

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

École Nationale Polytechnique d'Alger



Département d'Electrotechnique

Mémoire du projet de fin d'études pour l'obtention du diplôme
D'Ingénieur d'Etat en Electrotechnique

Thème

**Classification de la sévérité de pollution des isolateurs
haute tension par les méthodes Support Vecteur Machine et
l'Analyse Linéaire Discriminante**

Réalisé par :
Mr. ZEBIRI Karim
Mr. ZOUAGHI Ayyoub

Proposé et encadré par :
Pr A.MEKHALDI
M. MOULA

Promotion Juin 2013

Remerciements

*Nous tenons à remercier très vivement Monsieur le Professeur de l'Ecole Nationale Polytechnique **Abelouahab MEKHALDI** pour son aide et ses précieuses qualités humaines.*

*Nous tenons à remercier Monsieur **Belkacem MOULA** pour ses conseils, ses nombreux encouragements durant ce travail et l'aide qu'il nous a apporté durant l'élaboration de ce modeste travail.*

*Nous sommes très honorés que Monsieur le Professeur de l'Ecole National Polytechnique, **Madjid TEGUAR** à accepter la charge de juger ce mémoire. Son compétence en matière de Haute tension et réseaux électriques font que sa présence pour nous une véritable caution scientifique. Nous le remercions chaleureusement pour son dévouement.*

*Nous tenons à exprimer notre reconnaissance à Monsieur le Professeur de l'Ecole Nationale Polytechnique, **Ahmed BOUBAKEUR** pour la confiance et l'honneur qu'il nous accorde en acceptant de présider ce jury.*

Aux enseignants/chercheurs de L'Ecole National Polytechnique surtout ceux de la spécialité Electrotechnique pour les nombreux échanges scientifiques et les différents conseils et formations qui ont largement contribué à l'aboutissement des travaux présentés dans ce mémoire.

Dédicace :

Je dédie ce travail à mes parents, tout simplement les meilleurs parents du monde. Pour toutes les peines endurées, tous les sacrifices qui ont fait, pour faire de moi ce que je suis aujourd'hui.

Comme je le dédie à :

Mes frères : Zahir, Kamel et ma petite et aimable sœur Amel.

Chahinez qui a été toujours présente et qui m'a soutenu et encourager tout au long de mon cursus universitaire.

A Ayyoub bionomi et Khaled CHAOU que je remercie pour leurs aide durant l'élaboration de ce travail.

Tous mes amis et surtout: Brahim, Hamza, Fateh, Chokri, Yacine, Chouaib et Assia.

Toute la promotion de l'année 2013 de l'ENP, et tous ceux qui connaissent Karim de près ou de loin.

Karim

Je dédie ce modeste travail à :

Mes chers parents

Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être

Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours

Je prie Dieu puissant de vous protéger et vous accorder santé, bonheur et longue vie

Allah ykhalikoum liya

Ayyoub

SOMMAIRE

Introduction générale	1
------------------------------------	---

Chapitre I : Travaux antérieurs sur les SVM

Introduction	4
I.1. Évaluation de l'état de pollution de l'isolateur en utilisant le courant de fuite	4
I.2. Prévion de faillite en utilisant les supports vecteurs machine (SVM) avec le choix optimal des paramètres de fonction noyau.....	8
I.2.1. Collecte et prétraitement de données	8
I.2.2. Support vecteur machine	8
I.3. Machine à Vecteurs Supports Multi-Noyau pour la détection de points caractéristiques du visage	11
I.3.1. Méthodologie	11
I.3.2. Extraction des caractéristiques	12
I.3.3. Apprentissage Multi-Noyau	12
I.3.4. Bootstrap et Patches négatifs.....	12
I.3.5. Bases d'apprentissage	13
I.3.6. Evaluation de la procédure de bootstrap	14
I.3.7. Résultats expérimentaux et Comparaisons.....	15
I.4. Analyse de l'état des isolateurs des lignes aériennes de distribution en utilisant une combinaison machine à vecteurs de support (SVM) et la transformée en ondelettes	16
I.5. Optimisation des isolateurs supports utilisés dans les systèmes haute tension (HT) en utilisant la technique support vecteur machine.....	19
I.6. Classification basée sur la technique SVM pour la commande de Myoelectric appliquée à un organe supérieur	22
I.6.1. L'acquisition des données	22

I.6.2. Segmentation des données	22
I.6.3. Sélection des caractéristiques.....	23
I.6.4. Prétraitement	24
I.6.5. Classification.....	24
I.7. Discrimination de la source de décharge partielle en utilisant la méthode SVM.....	25
I.7.1. Acquisition et traitement des données.....	28
Conclusion	30

Chapitre II : Théorie des supports vecteurs machine (SVM)

Introduction au SVM	31
II.1. Types de classification	31
II.1.1. Classification supervisée	31
II.1.2. Classification non supervisée	32
II.2. Théorie de l'apprentissage	32
II.3. Les machines à vecteurs de support	33
II.4. Théorie de Vapnik-Chervonenkis	34
II.5. Principe de la technique SVM.....	38
II.6. Les SVM binaires.....	40
II.6.1. Notion de base	40
II.6.1.1. Hyperplan	40
II.6.1.2. Vecteur de support	40
II.6.1.3. Marge	41
II.6.2. Construction de l'hyperplan optimal	41
II.6.2.1. Cas des données linéairement séparables	41
II.6.2.2. Cas des données non-linéairement séparables	44

II.6.3. L’astuce du noyau (kernel trick)	46
II.6.3.1. Exemples de noyaux.....	47
II.7. SVM pour le cas multi-classes	48
II.7.1. Méthodes de décomposition	49
II.7.1.1. La stratégie un contre tous	49
II.7.1.2. La stratégie un contre un	51
II.8. Avantages et inconvénients des SVM	53
Conclusion	54

Chapitre III : Application des SVM pour la classification de la sévérité de pollution

Introduction	55
III.1. Rappel sur la transformée de Fourier	55
III.2. La transformée en ondelettes	56
III.2.1. L’ondelette d’analyse	57
III.2.2. La transformée en ondelettes continues (CWT)	58
III.2.2.2. Scalogramme	58
III.2.2.3. Calcul des coefficients de la CWT.....	59
III.2.3. L’analyse multi-résolution (AMR)	60
III.2.4. La transformée en ondelettes discrètes (DWT)	61
III.2.4.1. Niveau maximal de décomposition.....	61
III.2.4.2. Choix de l’ondelette analysante.....	61
III.3. Décomposition du signal du courant de fuite par DWT	62
III.4. Calcul des caractéristiques des coefficients de détails d’ondelette	64
III.4.1. Etude de la standard déviation des détails du signal du courant de fuite pour une pollution avec 100 g/l de Kaolin	64

III.4.2. Etude des maximums des détails du signal du courant de fuite pour une pollution avec 100 g/l de Kaolin.....	66
III.4.3. Etude des minimums des détails du signal du courant de fuite pour une pollution avec 100 g/l de Kaolin.....	68
III.4.4. Etude des moyennes des détails du signal du courant de fuite pour une pollution avec 100 g/l de Kaolin	70
III.5. Les Classes de pollution.....	72
III.6. Application des SVM pour la classification de sévérité de pollution des isolateurs	74
III.6.1. Composition du vecteur d'état.....	74
III.6.2. Entraînement du modèle	74
III.6.2.1. Base de données pour le cas de la pollution uniforme	74
III.6.2.2. Base de données pour le cas de la pollution non uniforme.....	74
III.6.3. Les différentes fonctions noyaux appliquées	76
III.7. Validation et résultat de classification	76
III.7.1. Cas de pollution uniforme	76
III.7.2. Cas de pollution non uniforme	78
III.8. Interprétation des résultats	80
III.8.1. Pollution uniforme.....	80
III.8.2. Pollution non uniforme	81
Conclusion	81

Chapitre IV : Application de la LDA pour la classification de la sévérité de pollution

Introduction.....	83
IV.1. Analyse discriminante	83
IV.2. La règle de décision de Bayes	84

IV.3. Les principales méthodes d'Analyse Discriminante.....	85
IV.3.1. Analyse Discriminante Quadratique (QDA)	85
IV.3.2. Analyse Discriminante Linéaire (LDA)	85
IV.3.3. Analyse Factorielle Discriminante (FDA)	86
IV.4. La classification avec l'analyse discriminante linéaire	88
IV.5. Application de LDA pour la classification de la sévérité de pollution des isolateurs	89
IV.5.1. Composition de la matrice d'apprentissage	89
IV.5.2. Validation et résultat de classification	89
IV.6. Comparaison entre SVM et LDA	91
Conclusion	92

Chapitre V : Caractérisation de l'état de surface des isolateurs par la technique des ondelettes

Introduction	93
V.1. Pollution uniforme	93
V.1.1 Influence de la conductivité volumique	93
V.1.2. Influence de la concentration de kaolin	98
Conclusion	102
Conclusion générale	103

Références bibliographiques

ملخص :

في العمل الحالي، يتم إجراء تصنيف شدة تلوث عوازل التوتر العالي بناءً على قواعد و نصوص IEC 71، ويكون ذلك باستغلال قواعد بيانات جمعت في مختبر التوتر العالي في المدرسة الوطنية المتعددة التقنيات على نماذج مسطحة. أولاً، سنركز على الأعمال الأخيرة التي تهتم بالتصنيف والتنبؤ باستغلال طريقة SVM في العديد من المجالات. ثم نطبق DWT على إشارات التيار المتسرب لاستخراج خصائص التفاصيل الخاصة بهم من أجل أداء تصنيف لشدة التلوث باستعمال طريقة SVM مع اثنين من الاستراتيجيات و باستخدام وظائف نواة مختلفة. ثانياً، سنقوم بتصنيف بواسطة LDA لمقارنة نتائجها مع نتائج التصنيف بطريقة SVM. وأخيراً، نقوم بتحليل إشارات التوتر المطبق باستعمال CWT لتقدير حالة سطح النموذج.

الكلمات الرئيسية:

شدة تلوث، تلوث منتظم، تلوث غير منتظم، التيار المتسرب، التنبؤ المتحكم فيه، التصنيف، تقنية الموجات المستمرة، تقنية الموجات المتقطعة، تحليل التمايز الخطي.

Résumé :

Dans le présent travail, la classification de la sévérité de pollution des isolateurs de haute tension est effectuée, suivant les recommandations de la CEI 71, en exploitant des bases de données établies au niveau du laboratoire de haute tension de l'Ecole Nationale Polytechnique sur des modèles plans. Dans un premier lieu, nous nous intéressons aux plus récents travaux qui traitent la classification et la prévision par la méthode SVM dans plusieurs domaines. Ensuite, nous appliquons la DWT sur les signaux du courant de fuite pour extraire les caractéristiques de leurs détails afin d'effectuer la classification de la sévérité de pollution par la méthode SVM avec deux stratégies en utilisant différentes fonctions noyaux. En deuxième lieu, nous faisons la classification par la LDA afin de la comparer avec SVM. Finalement, nous analysons les signaux de la tension appliquée par la CWT pour estimer l'état de surface du modèle.

Mots clés :

Sévérité de pollution, pollution uniforme, pollution non uniforme, courant de fuite, apprentissage supervisé, classification, support vecteur machine, transformée en ondelettes discrète, transformée en ondelettes continue, analyse discriminante linéaire.

Abstract:

In this work, the classification of the pollution severity of polluted insulators is carried out, according to IEC 60-1 recommendations, by using databases set at high voltage laboratory of Ecole Nationale Polytechnique on plan models. Firstly, we interest to the most recent work which treats classification and the forecast using SVM method in several fields. Then, we apply the DWT of the leakage current signals to extract the characteristics from their details in order to classify the severity of pollution using SVM method with two strategies and various kernel functions. After that, we perform the same classification using the LDA in order to compare it with SVM. Finally, we analyze the signals of the voltage applied using the CWT to estimate the surface quality of the model.

Key words:

Pollution severity, uniform pollution, not uniform pollution, leakage current, supervised training, classification, Support Vector Machine, Discrete Wavelet Transform, Continuous Wavelet Transform, Discriminant Linear Analysis.

Introduction

générale



Les lignes aériennes et les postes des réseaux de transport d'énergie électrique sont exposés à diverses contraintes. Parmi celles-ci, la pollution des isolateurs qui constitue l'un des facteurs de première importance dans la qualité et la fiabilité du transport d'énergie. En effet par temps de pluie ou de brouillard, les dépôts polluants se fixant sur les surfaces isolantes réduisent considérablement la résistivité superficielle et le contournement peut alors survenir [1].

Le phénomène de pollution des isolateurs constitue un sujet d'étude d'une importance primordiale. Cela provient du fait que la pollution des isolateurs constitue un sérieux problème pour l'exploitation des réseaux de haute tension. En effet, et suite à l'interaction entre l'air transportant des grains de poussières et l'isolateur, une couche étrangère se dépose sur les surfaces de celui-ci [2]. Une fois cette couche humidifiée, la tension de tenue peut diminuer considérablement, provoquant des incidents très graves.

Le contournement des isolateurs sous haute tension et en présence d'impuretés conductrices est lié à l'apparition des décharges électriques qui progressent à leurs surfaces conductrices jusqu'au court-circuit. Il peut être caractérisé par deux phases successives [3]:

- Initiation d'une décharge électrique instable à un certain seuil de tension à cause d'un assèchement local de la surface conductrice.
- Progression de la décharge électrique avec l'augmentation du courant de fuite sous des conditions de pollution favorables, pour former un arc de contournement stable qui modifie partiellement la conductivité du milieu.

Le choix des isolateurs à installer dans le réseau ne peut se faire que si la sévérité de pollution des différents sites concernés est connue. En effet, la mesure de cette sévérité est indispensable afin de dimensionner convenablement les isolateurs susceptibles d'assurer un service sans défaillance dans un site donné [4].

La mesure de la sévérité est généralement basée sur la détermination expérimentale de la conductance superficielle de la couche polluante ou la détermination au laboratoire du courant de fuite d'un isolateur pollué artificiellement [5].

