) sont combinées en une seule équation appelée le Lagrangien :

$$L(x, \lambda) = f(x) - \lambda g(x) \quad (\text{I.17})$$

On utilise cette équation et on cherche le point où :

$$\nabla L(x, \lambda) = 0 \quad (\text{I.18})$$

Lorsqu'on a plusieurs contraintes, la relation (I.17) devient :

$$L(x, \lambda) = f(x) - \sum_{i=1}^N \lambda_i g_i(x) \quad (\text{I.19})$$

- **Application de la méthode de Lagrange en SVM**

Il s'agit de faire rentrer les contraintes dans la fonction objective et de pondérer chacune d'entre elles par une variable duale

$$L(w, b, \lambda) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^N \lambda_i \{1_i (w^T x_i + b) - 1\} \quad (\text{I.20})$$

Les variables duales intervenant dans le Lagrangien sont appelées multiplicateurs de Lagrange. Notons que L doit être minimisé par rapport aux variables primales w et b et maximisé par rapport aux variables duales λ_i . Le point (minimal par rapport à une variable, maximal par rapport à l'autre) doit donc satisfaire les conditions nécessaires de stationnarité qui correspondent aux conditions Karush Kuhn et Tucker (KKT), nous trouvons :

$$\frac{\partial L(w,b,\lambda)}{\partial w} = 0 \quad (I.21)$$

$$\frac{\partial L(w,b,\lambda)}{\partial b} = 0 \quad (I.22)$$

Ce qui nous permet d'obtenir :

$$w = \sum_{i=1}^N \lambda_i l_i x_i \quad (I.23)$$

$$\sum_{i=1}^N \lambda_i l_i = 0 \quad (I.24)$$

Cette résolution donnera une valeur optimale pour w , mais rien pour b . Pour retrouver b , il suffit de se rappeler que pour les vecteurs supports l_s ($w^T \cdot x_s + b$) = 1. On en déduit donc que b est tel que :

$$\min l_k (w^T \cdot x_k + b) \quad (I.25)$$

I.4.1.3 Cas de SVM non linéaire

Dans la plupart des problèmes réels, les données ne sont pas linéairement séparables, il est donc nécessaire de contourner ce problème (difficile de séparer n'importe quel jeu de données par un simple hyperplan). Si par exemple les données des deux classes se chevauchent sévèrement, aucun hyperplan séparateur ne sera satisfaisant, il est nécessaire de projeter les points d'apprentissage dans un espace de dimension plus élevée.

On applique aux vecteurs d'entrée x une fonction de transformation non-linéaire φ , qu'on appelle fonction noyau.

On définit l'opération de re-description des points x de E vers E' par l'opération :

$$\varphi : E \rightarrow E' \quad (I.26)$$

$$x \rightarrow \varphi(x) \quad (I.27)$$

Dans ce nouvel espace E' , on va tenter d'entraîner le SVM, comme nous l'aurions fait dans l'espace E . Si les données y sont linéairement séparables, c'est gagné ! Par la suite, si l'on veut classer x , il suffira de classer $\varphi(x)$: on obtient ainsi un SVM fonctionnel.

- **L'astuce du noyau pour simplifier les calculs**

Quand on pose le problème d'optimisation quadratique dans l'espace E' , on s'aperçoit que les seules apparitions de φ sont de la forme $\varphi(x_i)^T \cdot \varphi(x_j)$. De même dans l'expression de la fonction de classification h' . Par conséquent, il n'y a pas besoin de connaître expressément E' ,

ni même φ : il suffit de connaître toutes les valeurs $\varphi(x_i)^T \cdot \varphi(x_j)$, qui ne dépendent donc que des x_i .

On appelle donc fonction noyau, la fonction $K: E \times E \rightarrow R$ définie de la façon suivante :

$$K(x, x') = \varphi(x)^T \cdot \varphi(x') \quad (\text{I.28})$$

A ce moment, le calcul de l'hyperplan séparateur dans E' ne nécessite ni la connaissance de E' , ni de φ , mais seulement de k .

Noyau symétrique semi-défini positif

Une fonction φ est dite symétrique si et seulement si :

$$\forall x, y \varphi(x, y) = \varphi(y, x) \quad (\text{I.29})$$

(Dans le cas où φ est à deux variables).

Une fonction symétrique est dite semi-définie positive si et seulement si, pour tous ensemble fini (x_1, \dots, x_n) , et pour tous réels (c_1, \dots, c_n) :

$$\sum c_i c_j K(x_i, x_j) \geq 0 \quad (\text{I.30})$$

Ainsi, il est possible d'utiliser n'importe quelle fonction noyau afin de réaliser une re-description dans un espace de dimension supérieure. La fonction noyau étant définie sur l'espace de description E (et non sur l'espace de re-description E' , de plus grande dimension), les calculs sont beaucoup plus rapides.

Voici une liste non exhaustive de noyaux couramment utilisés.

- Le noyau polynomial,

$$K(x, x') = (ax^T \cdot x' + \lambda)^d \quad (\text{I.31})$$

- Le noyau Gaussien,

$$k(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \quad (\text{I.32})$$

- Le noyau Laplacien,

$$k(x - x') = \exp\left(-\frac{\|x-x'\|}{\sigma}\right) \quad (\text{I.33})$$

- Le noyau rationnel,

$$k(x - x') = 1 - \frac{\|x - x'\|^2}{\|x - x'\|^2 + \sigma} \quad (1.34)$$

I.4.1.4 SVM pour le cas multi-classes

Les machines à vecteurs de support sont dans leur origine binaire. Cependant, les problèmes du monde réel sont dans la plupart des cas multi-classes.

Les méthodes des machines à vecteurs de support multi-classes, réduisent le problème multi-classes à une composition de plusieurs hyperplans bi-classes permettant de tracer les frontières de décision entre les différentes classes.

Ces méthodes décomposent l'ensemble d'exemples en plusieurs sous-ensembles représentant chacun un problème de classification binaire. Pour chaque problème un hyperplan de séparation est déterminé par la méthode SVM binaire. On construit lors de la classification, une hiérarchie des hyperplans binaires qui est parcourue de la racine jusqu'à une feuille pour décider de la classe d'un nouvel exemple [8].

On trouve dans la littérature plusieurs méthodes de décomposition, parmi ces méthodes on cite deux : méthode dite Un-Contre-Un (One-Versus-One), et l'autre Un-Contre-Tous (One-Versus-All).

L'idée est de cartographier les points de données vers un espace dimensionnel élevé afin d'obtenir une séparation linéaire mutuelle entre les deux classes. C'est ce qu'on appelle une approche Un-Contre-Un, qui décompose le problème multi-classes en plusieurs problèmes de classification binaire. Un classificateur binaire par paire de classes [9].

Une autre approche que l'on peut utiliser est Un-Contre-Tous. Dans cette approche, la répartition est définie sur un classificateur binaire par classe. Prenons un exemple de problème de classification des 3 classes ; vert, rouge et bleu, comme l'image suivante :

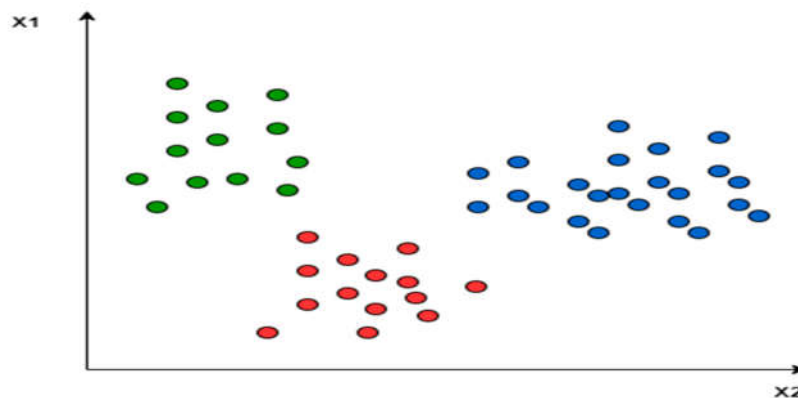


Figure I.3: Exemple d'un problème de classification multi-classes : 3 classes [9].

L'application des deux approches à cet ensemble de données se traduit par les éléments suivants :

- Dans l'approche Un-Contre-Un, nous avons besoin d'un hyperplan pour séparer entre les deux classes, négligeant les points de la troisième classe (figure I.4). Cela signifie que la séparation ne prend en compte que les points des deux classes dans la répartition actuelle. Par exemple, la ligne rouge-bleue tente de maximiser la séparation seulement entre les points bleus et rouges. Cela n'a rien à voir avec les points verts.

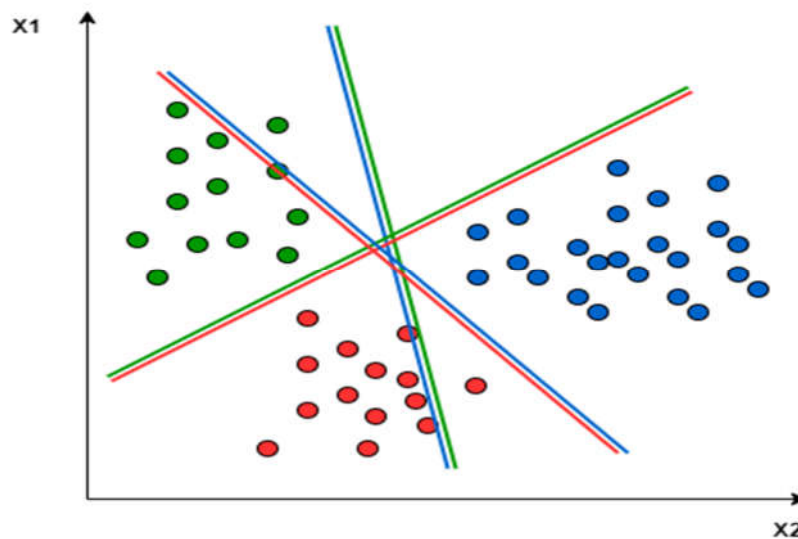


Figure I.4: Discrimination Un-Contre-Un [9].

- Dans l'approche Un-Contre-Tous, nous avons besoin d'un hyperplan pour séparer une classe et toutes les autres à la fois. Cela signifie que la séparation prend en compte tous les points, les divisant en deux groupes ; un groupe pour les points de classe et un groupe pour tous les autres points comme l'illustre la figure.

Par exemple, la ligne verte tente de maximiser la séparation entre les points verts et tous les autres points à la fois.

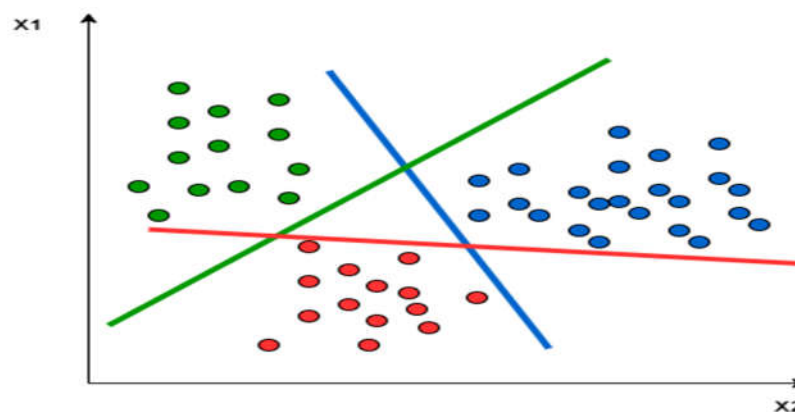


Figure I.5 : Discrimination par Un-Contre-Tous [9].

➤ Avantages des SVM [10]

Les SVM représentent plusieurs avantages, notamment ceux-ci :

- Elles ont une base théorique solide.
- Les SVM sont efficaces dans les espaces de grande dimension.
- Différentes fonctions noyau peuvent être spécifiées.

➤ Inconvénients des SVM

Malgré leurs performances, les SVM représentent aussi des faiblesses, notamment celles-ci :

- Elles utilisent des fonctions mathématiques complexes pour la classification.
- Les machines à vecteurs de supports demandent un temps énorme durant les phases d'apprentissage.

I.4.2 Classification par K-NN

L'algorithme des K Plus Proches Voisins (noté K-PPV) en anglais (K Nearest Neighbors : K-NN) est l'un des algorithmes d'apprentissage automatique les plus simples basés sur la technique d'apprentissage supervisé, il est basé sur la notion de proximité (voisinage) entre les instances et sur l'idée de raisonner à partir de cas similaires pour prendre une décision [11].

L'algorithme K-NN suppose la similitude entre le nouveau cas et les cas disponibles et place le nouveau cas dans la catégorie la plus similaire aux catégories disponibles.

L'algorithme K-NN stocke toutes les données disponibles et classe un nouveau point de données en fonction de la similitude. Cela signifie que lorsque de nouvelles données apparaissent, elles peuvent être facilement classées dans une catégorie de suite de points en utilisant l'algorithme K-NN.

K-NN est un algorithme non paramétrique, ce qui signifie qu'il ne fait aucune hypothèse sur les données sous-jacentes. Il est également appelé algorithme d'apprentissage paresseux, car il n'apprend pas immédiatement de l'ensemble d'apprentissage, mais stocke l'ensemble de données et, au moment de la classification, effectue une action sur l'ensemble de données [3].

L'algorithme K-NN à la phase d'apprentissage stocke simplement l'ensemble de données et lorsqu'il obtient de nouvelles données, il classe ces données dans une catégorie qui est très similaire aux nouvelles données.

I.4.2.1 Principe de fonctionnement

Le principe de l'algorithme K-NN est très simple, il se base sur un ensemble de données d'apprentissage, une fonction de distance et une fonction de vote de la classe la plus proche.

Ainsi, pour tout nouveau point observé, pour lequel on doit prendre une décision, l'algorithme recherche dans l'ensemble de données d'apprentissage les points les plus proches au point observé, et lui attribue la classe la plus fréquente dans ces voisins.

Dans le cas général, prendre plusieurs plus proches voisins, plutôt qu'un unique plus proche voisin permet une certaine robustesse face aux erreurs d'étiquetage [12].

La figure (I.6) représente le fonctionnement du K-NN.

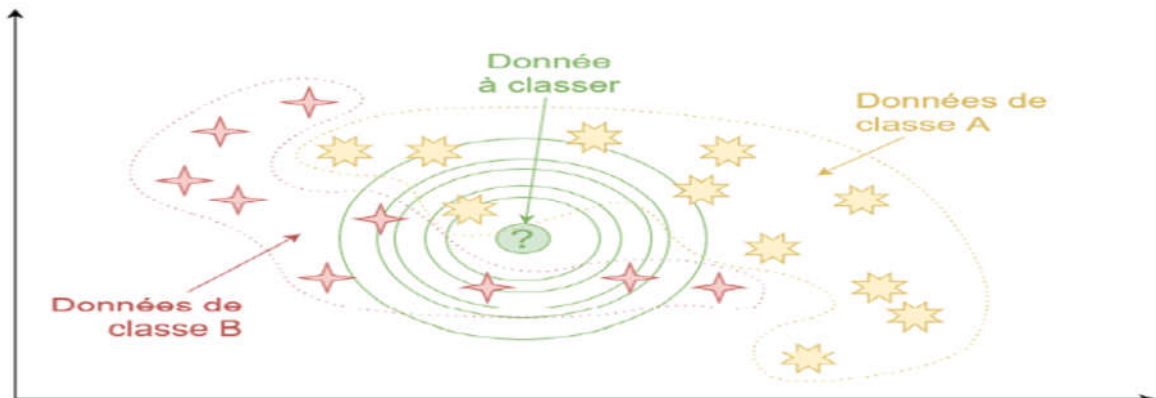


Figure I.6 : Fonctionnement du K-NN.

Pour appliquer cette méthode, le pseudo code suivant montre les étapes à suivre :

- **Pseudo code de l'algorithme K-NN**

Arguments :

X : L'instance non classée.

S : L'ensemble de toutes les instances (base d'apprentissage).

Y_i : Les classes des instances de S.

Y_{ij} : Les instances de S appartenant à la classe Y_i .

d : La métrique utilisée (distance).

Entrées : X, S, d, k.

Sorties : Y = classe de X.

Début :

1. Pour chaque instance Y_{ij} dans S
Calculer $d(X, Y_{ij})$
2. Classer les distances par ordre croissant.
3. Compter le nombre d'occurrences de chaque classe Y_i parmi les k plus proches selon l'ordre.
4. Y = la classe la plus fréquente.
5. Retourner Y.

Fin.

Cette méthode dépend donc des deux éléments suivants :

- 1- Le nombre de voisins retenus.
- 2- La mesure de similarité (la distance) entre deux instances.

Le résultat dépend du réglage de ces paramètres.

Pour le premier paramètre, on prend k comme un nombre de voisins tel que la valeur k est généralement impair.

Pour le deuxième paramètre, la méthode nécessite une distance pour mesurer la proximité entre l'instance à classer X et chacune des instances de l'ensemble d'apprentissage [11].

Parmi les distances qu'on peut utiliser, on trouve :

$$\text{La distance Euclidienne} \quad d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (\text{I.35})$$

$$\text{La distance de Minkowski} \quad d(x, y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p} \quad (\text{I.36})$$

$$\text{La distance de Manhattan} \quad d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (\text{I.37})$$

$$\text{La distance Cosinus} \quad \cos \theta = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\| \|\vec{y}\|} \quad (\text{I.38})$$

Où : x, y sont des vecteurs et $\|\vec{x}\|, \|\vec{y}\|$ sont leurs modules respectivement, p un paramètre. Lorsque les attributs sont numériques, la distance Euclidienne est généralement utilisée.

1.4.2.2 Sélection de la valeur de K dans l'algorithme K-NN

Voici quelques points à retenir lors de la sélection de la valeur de K dans l'algorithme K-NN : Il n'y a pas de moyen particulier de déterminer la meilleure valeur pour " K ", nous devons donc essayer certaines valeurs pour en trouver le meilleur. La valeur la plus préférée pour K est 5. Une valeur très faible pour K telle que $K = 1$ ou $K = 2$, peut être bruyante et entraîner des effets de valeurs aberrantes dans le modèle. Les grandes valeurs de K sont bonnes, mais cela peut rencontrer des difficultés.

➤ Avantages de l'algorithme K-NN

- C'est simple à mettre en œuvre.
- Il est robuste aux données d'apprentissage bruyantes.
- Cela peut être plus efficace si les données d'apprentissage sont volumineuses.

➤ Inconvénients de l'algorithme K-NN

- Il faut toujours déterminer la valeur de K qui peut être complexe un certain temps.
- Le coût de calcul est élevé en raison du calcul de la distance entre les points de données pour tous les échantillons d'apprentissage.

I.5 Conclusion

Nous avons présenté dans ce chapitre les notions essentielles sur l'apprentissage automatique et la classification en général et les méthodes SVM et K-NN en particulier. Ces dernières sont des méthodes de classification qui ont montrés de bonnes performances dans la résolution de différents problèmes. La vérification de ces performances dans le cas de la détection et la classification des pathologies vocales sera l'objectif du troisième chapitre.

Chapitre II

Paramètres et pathologies de la voix

II.1 Introduction

L'utilisation fréquente de la voix dans la communication et la vie quotidienne peut provoquer des changements (raucité, instabilité, ...), indiquant la présence de certaines pathologies qui nécessitent un diagnostic.

Dans ce chapitre, nous présenterons le mécanisme de la production de la parole, les différents types de pathologies, les paramètres de la voix utilisés pour le diagnostic (MFCC, Jitter, Shimmer, HNR, ...) ainsi que l'évaluation subjective et objective de ces troubles vocaux.

II.2 Mécanisme de production de la parole

Avant de parler sur le mécanisme de la production de la voix, on parle de la différence entre les mots « voix » et « parole » qui ne sont pas des synonymes, tel que la voix est le son produit par les cordes vocales lorsqu'elles vibrent sous l'influence de l'air pulmonaire. Par contre, la parole correspond aux modifications de ce son en fonction de la forme du conduit aérodigestif.

C'est-à-dire que la même vibration des cordes vocales (la même note) sera entendue comme des sons différents en fonction de la position de la langue, le rétrécissement du pharynx, etc. Donc, la voix est le support de la parole, « sans voix, pas de parole ».

Le processus de la production de la parole est un processus très complexe. Il existe plusieurs organes et muscles qui entrent dans la production des sons ou de la parole.