L'apprentissage supervisé a émergé ces dernières années comme un outil efficace pour la classification dans de nombreux domaines [6]. Les Machines à Vecteurs de Supports (SVM) sont une de ces techniques de classification, elles ont été appliquées à de très nombreux domaines (bioinformatique, recherche d'information, vision par ordinateur,

finance etc...) [7], et récemment pour le diagnostic des isolateurs en haute tension.

Le présent travail consiste à utiliser le signal du courant de fuite pour extraire des critères de jugement sur l'état de surface de l'isolateur pollué et servir d'entrées pour un classifieur SVM dont les sorties représentent la classe de pollution assignée à un isolateur pollué donné.

Dans le premier chapitre, nous présentons une synthèse bibliographique sur les principaux travaux qui traitent la classification et la prévision en utilisant les SVM. Nous avons présenté des applications des SVM dans différents domaines (économie, médecine, traitement des images et dans le domaine de la haute tension).

Dans le deuxième chapitre, nous présentons dans un premier temps les types de classification, la théorie d'apprentissage ainsi que la théorie de la méthode des Machines à Vecteurs de Supports qui représente un type d'apprentissage supervisé. Ensuite, nous nous intéressons à la classification basée sur les SVM dans le cas binaire ou on a deux classes et les différentes stratégies de classification dans le cas multi-classe. En outre, nous parlons aussi sur les fonctions noyaux (kernel function) afin de les exploiter dans notre modèle SVM par la suite.

Dans le troisième chapitre, dans une première partie nous donnons des rappels sur la transformée en ondelettes, plus particulièrement la décomposition du signal du courant de fuite en utilisant la transformée en ondelette discrète (DWT). Nous présentons par la suite les différentes classes associées à chaque niveau de pollution selon la CEI 60-01.

Dans une deuxième partie, nous présentons les résultats de classification obtenus par les modèles SVM construits et nous comparons les résultats obtenus par les différentes stratégies et avec les différentes fonctions noyaux pour obtenir le meilleur classifieur dans les deux cas de pollution uniforme et non uniforme.

Le quatrième chapitre présente la théorie de l'analyse discriminante plus particulièrement l'analyse discriminante linéaire (LDA). Nous utilisons cette dernière pour faire une classification à partir des caractéristiques des signaux du courant de fuite pour les deux cas de pollution uniforme et non uniforme. Ensuite, nous faisons une comparaison entre cette méthode et la méthode SVM.

Dans le cinquième chapitre, nous mettons en application l'analyse par la technique des ondelettes. Pour la pollution uniforme, la transformée en ondelette continue (CWT) est appliquée sur les signaux du courant de fuite pour un éventuel diagnostic de l'état de surface.

En dernier lieu, nous terminons par une conclusion générale qui représente une synthèse globale des principales observations au cours de l'application des différentes méthodes, ainsi que les perspectives qui permettront de réaliser une étude plus approfondie.

Chapitre I :

**Travaux antérieurs
sur les SVM**



Introduction :

Dans ce chapitre, nous présentons les plus récents travaux concernant la classification par la méthode Support Vecteur Machine (SVM). Ces travaux touchent différents domaines incluant la Haute Tension.

I.1. Évaluation de l'état de pollution de l'isolateur en utilisant le courant de fuite :

X.Jiang et al [9] ont proposé une méthode qui utilise la transformée en S et les supports vecteur machine (SVM) pour l'évaluation de l'état de sûreté des isolateurs pollués. Cette méthode est basée sur les essais de pollution artificielle sous différentes humidités relatives (RH) et différentes densités de dépôt de sel (SDD).

En premier lieu, l'approche utilise la transformée en S pour extraire l'amplitude I_h , la phase θ et la déformation harmonique totale THD du courant de fuite. Puis I_h , θ et THD sont utilisés comme paramètres caractéristiques à mettre dans le modèle SVM à construire. Par la suite, les paramètres de sortie de SVM sont classés dans quatre sous-ensembles pour l'évaluation de la sûreté des isolateurs pollués et pour finir, la performance du modèle SVM formé a été validée par 8 paires d'échantillons d'essai.

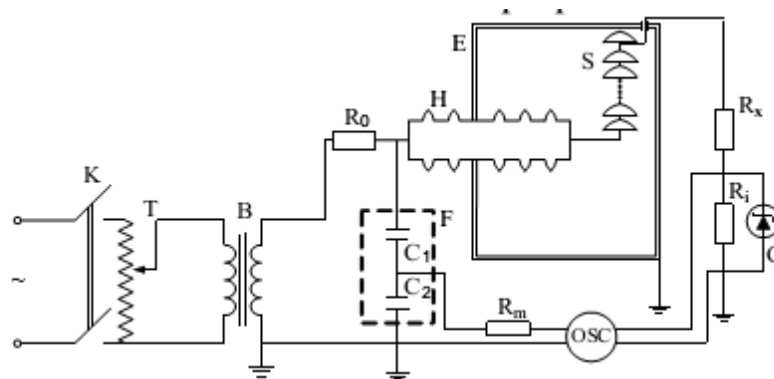


Figure I.1 : Circuit d'essais.

Les résultats d'essais prouvent que le courant de fuite devient inductif quand les arcs locaux forts se produisent, θ diminue avec l'augmentation de l'humidité relative RH (Figure I.4), et I_h n'a aucune relation avec le SDD quand le RH est faible, mais augmente avec l'augmentation du SDD quand le RH est élevé (Figure I.2). Le THD du courant de fuite augmente lentement avec l'augmentation du RH quand la pollution est légère. Quand la pollution est moyenne ou lourde, THD commence d'abord par diminuer puis augmente avec

L'augmentation de RH (Figure I.3). Le THD est minimum au RH environ de 80%.

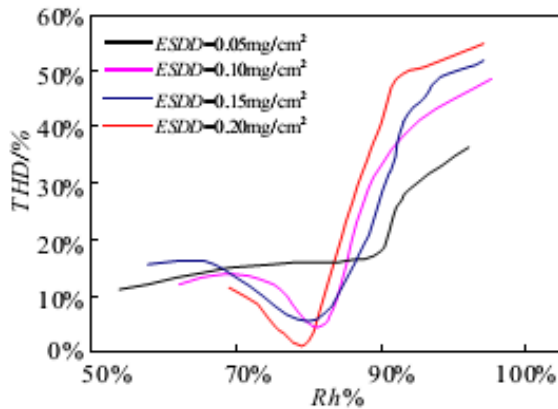


Figure I.3 : Rapport entre THD et RH.

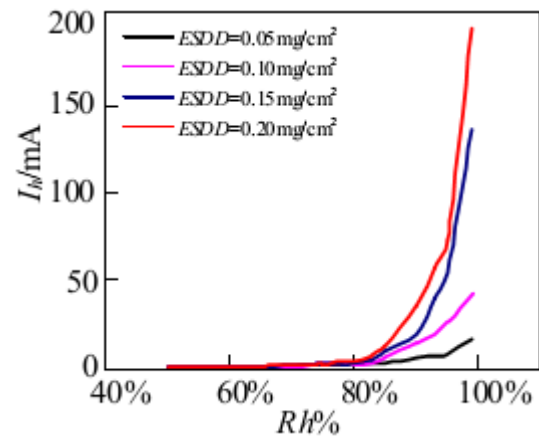


Figure I.2 : I_h sous différents SDD et RH.

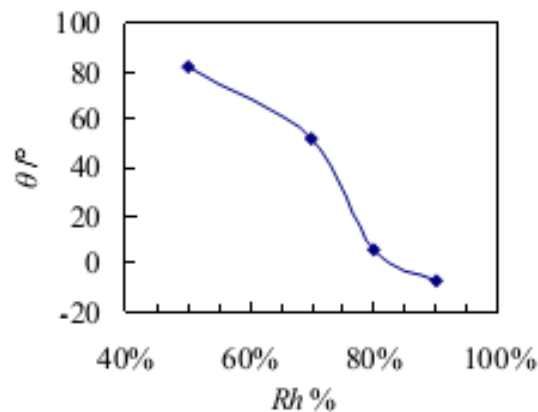


Figure I.4 : Relation entre le θ et RH.

Afin de développer un modèle SVM pour l'évaluation de la sûreté des isolateurs pollués, les chercheurs ont utilisé 68 paires d'entrée-sortie de données disponibles du travail expérimental. Parmi ces données, 60 paires étaient aléatoirement assignées pendant la formation du modèle, les 8 paires restantes étaient utilisées pour l'essai et la validation du modèle SVM formé.

Les caractéristiques de I_h , θ et THD sont calculées correctement à partir de la transformée en S et sont utilisés comme paramètres d'entrées pour le modèle SVM à construire. La sortie du SVM va être donnée par le degré d'appartenance à un état de sûreté d'un sous-ensemble donné A, B, C et D.

D'après des recherches sur les caractéristiques des formes d'onde du courant de fuite et

des différents comportements de la décharge sur une chaîne d'isolateurs de 5 éléments artificiellement pollués, les chercheurs ont constaté qu'il y avait sept phénomènes de décharges typiques avant la rupture totale de la chaîne [10, 11], suivant les indications du tableau I.1.

Stage	phénomène typique de décharge.
1	Aucune décharge évidente
2	étincelle faible (move)
3	décharge move sous forme de brosses
4	décharge locale courte
5	petit arc dense
6	décharge principale lumineuse
7	décharge principale rouge intense

Tableau I.1 : Phénomènes de décharge typiques.

Selon l'analyse du phénomène de décharge électrique des isolateurs, quatre sous-ensembles sont définis (quatre états de sûreté) qui sont: A= [sûr], B= [alarme légère], C= [alarme modérée] et D= [alarme sérieuse]. Où A ne se rapporte à aucune décharge ou seulement à une faible décharge sur la surface des isolateurs, dans ce cas, les isolateurs peuvent fonctionner sans risque, ainsi l'état de sûreté est sûr. B se rapporte à la partie fragmentaire du petit arc électrique sur la surface des isolateurs, l'exploitation sûre de l'isolateur ne serait pas accordée, ainsi l'état de sûreté correspond à une alarme légère. C se rapporte à l'arc court et continu sur la surface des isolateurs, dans cette condition, si la pollution s'aggrave, la rupture peut se produire, les systèmes d'alimentation doivent être examinés immédiatement, ainsi l'état de sûreté correspond à une alarme modérée. D se rapporte à l'arc principal intense sur la surface de l'isolateur, il s'agit d'un état critique, et dans ce cas le dispositif protecteur devrait agir instantanément, ainsi l'état de sûreté correspond à une alarme sérieuse.

La transformation d'une classification à une autre est un processus progressif, d'où l'utilisation de la fonction gaussienne qui a ce caractère progressif pour passer à 4 sous-ensembles. (Figure I.5)

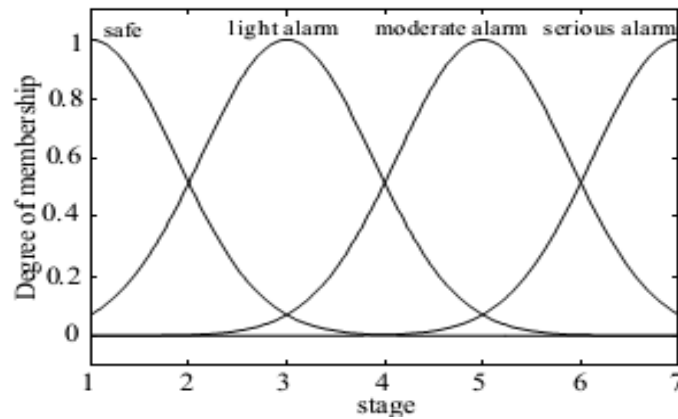


Figure I.5 : Fonctions d'appartenance des variables de sortie.

La performance du SVM a été vérifiée par l'essai sur 8 paires, l'état évalué de sûreté des isolateurs pollués a été comparé aux résultats d'essais et mentionné dans le Tableau I.2.

Test No.	1	2	3	4	5	6	7	8
$SDD/(mg/cm^2)$	0.05	0.05	0.10	0.10	0.15	0.15	0.20	0.20
$RH\%$	49%	97%	83%	99%	62%	98%	64%	99%
θ°	82°	-5.6°	4.3°	-8.9°	69°	-9.8°	68°	-13°
I_h/mA	0.41	0.17	11.31	35.12	0.38	121.05	0.42	190.21
$THD\%$	11.2%	30.3%	29.1%	48.8%	15.8%	42.9%	11.5%	56.7%
$A/B/C/D$	1.00/0.03/ 0.01/0.00	0.03/0.9/ 0.01/0.00	0.20/0.64/ 0.02/0.00	0.01/0.98/ 0.02/0.00	0.97/0.04/ 0.00/0.00	0.00/0.00/ 0.95/0.97	0.99/0.0/ 0.00/0.00	0.00/0.00/ 0.06/1.00
Verified results	Accord	Accord	Accord	Accord	Accord	Accord	Accord	Accord

Note: Group of string of insulator of 7 unit($t=26\pm 2$ °C)

Tableau I.2 : Vérification des résultats évalués avec les essais sur les 8 échantillons.

En conclusion, l'amplitude I_h , la déformation harmonique totale THD et la phase θ du courant de fuite ont été déterminées comme des paramètres caractéristiques d'entrées pour le modèle SVM et comme sortie le modèle utilisé fournit quatre sous-ensembles A, B, C et D évaluant l'état de sûreté des isolateurs pollués.

L'amplitude du courant de fuite I_h varie selon le SDD et l'humidité relative RH sur la surface des isolateurs. Quand RH est inférieur à 75%, le SDD n'a aucun effet sur l'amplitude du courant de fuite. Si le RH est compris entre 75% et 85%, I_h augmente avec l'augmentation de RH. Par contre, si RH est supérieur à 85%, l'amplitude du courant de fuite augmente d'une façon très rapide avec l'augmentation de RH.