Le fonctionnement de l'appareil phonatoire humain repose sur l'interaction de trois grandes classes d'organes :

Le niveau respiratoire : qui se compose du diaphragme, les poumons et les trachées, c'est la soufflerie qui fournit l'énergie et la quantité d'air nécessaire à la vibration des cordes vocales.

Le niveau phonatoire : le larynx qui contient les cordes vocales.

Le conduit vocal : formé des cavités résonantes supra-laryngées (pharynx, bouche, nez) où s'effectue l'articulation.

La figure (II.1) représente l'appareil phonatoire et ses différents éléments. Le trajet de l'air est indiqué en bleu.

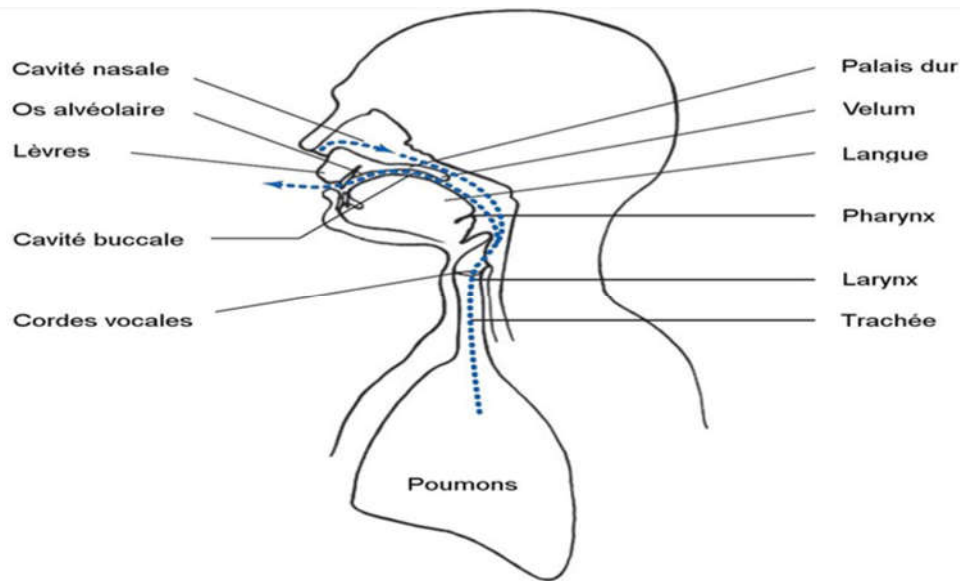


Figure II.1 : Illustration de l'appareil phonatoire et de ses différents éléments [13].

L'air expulsé des poumons, traverse l'appareil phonatoire comme un instrument à vent et crée la pression nécessaire à la génération d'un signal acoustique. Ce dernier traverse la trachée pour arriver dans le larynx où se trouvent les cordes vocales. L'air traversant le larynx met en vibration les cordes vocales.

La fréquence de vibration des cordes vocales est modulée en fonction de leur degré de contraction. Ces vibrations sont modifiées par le passage de l'air dans les différentes cavités qui composent le pharynx mais aussi dans les fosses nasales.

Ces résonateurs influent sur le son en atténuant certaines fréquences et en amplifiant d'autres. La forme et le volume de ces cavités, spécifiques au locuteur, modifient fortement le son produit.

Les organes modulateurs qui sont la langue, les lèvres et la mâchoire sculptent, produisent les phonèmes qui composent la parole. La position de ces différents organes est le mécanisme final qui permet la production de parole articulée. Certains organes impliqués dans la production de la parole se trouvent en partie visible extérieur (lèvres, langue, position des mâchoires) [14].

II.3 Les dysphonies

La dysphonie est une qualité perceptive de la voix qui indique que certains changements négatifs se sont produits dans les organes de phonation [15]. Plus précisément des changements dans les paramètres de la voix ou mal-fonctionnement de l'appareil phonatoire. La dysphonie peut également résulter de modifications morphologiques de l'anatomie du larynx,

principalement au niveau de la glotte, provoquées par des modifications de sa structure, la génération d'un excès de tissu biologique ou, au contraire, de défauts anatomiques provoqués par des interventions chirurgicales.

L'altération de la voix peut se manifester selon différentes modalités, en fonction de la cause du trouble. Par exemple, la voix peut être affaiblie, devenir rauque, être presque inaudible, etc. On distingue deux catégories des dysphonies : Les dysphonies d'origine morphologique et les dysphonies d'origine neurologique.

II.3.1 Dysphonies d'origine morphologique

On désigne sous ce terme les dysphonies consécutives aux changements anatomiques des structures impliquées dans la phonation, en particulier le larynx et les cordes vocales. Les dysphonies d'origine morphologique les plus citées sont :

- **Nodule**

Les nodules des cordes vocales sont des lésions bénignes. Ils sont le plus souvent bilatéraux. Ils se développent chez la femme adulte et chez l'enfant. Ces derniers sont des petits gonflements muqueux situés à la jonction du tiers antérieur et moyen de la corde vocale, où la vibration de la muqueuse est maximale. Ils sont de couleur blanche ou rosée, et d'une forme arrondie.

En phonation, la fermeture des cordes vocales est gênée par les nodules ce qui donne une voix voilée et d'une faible intensité et des difficultés à parler et de tenir une note longtemps. La figure (II.2) illustre un nodule de la corde vocale gauche (flèche noire).

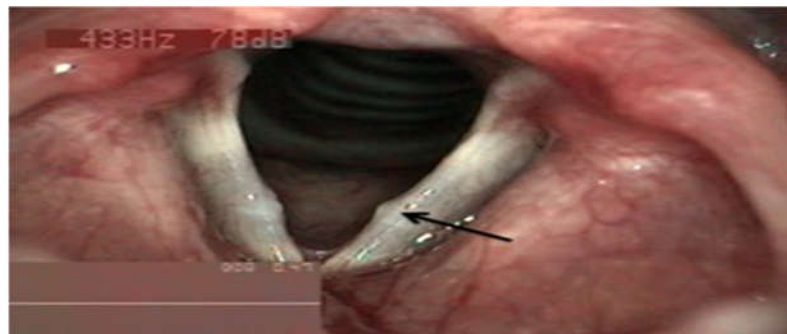


Figure (II.2) : Nodule de la corde vocale gauche (flèche noire) [16].

- **Polype**

Il se présente généralement sous la forme d'une petite masse arrondie développée à partir du pli vocal, mais on rencontre également des polypes volumineux, doublant ou triplant la corde vocale. La coloration du polype peut être rouge vif, ou bien pâle, grisâtre. Il se situe à des niveaux variés, mais le plus souvent on le trouve au tiers moyen de la corde vocale.

Contrairement aux nodules, les polypes apparaissent la plupart du temps unilatéralement, donc sur une seule corde vocale. Le polype survient quasiment toujours chez les hommes que chez les femmes et due au forçage de la voix, de la toux ou de la consommation excessive du tabac. La voix devient râpeuse, enrouée, poussée et forcée en intensité.

La figure (II.3) présente un polype sur la corde vocale droite en position d'ouverture des cordes vocales (A) et (B) en position de fermeture.

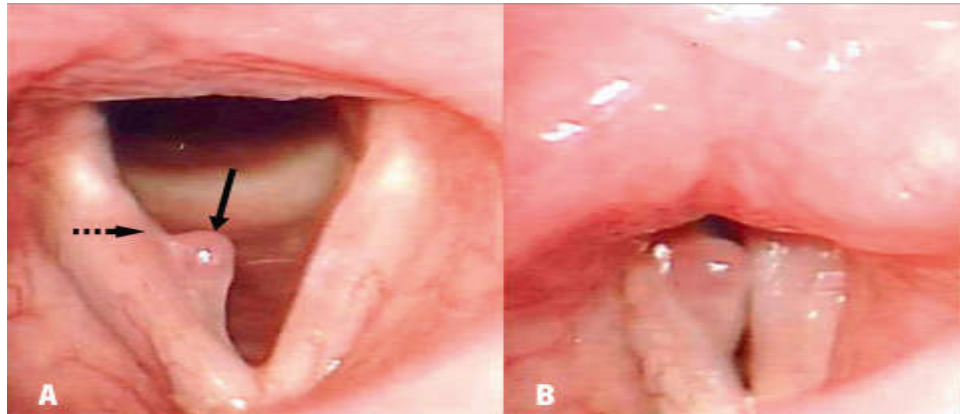


Figure (II.3) : Polype de la corde vocal droite [17].

- **Kyste muqueux**

Il s'agit d'une formation développée à partir d'une glande située dans la muqueuse des cordes vocales. Le kyste muqueux se développe dans l'épaisseur de la corde vocale. Cette lésion n'est pas liée au forçage vocal. Le contenu du kyste est un liquide muqueux, jaunâtre. Il déforme la corde vocale au niveau de sa face ou de son bord. La dysphonie est importante car le kyste gêne l'affrontement des cordes vocales, cette gêne est souvent variable dans le temps, et dépend de la quantité ou la présence ou l'absence du contenu de kyste (figure II.4).

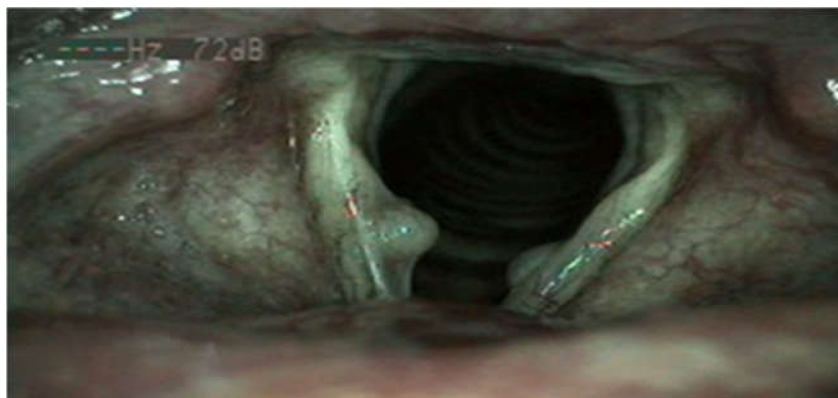


Figure II.4 : Kyste sur les cordes vocales [16].

- **Œdème de Reinke**

C'est une forme de laryngite chronique (appelée pseudomyxome des cordes vocales), les cordes ont un aspect soufflé en vessie de poisson, la muqueuse est très fine et enveloppe une matière d'aspect gélatineux de couleur ambrée (figure II.5). Le tabac est le principal facteur de risque de ce type de laryngite, associé au malmenage vocal. Le traitement des facteurs de risque peut stabiliser ou faire un peu régresser l'œdème, mais, une fois constitué, celui-ci a tendance à persister.

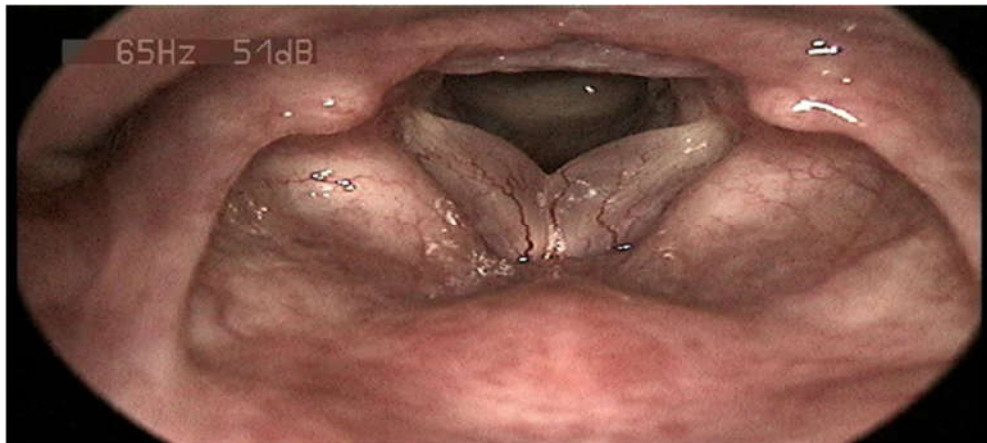


Figure II.5 : Œdème de Reinke.

- **Papillome corné**

Il a un aspect de tumeur exophytique (lésion qui apparaît vers l'extérieur) plus ou moins volumineuse, dure, caractérisée par des spicules, et une base d'implantation large et bien limitée. Il est de couleur grise ou blanchâtre (figure II.6). Souvent unique et limité à la partie antérieure de la glotte, il peut l'envahir en totalité et s'étendre à la partie supérieure du larynx et à la sous-glottes, réalisant alors une papillomatose cornée extensive.



Figure II.6 : Papillome corné.

- **Granulome**

C'est une lésion de la glotte postérieure ayant la forme d'une tuméfaction arrondie, souvent bilatérale, mais peut être unilatérale, se produisant au niveau des deux processus vocaux (cordes postérieures), plus grands que les nodules (figure II.7), touche principalement les hommes. Les causes sont abus vocal répété, maladie de reflux, intubation prolongée.

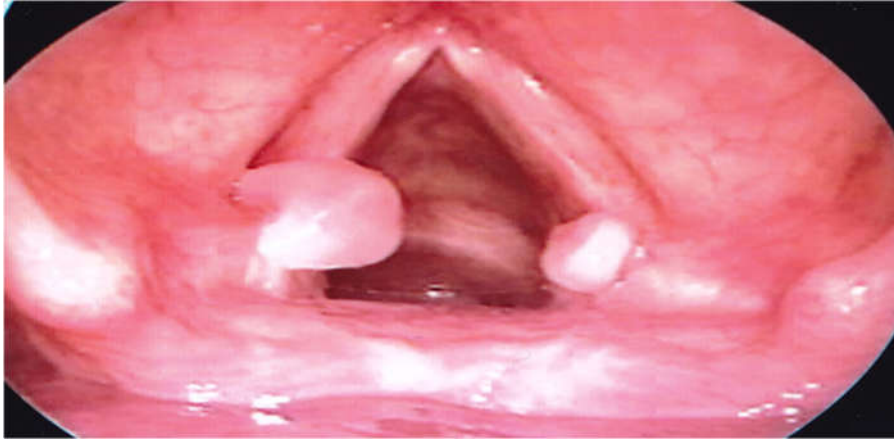


Figure II.7 : Granulome inflammatoire [17].

- **Cancer de la corde vocale**

Il est révélé par une dysphonie persistante (voix enrouée, d'une faible intensité et voilée) et tant qu'il reste localisé à la corde vocale il est indolore. C'est une lésion blanchâtre plus ou moins en relief, irrégulière (figure II.8).

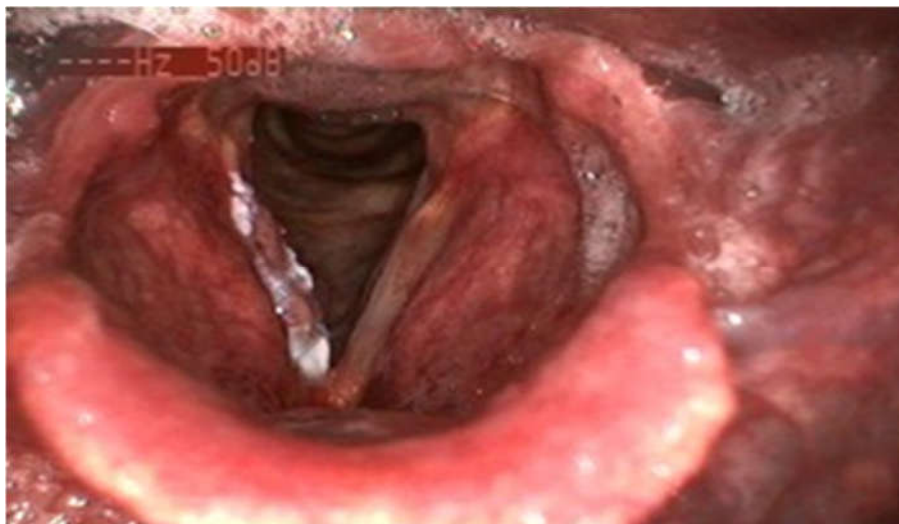


Figure II.8 : Cancer de la corde vocale droite [16].

- **Traumatismes chirurgicaux**

Les plus sérieuses modifications anatomiques du larynx sont provoquées par des traumatismes chirurgicaux à la suite de sectionnement d'un cancer ou de n'importe quelle autre lésion au niveau du larynx (nodule, polype, kyste, etc.). La voix est très abaissée, quoique fonction de la technique chirurgicale, avec parfois des désonorisations. Elle est grave, de faible intensité, mais compréhensible sauf dans le bruit. Le timbre est très rauque.

II.3.2 Dysphonies d'origine neurologique

On désigne sous ce terme, le groupe des dysphonies pour lequel l'examen laryngoscopique ne révèle aucune lésion des cordes vocales ou du larynx. Il s'agit principalement des dysphonies d'origine neurologique. Les plus citées dans cette catégorie sont :

- **Paralysie laryngée**

La paralysie laryngée est une atteinte neurologique qui se traduit par une immobilité d'une ou des deux cordes vocales (unilatérale ou bilatérale). La position de la corde paralysée est variable : elle peut être fixée en position médiane (adduction) ou bien en position intermédiaire, (semi-abduction) exceptionnellement, en position latérale (abduction complète).

Dans le cas unilatéral, la voix est monotone avec perte de mélodie et de nombreuses désonorisations. Elle est soufflée et rauque avec une fuite d'air importante, entraînant un essoufflement en fin de phrase et une voix projetée continue impossible, tandis que dans le cas bilatéral, la personne endure une aphonie totale.

- **Dysphonies Spasmodiques**

Il s'agit d'une dysphonie d'origine neurologique appartenant au groupe des dystonies focales des muscles abducteur et adducteur du larynx, c'est-à-dire ce n'est pas une maladie des organes de phonations (cordes vocal et larynx), mais elle est liée à la commande nerveuse de la mobilité du larynx, ou bien l'excitation des cordes vocales. C'est une affection rare et touche plus fréquemment les femmes que les hommes.

Les dysphonies spasmodiques causent des changements brutaux de la hauteur de la voix qui peut s'interrompre, repartir, glisser et trembler, elle peut avoir un timbre désagréable et devenir au pire une aphonie (des chuchotements non perceptibles). Le temps maximum de phonation sur une voyelle tenue est abaissé.

- **Dysarthrie parkinsonienne**

La maladie de Parkinson est un trouble progressif du système nerveux qui affecte le mouvement. Les troubles de la voix dans cette maladie associent une dysphonie (altération des qualités acoustiques de la voix) et une dysarthrie (trouble de l'articulation de la parole).

Cette dysphonie-dysarthrie touche les différents effecteurs de la chaîne parlée : la respiration, la phonation, les résonances (nasalisation), l'articulation et la prosodie (mélodie, intensité, durée).

- **Les dysphonies dysfonctionnelles (Hypotonie et hypertonie)**

Les dysphonies dysfonctionnelles au niveau de l'appareil phonatoire se manifestent soit par un excès de fonction (hypertonie), soit par une insuffisance de fonction (hypotonie).

L'hypertonie désigne l'exagération du tonus musculaire, ce dernier équivaut à la faculté permanente des muscles de se maintenir dans un état de tension dans le but de s'opposer à l'action de la gravité sur notre corps. Le tonus musculaire est à la fois indispensable pour se maintenir en équilibre et pour se mettre en mouvement. La situation inverse de l'hypertonie, est d'hypotonie qui correspond à un manque de tonus musculaire.

II.4 Paramètres de la voix

Le signal de parole porte plusieurs niveaux de description (acoustique, phonétique, morphologique, etc.) car il ne contient pas seulement des informations sur le texte parlé mais aussi des informations sur le locuteur, la langue, les émotions, etc. Donc, il peut être décrit et mesuré à l'aide des différents paramètres.

Parmi les paramètres qui nous permettent de décrire et mesurer ce signal, on trouve [18, 19] :

II.4.1 Paramètres acoustiques de la voix

- **La hauteur**

Elle est fixée par la fréquence fondamentale (F0, exprimée en Hz) de l'onde glottique qui correspond au nombre de cycles d'ouverture-fermeture des cordes vocales par seconde. Elle dépend de la longueur des cordes vocales, de leur masse et de leur tension, plus les cordes seront épaisses et lourdes, plus la F0 sera grave.

Les fréquences les plus basses (vibration lente) correspondent à des voix graves, les plus hautes (vibration rapide) à des voix aiguës.