On peut conclure aussi que quand la pollution est faible, THD augmente lentement avec l'augmentation de RH, mais il diminue au début puis augmente par la suite.

I.2. Prédiction de faillite en utilisant les supports vecteurs machine (SVM) avec le choix optimal des paramètres de fonction noyau :

Les problèmes de prise de décision dans le domaine financier sont d'une importance capitale pour les institutions financières dues au niveau élevé du risque des fausses décisions. H.Jae et al [14] ont appliqué la méthode des SVM à ces problèmes de décision et plus particulièrement la prédiction de faillite afin de fournir un nouveau modèle améliorant son exactitude de prédiction. Pour atteindre cet objectif, ces chercheurs ont employé un réseau-recherche (grid-search) en utilisant 5 fois la contre-vérification (cross-validation) pour découvrir les valeurs des paramètres optimales de la fonction noyau du modèle SVM.

En outre, pour évaluer l'exactitude de prédiction des SVM, ils ont comparé sa performance à celle de l'analyse discriminante multiple (MDA), de l'analyse de régression logistique (Logit) et des réseaux neurologiques de propagation de retour à trois couches (BPN).

I.2.1. Collecte et prétraitement de données :

La base de données utilisée dans cette étude a été obtenue à partir de la plus grande organisation de garantie de crédit de la Corée. Tous les cas de non-faillite sont des sociétés moyennes et des industries lourdes en 2002. Les cas de faillite sont également de la même industrie sauf que ces chercheurs ont rassemblé des cas de faillite additionnels des années de 2000, 2001 et de 2002, car le nombre de cas de faillite est plus petit que celui des cas de non-faillite. Les données expérimentales qui se composent d'un chiffre pair de nombre de faillite et de non-faillite, où 1888 sociétés incluant 944 cas de faillite et 944 de non-faillite placés dans un ordre aléatoire.

Dans les cas de SVM, MDA et Logit, l'ensemble des données est divisé en deux sous-ensembles : un ensemble de formation de 80% (1510) et un ensemble de vérification (holdout) de 20% (378) de toutes les données (1888). Pour le cas de BPN, l'ensemble des données est divisé en trois sous-ensembles : un ensemble de formation de 60% (1132), un ensemble de vérification de 20% (378), et un ensemble de validation de 20% (378).

I.2.2. Support vecteur machine:

Pour les SVM, il y a deux paramètres liés au noyau RBF Gaussien : C La limite supérieure et γ le paramètre du noyau qui jouent un rôle crucial dans la performance du SVM.

Récemment, Hsu et al [15] ont suggéré des guides pratiques sur les SVM en utilisant le réseau-recherche (grid-search) et la contre-vérification (cross-validation) pour identifier le choix optimal de C et de γ de sorte que le classifieur puisse exactement prévoir des données inconnues.

Dans cette étude, les chercheurs ont employé cette dernière méthode avec 5 fois la contre-vérification. Toutes les paires de (C, γ) ont été essayées et celle qui ont la meilleure exactitude de contre-vérification sont choisies. Dans ce cas, le choix optimal trouvé correspond à $C=2^{11}$ et $\gamma=2^{-7}$.

C	γ									
	2^{-3}	2^1	2^{-1}	2^{-3}	2^{-5}	2^{-7}	2^{-9}	2^{-11}	2^{-13}	2^{-15}
2^{-5}	47.2185	47.2185	74.4371	71.9205	57.2848	47.2185	47.2185	47.2185	47.2185	47.2185
2^{-3}	47.2185	62.9801	79.1391	77.2848	74.4371	67.8808	47.2185	47.2185	47.2185	47.2185
2^{-1}	48.1457	74.3709	81.2583	79.6689	77.7483	75.3642	68.1457	47.2185	47.2185	47.2185
2^1	62.5828	77.5497	81.0596	81.9868	79.2715	78.1457	75.6954	68.0795	47.2185	47.2185
2^3	62.5828	77.6159	80.9272	82.9139	81.1921	78.8079	78.0795	75.5629	67.8808	47.2185
2^5	62.5828	77.5497	79.8675	82.9139	82.5828	80.7285	78.4106	77.8808	75.5629	67.7483
2^7	62.5828	77.5497	79.7351	82.4503	82.8477	81.9205	79.7351	78.0795	77.6159	75.5629
2^9	62.5828	77.5497	79.7351	79.9338	82.7152	82.6490	80.6623	79.6026	78.0795	77.5497
2^{11}	62.5828	77.5497	79.7351	79.3377	82.1192	83.1126	81.8543	80.1325	79.3377	78.0132
2^{13}	62.5828	77.5497	79.7351	79.3377	81.2583	82.7152	82.5828	80.4636	79.8013	79.4702
2^{15}	62.5828	77.5497	79.7351	79.2053	79.9338	82.3179	82.8477	81.3245	79.6026	79.8013

Tableau I.3: Les résultats du réseau-recherche utilisant 5 fois la contre-vérification.

Avec les paramètres $C=2^{11}$ et $\gamma=2^{-7}$, ils ont exécuté les modèles de SVM dans le cas de différents noyaux RBF Gaussien, polynômial, linéaire et sigmoïde. Les résultats sont montrés sur le Tableau I.4.

Fonction noyau	C	γ	d	Exactitude de prévision (%)	
				Données d'entraînement	Données de vérification
Linéaire	2^{11}	N/A	N/A	80.2649	77.2487
RBF	2^{11}	2^{-7}	N/A	88.0132	83.0688
Polynomial ^a	2^{11}	2^{-7}	1	80.3311	77.2487
			2	86.6225	83.8624
			3	88.4768	82.0106
Sigmoid ^a	2^{11}	2^{-7}	N/A	72.6490	71.1640

Tableau I.4: Performances des modèles SVM avec différentes fonctions noyaux.

Le noyau polynômial donne des résultats un peu plus performants que le noyau RBF Gaussien quand le degré $d=2$ pour l'ensemble des vérifications et quand $d=3$ pour les données de formation (tableau I.4).

Le noyau polynômial a besoin de plus de paramètres que le noyau RBF Gaussien et il peut diverger à l'infini ou à zéro quand le degré « d » est élevé c'est pour ça qu'ils ont employé la fonction de base radiale (RBF Gaussien) comme fonction noyau du SVM pour cette étude.

On ce qui concerne les autres méthodes : BPN, MDA et Logit, les résultats de BPN selon l'ajustement de paramètre sont récapitulés dans le tableau I.5, les performances de MDA et Logit sont récapitulés dans le tableau I.6.

Learning epoch	Hidden nodes	Prediction accuracy (%)	
		Training data	Holdout data
100	8	83.1272	80.6878
	12	82.9505	80.4233
	16	81.5371	80.1587
	24	81.0954	79.8942
	32	83.5689	81.2169
200	8	84.8057	80.9524
	12	86.7491	81.2169
	16	82.6855	81.2169
	24	85.1590	81.2169
	32	85.2474	81.7460
300	8	84.8940	82.0106
	12	86.1307	81.4815
	16	87.0141	80.6878
	24	85.2474	82.5397
	32	88.1625	81.4815

Model	Prediction accuracy (%)	
	Training data	Holdout data
MDA	78.8079	79.1391
Logit	79.8676	78.3069

Tableau I.5: Performance de BPN.

Tableau I.6: Performance de MDA et Logit.

L'étape finale compare les performances de prévision de SVM, de BPN, de MDA, et de Logit pour les données de l'ensemble de formation et l'ensemble de vérification, et prouve que SVM surpasse BPN, MDA et Logit par 0.5, 4.8, et 3.9% respectivement. (Tableau I.7)

	SVM	BNN	MDA	Logit
Training data	88.0132	85.2474	78.8079	79.8676
Holdout data	83.0688	82.5397	79.1391	78.3069

Tableau I.7: Comparaison de l'exactitude de prévision de SVM, de BPN, de MDA, et de Logit en %.

SVM applique le principe structural de minimisation de risque (SRM), qui s'est avéré

supérieur au principe empirique traditionnel de minimisation de risque (MTC) utilisé par les réseaux neurologiques conventionnels d'où la méthode SVM qui donne des résultats plus performants par rapport aux autres méthodes.

On peut conclure aussi que cette étude qui applique SVM pour la prévision de faillite a montré sa puissance de prévision comparée aux méthodes existantes avec de bons résultats.

Dans cette étude, les chercheurs ont fait le choix optimal du paramètre C de limite supérieure et γ le paramètre du noyau à 2^{11} et 2^{-7} respectivement, en utilisant un réseau-recherche avec 5 fois la contre-vérification.

I.3. Machine à Vecteurs Supports Multi-Noyau pour la détection de points caractéristiques du visage :

V. Rapp et al [19] ont présenté une méthode robuste et précise pour détecter 17 points caractéristiques du visage sur des images expressives. Cette méthode introduit une nouvelle architecture multi-résolution basée sur les algorithmes multi-noyau. Les patches de faibles résolutions codent les informations globales du visage donnant lieu à une détection grossière mais robuste du point désiré. Les patches de grandes résolutions utilisent les détails locaux afin d'affiner cette localisation du point recherché.

L'apprentissage multi-noyau permet de combiner différents niveaux d'informations de façon pertinente. Ce système est entraîné à l'aide d'une procédure originale appelé bootstrap.

I.3.1. Méthodologie :

Le système présenté dans cette étude se décompose en deux étapes: dans un premier temps, chaque point caractéristique du visage est détecté indépendamment. Dans un second temps, ces détections sont validées lors d'un processus de validation statistique.

Ce détecteur utilise un classifieur de type SVM qui apprend à distinguer un point caractéristique des autres zones de l'image. Pour caractériser l'information locale autour d'un point, ils ont utilisé les niveaux de gris du voisinage. Sur ces niveaux de gris, sont extraits des patches de plusieurs résolutions introduisant différents niveaux d'informations sur la zone recherchée. La pondération de chacune de ces résolutions est apprise à l'aide des algorithmes multi-noyau appliqués aux machines à vecteurs supports, combinés avec une procédure originale de ces chercheurs « bootstrap » adaptée à ce problème.

I.3.2. Extraction des caractéristiques :

Le détecteur de points caractéristiques du visage proposé utilise des patches de différentes tailles en entrée du classifieur. Une taille trop petite ne prendra en compte que l'information locale et introduisant des détections fines, alors qu'une taille trop grande n'extraira que des informations grossières et peu précises.

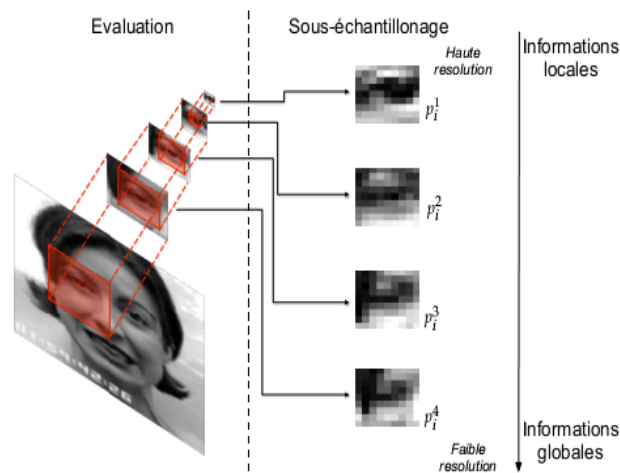


Figure I.6: Pyramide de patches multi-résolutions pour un pixel i . A chaque étape, différents niveaux d'informations sont mis en valeur.

I.3.3. Apprentissage Multi-Noyau :

Afin de procéder à la classification multi-noyau, les auteurs ont utilisé les machines à vecteurs supports avec une fonction noyau pour chaque résolution, ainsi, le système est capable de trouver la meilleure combinaison d'entrée afin de maximiser la marge qui sépare deux exemples de différentes classes. Pour cela ils ont utilisé les algorithmes SimpleMKL [20] pour l'apprentissage de leurs SVM multi-noyau.

I.3.4. Bootstrap et Patches négatifs :

La base d'apprentissage est divisée en trois bases différentes : A, B et C. A et B sont utilisés pour l'apprentissage, tandis que C est utilisé en cross-validation. La procédure de bootstrap a été implémentée de la façon suivante :

1. Apprentissage sur A et test B.
2. Récupération des fausses alarmes produites sur B et ajout de ces dernières dans l'ensemble

des exemples négatifs de B.

3. Validation sur C. Si le score de détection n'augmente plus, aller à l'étape 5. Sinon, aller à l'étape 4.

4. Inverser A et B et aller à l'étape 1.

5. Enchaînement des exemples positifs et négatifs de A et B et réalisation de l'apprentissage final.

Cette procédure originale de bootstrap ajoute des fausses alarmes à chaque itération sans mélanger les exemples de A et B. Cette stratégie permet de limiter les risques de surapprentissage : les fausses alarmes détectées sur les visages de A n'étant ajoutées que dans A. Ces fausses alarmes forcent la SVM, lors de l'itération suivante, à affiner l'hyperplan optimal.

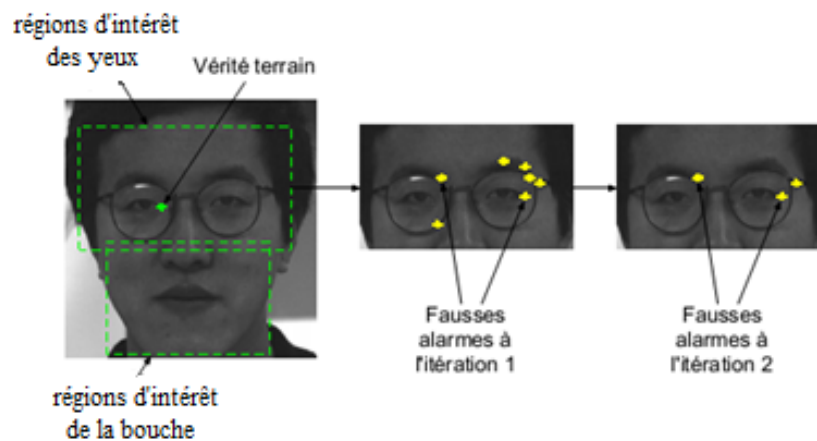


Figure I.7: Les fausses alarmes pertinentes sont ajoutées à l'ensemble de patche négatif à chaque itération.

I.3.5. Bases d'apprentissage :

Ce détecteur de points caractéristiques du visage a été entraîné à l'aide de deux bases différentes. La base de données Cohn-Kanade [21] contenant 486 séquences d'images de personnes faisant des expressions telles que la joie, la peur, la colère...etc. Les séquences d'images commencent par l'état neutre de l'expression pour finir avec le maximum de l'expression (apex). Pour l'apprentissage, Ces chercheurs ont utilisé 209 exemples choisis aléatoirement parmi les 486 séquences.

La base PIE (Pose, Illumination, Expression) de la CMU [22] présente plus de 40000 images de visages de 68 personnes. Chaque sujet a été capturé sous 13 différentes poses, 43

différentes illuminations, et 4 expressions différentes. Pour l'apprentissage de cette étude ils ont pris 108 exemples, choisis aléatoirement. Le détecteur a été entraîné sur ces deux bases, 317 visages au total.

Les résultats sont reportés dans le Tableau I.8

Fonctions noyaux	Linéaire	RBF	Polynomial n=5
Erreur moyenne	4.7%	4.3%	3.9%
Nb Vect. Supports	1579	851	612

Tableau I.8: Erreurs moyennes et nombres de vecteurs supports selon différentes fonctions noyaux.

L'utilisation de fonctions plus complexes qu'une simple fonction linéaire permet de diminuer le nombre de vecteurs supports. En effet, une fonction RBF ou polynomiale projetera les données dans un espace de dimension plus grande qu'une fonction linéaire. De fait, les données seront plus facilement séparables et le nombre de vecteurs supports utilisés sera moindre. Concernant l'erreur moyenne, cette projection dans un espace de plus grande dimension permet également d'améliorer les résultats.

I.3.6. Evaluation de la procédure de bootstrap :

Afin d'évaluer le processus de bootstrap ainsi que son impact sur les performances du système, ils ont effectué plusieurs tests. Les résultats de chacun de ces tests sont reportés dans le tableau I.9.

Méthode	Taux de bonne détection à 10%
MKL SVM + patches aléatoires	88%
MKL SVM + Bootstrap (1 it)	92%
MKL SVM + Bootstrap complet (6 it)	97%

Tableau I.9: Mesures de performance de différentes procédures de bootstrap.

Avec un ajout d'exemples négatifs aléatoirement sélectionnés ils ont obtenu 88% de bonnes détections. Une seule itération de bootstrap permet d'obtenir 92% de bonnes détections. Enfin, lorsque la procédure complète est exécutée, une amélioration des résultats de 5% pour obtenir donc 97% de bonnes détections. Ces évaluations ont été menées sur la base AR Face [23].

I.3.7. Résultats expérimentaux et Comparaisons :

Comme nous pouvons le voir à l'aide des scores de classification, tous les points de leur modèle sont détectés avec une grande précision, et ce même si la base de données inclut des images expressives.

On peut conclure que ce système se base sur l'exploitation d'informations à différentes résolutions. L'apprentissage multi-noyau permet de combiner ces différents niveaux d'informations de façon pertinente.

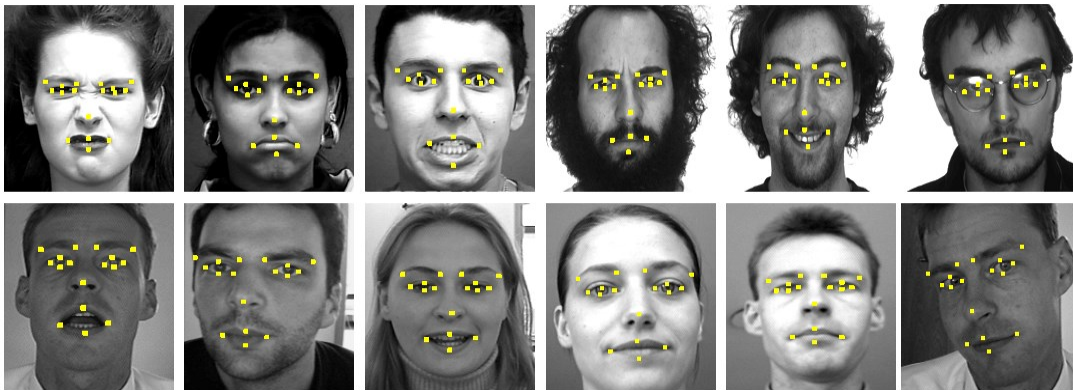


Figure I.8: Exemple de résultats sur les bases Cohn Kanade, AR Face.

Ce système est entraîné à l'aide d'une procédure originale de bootstrap. Les évaluations nous prouvent l'efficacité de ce bootstrap consistant à ajouter des exemples pertinents à la base d'apprentissage.

Enfin, en combinant les détections données par la SVM à une validation statistique, ce système corrige les exemples aberrants et assure des résultats robustes.

I.4. Analyse de l'état des isolateurs des lignes aériennes de distribution en utilisant une combinaison machine à vecteurs de support (SVM) et la transformée en ondelettes :

L'analyse de l'état des isolateurs des lignes aériennes de distribution en utilisant une combinaison machine à vecteurs de support (SVM) et l'analyse multi-résolution (MRA) semble être prometteuse pour faire face à la complexité du système. En Inde, les recherches réalisées par K.Tarakanath et al [24], se sont concentrés principalement sur l'application de la transformée en ondelettes basée sur l'extraction des caractéristiques de traitement d'image numérique et la SVM pour l'analyse de l'état des isolateurs. La plus grande contribution du document consiste à calculer les indices d'état pour les isolateurs à haute tension pour surmonter les difficultés liés aux applications de vidéo surveillance. La figure I.9 montre l'algorithme de surveillance de l'état des isolateurs et leur classification selon leurs états.

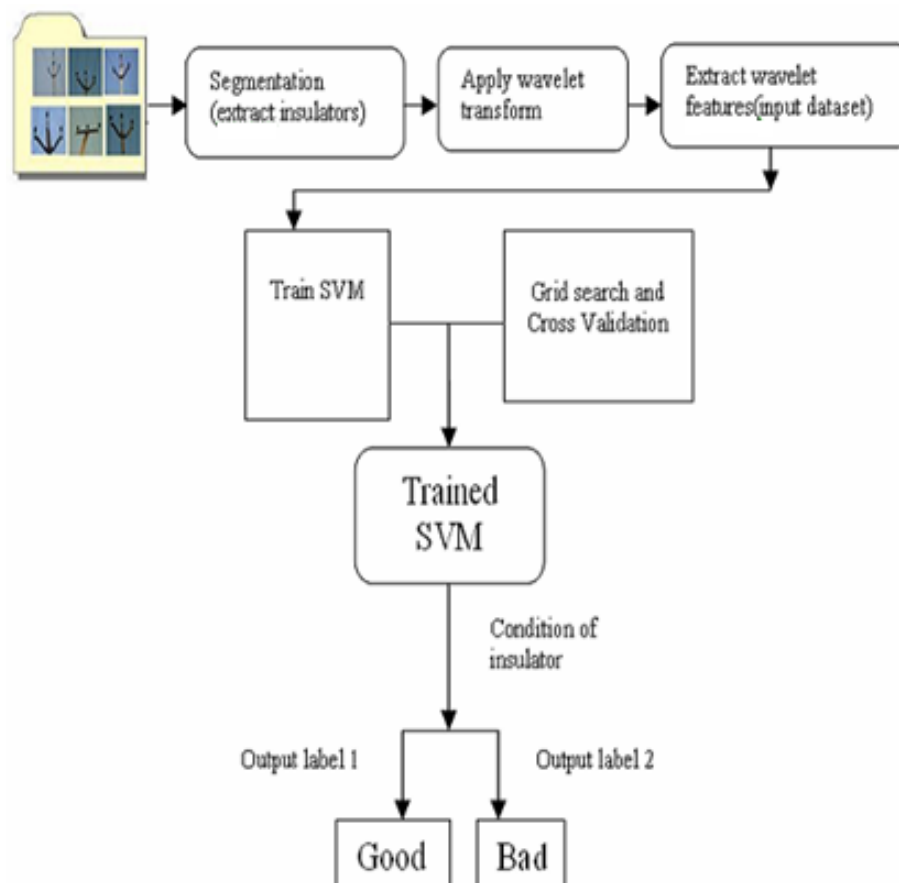


Figure I.9: Algorithme de surveillance de l'état des isolateurs.

Les caractéristiques d'ondelettes pour les isolateurs sains et endommagés sont présentées sous forme d'histogramme comme le montre les Figure I.10 et I.11.

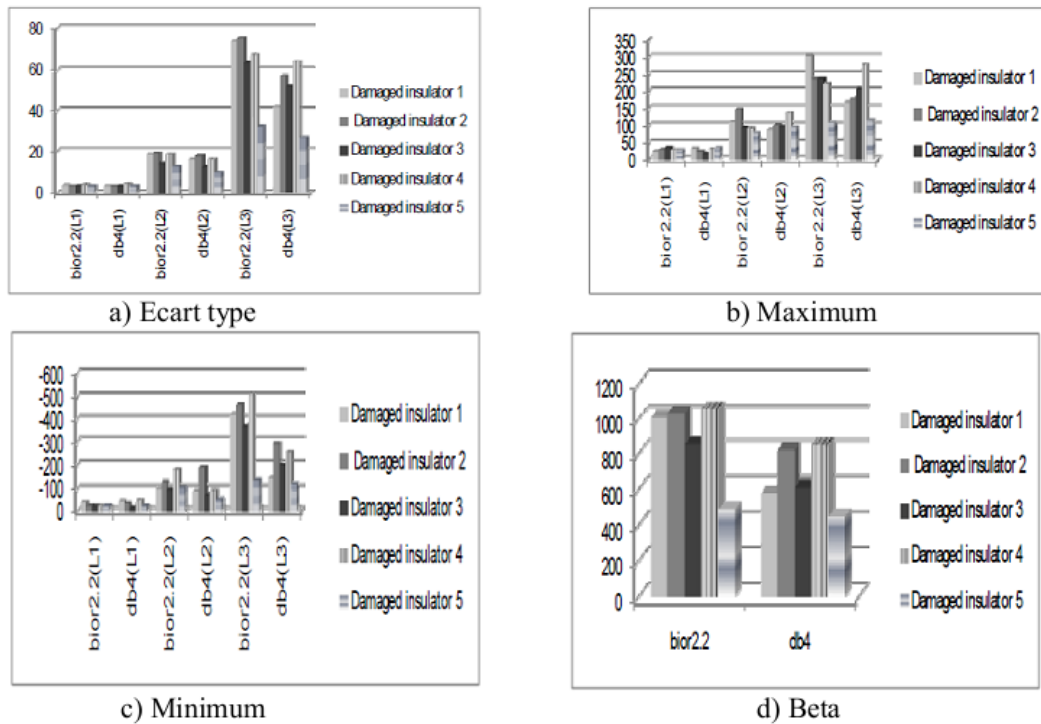


Figure I.10 : Caractéristiques d'ondelettes pour les isolateurs endommagés.

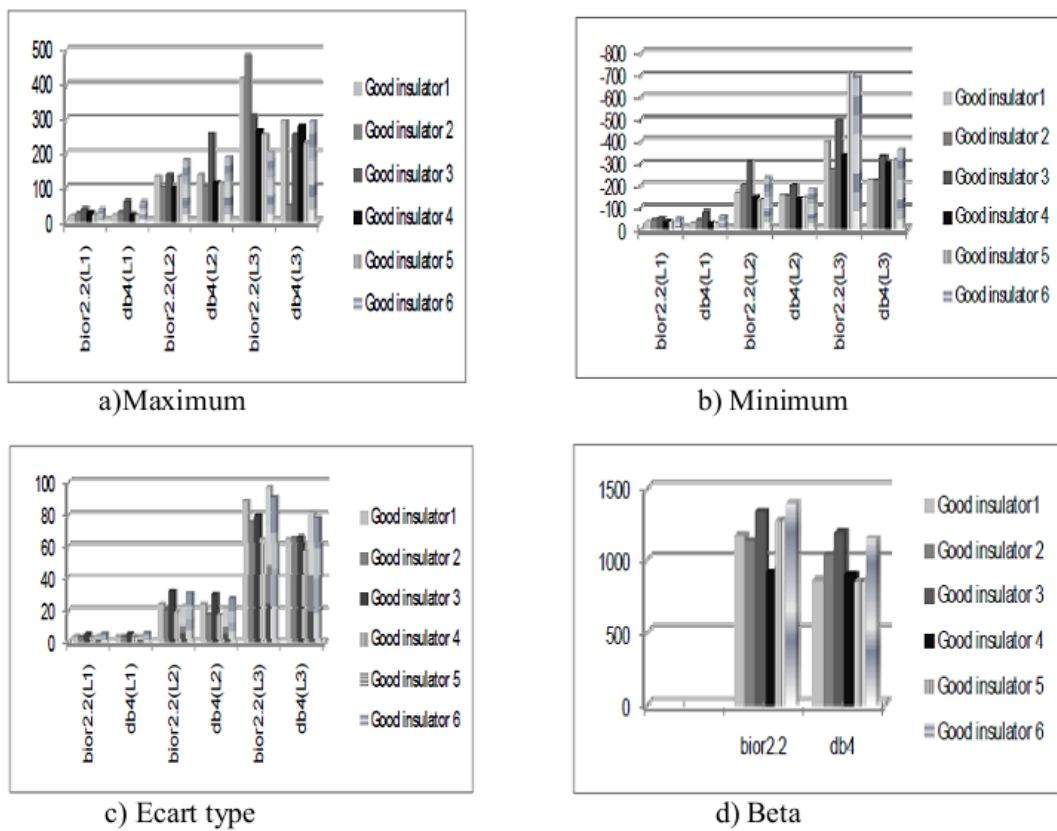


Figure I.11: Caractéristiques d'ondelettes pour les isolateurs sains.