- **L'intensité**

Elle se définit comme la puissance de la voix (forte ou faible voix) et est principalement contrôlée par la pression sous-glottique, liée à la pression de l'air expiré et à la tension d'accolement des cordes vocales ; plus la pression sous-glottique sera grande, plus l'intensité vocale le sera également.

Mesurée en décibel (dB), on distingue différentes plages d'intensité vocale [18] :

- 10 dB pour une voix chuchotée.
- (35-40) dB pour une voix calme.
- (50-70) dB pour une voix conversationnelle (la voix courante).
- (80-90) dB pour une voix criée.

- **Le timbre**

Il est tout comme les empreintes digitales, représente l'identité de la voix, puisqu'il nous permet d'identifier une personne à l'écoute de sa voix. Il dépend des propriétés physiologiques (masse et longueur) et biomécaniques des cordes vocales, et des caractéristiques anatomiques des cavités de résonance et dynamiques (mobilité du voile de palais et de la langue, tonicité des lèvres et des joues).

II.4.2 L'analyse cepstrale

Le son $s(n)$ est modélisé comme étant la convolution d'une certaine excitation $e(n)$ avec les conduits vocaux dont la réponse impulsionnelle $h(n)$, comme le montre la figure suivante.

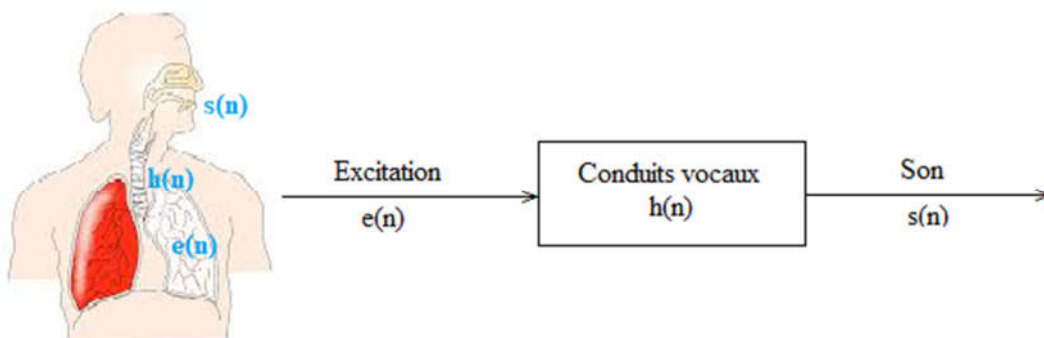


Figure II.9 : La convolution de l'excitation avec les conduits respiratoires.

Pour la plupart des applications du traitement de signal, nous avons seulement besoin que des spectres d'amplitude [20].

Dans l'analyse cepstrale, le spectre d'amplitude est représenté comme combinaison des fonctions de base cosinus avec des fréquences variables. Les coefficients cepstraux sont les amplitudes des fonctions de base.

Sachant que le vrai cepstre du signal est défini comme l'inverse de la Transformée de Fourier du logarithme du spectre d'amplitude.

- **Coefficients MFCC (Mel-scaled Frequency Cepstral Coefficients)**

Les coefficients cepstraux sur l'échelle Mel sont les paramètres les plus généralement utilisés dans la reconnaissance automatique de la parole, la détection et la classification des pathologies de la voix.

Ils sont basés sur la nature des perceptions auditives humaines qui ne peuvent pas percevoir les fréquences supérieures à 1 KHz sur une échelle linéaire [6].

Ces coefficients sont une représentation définie comme étant la transformée cosinus inverse du logarithme du spectre de l'énergie du segment de la parole. L'énergie spectrale est calculée en appliquant un banc de filtres uniformément espacés sur une échelle fréquentielle modifiée, appelée l'échelle Mel. L'échelle Mel redistribue les fréquences selon une échelle non linéaire qui simule la perception humaine des sons [21]. La figure ci-dessous montre les phases d'obtention des coefficients MFCC :

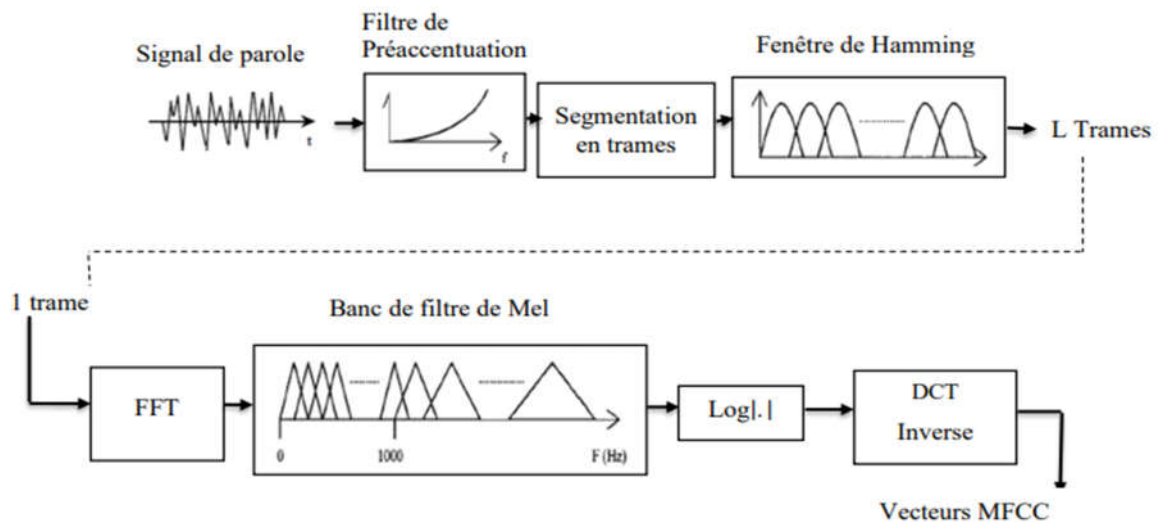


Figure II.10 : Phases d'obtention des coefficients MFCC [4].

➤ **Préaccentuation**

Après avoir filtré et échantillonner le signal vocal à une fréquence donnée, la préaccentuation est nécessaire, afin de relever les hautes fréquences qui sont moins énergétiques que les basses fréquences, en le passant dans un filtre numérique à réponse impulsionnelle finie (RIF) de premier ordre donnée comme suit :

$$H(z) = 1 - \alpha z^{-1} \quad \text{avec } 0.9 \leq \alpha \leq 1 \quad (\text{II.1})$$

Après ce processus, le signal devient :

$$s_1(n) = s(n) - \alpha s(n-1) \quad (\text{II.2})$$

➤ **Segmentation en trames**

Comme le signal vocal est un signal non stationnaire et les méthodes du traitement utilisées opèrent sur des signaux stationnaires, il est utile d'effectuer l'analyse de ce signal sur des trames successives de parole considérées stationnaires.

Donc, le signal pré-accentué $s_1(n)$ sera découpé en trames de N échantillons de parole dont chaque trame corresponde à environ 20 à 30 ms de parole.

➤ **Le Fenêtrage**

L'objectif de cette étape est de réduire les discontinuités aux frontières des trames produit par le processus de segmentation. Dans le domaine spectral, ces discontinuités se manifestent par des lobes secondaires. Donc, on parle d'une convolution du spectre à court terme avec la réponse d'amplitude de la fonction fenêtre.

La fonction fenêtre la plus utilisée dans le traitement de la parole est la fenêtre de Hamming, car elle entraîne un minimum de distorsion spectrale du signal de parole, par rapport aux autres fenêtres. Elle est définie par l'équation ci-dessous [4] :

$$h(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N}\right), & 0 \leq n \leq N - 1 \\ 0, & \text{ailleurs} \end{cases} \quad (\text{II.3})$$

Alors, le résultat du fenêtrage est le signal $s_2(n)$ donné par la formule (II.4) :

$$s_2(n) = s_1(n) \times h(n) \quad (\text{II.4})$$

➤ **La Transformée de Fourier Rapide (FFT)**

Ce processus permet de convertir chacune des trames précédentes du domaine temporel au domaine fréquentiel en utilisant la FFT (Fast Fourier Transform) qui est un algorithme rapide pour le calcul de la transformée de Fourier discrète (en anglais DFT : Discrete Fourier Transform).

➤ **Banc de filtres à l'échelle de Mel**

A cause de la nécessité d'un grand nombre de filtres dans la gamme des basses fréquences, un banc de filtres triangulaires espacés selon l'échelle Mel est utilisé.

Le but de ce banc de filtres est de simuler les filtres des bandes critiques du mécanisme d'audition.

L'échelle de Mel est linéaire pour les fréquences inférieures à 1KHz et logarithmique pour les fréquences supérieures (figure II.11).

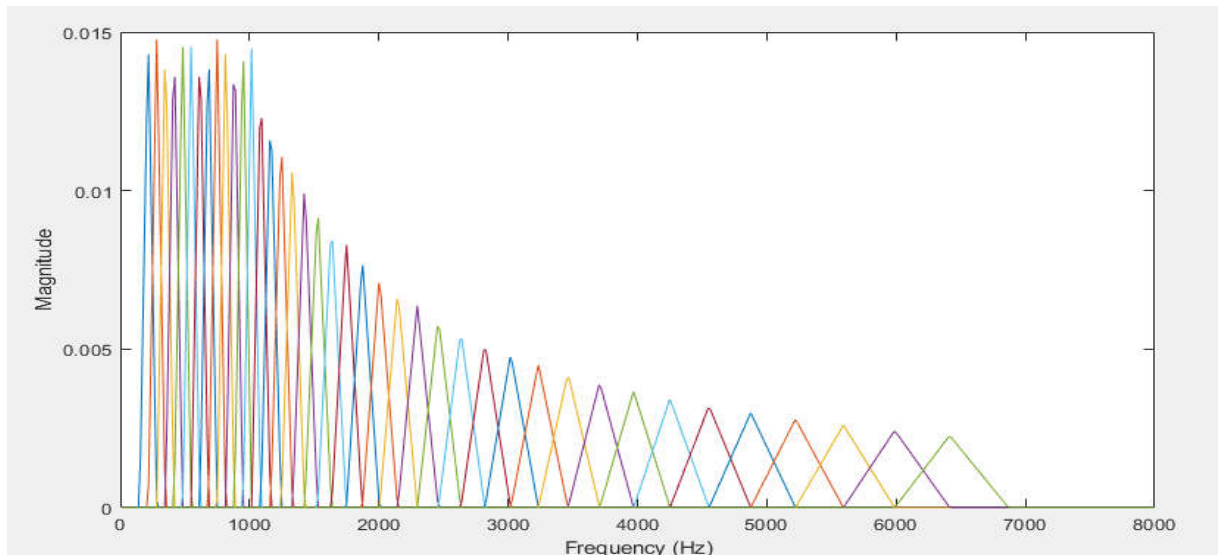


Figure II.11 : Approximation de l'échelle de Mel : Banc de filtres triangulaires [22].

Elle peut être approximée par l'équation suivante :

$$f_{Mel} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (\text{II.5})$$

Où f est la fréquence dans l'échelle linéaire et f_{Mel} la fréquence perçue.

➤ Calcul de logarithme

Les sorties du filtre triangulaire sont comprimées en utilisant le logarithme. Le passage dans le domaine log-spectral permet de déconvoluer le signal et de le compresser. Par conséquent, cette étape convertit simplement la multiplication, de l'amplitude de la transformée de Fourier, en addition [4].

➤ La Transformée en Cosinus Discrète (TCD) inverse

La DCT, pour Discrete Cosine Transform en anglais, inverse est appliquée au logarithme de la sortie du banc de filtres de Mel pour la reversion du log-Mel-spectre vers le domaine temporel.

Et donc, nous avons obtenu les coefficients MFCCs.

$$MFCC(i) = \sum_{j=1}^{N_f} S(j) \cos \left(i \left(j - \frac{1}{2} \right) \frac{\pi}{N_f} \right), \quad i = 1, 2, \dots, M \quad (\text{II.6})$$

Où N_f représente le nombre de filtres Mel utilisés.

M représente le nombre de coefficients MFCC.

$S(j)$ est le logarithme de l'énergie obtenue avec le filtre triangulaire d'indice j .

II.4.3 Le pitch

Le pitch de la voix humaine qui est l'un des attributs acoustiques est le phénomène prosodique le plus expressif. Il exprime la hauteur ou plus précisément la fréquence fondamentale (F0) perçue par l'oreille.

C'est grâce à lui ou plutôt à ses variations que nous pouvons discerner les différentes parties d'un message parlé ou d'une phrase, et aussi de distinguer des informations sur l'émotivité du locuteur et donc une meilleure segmentation des informations contenues dans le message.

Comme tous les paramètres cités précédemment, le pitch peut être utilisé pour la détection des voix saines et pathologiques.

Les trois phases principales de l'estimation du pitch sont les suivantes [4] :

Phase de prétraitement : réservée à la préparation de signal issu d'un microphone, choisir la durée des trames d'analyses et du recouvrement afin de moins compromettre la condition de stationnarité.

Phase de traitement : réservée à l'extraction de F0 par plusieurs algorithmes appartenant à l'une des trois catégories temporelles, fréquentielles et statistiques.

Phase de post-traitement : qui a pour rôle de diminuer les erreurs d'estimation qui sont :

- Erreur de voisement : la présence de F0 dans des zones non voisées et son absence dans des zones voisées.
- Erreur grossière : F0 correspond à une harmonique ou sous harmonique.
- Erreur fine : la valeur trouvée est située à plus ou moins de 10 % de la valeur réel

II.5 Evaluation des troubles de la voix

L'évaluation de la voix est fondée d'une part sur une évaluation perceptive à l'oreille, de la qualité de la voix et, d'autre part, sur une analyse instrumentale des paramètres ou des mesures physiques du son obtenue à l'aide des microphones et capteurs, des logiciels et des techniques de traitement du signal.

L'évaluation perceptive, bien qu'elle soit considérée comme la base de la consultation phoniatrique, possède une certaine subjectivité ; l'analyse instrumentale a pour avantage d'être objective.

C'est pourquoi on intègre les deux méthodes d'évaluation, pour pouvoir évaluer la dysphonie selon plusieurs angles d'observation.

II.5.1 Evaluation subjective

On distingue deux formes d'évaluation subjective de la voix : Celle du spécialiste : c'est l'analyse perceptive (ou perceptuelle), celle du patient : c'est l'auto-évaluation par questionnaire.

Pour l'analyse perceptive, le spécialiste doit juger de la qualité de la voix et de ses caractéristiques perceptives. Il peut pour cela utiliser différentes échelles dont la plus utilisée est l'échelle GRBASI qui évalue six paramètres :

G : grade (Grade), pour l'impression générale de sévérité de la dysphonie.

R : raucité (Roughness), pour la raucité qui réfère à l'impression d'irrégularité de vibration.

B : souffle (Breathiness), pour le souffle qui évalue la présence de fuite d'air altérant la voix, par manque d'accolement glottique phonatoire.

A : sensation de faiblesse (Asthenia), pour l'asthénie qui traduit le manque de puissance, la faiblesse de la voix.

S : sensation de forçage (Strain) pour le serrage, l'impression d'effort et de tension, notamment lorsqu'il y a un forçage vocal.

I : (Instability), pour mesurer l'instabilité qui correspond aux variations de la qualité de la voix d'un instant à l'autre.

L'évaluation perceptive est essentielle puisque l'oreille est l'outil le plus adéquat et le plus accessible pour juger la qualité de la voix. Elle est cependant insuffisante du fait des biais qu'elle présente.

Lors de l'utilisation d'un groupe d'écoute, la différence de scores inter-évaluateur (jugement de deux évaluateurs sur une même voix) et même intra-évaluateur (jugement d'un même évaluateur à des temps différents) est significativement réduite, mais ce dernier est difficile à mettre en œuvre en pratique clinique courante. Les évaluations dites objectives compensant cette importante subjectivité [23].

Le patient évalue parallèlement la qualité de sa voix en se basant sur un questionnement d'auto-évaluation. Ce questionnement est l'occasion de l'amener à s'interroger sur l'importance respective de l'altération acoustique et des conséquences de celle-ci.

Le Voice Handicap Index (VHI) est l'auto-questionnaire évaluant quantitativement le handicap vocal le plus répandu. Il distingue l'impact de la dysphonie sur les sphères

fonctionnelle (impact dans les activités de la vie quotidienne), émotionnelle (réponse affective) et physique (perception que le patient a des caractéristiques physiques de sa voix).

II.5.2 L'évaluation objective

L'évaluation objective de la voix qualifie mais surtout quantifie la fonction vocale et ses éventuels dysfonctionnements à partir des mesures acoustiques et aérodynamiques en extrayant des indices acoustiques pertinents, ces derniers permettent de déterminer les caractéristiques de la voix afin d'avoir un aperçu de l'état du larynx du patient. Il est nécessaire de combiner différentes mesures complémentaires afin de prendre en compte l'aspect multidimensionnel de la production vocale.

Les mesures acoustiques révèlent les caractéristiques audibles de la dysphonie. Il s'agit principalement des mesures de la fréquence et de l'intensité, de leur stabilité, ainsi que de l'analyse du spectre du son. Les études de caractérisation des troubles de la voix appuient sur le développement de méthodes analytiques pour estimer les dyspériodicités vocales dans le signal de parole. Les indices les plus utilisées sont :

II.5.2.1 Indices de perturbation du pitch

- **Le Jitter**

Il correspond aux perturbations à court terme de la fréquence fondamentale F0 (pitch) d'un cycle glottique à l'autre (figure II.12), il est mesuré sur un /a/ tenu et est considéré le meilleur corrélat acoustique de la raucité. Donc, c'est un paramètre acoustique très important pour la détection d'un désordre vocal.

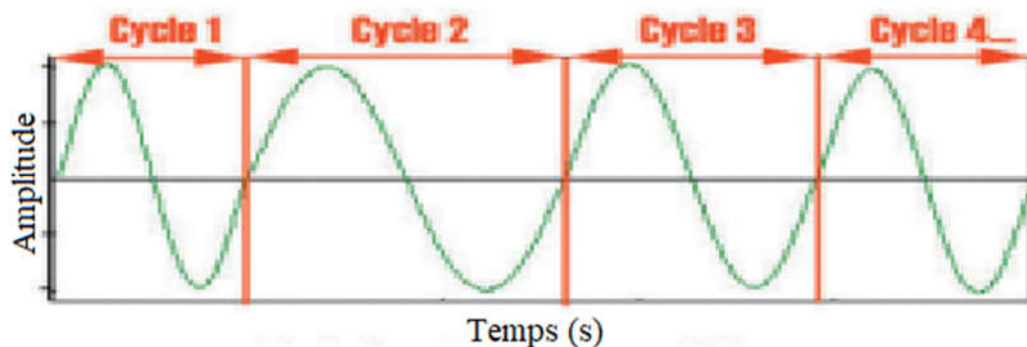


Figure II.12: Variations de fréquence (Jitter) [24].

Il existe plusieurs représentations du Jitter.

- Le Jitter local absolu (en seconde), appelé aussi Jitta, est la différence moyenne absolue entre des périodes consécutives.

$$Jitta = \frac{1}{N-1} \sum_{i=2}^N |T_i - T_{i-1}| \quad (\text{II.7})$$

Où T_i est la période fondamentale de la i ème trame avec $T_i=1/F_i$, N est le nombre de trames. Le seuil normal/ pathologique du Jitter est 83.2 μ s.

- Le Jitter local relatif (%) ou le Jitter est la différence moyenne absolue entre des périodes consécutives, divisée par la période moyenne.

$$Jitter(\%) = \frac{\frac{1}{N-1} \sum_{i=2}^N |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i} \quad (\text{II.8})$$

Le seuil normal/ pathologique du Jitter (%) est 1.04 %, selon le manuel du logiciel PRAAT qui est l'un des plus importants logiciels téléchargeables gratuitement sur internet pour l'analyse acoustique de la parole et la manipulation de sons [25].

- Le Jitter RAP pour Relative Average Perturbation en anglais, est la différence moyenne absolue entre une période et la moyenne de celle-ci et de ses deux voisines, divisée par la période moyenne.

$$Jitter\ RAP = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| T_i - \frac{1}{3}(T_{i-1} + T_i + T_{i+1}) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (\text{II.9})$$

Le seuil normal/ pathologique du Jitter RAP est 0.68 %.

- Le Jitter PPQ5 (five-point Period Perturbation Quotient) est la différence moyenne absolue entre une période et la moyenne de celle-ci et de ses quatre voisines les plus proches, divisée par la période moyenne.

$$PPQ5 = \frac{\frac{1}{N-4} \sum_{i=3}^{N-2} \left| T_i - \frac{1}{5}(T_{i-2} + T_{i-1} + T_i + T_{i+1} + T_{i+2}) \right|}{\frac{1}{N} \sum_{i=1}^N T_i} * 100 \quad (\text{II.10})$$

Le seuil normal/ pathologique du Jitter PPQ5 est 0.84 %.

II.5.2.2 Indices de perturbations d'amplitude

- **Le Shimmer**

Il correspond aux perturbations à court terme de l'amplitude d'un cycle glottique à l'autre. Comme le Jitter, il est évalué sur un /a/ tenu.

La figure suivante montre ces perturbations d'amplitudes.

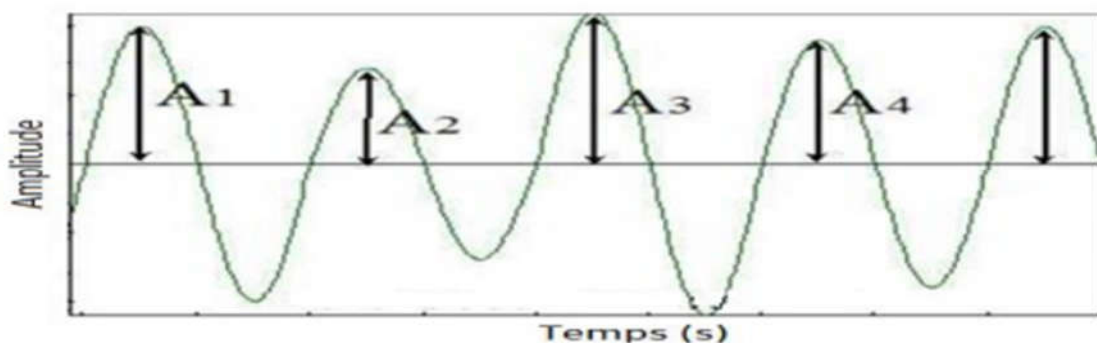


Figure II.13 : Variations d'amplitude (Shimmer) [24].

Dans la littérature [24], plusieurs variantes du Shimmer existent :

- Le Shimmer local, appelé aussi Shim est la différence absolue moyenne entre les amplitudes de deux périodes consécutives, divisée par l'amplitude moyenne.

$$\text{Shimmer local} = \frac{\frac{1}{N-1} \sum_{i=2}^N |A_i - A_{i-1}|}{\frac{1}{N} \sum_{i=1}^N A_i} \quad (\text{II.11})$$

Où A_i sont les données d'amplitudes crête à crête extraites, et N est le nombre de périodes extraites.

Le seuil pathologique de Shimmer est 3.81 %.

- Le Shimmer local dB, ou ShdB est le logarithme absolu moyen en base 10 de la différence entre les amplitudes de deux périodes consécutives, multiplié par 20.

$$\text{Shimmer local dB} = \frac{1}{N-1} \sum_{i=1}^{N-1} \left| 20 \log \left(\frac{A_{i+1}}{A_i} \right) \right| \quad (\text{II.12})$$

Le seuil pathologique de ShdB est 0.35 dB.

- Le Shimmer APQ3 pour three-point Amplitude Perturbation Quotient en anglais, est la différence absolue moyenne entre l'amplitude d'une période et la moyenne des amplitudes de ses voisines, divisée par l'amplitude moyenne.

$$\text{APQ3} = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| A_i - \frac{1}{3} (A_{i-1} + A_i + A_{i+1}) \right|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (\text{II.13})$$

- Le Shimmer APQ5 pour five-point Amplitude Perturbation Quotient en anglais, est la différence absolue moyenne entre l'amplitude d'une période et la moyenne des

amplitudes de cette période et de ses quatre périodes voisines les plus proches, divisée par l'amplitude moyenne.

$$APQ5 = \frac{\frac{1}{N-4} \sum_{i=3}^{N-2} |A_i - \frac{1}{5}(A_{i-2} + A_{i-1} + A_i + A_{i+1} + A_{i+2})|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (\text{II.14})$$

- Le Shimmer APQ11 pour 11-point Amplitude Perturbation Quotient en anglais ou APQ, est la différence absolue moyenne entre l'amplitude d'une période et la moyenne des amplitudes de celle-ci et de ses dix périodes voisines les plus proches, divisée par l'amplitude moyenne.

$$APQ11 = \frac{\frac{1}{N-10} \sum_{i=6}^{N-5} |A_i - \frac{1}{11}(A_{i-5} + \dots + A_{i-1} + A_i + A_{i+1} + \dots + A_{i+5})|}{\frac{1}{N} \sum_{i=1}^N A_i} * 100 \quad (\text{II.15})$$

Le seuil pathologique d'APQ est 3.07 %.

II.5.2.3 HNR (Harmonic to Noise Ratio)

Le rapport harmonique sur bruit (HNR) mesure le rapport entre l'énergie harmonique qui résulte des interruptions quasi périodiques du passage de l'air entre les cordes vocales, et l'énergie du bruit qui correspond aux inter-harmoniques c'est-à-dire le passage turbulent et continu de l'air à travers la glotte [26].

Selon le manuel du PRAAT, on connaît un cas pathologique lorsqu'un locuteur a un HNR inférieur à 40 dB pour [i], car sa voix a baissé au-dessus de 2000 Hz.

Dans PRAAT, il est estimé par la relation suivante :

$$HNR_{dB} = 10 * \log_{10} \left(\frac{ACv(T)}{ACv(0) - ACv(T)} \right) \quad (\text{II.16})$$

Où :

ACv (0) est le coefficient d'autocorrélation à l'origine qui est aussi l'énergie du signal.

ACv (T) est la composante d'autocorrélation correspondante à la période du pitch.

La différence (ACv (0) – ACv (T)) peut être supposée comme l'énergie du bruit.

II.5.2.4 NHR (Noise to Harmonic Ratio)

A l'inverse du HNR, le NHR est le rapport bruit sur harmonique, il mesure le rapport moyen entre le bruit dans la bande de fréquence [1500-4500] Hz et l'énergie harmonique dans la bande [70-4500] Hz.

Sa valeur est autour de 0.1, et augmente avec la sévérité des dysphonies.

II.6 Conclusion

Un travail sur la classification des voix pathologiques nécessite la connaissance du mécanisme de production de la parole. Dans ce chapitre, nous avons présenté quelques notions de ce mécanisme, les différents types de dysphonies et leurs évaluations subjectives et objectives. De plus, nous avons parlé des paramètres utilisés pour la classification à savoir les MFCC, le Jitter et ses variantes, le Shimmer et ses variantes et le HNR ainsi que les étapes d'obtention de ces paramètres.

Chapitre III

Implémentation, résultats et interprétations de la classification

III.1 Introduction

Les techniques d'apprentissage automatique utilisées pour l'identification et la classification des pathologies de la voix ont été implémentées sur des fichiers des bases de données de la parole.

Ce chapitre sera consacré aux conditions des implémentations, les mesures de performance, les résultats de la classification et leurs interprétations.

III.2 Bases de données utilisées pour la classification des voix saines et pathologiques

Pour la classification des voix saines et pathologiques, plusieurs bases de données ont été utilisées. Il existe des bases de données publiques qui peuvent être librement ou commercialement distribuées, et des bases de données privées qui ne sont pas disponibles.

Parmi les bases de données publiques, on trouve la base MEEI pour Massachusetts Eye and Ear Infirmary, et la base SVD pour Saarbrücken Voice Database.

III.2.1 Base de données SVD

La SVD est une base de données allemande gratuitement téléchargeable [27], enregistrée par l'institut de phonétique de l'université de Saarland.

Cette base de données est une collection d'enregistrements de voix de plus de 2000 patients (normal ou souffre d'une pathologie (environ 71 pathologies)).

Elle contient des voyelles soutenues /a/, /i/ et /u/ produit à un ton normal, bas, haut et bas-haut-bas, avec une phrase parlée en allemand « Guten Morgen, Wie geht es Ihnen ? » qui signifie « Bonjour, Comment vas-tu ? », de durée allant de 1 à 3 secondes.

Tous les enregistrements ont été échantillonnés à 50 KHz avec une résolution de 16 bits.

III.2.2 Base de données MEEI

La base de données MEEI est l'une des bases de données les plus populaires dans le domaine de la pathologie vocale.

Elle a été développée et commercialisée par le laboratoire « MEEI Voice and Speech Kay Elemetrics Corp » [28].

Cette base de données est une collection d'enregistrements de voix (environ 1400 voix). Elle contient la voyelle tenue /ah/ (53 normal et pathologique 657), et la phrase « le passage de l'arc-en-ciel » (53 normal et 662 pathologiques).

Toutes les voyelles normales et 77 voyelles pathologiques sont échantillonnées à 50 KHz, tandis que les 580 voyelles pathologiques restantes sont échantillonnées à 25 KHz.

III.2.3 Base de données utilisée

Dans ce mémoire, un ensemble de 150 enregistrements de la voyelle tenue /ah/ de la base de données MEEI a été sélectionné.

Ces enregistrements se répartissent en six groupes, un groupe normal et cinq pathologiques, où chaque groupe contient un nombre précis de fichiers comme il est mentionné dans le tableau suivant.

Sujets	Type de pathologie	Nombre de fichiers	Total
Pathologique	Kératose (Keratosi)	25	97
	Paralysie (Paralysis)	20	
	Nodules des cordes vocales (Vocal fold nodules)	19	
	Polypes des cordes vocales (Vocal fold polyp)	20	
	Dysphonie spasmodique d'adduction (Adductor spasmodic dysphonia)	17	
Normal	-	53	53

Tableau III.1 : Distribution des fichiers normaux et pathologiques utilisés.

La fréquence d'échantillonnage de ces fichiers est de 25 KHz et une résolution de 16 bits, sachant que les fichiers avec des fréquences d'échantillonnages de 50 KHz et de 10 KHz seront ré-échantillonnés à la fréquence de 25 KHz.

III.3 Conditions et Paramètres de simulation

III.3.1 Conditions d'implémentation

Pour nos simulations de la classification des fichiers étudiés, nous avons utilisé l'application Classification Learner qui est l'une des applications du logiciel MATLAB version 2019 (Matlab R2019b) qui contient plusieurs applications et outils pour le calcul technique et l'analyse des signaux à haute performances.

III.3.2 Paramètres acoustiques utilisés

Lorsqu'on parle de la classification des voix saines et pathologiques, on parle de certaines caractéristiques représentatives de la parole qui sont présentées dans le chapitre précédent, qui nous a permis de classer la voix soit normale soit pathologique et le type de la pathologie selon des seuils précis.

Dans notre travail, nous avons utilisé les coefficients MFCC, les paramètres Jitter Rap et Jitter Local, pour les variations du Pitch ; Shimmer Local, Shimmer Local dB, Shimmer APQ3, Shimmer APQ5 et Shimmer APQ11 pour les variations d'amplitude et le HNR pour les mesures du bruit.

Les paramètres	Jitter local absolu	Jitter local relatif	Jitter rap	Jitter ppq5
Le seuil normal/pathologique	83.2 μ s	1.04 %	0.68 %	0.84 %

Tableau III.2 : Les seuils normal/ pathologique du Jitter selon le logiciel PRAAT.

Les paramètres	Shimmer local	Shimmer local dB	Shimmer apq
Le seuil normal/pathologique	3.81 %	0.35 dB	3.07 %

Tableau III.3 : Les seuils normal/ pathologique du Shimmer selon le logiciel PRAAT.

III.3.3 Extraction des paramètres

Après avoir sélectionné les 150 fichiers de la base de données utilisée, nous passons à l'étape d'extraction des paramètres, les 13 coefficients MFCC sont obtenus en utilisant un programme MATLAB qui nous a permis de déviser chaque fichier de tous les sous-répertoires (adductor spasmodic dysphonia, keratosis, normal, paralysis, vocal fold nodules et vocal fold polyp) en trames de 30 ms et de calculer automatiquement ces 13 coefficients de chaque trame. En plus, des paramètres de variation du Pitch, d'amplitude et du HNR qui sont obtenus en utilisant un script programmé sous le logiciel PRAAT qui nous a permis d'avoir les valeurs du Jitter Rap, Jitter Local, du Shimmer et ses différentes représentations, ainsi que le HNR de chaque fichier.

Enfin, en utilisant MATLAB, ces 21 paramètres extraits (13 coefficients MFCC et 8 paramètres extraits par PRAAT) seront rassemblés dans un tableau de 22 colonnes, 21 colonnes pour les paramètres et une colonne pour l'étiquette (Label) de chaque fichier (soit normal soit pathologique), et un nombre de lignes égal au nombre total de trames de tous les fichiers des sous-répertoires.

III.4 Validation

Lors de l'évolution de tout algorithme d'exploitation de données, on divise les données aléatoirement en deux parties, la plus grande partie est pour l'apprentissage et la plus petite est pour le test, d'une façon pour que les deux bases soient d'une part, représentatives de l'ensemble de données globale.

La validation estime les performances du modèle sur les nouvelles données (données de test) par rapport aux données d'apprentissage et nous aide à choisir le meilleur modèle et elle protège contre le sur-apprentissage. Diverses techniques ont été développées pour réaliser la validation, on cite dans ce qui suit deux méthodes qui sont la validation simple et la validation croisée.

III.4.1 Validation simple

Parmi les différentes stratégies de ré-échantillonnage des données, l'une des plus simples est la validation simple (hold-out). La validation simple se produit lorsqu'on divise notre ensemble de données en un ensemble d'apprentissage et un ensemble de test.

L'ensemble d'apprentissage est ce sur quoi le modèle est entraîné, et l'ensemble de test est utilisé pour voir dans quelle mesure ce modèle fonctionne sur des données invisibles. Une division courante lors de l'utilisation de cette méthode consiste à utiliser 80 % des données pour l'apprentissage et les 20 % restants pour les tests (figure III.1).

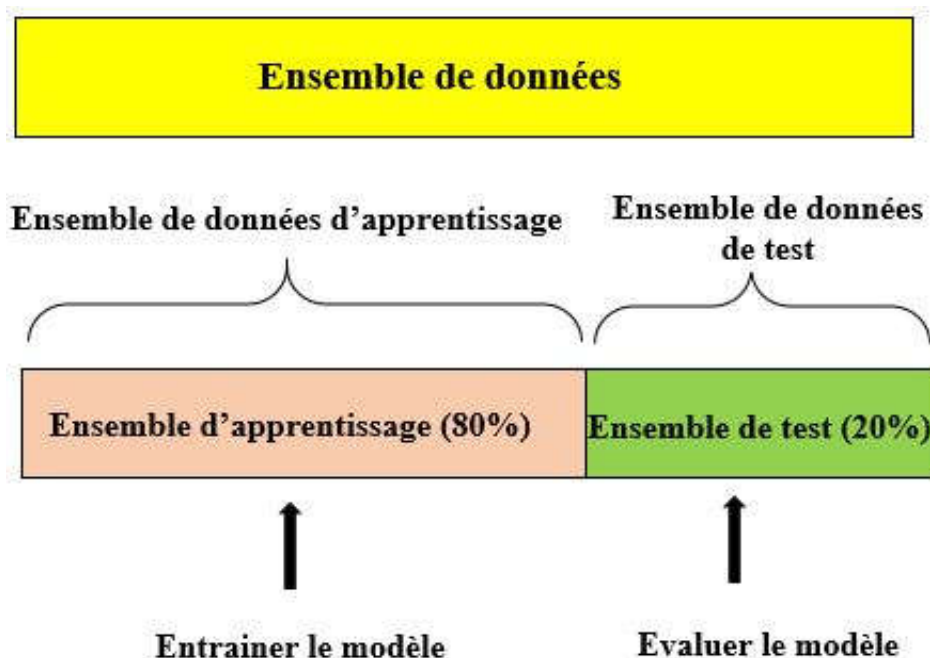


Figure III.1 : Méthode de validation simple.

III.4.2 Validation croisée

La validation croisée est une procédure de ré-échantillonnage utilisée pour évaluer les modèles d'apprentissage automatique sur un échantillon de données limité. La procédure a un seul paramètre appelé k qui fait référence au nombre de groupes dans lesquels un échantillon de données donné doit être divisé.

La validation croisée est principalement utilisée dans l'apprentissage automatique appliqué pour estimer la compétence d'un modèle d'apprentissage automatique sur des données invisibles [29].

La procédure d'une validation croisée k -fold est comme suit :

- Partitionner l'ensemble d'apprentissage disponible en k sous-ensembles disjoints de taille approximativement égale.
- Le modèle est entraîné à l'aide de $(k - 1)$ sous-ensembles, ensemble représentant l'ensemble d'apprentissage
- Ensuite, le modèle est appliqué au sous-ensemble restant, qui est désigné comme l'ensemble de validation, et la performance est mesurée.

Cette procédure est répétée jusqu'à ce que chacun des k sous-ensembles ait servi d'ensemble de validation. La figure (III.2) montre les étapes de cette méthode avec $k=5$.

Iteration 1	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4	Echantillon 5
Itération 2	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4	Echantillon 5
Itération 3	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4	Echantillon 5
Itération 4	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4	Echantillon 5
Itération 5	Echantillon 1	Echantillon 2	Echantillon 3	Echantillon 4	Echantillon 5

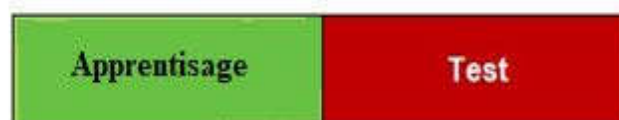


Figure III.2: Etapes de la méthode de validation croisée ($k=5$).

III.5 Mesures de performances

Pour évaluer les performances des classificateurs des pathologies de la voix, nous avons pris en compte certains indicateurs tels que la sensibilité, la spécificité et l'efficacité. Ces mesures sont utilisées dans la majorité des travaux de classification en général, c'est pourquoi nous les avons cités dans notre travail, en commençant tout d'abord par introduire la notion de la matrice de confusion.

- **Matrice de confusion**

Une fois le modèle est développé, la phase suivante consiste à calculer ses performances à l'aide d'un outil nommé la matrice de confusion. Une matrice de confusion est organisée sous forme d'un tableau avec 4 combinaisons différentes de classes prédites et réelles.

Le Tableau (III.4) représente une matrice de confusion à deux dimensions.

		Classes Prédites	
		Décision Positive	Décision Négative
Classes Réelles	Etiquette Positive	Vrais Positifs (VP)	Faux Négatifs (FN)
	Etiquette Négative	Faux Positifs (FP)	Vrais Négatifs (VN)

Tableau III.4 : Matrice de confusion.

Etiquette positive : échantillon vocal pathologique.

Etiquette négative : échantillon vocal sain.

Décision Positive : échantillon vocal prédit comme pathologique.

Décision Négative : échantillon vocal prédit comme sain.

Vrai Positif (VP) : l'échantillon vocal est pathologique et l'algorithme le reconnaît.

Vrai Négatif (VN) : l'échantillon vocal est sain et l'algorithme le reconnaît.

Faux Positif (FP) : l'échantillon vocal est sain mais l'algorithme le reconnaît comme pathologique.

Faux Négatif (FN) : l'échantillon vocal est pathologique mais l'algorithme le reconnaît comme sain.

La sensibilité : ('Sensitivity' en anglais), est le rapport entre les échantillons pathologiques réellement identifiés et le nombre total d'échantillons pathologiques, appelée aussi le taux des vrais positifs. Elle mesure la capacité d'un classificateur à détecter les individus malades.

$$\text{Sensibilité} = \frac{VP}{VP+FN} \quad (\text{III.1})$$

La spécificité : ('specificity' en anglais), est le rapport entre les échantillons normaux vraiment classés et le nombre total d'échantillons normaux. Elle mesure la capacité d'un classificateur à détecter les individus sains.

$$\text{spécificité} = \frac{VN}{VN+FP} \quad (\text{III.2})$$

L'efficacité : ('accuracy' en anglais), est le rapport entre les échantillons correctement détectés et le nombre total d'échantillons.

$$\text{efficacité} = \frac{VP+VN}{VP+VN+FP+FN} \quad (\text{III.3})$$

Taux de faux positifs (TFP) : proportion de négatifs détectés comme des positifs par le test (1-Spécificité).

$$\text{TFP} = \frac{FP}{FP+VN} \quad (\text{III.4})$$

Taux de faux négatifs (TFN) : proportion de positifs détectés comme des négatifs par le test (1-Sensibilité).

$$\text{TFN} = \frac{FN}{VP+FN} \quad (\text{III.5})$$

- **Courbe ROC et mesure AUC [30]**

La courbe ROC (Receiver Operating Characteristic) est une mesure de la qualité d'un algorithme de classification évaluée en traçant une courbe représentant la sensibilité par rapport à la complémentarité de la spécificité (1-spécificité) et en mesurant l'aire sous cette courbe (AUC). Son but est de déterminer le seuil optimal de la sensibilité et de la spécificité du classificateur.