Good	SD1	SD2	SD3	BETA (β)	Actual Health of insulator(I) = ($\beta / 1200$)*100	Condition Indices	SVM Output	Predicted Class
	3.825	23.63	88.7	1166.55	97.2	Healthy	Healthy	Yes
	3.407	20.86	74.87	1125.07	93.8	Healthy	Healthy	Yes
	6.575	30.7	90.77	1431.78	100	Healthy	Healthy	Yes
	5.663	27.57	86.12	1200.37	100	Healthy	Healthy	Yes
	5.094	31.6	80.01	1343.47	100	Healthy	Healthy	Yes
	3.851	22.54	96.92	1280.33	100	Healthy	Healthy	Yes
	3.878	24.12	86.97	1235.75	100	Healthy	Healthy	Yes
	5.541	30.8	91	1398.59	100	Healthy	Healthy	Yes
	5.941	29.8	87.9	1454.62	100	Healthy	Healthy	Yes
	5.1	30.12	91.25	1640	100	Healthy	Healthy	Yes
Damaged	3.083	17.8	57.12	823.74	68.6	Marginal	Marginal	Yes
	4.173	16.07	64	847.85	70.7	Marginal	Marginal	Yes
	3.409	14.017	63.44	853.3	71.1	Marginal	Marginal	Yes
	3.528	18.48	73.67	1008.31	84	Marginal	Marginal	Yes
	4.673	21.41	55.16	918.3	76.5	Marginal	Marginal	Yes
	3.293	9.837	27	450.68	37.6	Risky	Risky	Yes
	3.306	9.497	31	517.52	43.1	Risky	Risky	Yes
	4.97	16.75	41.12	678.18	56.5	Risky	Risky	Yes
	3.313	16.12	41.73	576.44	48	Risky	Risky	Yes
	3.229	12.55	51.68	611.22	50.9	Risky	Risky	Yes

Tableau I.10 : Indices de l'état des isolateurs en utilisant SVM [24].

Les éléments importants suivants sont extraits pour analyser l'état des isolateurs. Le β et les écarts type $\sigma L1$, $\sigma L2$, $\sigma L3$ sont obtenus à partir des détails des trois premiers niveaux de décomposition par la DWT. Les indices de l'état des isolateurs sont aussi obtenus en utilisant SVM comme indiqué dans le tableau I.10. Par l'observation de plusieurs isolateurs endommagés et sains, nous avons observé que la valeur Beta (β) joue un rôle important pour distinguer entre les isolateurs parfaits et endommagés.

La principale limite de la méthode proposée est l'installation de caméras qui ont deux degrés de liberté (horizontale et verticale). Aussi la technique proposée ne peut pas prendre en compte les défauts non-visuels tels que les fissures internes, les activités à fortes décharges partielles et le courant de fuite élevé en raison de la contamination. Cependant, l'algorithme pour la surveillance des isolateurs d'un système de distribution à haute tension en utilisant la vidéo surveillance (VS) avec l'aide de SVM, démontre l'efficacité pour des applications en temps réel pour l'approvisionnement en alimentation fiable.

I.5. Optimisation des isolateurs supports utilisés dans les systèmes haute tension (HT) en utilisant la technique support vecteur machine :

Les chercheurs S.Banerjee et al [12] ont utilisé la méthode SVM pour l'optimisation du champ électrique le long des isolateurs supports utilisés dans les systèmes HT.

Pour illustrer l'efficacité de SVM dans l'optimisation des problèmes de champ électrique, les chercheurs ont considéré deux exemples : le premier est un exemple simple qui se compose d'un isolateur cylindrique de porcelaine soutenant une électrode axisymétrique comme représenté dans la Figure I.12. Pour cet exemple, le potentiel entre l'électrode de phase et la terre est maintenu à 1 kilovolt. Le deuxième exemple illustré est un isolateur avec un noyau solide de porcelaine avec deux brides comme représenté sur la Figure I.13. Le potentiel pour cet exemple est maintenu à 100 kilovolts.

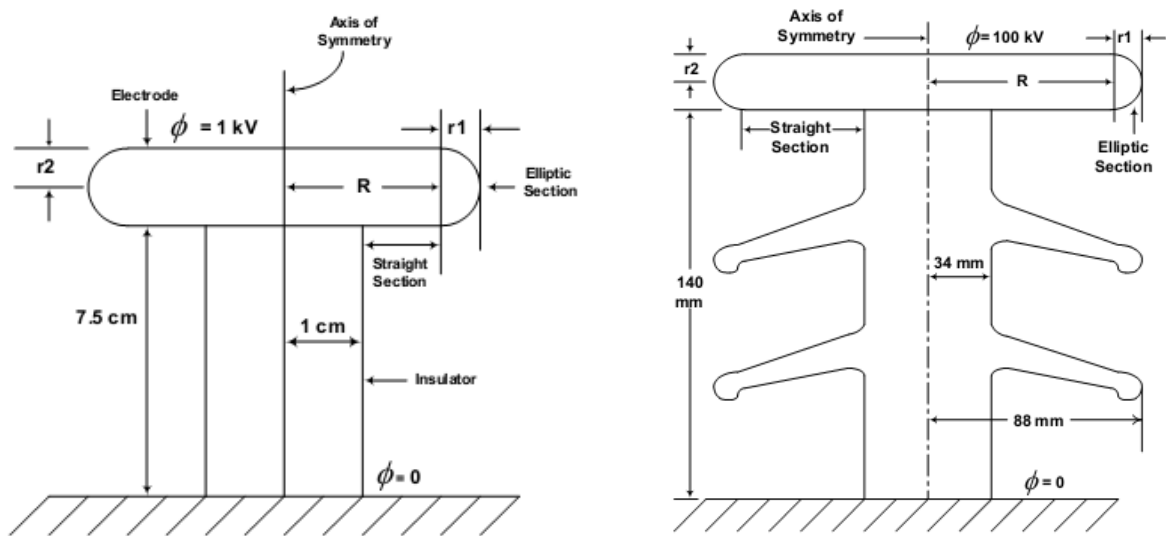


Figure I.12 : Isolateur utilisé dans l'exemple1. Figure I.13 : Isolateur utilisé dans l'exemple2.

La formation et les essais pour le SVM ont été préparés par calcul de champ électrostatique en utilisant la méthode indirecte d'élément de frontière (BEM).

La superficie de l'électrode et la distribution de l'effort le long de la surface de l'isolateur varient avec les paramètres R , r_1 et r_2 .

Dans cette étape, les chercheurs ont fait varier la valeur $R+r_1$, et pour chaque combinaison des paramètres R , r_1 et r_2 , ils ont calculé l'effort maximum le long de la surface de l'isolateur et la superficie de l'électrode.

Pour les deux exemples, le modèle SVM est formé pour la surface de l'électrode de phase et l'effort résultant maximum le long de la surface de l'isolateur, avec le rayon R comme vecteur de modèle d'entrée et r_2 comme vecteur de modèle de sortie.

Pour le premier exemple : les valeurs de $R+r_1$ sont: 10 cm, 11cm, 12cm.

Input Pattern Vectors			Output Pattern Vector	Input Pattern Vectors			Output Pattern Vector	Input Pattern Vectors			Output Pattern Vector
Surface Area (cm ²)	Maximum Stress (kV/cm)	R (cm)	r ₂ (cm)	Surface Area (cm ²)	Maximum Stress (kV/cm)	R (cm)	r ₂ (cm)	Surface Area (cm ²)	Maximum Stress (kV/cm)	R (cm)	r ₂ (cm)
777.9	0.143260	4.0	3.0	1086.2	0.137222	6.0	3.0	925.8	0.139886	5.0	3.0
875.8	0.142303	5.0	4.0	1204.8	0.137511	7.0	4.0	1034.0	0.138853	6.0	4.0
996.1	0.140749	6.0	5.0	1350.2	0.136588	8.0	5.0	1166.8	0.138773	7.0	5.0
835.5	0.139505	7.0	3.0	1154.9	0.135846	9.0	3.0	988.9	0.137070	8.0	3.0
1085.1	0.138385	8.0	5.0	1455.9	0.135960	10.0	5.0	1264.2	0.136731	9.0	5.0
932.8	0.139768	7.0	4.0	1272.8	0.135972	9.0	4.0	1096.5	0.137248	8.0	4.0
868.5	0.138189	8.0	3.0	1194.4	0.135885	10.0	3.0	1025.1	0.136598	9.0	3.0
964.7	0.142976	5.0	5.0	1312.9	0.137796	7.0	5.0	1132.5	0.139297	6.0	5.0
855.2	0.144402	4.0	4.0	1180.3	0.137716	6.0	4.0	1011.5	0.140619	5.0	4.0
811.3	0.139854	6.0	3.0	1126.1	0.136188	8.0	3.0	962.4	0.138189	7.0	3.0
792.7	0.141548	5.0	3.0	1103.9	0.137181	7.0	3.0	942.1	0.138349	6.0	3.0
950.9	0.140339	6.0	4.0	1234.8	0.136400	8.0	4.0	1061.6	0.138503	7.0	4.0
1035.0	0.139990	7.0	5.0	1396.4	0.136070	9.0	5.0	1209.4	0.137405	8.0	5.0
975.0	0.138292	8.0	4.0	1322.9	0.135923	10.0	4.0	1142.7	0.136669	9.0	4.0
938.7	0.145417	4.0	5.0	1281.9	0.138153	6.0	5.0	1104.0	0.141285	5.0	5.0

(a) (b) (c)
 Tableau I.11 : modèle vecteur pour (a) $R+r_1=10cm$ (b) $R+r_1=11cm$ (c) $R+r_1=12cm$

Les tableaux I.11 représente les données formées pour l'exemple 1 avec les trois valeurs de $R+r_1$. Les résultats d'essais sont représentés dans le tableau suivant :

Input Parameters			Output Parameter	Input Parameters			Output Parameter	Input Parameters			Output Parameter
Desired Surface Area (cm ²)	Desired Stress (kV/cm)	R (cm)	r ₂ (cm)	Desired Surface Area (cm ²)	Desired Stress (kV/cm)	R (cm)	r ₂ (cm)	Desired Surface Area (cm ²)	Desired Stress (kV/cm)	R (cm)	r ₂ (cm)
810.0	0.13819	5.0	3.07	950.0	0.1366	6.0	3.31	1090	0.1359	6.0	3.42
825.0	0.13819	5.0	3.25	1000.0	0.1366	8.0	3.16	1100	0.1359	7.0	3.16
850.0	0.13819	5.0	3.58	1050.0	0.1366	7.0	4.36	1125	0.1359	8.0	3.05
900.0	0.13819	5.0	4.15	1100.0	0.1366	5.0	4.99	1150	0.1359	9.0	3.00
950.0	0.13819	7.0	4.11	1150.0	0.1366	6.0	5.00	1175	0.1359	7.5	3.61
1000.0	0.13819	6.0	5.00	1200.0	0.1366	7.0	5.00	1200	0.1359	10.0	3.02
1050.0	0.13819	8.0	5.00	950.0	0.1366	7.0	3.94	1250	0.1359	8.0	4.30

(a) (b) (c)
 Tableau I.12: les essais pour (a) $R+r_1=10cm$ (b) $R+r_1=11cm$ (c) $R+r_1=12cm$

Pour le deuxième exemple : les valeurs de $R+r_1$ sont: 90mm, 100mm, 110mm.

Input Pattern Vectors			Output Pattern Vector	Input Pattern Vectors			Output Pattern Vector	Input Pattern Vectors			Output Pattern Vector
Surface Area (mm ²)	Maximum Stress (kV/mm)	R (mm)	r2 (mm)	Surface Area (mm ²)	Maximum Stress (kV/mm)	R (mm)	r2 (mm)	Surface Area (mm ²)	Maximum Stress (kV/mm)	R (mm)	r2 (mm)
58794.0	1.32322	87.0	9.0	71614.7	1.25402	97.0	9.0	85692.0	1.19955	107.0	9.0
55933.7	1.33325	85.0	7.0	68436.0	1.26022	95.0	7.0	82194.9	1.20757	105.0	7.0
57178.8	1.33217	83.0	9.0	69822.5	1.26097	93.0	9.0	83722.9	1.20598	103.0	9.0
53380.0	1.34487	81.0	5.0	65597.1	1.27045	91.0	5.0	79070.9	1.20885	101.0	5.0
54442.4	1.34441	78.0	7.0	66780.3	1.27054	88.0	7.0	80374.9	1.21317	98.0	7.0
52795.6	1.35268	75.0	5.0	64948.1	1.27687	85.0	5.0	78357.2	1.21518	95.0	5.0
54935.9	1.34014	81.0	7.0	67328.3	1.26721	91.0	7.0	80977.3	1.21074	101.0	7.0
53685.8	1.34331	83.0	5.0	65936.8	1.26823	93.0	5.0	79444.4	1.21210	103.0	5.0
56631.7	1.33598	81.0	9.0	69215.3	1.26387	91.0	9.0	83055.5	1.20815	101.0	9.0
53042.6	1.34831	78.0	5.0	65222.5	1.27223	88.0	5.0	78658.9	1.21176	98.0	5.0
55499.5	1.34541	75.0	9.0	67958.4	1.27106	85.0	9.0	81674.0	1.21379	95.0	9.0
56712.2	1.32871	87.0	7.0	69300.2	1.25917	97.0	7.0	83144.7	1.20381	107.0	7.0
54099.3	1.34009	85.0	5.0	66396.0	1.27657	95.0	5.0	79949.3	1.21080	105.0	5.0
55369.0	1.33693	83.0	7.0	67809.2	1.26474	93.0	7.0	81506.0	1.20895	103.0	7.0
55992.6	1.34096	78.0	9.0	68505.9	1.26768	88.0	9.0	82275.8	1.21110	98.0	9.0
54070.3	1.34830	75.0	7.0	66367.1	1.27332	85.0	7.0	79920.5	1.21549	95.0	7.0
54700.2	1.34003	87.0	5.0	67063.2	1.26342	97.0	5.0	80682.9	1.21326	107.0	5.0
57872.7	1.32811	85.0	9.0	70592.5	1.25804	95.0	9.0	84568.9	1.20384	105.0	9.0