- **Air sous la courbe (AUC)**

L'AUC (Area Under Curve) peut être interprétée comme la valeur moyenne de sensibilité pour toutes les valeurs possibles de spécificité. La surface sous la courbe est une mesure de la qualité globale du classificateur. Elle varie entre 0.5 à 1. Des valeurs d'AUC plus grandes indiquent de meilleures performances du classificateur.

Le maximum (AUC = 1) signifie que l'algorithme est parfait dans la classification entre les voix malades et non malades. D'autre part, AUC = 0 signifie que l'algorithme classe à tort tous les sujets atteints de maladies comme négatifs et tous les sujets sains comme pathologiques.

La figure (III.3) représente un exemple d'une courbe ROC et de l'AUC d'un classificateur entraîné dans l'application « Classification Learner » sur MATLAB. On a en abscisse le taux de faux positifs et en ordonnée le taux de vrais positifs.

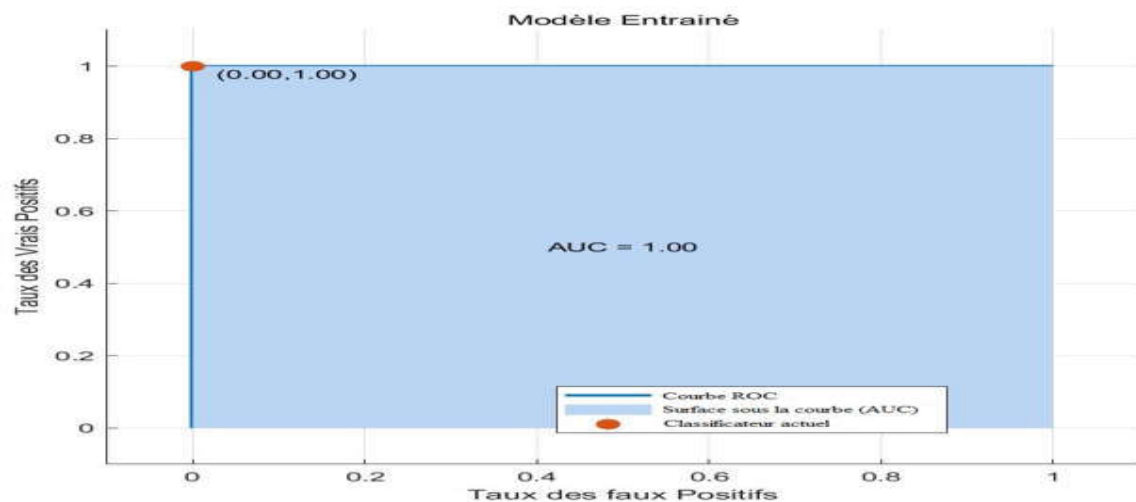


Figure III.3 : Exemple d'une courbe ROC et de l'AUC d'un classificateur entraîné.

Le marqueur sur le tracé montre les performances du classificateur entraîné. Il affiche aussi les valeurs du taux de faux positifs (TFP) et du taux de vrais positifs (TVP) du classificateur. Par exemple, un taux de faux positifs (TFP) de 0,1% indique que le classificateur attribue de manière incorrecte 10 % des observations à la classe positive. Un taux de vrais positifs (TVP) de 0,8 % indique que le classificateur actuel attribue correctement 80 % des observations à la classe positive.

Un résultat parfait sans points mal classés est un angle droit en haut à gauche du tracé. Un mauvais résultat qui n'est pas meilleur que le hasard est une ligne à 45 degrés.

III.6 Implémentations et résultats

- **Prise en main de l'application**

Dans l'onglet des applications sur Matlab, on choisit l'application « Classification Learner », la figure (III.4) illustre son interface graphique.

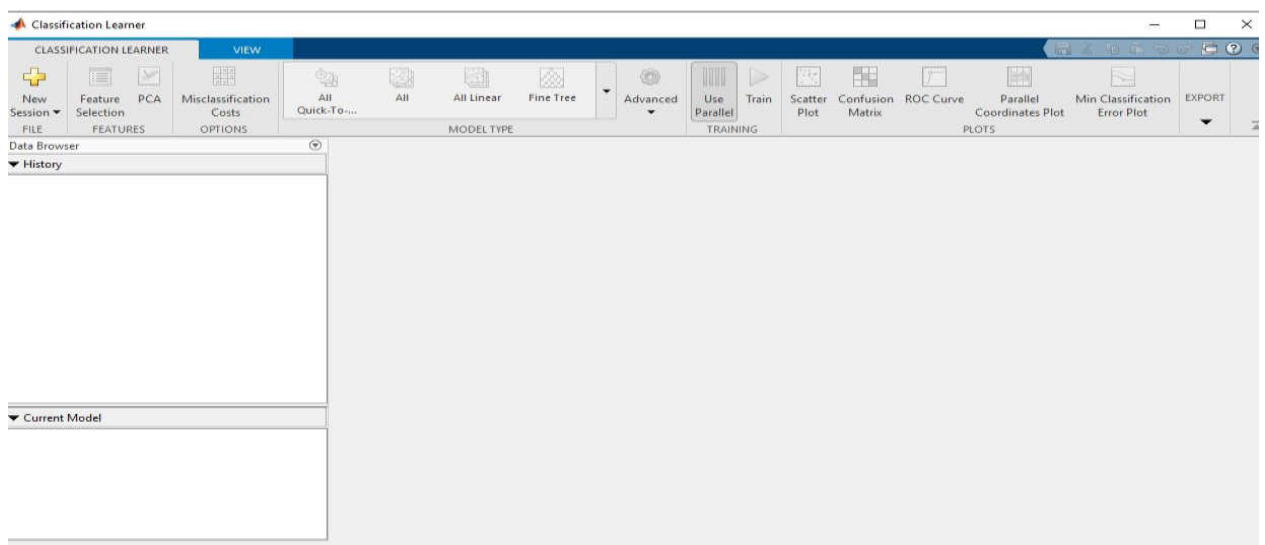


Figure III.4 : Interface graphique de l'application « Classification Learner ».

Pour commencer une nouvelle session, on clique sur « New Session » et on sélectionne depuis l'espace de travail de Matlab « From Workspace ». Une fenêtre qui nous permet d'exporter des variables disponibles sur l'espace de travail s'affiche (figure III.5).

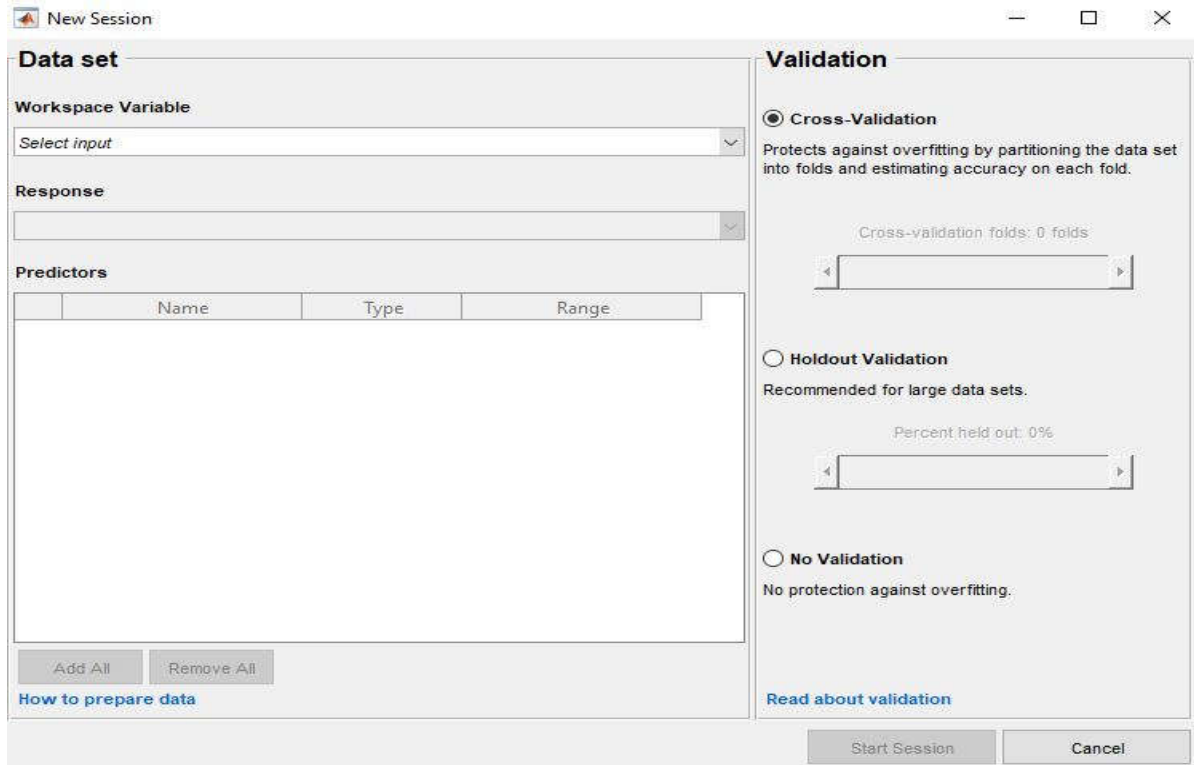


Figure III.5 : Interface de la nouvelle session.

Ensuite, on choisit les réponses, les prédicteurs et le type de validation comme le montre la figure (III.6).

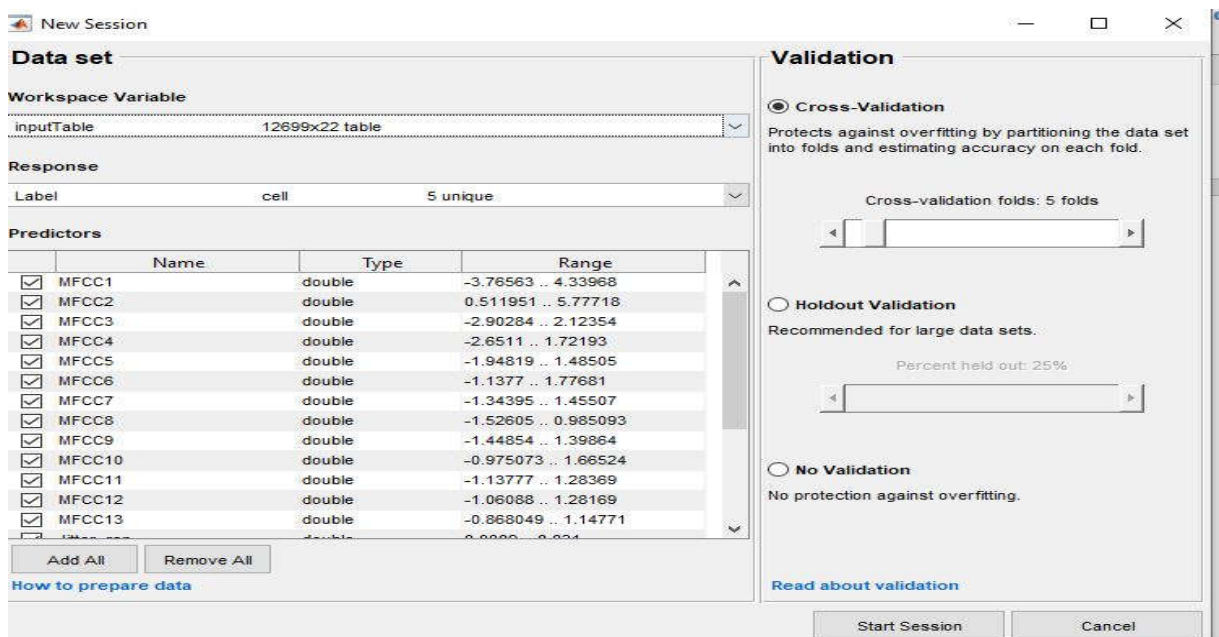


Figure III.6 : Interface de la nouvelle session après sélection.

En cliquant sur « Start Session », une nouvelle session commence (figure III.7).

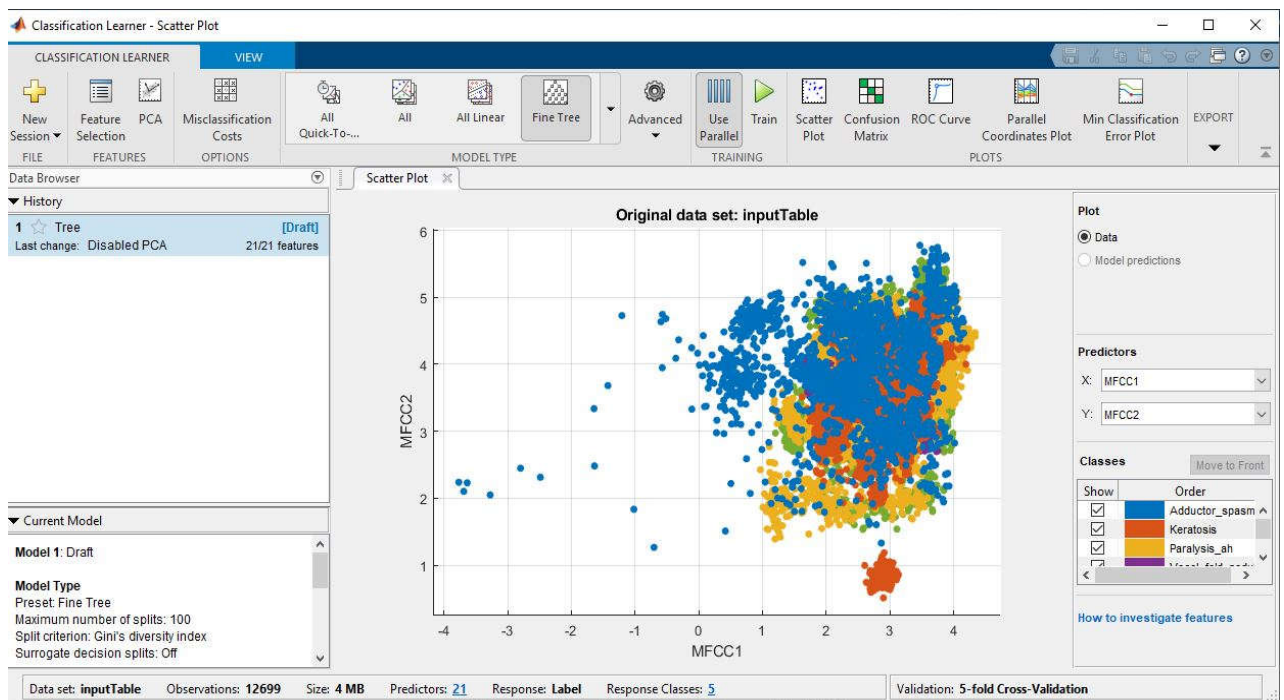


Figure III.7 : Interface de l’application avant apprentissage.

Pour commencer l’apprentissage, on choisit un modèle de classification dans l’onglet type de modèle puis on clique sur « Train ».

Parmi les méthodes de classification disponibles sur l’application MATLAB « Classification Learner », nous utiliserons les méthodes SVM et K-NN qui ont, par défaut, les propriétés présentées dans les tableaux suivants.

Modèle	K-NN Fine	K-NN Moyen	K-NN Grossier	K-NN Cosinus	K-NN Pondéré	K-NN Cubique
Nombre de voisins	1	10	100	10	10	10
Distance utilisée	Euclidienne	Euclidienne	Euclidienne	Cosinus	Euclidienne avec pondération	Minkowsky

Tableau III.5 : Propriétés par défaut des modèles K-NN.

Modèle	SVM	SVM	SVM	SVM	SVM	SVM
Type de noyau	Linéaire	Quadratique	Cubique	Gaussien fine	Gaussien moyen	Gaussien grossier
Echelle du noyau	automatique	automatique	automatique	1,10	4,60	18

Tableau III.6 : Propriétés par défaut des modèles SVM.

III.6.1 Résultats et interprétations de la détection

III.6.1.1 Résultats de la détection par la validation simple

Commençons par une validation simple en divisant l'ensemble des échantillons en deux parties aléatoires, une partie pour l'apprentissage (80%) et une partie pour le test (20%).

Les résultats de nos implémentations présentés dans le tableau ci-dessous sont exprimés en termes d'efficacité, spécificité, sensibilité, AUC le temps d'apprentissage pour chaque modèle où la valeur la plus élevée est indiquée en gras dans chaque colonne.

Méthode	Modèle	Efficacité (%)	Spécificité (%)	Sensibilité (%)	AUC	Temps d'apprentissage (min)
K-NN	K-NN Fine	100	100	100	1	1,93
	K-NN Moyen	100	100	100	1	1,28
	K-NN Grossier	99,60	99	100	1	2,02
	K-NN Cosinus	100	100	100	1	1,98
	K-NN Pondéré	100	100	100	1	1,77
SVM	Linéaire	92	87	95	0,95	2,12
	Quadratique	99,80	100	100	1	3,15
	Cubique	100	100	100	1	3,11
	Gaussien Fine	100	100	100	1	3,63
	Gaussien Moyen	100	100	100	1	3,89
	Gaussien Grossier	94,50	90	97	0,98	4,43

Tableau III.7: Mesures des performances pour la détection voix normale/ voix pathologique par validation simple 20% avec K-NN et SVM.

Nous observons que tous les modèles des méthodes SVM et K-NN assurent des bonnes performances en terme d'efficacité, spécificité, sensibilité et AUC.

Tous les modèles K-NN ont une AUC de 1, et une efficacité, spécificité et sensibilité de 100% sauf le modèle K-NN grossier qui a une efficacité de 99,60% et une spécificité de 99%.

Pour la méthode SVM, nous observons que les trois modèles cubique, gaussien fine et gaussien moyen ont une efficacité, spécificité et sensibilité de 100% et une AUC de 1, suivis par le modèle SVM quadratique qui a une efficacité de 99,80%.

Concernant les modèles SVM linéaire et gaussien grossier, ils sont aussi des bons classificateurs mais avec un peu moins de performances.

En comparant le temps d'apprentissage, nous remarquons que l'apprentissage des modèles SVM est plus lent par rapport l'apprentissage des modèles K-NN.

Le meilleur modèle des deux méthodes qui assure les bonnes performances le plus rapidement possible est le K-NN moyen avec un temps d'apprentissage = 1,28 min, et le SVM cubique avec un temps d'apprentissage = 3,11 min.

Ces résultats sont obtenus en utilisant un ordinateur Toshiba doté d'un processeur : Intel(R) Core(TM) i3-3110M CPU @ 2.40 GHz, une mémoire (RAM) de 4,00 Go et un système d'exploitation 64 bits, processeur \times 64.

III.6.1.2 Résultats de la détection par la validation croisée

Nous nous intéressons maintenant aux résultats de la détection voix normale / voix pathologique dans le cas d'une validation croisée (k-fold cross-validation) avec $k_v=5$ et $k_v=10$.

➤ $K_v=5$

Méthode	Modèle	Efficacité (%)	Spécificité (%)	Sensibilité (%)	AUC
K-NN	K-NN Fine	100	100	100	1
	K-NN Moyen	100	100	100	1
	K-NN Grossier	99,50	99	100	1
	K-NN Cosinus	100	100	100	1
	K-NN Pondéré	100	100	100	1
SVM	Linéaire	91,60	86	95	0,95
	Quadratique	99,80	100	100	1
	Cubique	100	100	100	1
	Gaussien Fine	100	100	100	1
	Gaussien Moyen	100	100	100	1
	Gaussien Grossier	94,10	88	98	0,97

Tableau III.8: Mesures des performances pour la détection voix normale/ voix pathologique par validation croisée $k_v=5$ (K-NN et SVM).

Nous observons que les modèles K-NN (fine, moyen, cosinus et pondéré) offrent des meilleures performances 100 % en termes d'efficacité, spécificité et sensibilité, et 1 pour l'AUC. Par contre, le modèle K-NN grossier est moins performant avec une efficacité et spécificité de 99,50 % et 99 % respectivement, par rapport aux modèles cités avant.

Dans le cas des SVM, les meilleures performances sont obtenues par les modèles cubique, gaussien fine et gaussien moyen dont chacun a une efficacité, spécificité et sensibilité de 100 % et une AUC de 1, suivis par le modèle SVM quadratique avec une efficacité de 99,80%.

Par contre, le modèle SVM linéaire offre, 91,60% d'efficacité, 86 % de spécificité, 95% de sensibilité et 0,95 pour l'AUC et qui sont meilleures que les performances obtenues par les méthodes de classification non basées sur Machine Learning citées dans la littérature.

➤ **Kv=10**

Méthode	Modèle	Efficacité (%)	Spécificité (%)	Sensibilité (%)	AUC
K-NN	K-NN Fine	100	100	100	1
	K-NN Moyen	100	100	100	1
	K-NN Grossier	99,60	99	100	1
	K-NN Cosinus	100	100	100	1
	K-NN Pondéré	100	86	100	1
SVM	Linéaire	91,70	100	95	0,95
	Quadratique	99,80	100	100	1
	Cubique	100	100	100	1
	Gaussien Fine	100	100	100	1
	Gaussien Moyen	100	100	100	1
	Gaussien Grossier	94,20	89	98	0,97

Tableau III.9: Mesures des performances pour la détection voix normale/ voix pathologique par validation croisée kv=10 (K-NN et SVM).

Dans le tableau (III.9) chaque colonne indique la valeur la plus élevée en gras de l'efficacité, la spécificité, la sensibilité et l'AUC pour une valeur kv=10 pour tous les modèles du SVM (linéaire, quadratique, cubique, gaussien fine, gaussien moyen, gaussien grossier) et du K-NN (fine, moyen, grossier, cosinus, pondéré).

Comme nous pouvons le voir la valeur du AUC est égale à 1 pour toutes les variantes du K-NN et SVM sauf pour le SVM linéaire et gaussien grossier (0.95 ,0.97) respectivement.

Une autre constatation est que les 3 variantes du K-NN (fine, moyen et cosinus) assurent de bons résultats en termes d'efficacité, spécificité et sensibilité (100%) chacune, alors que le modèle K-NN grossier offre une efficacité de 99.60%, une spécificité de 99 %, et une sensibilité maximale de 100 %.

Nous remarquons aussi que la méthode de classification SVM avec des noyaux cubique, gaussien fine et gaussien moyen offrent les meilleures performances en termes d'efficacité, sensibilité et spécificité, (100%) chacune.

On remarque que la plus faible efficacité est de 91.70% obtenue avec un SVM avec un noyau linéaire.

Vu les résultats des deux tableaux (III.8) et (III.9), les deux valeurs de kv=5 et kv=10 assurent presque les mêmes performances de la validation croisée.

III.6.1.3 Résultats en fonction des paramètres choisis

Nous avons étudié aussi l'influence des paramètres choisis sur les performances de notre système de détection voix normale/ voix pathologique. Nous avons sélectionné la variante du classificateur SVM avec noyau cubique et la variante K-NN moyen et une validation simple.

L'utilisation de tous les 21 paramètres donne des performances parfaites en termes d'efficacité, sensibilité, spécificité et l'AUC, comme le montre le tableau suivant, et l'utilisation des 13 paramètres MFCC uniquement n'influencent pas les performances des deux classificateurs.

Par contre, lors de l'utilisation des 8 différents paramètres du Jitter, Shimmer et du HNR sans les paramètres MFCC, le classificateur K-NN a des performances acceptables, mais moins par rapport aux deux variantes précédentes. Alors, que les performances du classificateur SVM ont clairement diminuées avec une efficacité de 61%, une spécificité très faible de 23%, une sensibilité acceptable de 84% et une AUC de 0,64.

La première variante a les plus hautes précisions car elle combine les avantages de tous les paramètres.

Modèle	Paramètres sélectionnés	Efficacité (%)	Spécificité (%)	Sensibilité (%)	AUC
K-NN Moyen	Tous les paramètres	100	100	100	1
	MFCC uniquement	100	100	100	1
	Jitter, Shimmer ,HNR	98,90	97	100	0,99
SVM Cubique	Tous les paramètres	100	100	100	1
	MFCC uniquement	99,9	100	100	1
	Jitter, Shimmer ,HNR	61	23	84	0,64

Tableau III.10 : Mesures des performances pour la détection voix normale/voix pathologique avec sélection de paramètres.

Les figures (III.8) et (III.9) représentent respectivement les courbes ROC pour un classificateur SVM avec un noyau gaussien moyen pour les deux cas MFCC uniquement et Shimmer, Jitter et HNR uniquement.

On observe que la courbe ROC dans le premier cas représente un résultat parfait par contre dans le deuxième cas la courbe est déformée, il est remarquable que le résultat est moins intéressant.

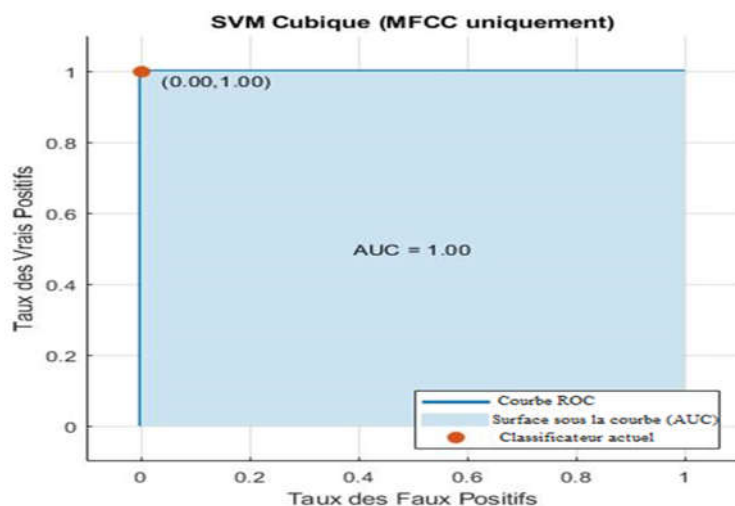


Figure III.8: Courbe ROC pour un classificateur SVM cubique en sélectionnant les MFCC uniquement.

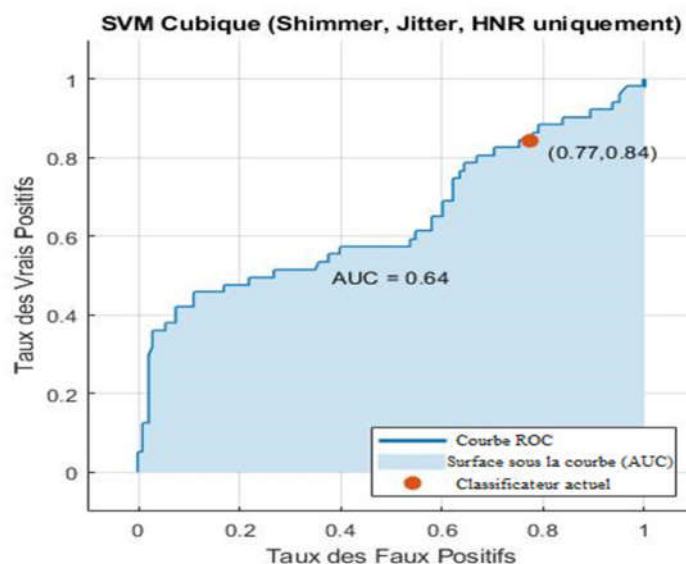


Figure III.9 : Courbe ROC pour un classificateur SVM cubique en sélectionnant les paramètres Shimmer, Jitter et HNR uniquement.

III.6.2 Résultats et interprétations de la classification des voix pathologiques

Dans notre mémoire, on a proposé un système à deux niveaux pour la détection et la classification du type de pathologie vocale. Au premier niveau, le système identifie si la voix est normale ou pathologique, on parle alors d'une classification binaire. Dans le cas d'une voix normale aucune classification supplémentaire n'est nécessaire. Si la voix est classée comme pathologique, on passe au deuxième niveau qui a pour but à identifier le type particulier de pathologie.

La figure (III.10) illustre les étapes suivies dans notre système.

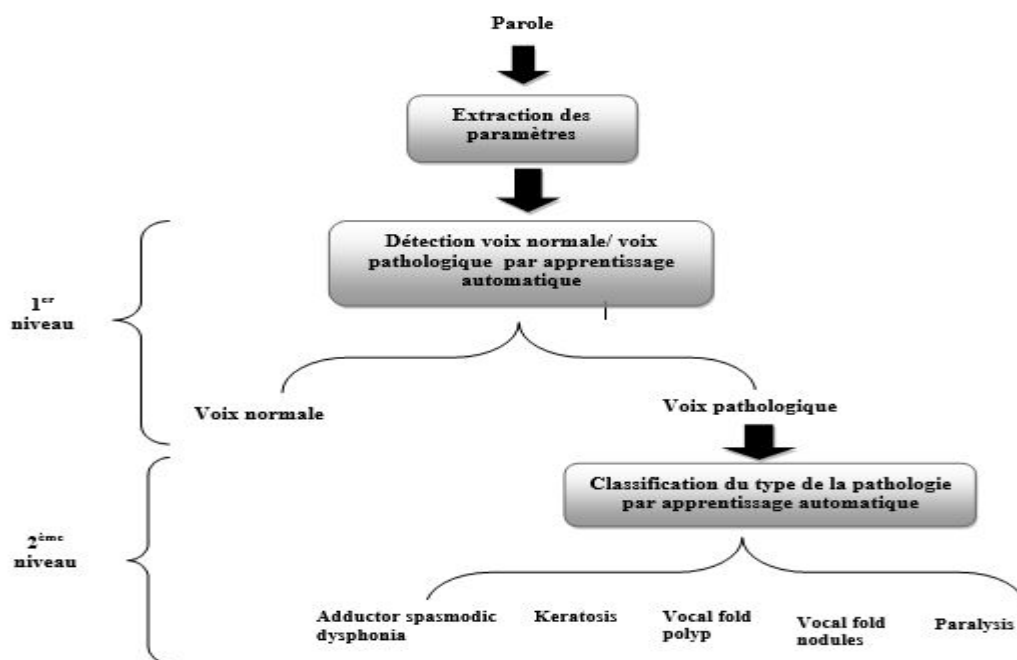


Figure III.10 : Système proposé pour l’identification des troubles de la voix.

Les expériences, tests et résultats du premier niveau ont été présentés et interprétés dans la section précédente. Cette partie sera consacrée seulement au deuxième niveau réservé à la classification des pathologies par les différentes variantes des classificateurs SVM et K-NN.

III.6.2.1 Résultats de la classification par la validation simple

Méthode	Modèle	Efficacité (%)	Spécificité (%)	Sensibilité (%)	AUC
K-NN	K-NN Fine	93,10	100	100	1
	K-NN Moyen	95,10	100	100	1
	K-NN Grossier	95,40	100	99	1
	K-NN Cosinus	95	100	100	1
	K-NN Cubique	95,30	100	100	1
	K-NN Pondéré	93,10	100	100	1
SVM	Linéaire	76,10	97	82	0,94
	Quadratique	95,50	100	100	1
	Cubique	94,60	100	100	1
	Gaussien Fine	94,40	100	100	1
	Gaussien Moyen	95,70	100	100	1
	Gaussien Grossier	78,70	96	83	0,94

Tableau III.11: Mesures des performances pour la classification des pathologies par validation simple 20% (K-NN et SVM).

Dans le tableau (III.11), nous remarquons que tous les modèles K-NN offrent des performances parfaites (1 en terme d'AUC et 100 % en termes de spécificité et sensibilité), le modèle K-NN grossier a une sensibilité de 99 % et assure la meilleure efficacité 95,40 %.

Pour la méthode SVM, les meilleures performances en termes de spécificité, sensibilité et AUC sont obtenues par les modèles quadratique, cubique, gaussien fine et gaussien moyen. En termes d'efficacité, la valeur la plus élevée est obtenue par le modèle gaussien moyen (95,70 %).

III.6.2.2 Résultats de la classification par la validation croisée

➤ Kv=5

Méthode	Modèle	Efficacité (%)	Spécificité (%)	Sensibilité (%)	AUC
K-NN	K-NN Fine	92,70	100	100	1
	K-NN Moyen	94,70	100	100	1
	K-NN Grossier	95,20	100	99	1
	K-NN Cosinus	94,70	100	100	1
	K-NN Cubique	94,80	100	100	1
	K-NN Pondéré	92,70	100	100	1
SVM	Linéaire	75,80	97	80	0,95
	Quadratique	95	100	100	1
	Cubique	94,20	100	100	1
	Gaussien Fine	93,50	100	100	1
	Gaussien Moyen	95,50	100	100	1
	Gaussien Grossier	78,30	96	80	0,95

Tableau III.12: Mesures des performances pour la classification des pathologies par validation croisée kv=5 (K-NN et SVM).

Comme nous pouvons le voir, presque tous les modèles sont des classificateurs parfaits, ils offrent une spécificité et une sensibilité de 100 %, et une AUC de 1.

En termes d'efficacité, la valeur maximale est 95,20 % pour la méthode K-NN obtenue par le modèle K-NN grossier, et 95,50 % pour la méthode SVM obtenue par le modèle gaussien moyen SVM.

Une autre constatation est que les deux modèles SVM linéaire et SVM gaussien grossier sont moins performants avec une efficacité de 75,80 % et 78,30 % respectivement.

➤ **Kv=10**

Méthode	Modèle	Efficacité (%)	Spécificité (%)	Sensibilité (%)	AUC
K-NN	K-NN Fine	92.10	100	100	1
	K-NN Moyen	94.70	100	100	1
	K-NN Grossier	94.90	100	100	1
	K-NN Cosinus	94.60	100	100	1
	K-NN Pondéré	92.10	100	100	1
	K-NN Cubique	94.70	100	100	1
SVM	Linéaire	76	98	80	0.95
	Quadratique	94.60	100	100	1
	Cubique	93.70	100	100	1
	Gaussien Fine	93	100	100	1
	Gaussien Moyen	95.30	100	100	1
	Gaussien Grossier	79.40	96	84	0.95

Tableau III.13 : Mesures des performances pour la classification des pathologies par SVM et K-NN avec une validation croisée kv=10.

Dans le tableau (III.13), la méthode K-NN offre des résultats acceptables avec une valeur maximale d'efficacité (94.90%) atteinte avec le modèle K-NN grossier. En ce qui concerne la sensibilité, la spécificité et l'AUC, toutes les variantes du K-NN assurent d'excellentes performances.

Une autre constatation, est que le classificateur SVM avec un noyau gaussien fine donne la meilleure efficacité (95.30%) et une AUC égale à 1, une sensibilité et spécificité de 100% chacune.

Les plus faibles performances en terme d'efficacité, spécificité, sensibilité et AUC sont obtenues par les deux variantes SVM avec des noyaux linéaire et gaussien grossier.

En scrutant les tableaux (III.11), (III.12) et (III.13), nous observons que les résultats de la classification pour les deux variantes K-NN et SVM avec une validation simple sont un petit peu meilleurs que ceux obtenus par une validation croisée soit avec kv=5 soit kv=10.

III.6.3 Résultats de la classification par la méthode Un-contre-Tous

La méthode de la classification multi-classes par SVM est basée sur deux approches : Un-Contre-Un (One-Versus-One) et Un-Contre-Tous (One-Versus-All). Les résultats de la classification par SVM obtenus et présentés dans les tableaux précédents sont ceux de l'approche par défaut dans l'application « Classification Learner » de Matlab, qui est l'approche Un-Contre-Un de la classification multi-classes.

Les résultats de la classification multi-classes basés sur l'approche Un-Contre-Tous sont illustrés dans le tableau ci-dessous, avec une validation simple, validation croisée (kv=5) et (kv=10).

Validation utilisée	Modèle	Efficacité (%)	Spécificité (%)	Sensibilité (%)	AUC
Validation simple 20 %	Linéaire	62,10	92	66	0,91
	Quadratique	94,80	100	100	1
	Cubique	93,60	100	100	1
	Gaussien fine	93,50	100	100	1
	Gaussien Moyen	95,10	100	100	1
	Gaussien Grossier	75,70	98	76	0,97
Validation croisée Kv=5	Linéaire	62,50	93	68	0,91
	Quadratique	95,20	100	100	1
	Cubique	94,30	100	100	1
	Gaussien fine	93,80	100	100	1
	Gaussien Moyen	95,40	100	100	1
	Gaussien Grossier	74,30	97	78	0,98
Validation croisée Kv=10	Linéaire	62,60	93	68	0,91
	Quadratique	95	100	100	1
	Cubique	93,60	100	100	1
	Gaussien fine	93,10	100	100	1
	Gaussien Moyen	95,10	100	100	1
	Gaussien Grossier	75,10	97	79	0,98

Tableau III.14: Mesures des performances pour la classification des pathologies par la méthode multi-classes Un-contre-Tous.

Nous remarquons que la valeur maximale de l'efficacité qui est autour de 95% est obtenue par le modèle gaussien moyen pour les trois types de validation, et la plus faible par le modèle SVM avec un noyau linéaire. En ce qui concerne la spécificité, la sensibilité et l'AUC, quel que soit le type de validation, les modèles quadratique, cubique, gaussien fine et gaussien moyen assurent des performances parfaites.

En comparant les résultats des deux approches de la classification multi-classes : les résultats de la SVM dans les tableaux (III.11), (III.12) et (III.13) pour l'approche Un-Contre-Un versus les résultats du tableau (III.14), l'approche Un-Contre-Un est favorisée.

III.7 Conclusion

Les simulations, tests et résultats présentés dans ce chapitre confirment l'efficacité des techniques d'apprentissage automatique par rapport aux méthodes classiques d'apprentissage, plus précisément les algorithmes de classification (SVM et K-NN) pour la détection et la classification des voix pathologiques où nous avons étudié l'influence des divers paramètres et choix sur les mesures de performance.

Chapitre IV

Optimisation des hyperparamètres

IV.1 Introduction

Les algorithmes d'apprentissage automatique ont été largement utilisés dans diverses applications et domaines. Pour adapter un modèle d'apprentissage automatique à différents problèmes, ses hyperparamètres doivent être réglés.

La sélection de la meilleure configuration d'hyperparamètres pour les modèles d'apprentissage automatique a un impact direct sur les performances du modèle. Cela nécessite souvent une connaissance approfondie des algorithmes d'apprentissage automatique et des techniques d'optimisation d'hyperparamètres appropriées.

Dans ce chapitre, nous allons présenter quelques techniques d'optimisations et tester leur impact sur nos meilleurs modèles entraînés dans le chapitre précédent.

IV.2 Définitions

Les modèles d'apprentissage automatique sont composés de deux types de paramètres différents :

- **Hyperparamètres** : sont tous les paramètres qui peuvent être définis arbitrairement par l'utilisateur avant de commencer l'apprentissage (par exemple la fonction noyau du SVM, valeur de K dans le K-NN).
- **Paramètres du modèle** : sont plutôt appris pendant l'apprentissage du modèle (par exemple les vecteurs de support du SVM).

Une définition des notions suivantes est nécessaire pour mieux comprendre la technique d'optimisation :

- **L'optimisation des hyperparamètres** : choisir un ensemble d'hyperparamètres optimaux pour un algorithme d'apprentissage donné.
- **Score** : la valeur de l'hyperparamètre que nous voulons régler (minimiser ou maximiser).
- **Fonction objective** : qui prend un hyperparamètre et génère un score.

IV.3 Techniques d'optimisation des hyperparamètres

Le processus de recherche des hyperparamètres les plus optimaux en apprentissage automatique est appelé optimisation des hyperparamètres. Les algorithmes courants incluent : optimisation bayésienne, recherche de grille, recherche aléatoire, etc.

IV.3.1 Optimisation par recherche de grille

La recherche de grille est l'algorithme le plus simple pour l'optimisation des hyperparamètres. La méthode consiste à diviser le domaine des hyperparamètres en une grille et ensuite essayer toutes les combinaisons de valeurs de cette grille, en calculant certaines métriques de performance à l'aide de la validation croisée [31].

Le point de la grille qui maximise la valeur moyenne en validation croisée, est la combinaison optimale de valeurs pour les hyperparamètres. La figure suivante illustre un exemple de cette technique.