(a)

(b)

(c)

Tableau I.13: modèle vecteur pour : (a) $R+r_1=90mm$ (b) $R+r_1=100mm$ (c) $R+r_1=110mm$

Les résultats pour le deuxième exemple sont représentés dans les tableaux suivants :

Input Parameters				Output Parameter	Input Parameters				Output Parameter	Input Parameters				Output Parameter	
Desired Surface Area (mm ²)	Desired Stress (kV/mm)	R (mm)	r2 (mm)	Desired Surface Area (mm ²)	Desired Stress (kV/mm)	R (mm)	r2 (mm)	Desired Surface Area (mm ²)	Desired Stress (kV/mm)	R (mm)	r2 (mm)	Desired Surface Area (mm ²)	Desired Stress (kV/mm)	R (mm)	r2 (mm)
53000	1.325	75.0	5.457	65000	1.255	85.0	5.252	78700	1.2	95.0	5.030	53000	1.325	75.0	5.457
54500	1.325	79.0	6.804	66000	1.255	85.0	6.176	79000	1.2	98.0	5.042	54500	1.325	79.0	6.804
55000	1.325	80.0	7.095	66500	1.255	86.0	6.253	80500	1.2	107.0	5.163	55000	1.325	80.0	7.095
55500	1.325	77.0	9.000	67000	1.255	88.0	6.348	81500	1.2	95.0	8.780	55500	1.325	77.0	9.000
56000	1.325	81.0	7.791	68000	1.255	93.0	6.700	82500	1.2	100.0	8.691	56000	1.325	81.0	7.791
57000	1.325	83.0	8.984	69000	1.255	91.0	8.990	83500	1.2	103.0	8.768	57000	1.325	83.0	8.984
58000	1.325	85.0	9.000	70000	1.255	95.0	7.910	84500	1.2	105.0	8.900	58000	1.325	85.0	9.000

(a)

(b)

(c)

Tableau I.14: les essais pour : (a) $R+r_1=90mm$ (b) $R+r_1=100m$ (c) $R+r_1=110mm$

Pour calculer une erreur entre le y_k de la cible et la valeur obtenue à partir de SVM.

L'équation de SVM pour la régression est comme suit :

$$\text{Minimiser : } F(w) = 0.5 (w \cdot w^T) + \zeta \sum_k \xi_k$$

Le modèle SVM construit utilise la fonction noyau linéaire dans les deux cas étudiés.

L'exactitude de la formation est mesurée en termes d'erreur absolue moyenne (MAE).

Le tableau suivant présente la valeur optimale du ζ qui correspond à la valeur minimale de l'erreur absolue moyenne (MAE) pour différentes valeurs de $R+r_1$.

R+a (cm)	Optimum ζ	% MAE	R+a (mm)	Optimum ζ	% MAE
10	7	1.212043	90	13	0.804980
11	34	0.131239	100	9	0.517647
12	17	0.402551	110	26	0.297714

Tableau I.15: la valeur optimale du ζ pour différentes valeurs de $R+r_1$ pour l'exemple 1 (à gauche) et l'exemple 2 (droite).

En conclusion, le degré d'exactitude, mesuré en terme de MAE, réalisé à base du SVM, montre clairement l'efficacité d'utiliser cet outil pour optimiser le champ électrique le long de la surface des isolateurs des systèmes HT.

I.6. Classification basée sur la technique SVM pour la commande de Myoelectric appliquée à un organe supérieur :

Dans cet article, les chercheurs M.Asghari Oskoei et Huosheng Hu [15] proposent et évaluent l'application de SVM pour classifier les mouvements des organes supérieurs en utilisant les signaux myoelectric (les propriétés électriques du muscle). Ce travail présente une méthode pour ajuster les paramètres de SVM avant classification, et traite deux techniques pour améliorer la performance de la commande: la segmentation recouverte et le vote de majorité.

Le modèle SVM est ensuite comparé avec deux classifieurs connus : analyse discriminante linéaire (LDA) et les réseaux neurologiques multicouches (MLP).

I.6.1. L'acquisition des données :

Un MES (Surface myoelectric signals) de quatre canaux a été placé dans quatre endroits sur un avant-bras en utilisant des électrodes actives bipolaires.

Des données ont été collectées de 11 modèles en bonne santé. Chaque modèle a exécuté cinq mouvements de l'organe, et s'appuyé pour fournir six états distincts (classes). Deux séquences de six mouvements s'appellent un bloc. Quatre blocs de données ont été recueillis des modèles en chaque session, l'exactitude de la classification a été calculée suivant une méthode quadruple de contre-vérification (cross-validation).

I.6.2. Segmentation des données :

Un segment est une séquence des données limitées dans le temps, qui est utilisée pour

estimer les caractéristiques d'un signal. Un segment des données de MES avec une durée de 200 ms (ou plus) contient assez d'informations pour estimer un état de mouvement de la main.

Dans cette partie, deux types de segmentation sont utilisées : le premier est la segmentation disjointe qui est associée à la longueur de segment et la deuxième c'est la segmentation recouverte, elle est associée à la longueur et à l'incrément.

L'incrément est l'intervalle de temps entre deux segments consécutifs. Il devrait être plus petit que la longueur de segment, et plus grand que la durée de la transformation.

I.6.3. Sélection des caractéristiques :

Dans ce travail, les caractéristiques impliquées du domaine temporel (TD) sont : la valeur absolue moyenne (MAV), la racine carrée moyenne (RMS), la longueur d'onde (WL), la variance (VAR), passages à zéro (ZCs), changement de signe de pente (SSC), amplitude de William (WAMP), et deux types des valeurs absolues moyennes modifiées (MAV1 et MAV2). Les caractéristiques considérées du domaine fréquentiel (FD) sont : des spectres de puissance (PS), des coefficients autorégressifs d'ordre 2 et 6 (AR2 et AR6), et la moyenne et la médiane des fréquences de signal (FMN, FMD).

En outre, quatre groupes des caractéristiques (multifeatures) sont traités dans ce travail. Le premier groupe s'appelle le groupe des caractéristiques du TD (il a été utilisé par Hudgins et autres [15]), il contient MAV, WL, ZC, et SSC. Le deuxième groupe (recommandé par Huang et autres [16]), se compose du RMS et de l'AR6. Le troisième groupe est constitué par MAV et WL. Le quatrième groupe contient RMS et AR2.

La dimension d'un espace des caractéristiques dépend de leurs types et de leurs nombre; par exemple, les caractéristiques simples MAV et AR6, aussi bien que les caractéristiques groupées RMS+ AR6, produisent 4, 24, et 28 vecteurs dimensionnels respectivement.

Ces chercheurs ont effectué une expérience afin d'étudier la relation entre les caractéristiques (individuels ou en groupe) et la longueur de segment.

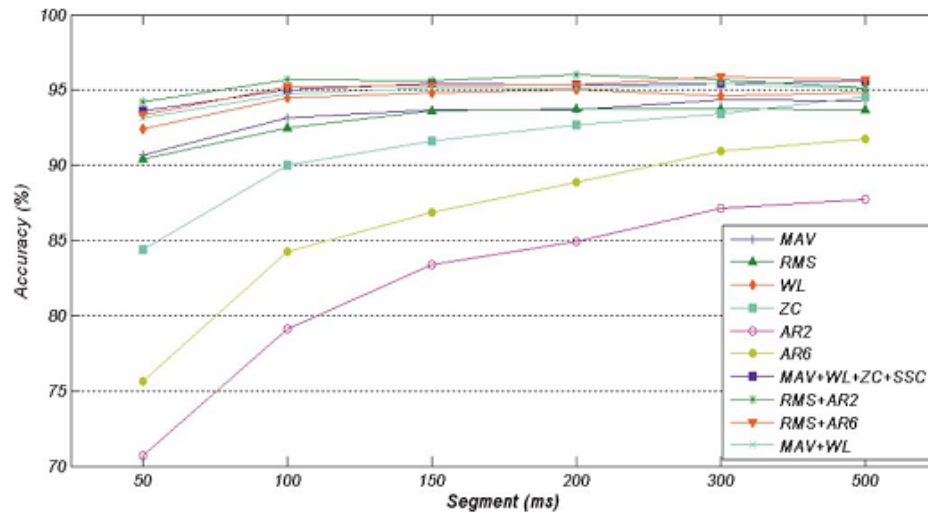


Figure I.14 : La performance de classification de différentes caractéristiques en diminuant la longueur du segment.

La Figure I.14 représente l'exécution de la classification pour différentes longueurs de segment. Les caractéristiques simples (MAV, WL, RMS, ZC, AR2, et AR6) et les caractéristiques multiples (MAV+ WL+ ZC+ SSC, RMS+ AR2, RMS + AR6, et MAV+WL) ont été classifiées en utilisant un SVM avec le noyau RBF et les paramètres pré réglés.

On peut voir que pour toutes les caractéristiques simples, plus la longueur du segment est petite, plus l'exactitude est faible; c'est parcequ'un segment plus court rapporte plus de polarisation et de désaccord dans l'évaluation de caractéristiques. Les caractéristiques simples du TD (MAV, WL, RMS, et ZC) sont plus stables aux changements de la longueur de segment que les caractéristiques simples de FD (AR2 et AR6).

On peut voir aussi que les caractéristiques multiples, qui incluent une combinaison de différentes caractéristiques simples sont plus stables aux changements de la longueur de segment que les caractéristiques simples entières.

Cette expérience suggère l'utilisation des caractéristiques individuelles du TD (MAV, WL, et RMS) ou de toutes les caractéristiques multiples pour les arrangements de commande rapides dans lesquels un contrôleur myoelectric devrait répondre dans un temps réduit.

I.6.4. Prétraitement :

L'étude effectuée par Englehart et autres [18] montre que l'application de la segmentation recouverte avec le vote de majorité (MV) comme un mécanisme de prétraitement améliore l'exactitude et empêche la dégradation, en diminuant la taille du

segment. L'expérience suivante examine cette idée quand on applique la méthode SVM pour la classification, et aussi étudie la performance de la classification avec et sans l'utilisation de MV.

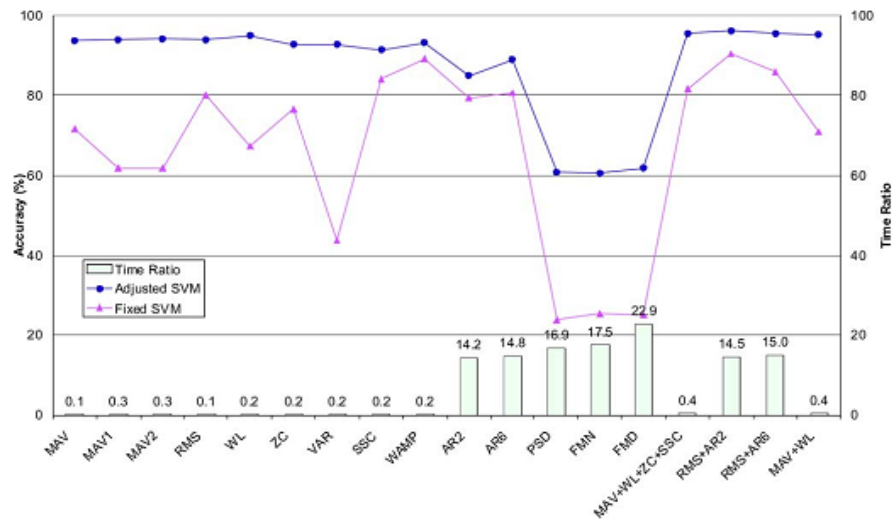


Figure I.15: La performance de classification et le rapport de temps de calcul pour 14 caractéristiques individuelles et 4 caractéristiques multiples.

La Figure I.15 montre la différence considérable entre l'exactitude de la classification SVM avant et après l'ajustement des paramètres, ainsi que le rapport de temps de calcul pour chaque caractéristique.

On peut voir également que les caractéristiques du domaine fréquentiel (FD) imposent un temps de calcul relativement élevé par rapport à celles des caractéristiques du temporel (TD).

On peut dire que la caractéristique multiple RMS + AR2 est la meilleure pour une durée de segment de 200ms. Les caractéristiques du TD (MAV + WL+ ZC+ SSC ou MAV + WL) sont les meilleures options pour la classification de MES à cause de leur stabilité de l'exécution dans diverses tailles de segment. Parmi les caractéristiques individuelles, WL est le meilleur.

I.6.5. Classification :

Dans cette étape on compare les différents noyaux SVM : RBF, linéaire, polynôme, et sigmoïde avec deux classificateurs bien connus, LDA et MLP. La comparaison effectuée est basée sur les différentes caractéristiques choisies (unique ou multi caractéristiques) avec une

longueur de segment de 200 ms.

Les chercheurs ont utilisé Grid-search afin d'ajuster les paramètres de SVM (C , r et γ) [17]. La disposition du MLP a été également ajustée avant utilisation. Plusieurs dispositions ont été examinées et celles qui ont donné le meilleur résultat ont été choisies.

L'expérience suivante consiste à comparer l'exactitude du SVM à celle de LDA et MLP. Des résultats expérimentaux sont affichés dans la *Figure I.16* (à gauche).