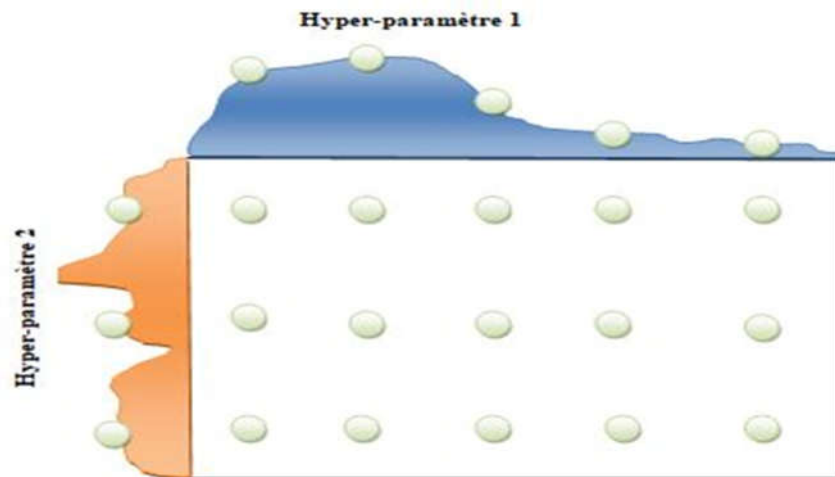


Figure IV.1: Exemple de recherche de grille.

IV.3.2 Optimisation par recherche aléatoire

La recherche aléatoire est similaire à la recherche par grille, mais au lieu d'utiliser tous les points de la grille, elle ne teste qu'un sous-ensemble sélectionné au hasard de ces points. Plus ce sous-ensemble est petit, plus l'optimisation est rapide mais moins précise. Plus cet ensemble de données est grand, plus l'optimisation est précise mais plus proche d'une recherche par grille. La figure (IV.2) illustre un exemple de cette technique.

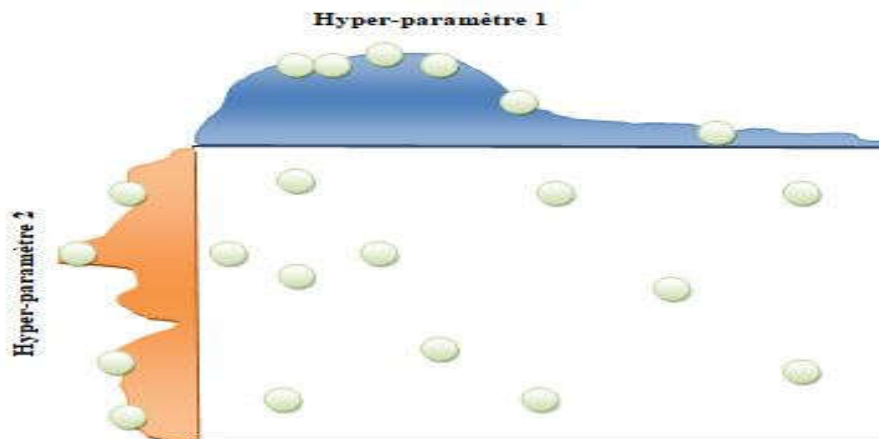


Figure IV.2: Exemple de recherche aléatoire.

IV.3.3 Optimisation Bayésienne

L'optimisation Bayésienne est l'une des méthodes efficaces qui ont été largement utilisées pour le réglage des hyperparamètres dans le domaine de l'apprentissage automatique, son concept sous-jacent est très simple.

Contrairement à la recherche aléatoire ou la recherche de grille, l'optimisation bayésienne garde une trace des résultats d'évaluation passés qu'elle utilise pour construire un modèle probabiliste de substitution, qui est plus facile à optimiser que la fonction objective, mappant les hyperparamètres à une probabilité d'un score sur la fonction objective :

$$P(\text{score}/\text{hyperparameters})$$

Le principe de cette approche est le suivant :

1. Construire un modèle de probabilité de substitution de la fonction objective.
2. Trouver les hyperparamètres qui fonctionnent le mieux sur le substitut.
3. Appliquer ces hyperparamètres à la vraie fonction objective.
4. Mettre à jour le modèle de substitution intégrant les nouveaux résultats.
5. Répéter les étapes 2 à 4 jusqu'à ce que le nombre maximal d'itérations ou de temps soit atteint.

L'idée de base de l'optimisation bayésienne est de passer un peu plus de temps à sélectionner les prochains hyperparamètres qui est sans conséquences afin de faire moins d'appels à la fonction objective.

De cette façon, elle peut faire un choix éclairé des prochains hyperparamètres. À un niveau élevé, le modèle de substitution se rapproche de la fonction objective.

IV.4 Résultats d'optimisation des hyperparamètres

Pour l'optimisation des hyperparamètres des modèles étudiés dans les chapitres précédents, nous avons sélectionnés les deux modèles de l'application « Classification Learner » K-NN optimisé et SVM optimisé dans le cas d'une validation croisée avec $k_v=10$.

IV.4.1 K-NN Optimisé

Commençons par le modèle K-NN optimisé, nous allons optimiser les quatre hyperparamètres qui sont le nombre de voisins, la distance métrique, le poids de distance et la standardisation des données avec les trois méthodes d'optimisation (recherche aléatoire, recherche de grille et optimisation bayésienne).

➤ Par recherche aléatoire

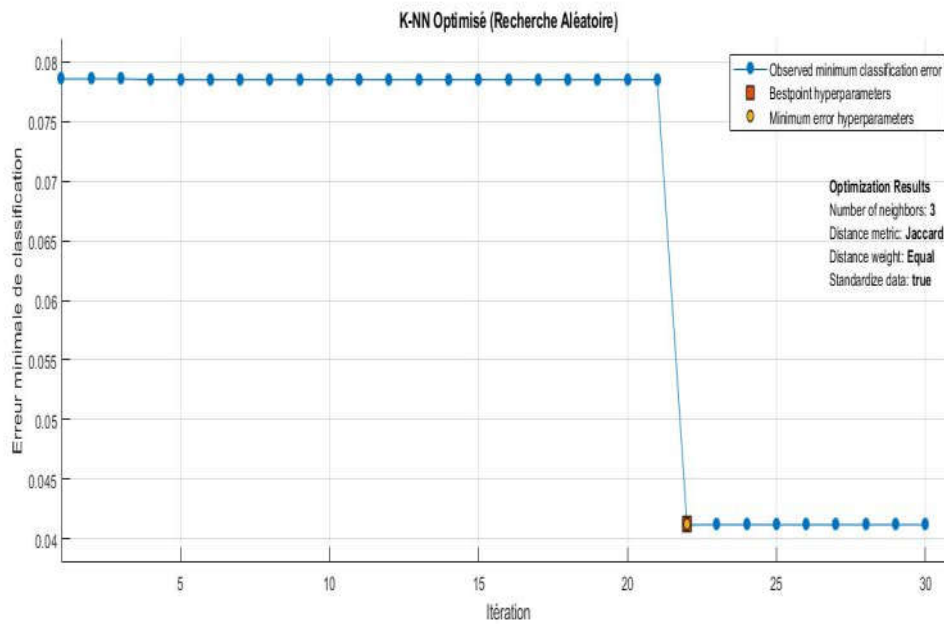


Figure IV.3 : Modèle K-NN optimisé par recherche aléatoire.

La figure ci-dessus représente la courbe d'erreur minimale de classification de la méthode K-NN en fonction d'itérations.

Chaque point bleu correspond à l'erreur de classification minimale calculée jusqu'à cette itération par le processus d'optimisation.

Le point jaune indique l'itération qui correspond aux hyperparamètres qui produisent l'erreur minimale de classification (0.041183 à l'itération 22).

Le carré rouge indique l'itération qui correspond aux hyperparamètres optimisés (l'itération 22) et les valeurs de ces hyperparamètres qui sont affichées en haut droit sous la légende de la courbe.

- **Résultats d'optimisation**

Nombre de voisins :	3
Distance métrique :	Jaccard
Poids de distance :	égal
Standardisation des données :	vrai
Temps d'apprentissage :	23,31 min

➤ Par recherche de grille

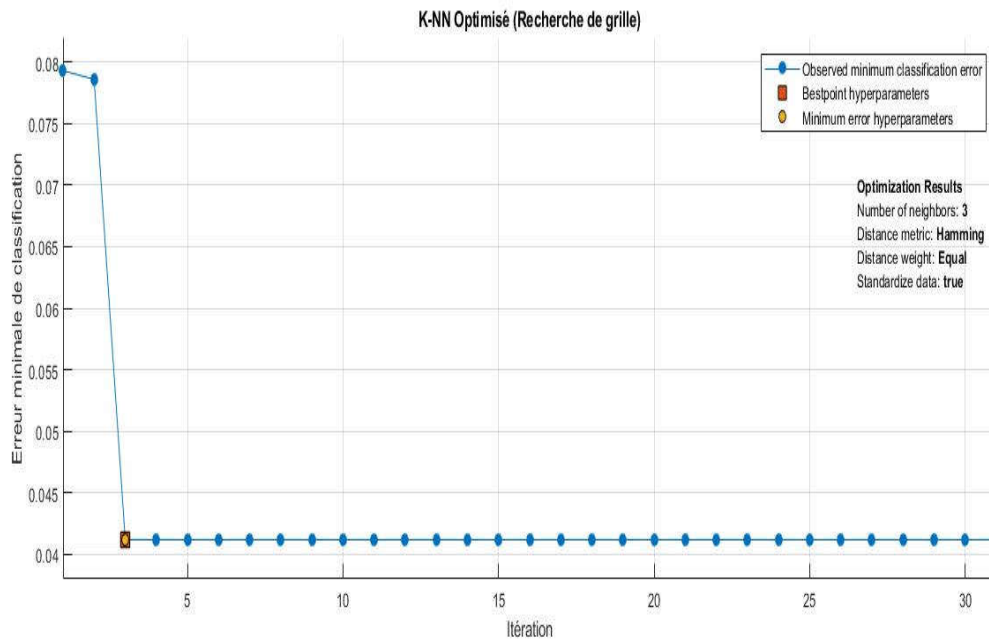


Figure IV.4 : Modèle K-NN optimisé par recherche de grille.

La figure ci-dessus représente la courbe d'erreur minimale de classification de la méthode K-NN en fonction d'itérations.

Chaque point bleu correspond à l'erreur de classification minimale calculée jusqu'à cette itération par le processus d'optimisation.

Le point jaune indique l'itération qui correspond aux hyperparamètres qui produisent l'erreur minimale de classification (0,041183 à l'itération 3).

Le carré rouge indique l'itération qui correspond aux hyperparamètres optimisés (l'itération 3) et les valeurs de ces hyperparamètres qui sont affichées en haut droit sous la légende de la courbe.

• Résultats d'optimisation

Nombre de voisins :	3
Distance métrique :	Hamming
Poids de distance :	égal
Standardisation des données :	vrai
Temps d'apprentissage :	26 min

➤ Par optimisation Bayésienne

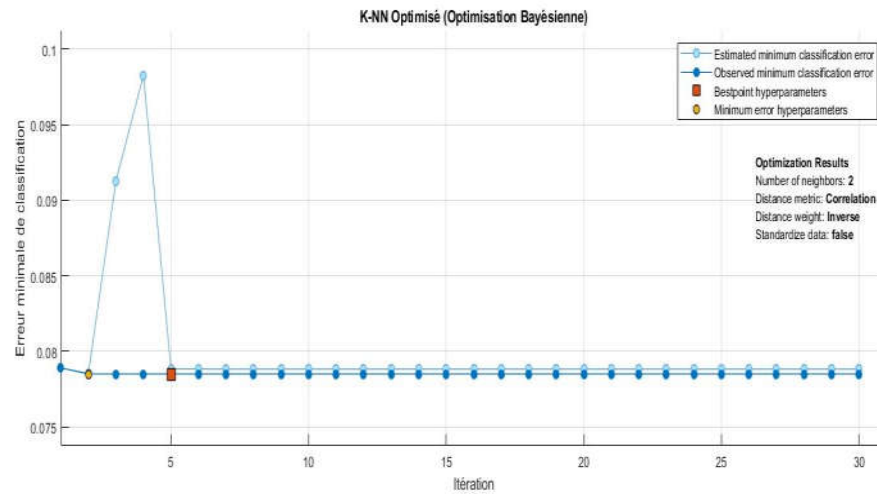


Figure IV.5 : Modèle K-NN optimisé par Optimisation Bayésienne.

La figure ci-dessus représente la courbe d'erreur minimale de classification de la méthode K-NN en fonction d'itérations.

Chaque point bleu-clair correspond à l'erreur de classification minimale estimée par le processus d'optimisation lors de l'examen des valeurs de l'hyperparamètre à cette itération.

Chaque point bleu-foncé correspond à l'erreur de classification minimale calculée jusqu'à cette itération par le processus d'optimisation.

Le point jaune indique l'itération qui correspond aux hyperparamètres qui produisent l'erreur minimale de classification (0,078504 à l'itération 2).

Le carré rouge indique l'itération qui correspond aux hyperparamètres optimisés (l'itération 5) et les valeurs de ces hyperparamètres qui sont affichées en haut droit sous la légende de la courbe.

• Résultats d'optimisation

Nombre de voisins :	2
Distance métrique :	Corrélation
Poids de distance :	inverse
Standardisation des données :	faux
Temps d'apprentissage :	11,71 min

En comparant le temps d'apprentissage et la valeur de l'erreur minimale de classification des trois techniques d'optimisation, nous remarquons que l'optimisation par recherche aléatoire est la meilleure méthode d'optimisation des hyperparamètres du K-NN.

IV.4.2 SVM Optimisé

Pour le modèle SVM Optimisé, nous allons optimiser les trois hyperparamètres qui sont la méthode multi-classes, l'échelle du noyau et la fonction noyau avec les trois méthodes d'optimisation.

➤ Par recherche aléatoire

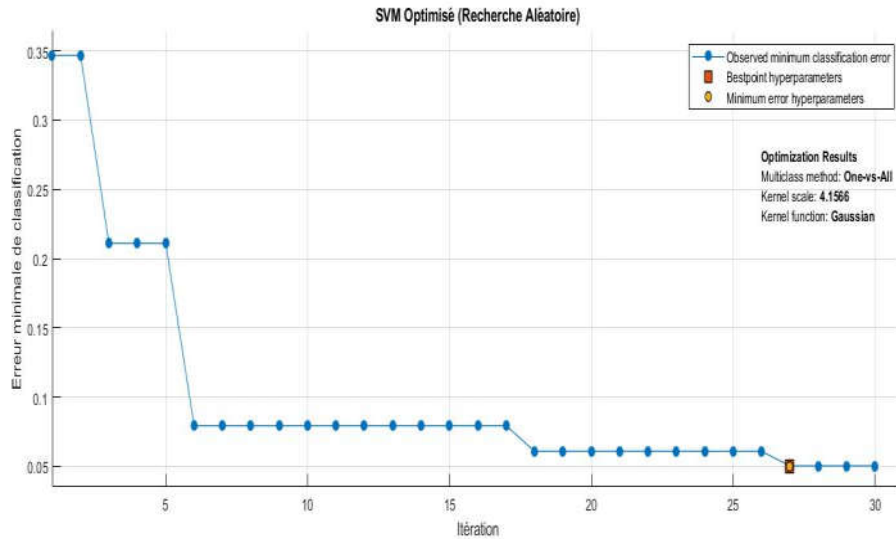


Figure IV.6 : Modèle SVM optimisé par recherche aléatoire.

La figure ci-dessus représente la courbe d'erreur minimale de classification de la méthode SVM en fonction d'itérations.

Chaque point bleu correspond à l'erreur de classification minimale calculée jusqu'à cette itération par le processus d'optimisation.

Le point jaune indique l'itération qui correspond aux hyperparamètres qui produisent l'erreur minimale de classification (0,050236 à l'itération 27).

Le carré rouge indique l'itération qui correspond aux hyperparamètres optimisés (l'itération 27) et les valeurs de ces hyperparamètres qui sont affichées en haut droit sous la légende de la courbe.

• Résultats d'optimisation

Méthode multi-classes : Un-contre-Tous
 Echelle du noyau : 4,1566
 Fonction noyau : Gaussien
 Temps d'apprentissage : 9,98 h

➤ Par recherche de grille

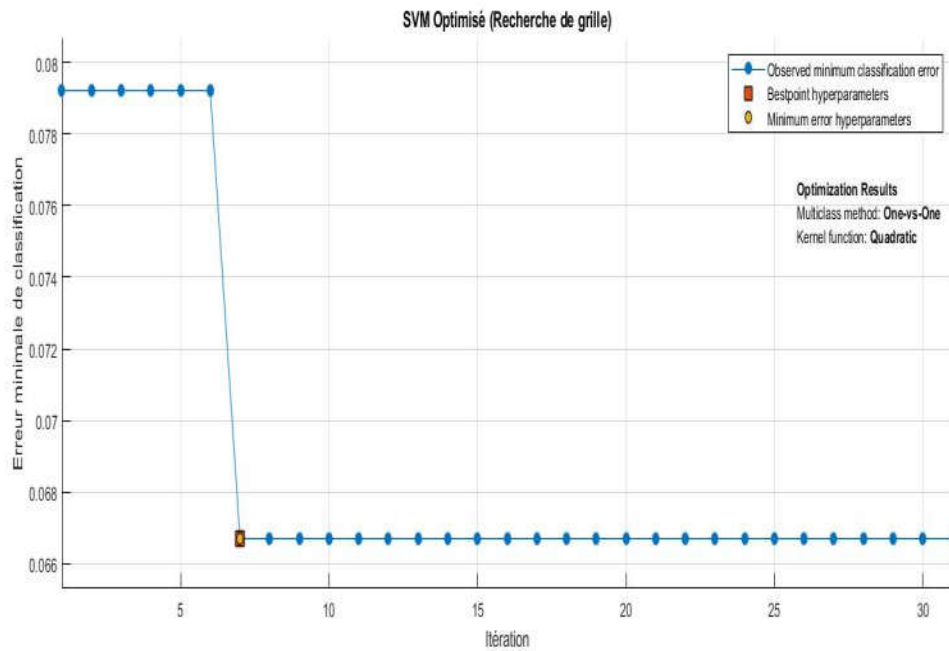


Figure IV.7 : Modèle SVM optimisé par recherche de grille.

La figure ci-dessus représente la courbe d'erreur minimale de classification de la méthode SVM en fonction d'itérations.

Chaque point bleu correspond à l'erreur de classification minimale calculée jusqu'à cette itération par le processus d'optimisation.

Le point jaune indique l'itération qui correspond aux hyperparamètres qui produisent l'erreur minimale de classification (0,066692 à l'itération 7)

Le carré rouge indique l'itération qui correspond aux hyperparamètres optimisés (l'itération 7) et les valeurs de ces hyperparamètres qui sont affichées en haut droit sous la légende de la courbe.

- **Résultats d'optimisation**

Méthode multi-classes : Un-contre-Un
 Echelle du noyau : 1
 Fonction noyau : Quadratique
 Temps d'apprentissage : 5,23 h

➤ Par optimisation Bayésienne

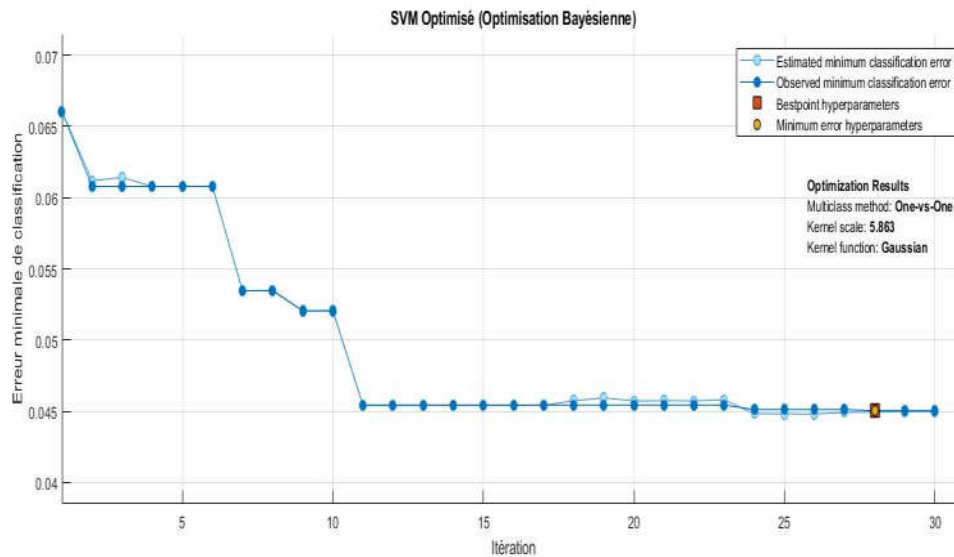


Figure IV.8 : Modèle SVM optimisé par Optimisation Bayésienne.

La figure ci-dessus représente la courbe d'erreur minimale de classification de la méthode SVM en fonction d'itérations.

Chaque point bleu-clair correspond à l'erreur de classification minimale estimée par le processus d'optimisation lors de l'examen des valeurs de l'hyperparamètre à cette itération.

Chaque point bleu-foncé correspond à l'erreur de classification minimale calculée jusqu'à cette itération par le processus d'optimisation.

Le point jaune indique l'itération qui correspond aux hyperparamètres qui produisent l'erreur minimale de classification (0,045042 à l'itération 28).

Le carré rouge indique l'itération qui correspond aux hyperparamètres optimisés (l'itération 28) et les valeurs de ces hyperparamètres qui sont affichées en haut droit sous la légende de la courbe.

• Résultats d'optimisation

Méthode multi-classes : Un-contre-Un
 Echelle du noyau : 5,863
 Fonction noyau : Gaussien
 Temps d'apprentissage : 2,70 h

En comparant le temps d'apprentissage et la valeur de l'erreur minimale de classification des trois techniques d'optimisation, nous observons que l'optimisation bayésienne est la meilleure méthode d'optimisation des hyperparamètres du SVM.

Dans ce qui suit, nous allons évaluer les meilleurs résultats de l'optimisation en les comparant avec ceux des deux meilleurs classificateurs des méthodes K-NN et SVM dans le cas de la classification du type de pathologie avec une validation croisée $k=10$ présentés dans le tableau (III.13).

Commençons par le réglage des hyperparamètres des modèles K-NN grossier et SVM gaussien moyen.

Dans l'onglet type du modèle, on clique sur « Advanced » et on choisit « Tune » (figure IV.9).

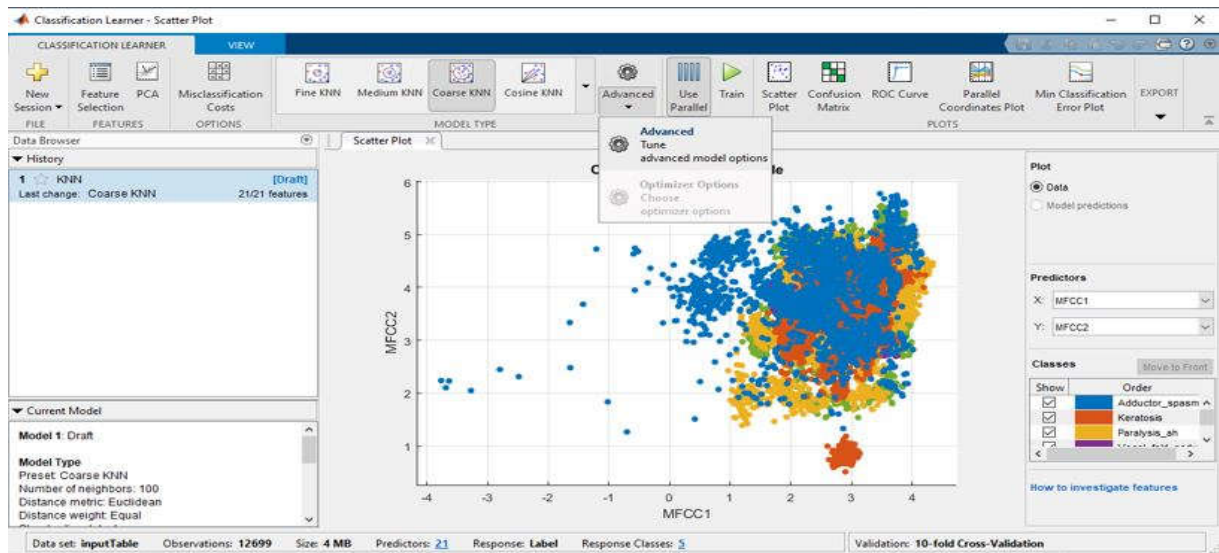


Figure IV.9 : Interface indiquant l'option du réglage des paramètres.

Pour le K-NN grossier, on utilise les résultats obtenus par la méthode d'optimisation par recherche aléatoire, le réglage est illustré dans la figure suivante :



Figure IV.10 : Réglage des hyperparamètres pour le classificateur K-NN grossier.

Pour le SVM gaussien moyen, on utilise les résultats de l'optimisation bayésienne comme l'illustre la figure (IV.11).

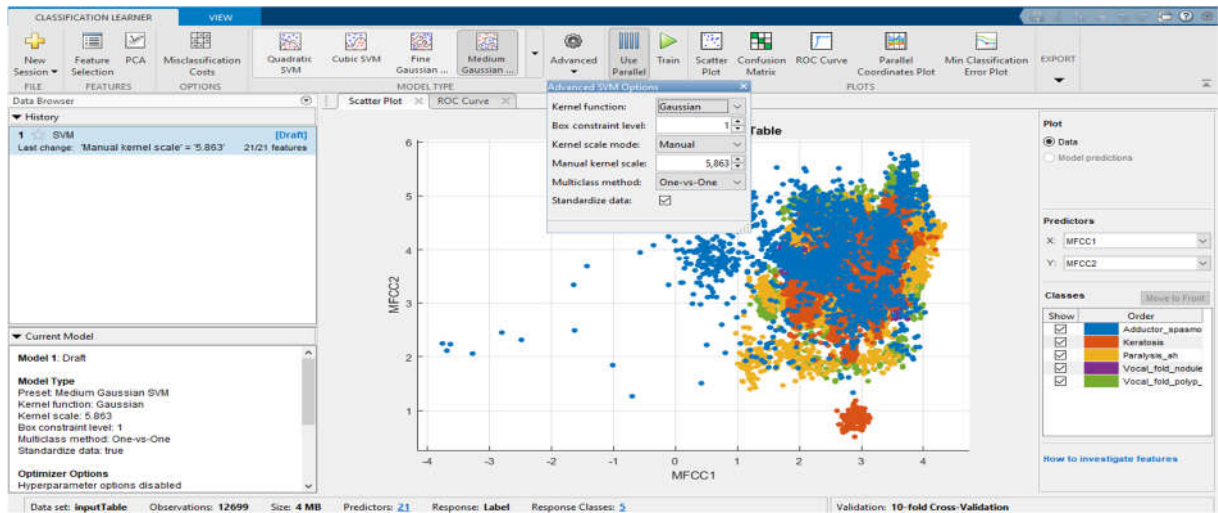


Figure IV.11 : Réglage des hyperparamètres pour le classificateur SVM gaussien moyen.

Les résultats du réglage des hyperparamètres des deux classificateurs K-NN grossier et SVM gaussien moyen sont représentés dans le tableau suivant où la valeur la plus élevée est indiquée en gras dans chaque colonne :

Modèle		Efficacité (%)	Spécificité (%)	Sensibilité (%)	AUC
K-NN grossier	Avant réglage	94,9	100	100	1
	Après réglage	95,9	100	100	1
SVM gaussien moyen	Avant réglage	95,3	100	100	1
	Après réglage	95,5	100	100	1

Tableau IV.1 : Comparaison entre les performances du K-NN grossier et du SVM gaussien moyen avant et après le réglage de leurs hyperparamètres.

Nous observons que le réglage des hyperparamètres en utilisant les hyperparamètres optimaux obtenus par le processus d’optimisation apporte une augmentation des performances pour les deux classificateurs K-NN grossier et SVM gaussien moyen plus précisément dans leurs efficacités.

IV.5 Conclusion

Nous avons présenté dans ce chapitre la technique d’optimisation des hyperparamètres de la méthode SVM et K-NN par recherche aléatoire, recherche de grille et l’optimisation bayésienne qui est considérée, à partir des résultats de l’application « Classification Learner », la méthode la plus rapide parmi les trois.

Nous avons montré que le temps d’apprentissage du K-NN optimisé est plus faible par rapport à celui du SVM optimisé.

L’optimisation est effective pour l’obtention des bonnes performances en une seule étape au lieu de vérifier tous les modèles en changeant les hyperparamètres manuellement.

Conclusion générale

Conclusion générale

L'identification des troubles de la voix à l'aide des techniques d'apprentissage automatique était l'objectif de ce modeste travail. Deux méthodes de classifications ont été proposées pour étudier leurs efficacités dans la détection et la classification de certaines pathologies de la voix : SVM et K-NN.

Le système proposé peut être utilisé par les orthophonistes pour faire une évaluation objective de la voix du patient à partir des mesures acoustiques et aérodynamiques : MFCC, Jitter, Shimmer et HNR.

A partir des résultats obtenus, nous avons remarqué que les techniques d'apprentissage automatique SVM et K-NN, évaluées sur la base de données MEEI qui est utilisée dans la plupart des études de recherche connexes, peuvent détecter efficacement les voix pathologiques et normales et également classer le type de la pathologie.

Nous avons montré que les méthodes de classification SVM et K-NN disponibles dans l'application « Classification Learner » assurent des bonnes performances pour la détection avec une efficacité de 100 % en utilisant la validation simple et la validation croisée ($k_v=5$ et $k_v=10$).

Les résultats obtenus prouvent que les paramètres choisis MFCC, Jitter, Shimmer et le HNR sont efficaces pour la discrimination des pathologies de la voix, lorsqu'ils sont évalués individuellement ou combinés.

Pour la classification, les meilleures performances de la méthode SVM sont obtenues par le classificateur avec un noyau gaussien dans le cas d'une validation simple : une efficacité = 95,7 % pour la méthode multi-classes Un-contre-Un, et 95,1 % pour la méthode Un-contre-Tous, une spécificité = 100 %, une sensibilité = 100 et un AUC =1 pour les deux méthodes.

Pour ce qui est des résultats de la méthode K-NN, le classificateur K-NN grossier assure les meilleures performances : une efficacité = 95,4 %, une spécificité = 100 % et un AUC =1, dans le cas d'une validation simple. Pour la validation croisée : une efficacité = 95,2 % (pour $k_v=5$) et 94,9 % (pour $k_v=10$), une spécificité = 100 %, une sensibilité = 99 % et AUC = 1 pour les deux valeurs de k_v .

Nous avons appliqué les trois techniques d'optimisation des hyperparamètres disponibles dans l'application « Classification Learner » sur les deux classificateurs : SVM et K-NN. Les résultats peuvent être utilisés pour adapter ces deux classificateurs à d'autres problèmes d'apprentissage automatique.

Il reste de nombreux aspects pouvant faire l'objet d'une étude plus poussée. En premier lieu, nous proposons d'étudier ce système sur d'autres bases de données. Plusieurs autres points pourraient être améliorés, par exemple, son exploitation sur une carte graphique ou un système d'évaluation en ligne. Il serait également envisageable de démarrer ces études par les résultats de l'optimisation des hyperparamètres.

Liste des références

Liste des références

- [1] Claire Rémy, *L'intelligence artificielle*, DUNOD, Paris, France, 158p, 1994.
- [2] A. Cornuéjols, L. Miclet et Y. Kodratoff, *Apprentissage artificiel : Concepts et algorithmes*, EYROLLES, Paris, France, 591p, 2003.
- [3] Laura Verde et al., "Voice disorder identification by using machine learning techniques," *IEEE Access*, vol. 6, pp. 16246-16255, 2018.
- [4] L. Zeriouel et M. Benhabiles, "Détection et classification des pathologies de la voix par SVM," Mémoire de Master, Systèmes des Télécommunications, Département d'Electronique, Université Mohammed Seddik BENYAHIA-Jijel, Algérie, 2020.
- [5] Nedjem Eddine Ayat, "Sélection automatique de modèle dans les machines à vecteurs de support : Application à la reconnaissance d'images de chiffres manuscrits," Thèse de Doctorat en Génie Electrique, Université du Québec-Montréal, Canada, 2004.
- [6] Malak Al Mojaly et al. , "Detection and classification of voice pathology using feature selection," in *The 11th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA)*, Doha, Qatar, 2014, pp. 571-577.
- [7] Souad Khellat Kihel, "Les séparateurs à vaste marge Bi-classes," Mémoire de Master, Département d'Informatique, Université des Sciences et de la Technologie d'Oran, Algérie, 2012.
- [8] Nadia Marref, "Apprentissage incrémental et machines à vecteurs supports," Mémoire de Magistère, Informatique industrielle, Département d'informatique, Université HADJ LAKHDAR, Batna, Algérie, Décembre 2013.
- [9] <https://www.baeldung.com/cs/svm-multiclass-classification> (Consulté 02/05/2021)
- [10] Ali Labiad, "Sélection des mots clés basée sur la classification et l'extraction des règles d'association," Mémoire de Maitrise, Mathématique et Informatique appliquée, Université du Québec, Montréal, Canada, 2017.
- [11] Amel Haliche, "Classification supervisée à base de KNN avec pondération d'attributs par l'algorithme génétique," Mémoire de Magistère, Systèmes Intelligents et Ingénierie des Logiciels, Département d'Informatique, Université des Sciences et de la Technologie Houari Boumediene, Alger, Algérie, 2015.

- [12] A. Mekhbi et Ch. Menouar, “*Application des techniques de data mining et de machine learning pour la détection d’intrusions,*” Mémoire de Master, Informatique Légale et Multimédia, Département d’Informatique, Université Mohammed Seddik BENYAHIA-Jijel, Algérie, 2018.
- [13] Léo Varnet, “*Identification des indices acoustiques utilisés lors de la compréhension de la parole dégradée,*” Thèse de Doctorat, Neurosciences, Université Claude Bernard Lyon 1, Lyon, France, 2015.
- [14] Raoudha Mabrouki, “*Authentification biométrique par le biais du signal parole et de l’iris,*” Mémoire de Master Professionnel, Sciences et Technologies, Sécurité des Systèmes Informatiques Communicants et Embarqués, Département d’Informatique, Institut Supérieur d’Informatique, Tunisie, 2015.
- [15] Zuzana Dankovičová et al. , “Machine Learning Approach to dysphonia detection,” *Applied Sciences*, vol. 8, n°10, c.1927, 2018.
- [16] <http://www.fo-rothschild.fr/soins/orl/informations-medicales/troubles-voix.html>
(Consulté 14/05/2021)
- [17] <https://www.revmed.ch/RMS/2016/RMS-N-533/Lesions-benignes-courantes-des-cordes-voicales-criteres-diagnostiques-et-prise-en-charge> (Consulté 14/05/2021)
- [18] Corinne Fredouille, “*Traitement automatique et troubles de la voix et de la parole : Champs d’application, contraintes et limites,*” Mémoire d’HDR, Informatique et Langage, Université d’Avignon et du Pays de Vaucluse, France, 2016.
- [19] Tchinta Gunawan, “*Histoire de rééducation : entre mythe et réalité. Suivi longitudinal de dix patients adultes dysphoniques,*” Mémoire du Certificat de Capacité d’Orthophonie, Institut d’Orthophonie Gabriel Decroix, Université de Lille2, France, 2010.
- [20] Charles Pelletier, “*Classification des sons respiratoires en vue d’une détection automatique des sibilantes,*” Mémoire de Maîtrise, Ingénierie, Université du Québec, Montréal, Canada, 2006.
- [21] Jorge Martinez, Hector Perez and Enrique Escamilla, “Speaker recognition using Mel Frequency Cepstral Coefficients (MFCC) and Vector Quantization (VQ) techniques,” in *22nd International Conference on Electronics Communications and Computing*, United States February 2012, pp. 248-251.
- [22] <https://ww2.mathworks.cn/help/audio/ug/speaker-identification-using-pitch-and-mfcc.html> (Consulté 23/05/2021)

- [23] Clémentine Mertz, “*Étude des corrélations entre les évaluations subjectives et objectives de la voix chez des patients présentant une paralysie laryngée bilatérale,*” Mémoire du Certificat de Capacité d’Orthophonie, Département d’Orthophonie, Université de Rouen Normandie, France, 2018.
- [24] Alain Ghio, “Bilan instrumental de la dysphonie,” Rééducation orthophonique, Ortho édition, pp.9-29, 2013.
- [25] www.praat.org
- [26] [CHAPITRE 3:L’EVALUATION SUBJECTIVE ET OBJECTIVE DE LA VOIX \(institut-numerique.org\)](http://institut-numerique.org) (Consulté 23/05/2021)
- [27] <http://www.stimmdatenbank.coli.uni-saarland.de/> (Consulté 25/05/2021)
- [28] Kay Elemetrics Corp, Disordered Voice Database Model 4337, Verl. 03, Massachusetts Eye and Ear Infirmary Voice and Speech Lab, 2002.
- [29] <https://machinelearningmastery.com/k-fold-cross-validation/> (Consulté 26/05/2021)
- [30] K. Hajian-Tilaki, “Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation,” *Caspian J. Internal Med.*, vol. 4, no. 2, pp. 627-635, 2013.
- [31] <https://www.yourdatateacher.com/2021/05/19/hyperparameter-tuning-grid-search-and-random-search/> (Consulté 12/06/2021)

Abstract:

In this thesis, we focus on machine learning techniques used for voice disorder identification. We present the two supervised machine learning methods: Support Vector Machine (SVM) and K-Nearest Neighbors (K-NN) to detect normal and pathological voice from 150 files in the MEEI database and classify five different voice pathologies. We use MFCC, Jitter and its variants, Shimmer and its variants and HNR as parameters for voice discrimination. We study these two classifiers by MATLAB Classification Learner application using simple validation and cross validation ($kv = 5$ and $kv = 10$), and changing the multi-class method for SVM. We conclude this work by optimizing the hyper-parameters of both methods by applying the three techniques available in the Classification Learner application (grid search, random search and Bayesian optimization).

Keywords: Machine learning, Classification; SVM; K-NN; Voice pathologies; MFCC; Jitter; Shimmer; HNR; Learner Classification; Optimization.

Résumé :

Dans ce mémoire, nous nous intéressons aux techniques d'apprentissage automatique utilisées pour l'identification des troubles de la voix. Nous présentons les deux méthodes d'apprentissage automatique supervisé Machine à Vecteurs de Support (SVM) et K-Plus Proches Voisins (K-NN) pour détecter la voix normale et pathologique de 150 fichiers de la base de données MEEI et classer cinq différentes pathologies vocales. Nous utilisons les MFCC, Jitter et ses variantes, Shimmer et ses variantes et HNR comme paramètres pour la discrimination de la voix. Nous étudions ces deux classificateurs par l'application MATLAB Classification Learner en utilisant la validation simple et la validation croisée ($kv = 5$ et $kv = 10$), et en changeant la méthode multi-classes pour le SVM. Nous terminons ce travail par l'optimisation des hyper-paramètres des deux méthodes en appliquant les trois techniques disponibles dans l'application Classification Learner (recherche de grille, recherche aléatoire et optimisation Bayésienne).

Mots clés : Apprentissage automatique, Classification ; SVM ; K-NN ; Pathologies vocales ; MFCC ; Jitter ; Shimmer ; HNR ; Classification Learner ; Optimisation.

ملخص:

في هذه الأطروحة، نحن مهتمون بتقنيات التعلم الآلي المستخدمة لتحديد اضطرابات الصوت. نقدم طريقتين من طرق التعلم الآلي الخاضع للإشراف آلات المتجهات الداعمة (SVM) وخوارزمية K أقرب الجيران للكشف عن الصوت الطبيعي والمريض لـ 150 ملفاً من قاعدة البيانات MEEI وتصنيف خمسة أمراض مختلفة. نقوم باستعمال المعلمات MFCC، Jitter، ومتغيراته، Shimmer ومتغيراته و HNR كمعلمات لتمييز الأصوات عن بعضها. ندرس هذين المصنفين بواسطة التطبيق MATLAB Classification Learner باستخدام التحقق البسيط والتحقق المتقاطع ($kv=5$ و $kv=10$)، وتغيير طريقة الفئات المتعددة بالنسبة لـ SVM. ننهي هذا العمل بتحسين المعلمات الفائقة لكلتا الطريقتين من خلال تطبيق التقنيات الثلاثة المتاحة في التطبيق Classification Learner (البحث الشبكي، البحث العشوائي وتحسين بايس).
كلمات مفتاحية: التعلم الآلي، التصنيف، آلات المتجهات الداعمة، خوارزمية أقرب الجيران، الأمراض الصوتية، المعلمات MFCC، Jitter، Shimmer، HNR، تصنيف المتعلم، التحسين.