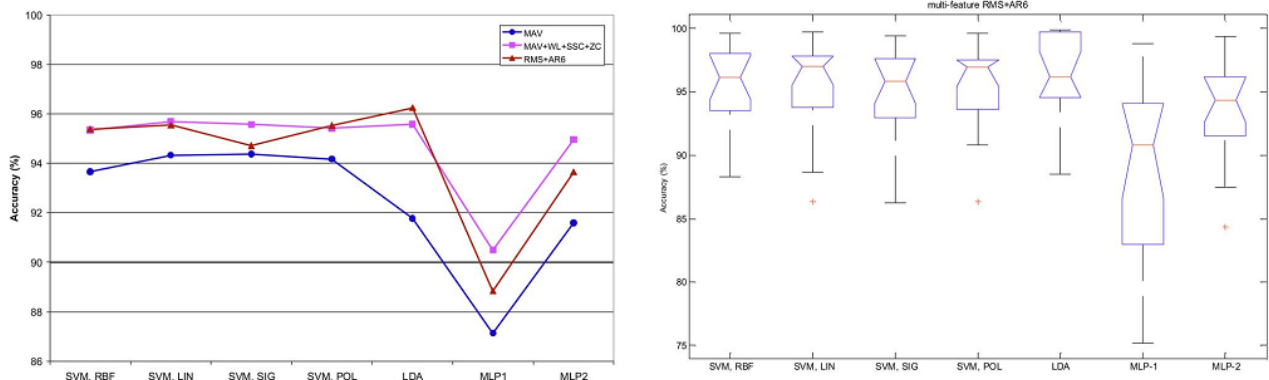


Figure I.16 : comparaison des différentes fonctions noyaux de SVM, LDA et MLP avec une et deux couches cachées.

Les classificateurs ont été examinés pour la caractéristique MAV, et les deux caractéristiques multiples MAV + WL+ SSC + ZC, et RMS + AR6.

On constate que l'exactitude moyenne pour toutes les fonctions noyaux est approximativement égale à $95.5 \pm 3.8\%$. Le fonctionnement des différents noyaux est presque identique, ceci peut être expliqué par la linéarisation des frontières entre les classes. Pour LDA, l'exactitude moyenne est de $94.5 \pm 4.9\%$. L'exactitude du MLP avec une couche cachée chute approximativement de 6%, alors que le MLP avec deux couches cachées donne une précision presque comme celle obtenue en utilisant SVM ou LDA.

D'après cette étude on peut conclure qu'un SVM offre la performance de classification qui surpasse les autres classificateurs. Les caractéristiques multiples du TD (MAV + WL+ ZC+SSC) surpassent les autres caractéristiques, en raison de leur taux d'exactitude relativement élevé, de son stabilité pour diverses durées de segment et leur simplicité informatique. Parmi les caractéristiques individuelles, WL est le meilleur à cause de son exactitude élevée et sa stabilité avec les différentes méthodes de segmentation.

Le système MV ne fournit aucune amélioration remarquable d'exactitude, mais donne un rendement plus stable.

I.7. Discrimination de la source de décharge partielle en utilisant la méthode SVM :

Les chercheurs L. Hao et al [13] ont utilisé la méthode SVM pour la discrimination des sources des décharges partielles (DP), ils ont utilisé trois modèles artificiels pour simuler les sources typiques de DP qui peuvent exister dans l'appareillage des systèmes d'alimentation.

Dans cette étude, ils ont utilisé un système de mesure de DP de grande largeur de bande. Ce système se compose d'une sonde RFCT (radio frequency current transducer), d'un oscilloscope de mémoire numérique et d'un PC de haute performance pour faciliter l'identification automatique de source de DP.

Le circuit d'essai est représenté sur le schéma suivant :

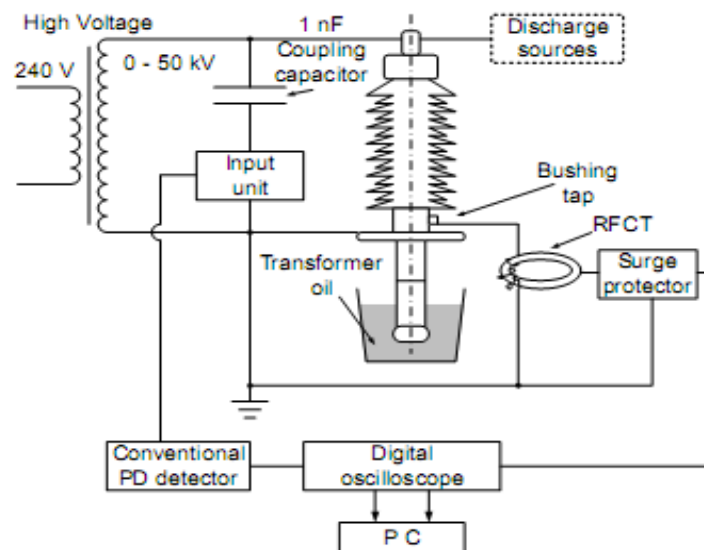


Figure I.17: Diagramme schématisé du circuit de mesure de DP.

Remarque : Dans la figure I.8 (conventional PD detector) est appelé aussi détecteur conventionnel de Robinson. Il a été employé pour le calibrage, pour mesurer la charge apparente et pour produire des données de formation du modèle SVM.

Les trois modèles artificiels des sources de DP étudiés sont : décharge par effet couronne avec la terre isolée (a), décharge extérieure en air (b) et la décharge interne en huile (c). Les trois modèles sont représentés dans la figure I.18 :

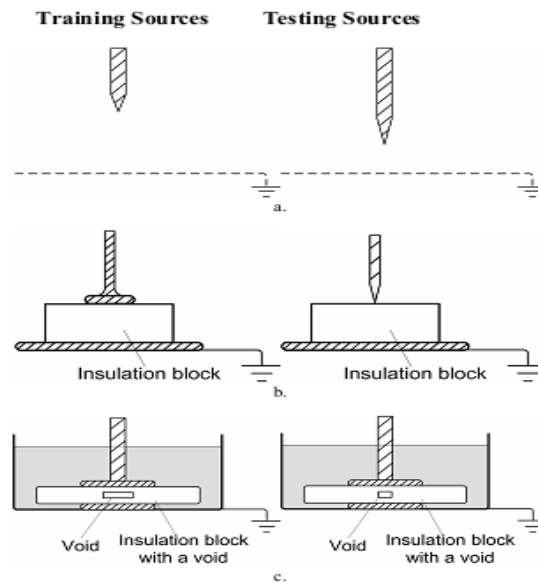


Figure I.18: Les sources de DP artificielles.

I.7.1. Acquisition et traitement des données :

Les signaux générés par le détecteur de Robinson et les signaux à la sortie du RFCT sont rassemblés pour former les données de formation et de validation respectivement. (Tableau I.16).

Data type	Training data	Testing data
Sensor	Robinson detector	RFCT
Data quantity	500 cycles × 3 PD sources × 2 applied voltages	20 cycles × 3 PD sources × 2 applied voltages

Tableau I.16 : Les données de formation et de validation du modèle SVM.

Les chercheurs ont appliqué l'analyse par la transformée en ondelettes pour traiter les données de mesure obtenues par la RCFT.

La figure I.19 montre les spectres d'énergies moyens des coefficients de décomposition d'ondelette des trois sources de DP.

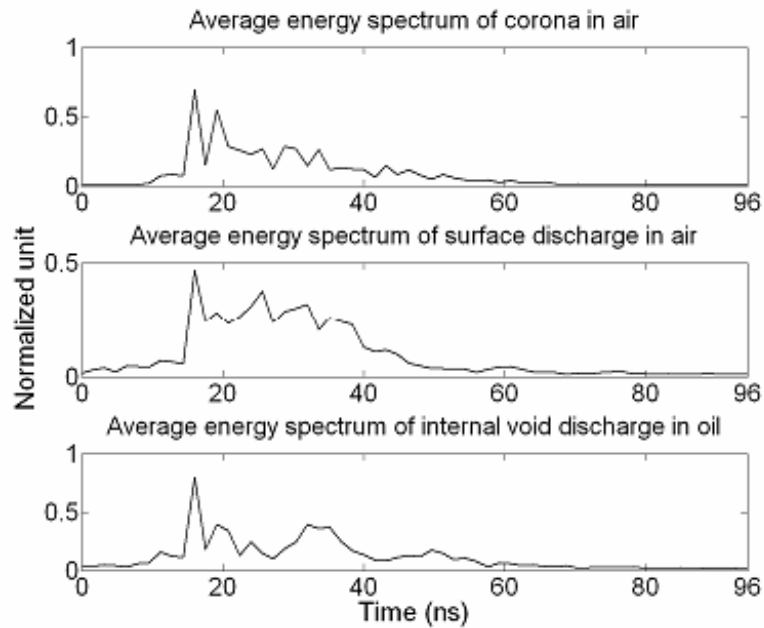


Figure I.19: les spectres d'énergies moyens des coefficients de décomposition d'ondelette des trois sources de DP.

En ce qui concerne le détecteur conventionnel de Robinson, les données sont traitées pour produire des modèles résolus échelonnés de ϕ -q-n, comme il est montré sur la figure I.20.

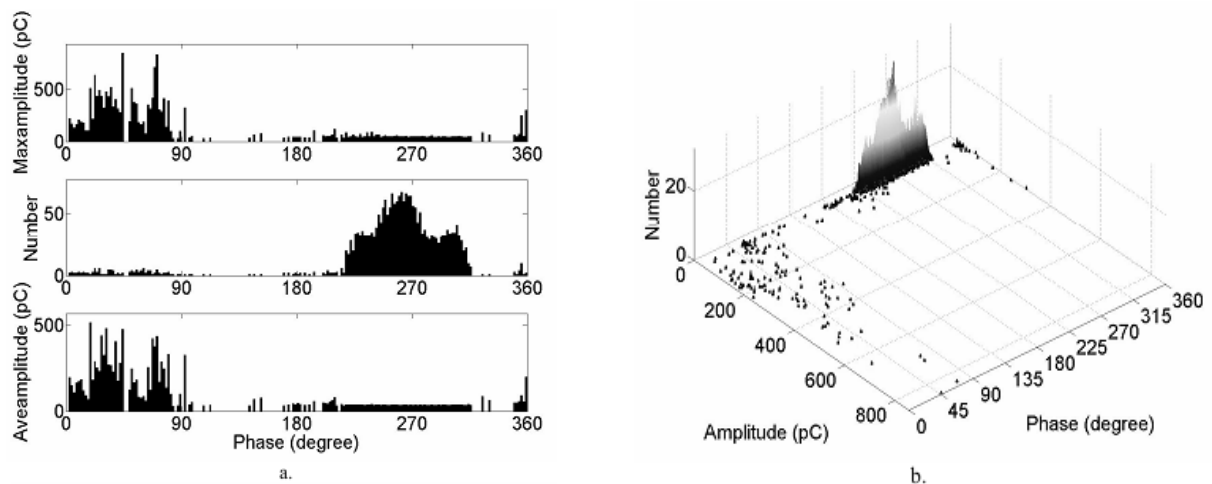


Figure I.20 : l'analyse de trois sources de DP simulée.

a. Histogrammes de phase résolue en 2D

b. 3D - modèle de ϕ -q-n

La procédure utilisée par ces chercheurs pour l'entraînement et la validation du modèle SVM est résumée dans la figure I.21 :

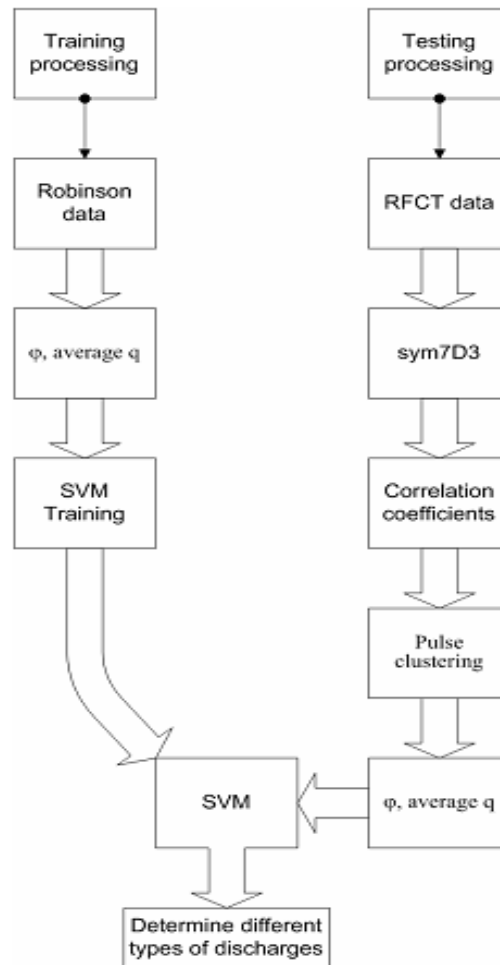


Figure I.21: Procédure de la classification par SVM.

Conclusion :

La méthode développée dans cette étude peut distinguer entre les trois sources de DP mais ne peut pas confirmer la présence de l'activité interne de DP. L'utilisation de plus d'un paramètre d'entrée (par exemple la phase et les coefficients de la transformée en ondelette) donnent une bonne performance pour l'identification de source multiple de DP.

Chapitre II :

Théorie des

Supports Vecteurs

Machine (SVM)



Introduction au SVM :

La classification est une opération de structuration qui vise à organiser un ensemble d'observations en groupes homogènes et contrastés afin de faciliter l'analyse des informations et d'effectuer des prédictions. Le but principal des différentes méthodes de classification est l'identification des classes auxquelles appartiennent des objets à partir de certains paramètres descriptifs. Elles s'appliquent à un grand nombre d'activités humaines et conviennent en particulier au problème de la prise de décision automatisée. La procédure de classification sera extraite automatiquement à partir d'un ensemble d'exemples. Un exemple consiste en la description d'un cas avec la classification correspondante. Un système d'apprentissage doit alors, à partir de cet ensemble d'exemples, extraire une procédure de classification, il s'agit en effet d'extraire une règle générale à partir des données observées. La procédure générée devra classer correctement les exemples de l'échantillon et avoir un bon pouvoir prédictif pour classer correctement de nouvelles descriptions [24, 29].

Les méthodes utilisées pour la classification sont nombreuses: la méthode des machines à vecteurs de support (Séparateurs à Vastes Marges (SVM)), les réseaux de Neurones, etc. Nous présentons dans la suite de ce chapitre une étude détaillée sur la méthode SVM. Cette méthode a montré son efficacité dans plusieurs domaines d'applications tels que le traitement d'image, la catégorisation de textes, le diagnostic médicale et comme dans notre cas, dans le domaine de la Haute Tension.

II.1. Types de classification :

Il existe ainsi deux grandes familles de méthodes de classification [30]:

II.1.1. Classification supervisée :

Le classifieur doit être capable de prédire la classe d'appartenance ω_j d'un objet de vecteur de description x_i avec un minimum d'erreur. Pour cela, il estimera la probabilité d'appartenance à la classe ω_i du prototype x_i notée $p(\omega_i / x_i)$.

On considère une machine devant classer L objets d'un ensemble D . Chaque échantillon est donc représenté par un doublet $\{x_i, y_i\}$ où x_i est le vecteur représentant le $i^{\text{ème}}$ objet dans l'espace de description (typiquement \mathbb{R}^n) et y_i est l'étiquette attribuée à la classe d'appartenance de l'objet (c'est-à-dire si x_i appartient à ω_i alors $j = y_i$). On cherche donc à définir la loi de probabilité $p(x, y)$, qui pour chaque x retourne la probabilité qu'il

appartienne à la classe d'étiquette y .

II.1.2. Classification non supervisée :

Dans ce cas, il n'existe pas d'expert capable de définir la classe de chaque objet. Ces méthodes de classification ne peuvent pas se baser sur l'estimation de probabilités puisqu'elles ne sont ni définissables ni estimables dans la mesure où on ne peut connaître l'appartenance d'un objet à une classe. Ces méthodes essaient donc de découper la représentation ponctuelle dans l'espace de description en parties homogènes dites clusters par le biais d'agrégations. La principale difficulté de ces méthodes réside dans la détermination du critère d'homogénéité permettant de former les clusters.

Les SVM se situent dans le groupe de classification supervisés puisque on utilise une base d'exemples pour obtenir la règle de classification.

II.2. Théorie de l'apprentissage : [24, 25].

L'apprentissage est défini comme l'acquisition de connaissances et compétences permettant la synthèse d'information. Un algorithme d'apprentissage va permettre de passer d'un espace d'exemples X à un espace dit des hypothèses H .

L'objectif de l'apprentissage à partir d'exemples étiquetés appelé aussi apprentissage supervisé est de construire une fonction qui permet d'approcher au mieux une fonction inconnue qui génère des données aléatoires, indépendantes et identiquement distribuées et dont nous ne disposons que de quelques exemples.

Un système d'apprentissage à partir d'exemples est composé de trois modules principaux :

- Un générateur qui génère des données aléatoires appelées les vecteurs d'entrée. Ces vecteurs sont indépendants et identiquement distribués suivant une distribution de probabilité inconnue $P(x)$.
- Un superviseur qui associe pour chaque vecteur d'entrée x une sortie y (la classe) suivant une distribution de probabilité également inconnue $P(x, y)$.
- Une machine d'apprentissage qui permet d'implémenter une famille de fonctions $f_{\alpha}(x)$; $\alpha \in A$, ou A est un ensemble de paramètres. Ces fonctions doivent produire pour chaque vecteur d'entrée x une sortie \tilde{y} la plus proche possible de la sortie y du superviseur.

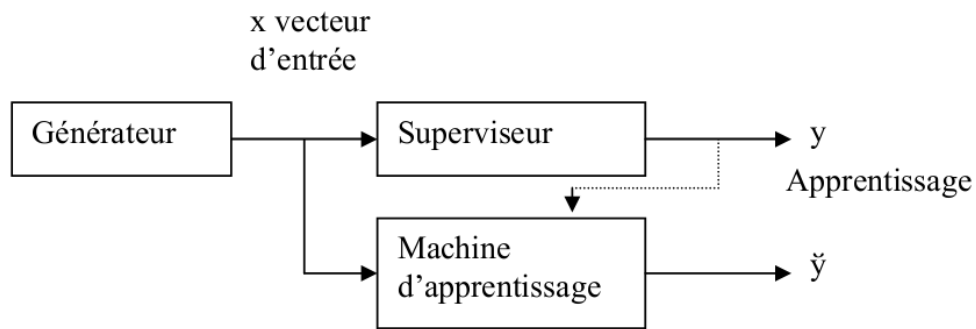


Figure II.1: les modules d'un système d'apprentissage.

Le problème d'apprentissage à partir d'exemples apparaît dans plusieurs domaines comme la prédiction, la classification, la reconnaissance de formes, la reconnaissance d'objets, etc.

Il y a trois types d'apprentissage :

- **L'apprentissage supervisé** : on dispose d'un nombre fini d'exemples d'une tâche à réaliser, sous forme de paires (entrée, sortie désirée), et on souhaite obtenir, d'une manière automatique, un système capable de trouver de façon relativement fiable la sortie correspondant à toute nouvelle entrée qui pourrait lui être présentée.

- **L'apprentissage non-supervisé** : Dans l'apprentissage non-supervisé il n'y a pas de notion de sortie désirée, on dispose seulement d'un nombre fini de données d'apprentissage, constituées "d'entrées", sans qu'aucun label n'y soit rattaché.

- **L'apprentissage par renforcement** : La particularité et la difficulté du cadre de l'apprentissage par renforcement est que les décisions prises par l'algorithme influent sur l'environnement et les observations futures.

II.3. Les machines à vecteurs de support:

Les machines à vecteur de Support ou séparateurs à vaste marge (SVM) sont de nouvelles techniques discriminantes dans la théorie de l'apprentissage statistique supervisé. Elles ont été proposées en 1995 par V. Vapnik dans son livre « The nature of statistical learning theory » [35]. Elles permettent d'aborder plusieurs problèmes divers et variés comme la régression, la classification, la prévision... etc. Depuis leur introduction dans le domaine de la reconnaissance de formes, plusieurs travaux ont pu montrer l'efficacité de cette technique.

SVM est une méthode particulièrement bien adaptée pour traiter des données de très haute dimension telles que les textes et les images. Le succès de cette méthode est

justifié par les solides bases théoriques qui la soutiennent [24, 30].

Les SVM sont des méthodes d'apprentissage automatique, c'est-à-dire à partir d'un échantillon d'apprentissage, on va créer un modèle applicable sur des échantillons à tester, par contre elles ne sont pas adaptées au cas d'une discrimination à plus de deux classes. Par rapport aux techniques classiques d'apprentissage, les SVM ne dépendent pas de la dimension de l'espace de représentation des données. Grâce à l'usage d'une fonction noyau, elles permettent une classification non linéaire comme nous le verrons par la suite [28, 29].

L'idée des hyperplans à marge maximale a été explorée dès 1963 par V.Vapnik et A. Lerner, et en 1973 par Richard Duda et Peter Hart dans leur livre « Pattern Classification ». Les fondations théoriques des SVM ont été explorées par V.Vapnik et ses collègues dans les années 70 avec le développement de la Théorie de Vapnik-Chervonenkis, et par Valiant [34].

L'idée des fonctions noyaux n'est pas non plus nouvelle: le théorème de Mercer date de 1909, et l'utilité des fonctions noyaux dans le contexte de l'apprentissage artificiel a été montrée dès 1964 par Aizermann, Bravermann et Rozoener [34].

Ce n'est toutefois qu'en 1992 que ces idées seront bien comprises et rassemblées par Boser, Guyon et Vapnik dans un article, qui est l'article fondateur des séparateurs à vaste marge. L'idée des variables ressorts, qui permet de résoudre certaines limitations pratiques importantes, ne sera introduite qu'en 1995. À partir de cette date, qui correspond à la publication du livre de V.Vapnik : « The nature of statistical learning theory » [35], les SVM gagnent en popularité et sont utilisés dans de nombreuses applications [34].

II.4. Théorie de Vapnik-Chervonenkis:

Une manière de contrôler la complexité d'une classe de fonctions est donnée par la théorie de Vapnik-Chervonenkis (VC) et le principe de minimisation du risque structurel. Les travaux de Vapnik et Chervonenkis sur cette théorie, depuis les années 1970, font surgir une condition nécessaire pour avoir la convergence uniforme dépendant de ce qu'ils appellent la capacité h de la classe F de fonctions dite aussi la dimension de Vapnik-Chervonenkis (dim VC), qui est une mesure de sa complexité. Grossièrement, la dimension de VC mesure combien d'échantillons de l'ensemble d'apprentissage peuvent être séparés par toutes les classifications possibles issues des fonctions de la classe [31].

La dimension de Vapnik-Chervonenkis h d'une classe de fonctions F en classification binaire est définie comme étant le cardinal maximal d'un sous-ensemble $A \subset X$ tel qu'on

puisse toujours trouver une fonction $f \in F$ qui classe parfaitement tous les éléments de A quelles que soient leurs étiquettes. En d'autres termes :

$$h = \max \{|A| / \forall \theta_i \in \{-1, +1\}^{|A|}, \exists f \in F \text{ telle que } \forall x_i \in A, f(x_i) = \theta_i\} \quad (\text{II.1})$$

Cette définition est assez complexe, et souvent il est plus commode de comprendre la dim VC à contrario : si le nombre des éléments de A est plus grand que la dim VC, alors il existe des dichotomies qui ne sont pas réalisées par F .

Par exemple, si on se place dans le plan, on est toujours capable de séparer trois points (non alignés) quel que soit leur étiquetage en utilisant des lignes. Par contre on est incapable de le faire pour quatre points. (Figure II.2).

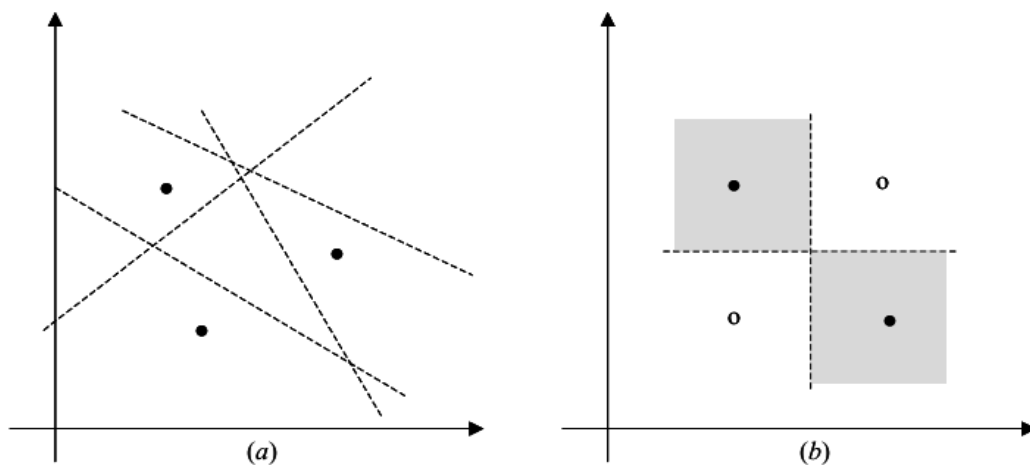


Figure II.2: Le problème du XOR. (a) Trois points non alignés du plan sont linéairement séparables quelque soit leur étiquetage. (b) Cette répartition de quatre points est non linéairement séparable.

Ainsi, la dimension de Vapnik-Chervonenkis pour l'ensemble des droites sur \mathbb{R}^2 est égale à trois. Pour certaines classes de fonctions F on peut calculer explicitement la valeur de dim VC sinon on la majore. Le théorème suivant donne la dim VC pour la classe des hyperplans définis dans \mathbb{R}^p [31].

Théorème 1 : (Vapnik 1998) : Soit F l'ensemble des hyperplans dans \mathbb{R}^p ,

$$F = \{x \rightarrow \text{sign}(\langle w \cdot x \rangle + b), w \in \mathbb{R}^p, b \in \mathbb{R}\}. \quad (\text{II.2})$$

La dimension de VC pour F est égale à $p+1$.

La minimisation du risque structurel (MRS) consiste à choisir la classe F_i (et la fonction f_i) de sorte à ce qu'une borne supérieure de l'erreur de généralisation puisse être minimisée grâce au théorème suivant :

Théorème 2 : Soient h la dimension de VC de la classe de fonctions F , $R_{emp}[f]$ le risque empirique définit par l'équation :

$$R_{emp}[f] = \frac{1}{N_x} \sum_{i=1}^{N_x} L[f(x_i), y_i] \quad (\text{II.3})$$

Avec la fonction perte 0/1 (i.e. $L[f(x), y] = H(-y f(x))$), où H désigne la fonction de Heaviside. Pour tout $\delta > 0$ et $f \in F$, l'inégalité bornant le risque :

$$R[f] \leq R_{emp}[f] + \sqrt{\frac{h(\ln \frac{2N_x}{h} + 1) - \ln(\delta/4)}{N_x}} \quad (\text{II.4})$$

Et varie avec une probabilité d'au moins de $(1-\delta)$ pour $N_x > h$. Tel que N_x est la taille de l'échantillon.

Cette borne n'est qu'un exemple et des formulations du même type ont été démontrées pour d'autres fonctions perte et d'autres mesures de complexité. Le but recherché ici est de minimiser l'erreur de généralisation $R[f]$ en obtenant un faible risque empirique $R_{emp}[f]$ tout en gardant la plus petite classe de fonctions possible [31].

L'inégalité (II.4) fait apparaître deux cas extrêmes:

- Une très petite classe de fonctions (par exemple F_l) fait décroître rapidement le terme de complexité (celui en racine carrée), mais le risque empirique demeure grand.
- Une très grande classe de fonctions (par exemple F_k) implique un risque empirique petit, mais le terme de complexité explose.

La courbe croissante, appelée confiance, correspond à la borne supérieure du terme de complexité. Les comportements du terme de complexité et de l'erreur empirique sont clairement opposés. On recherche donc le meilleur compromis entre complexité et erreur empirique. (Figure II.3)

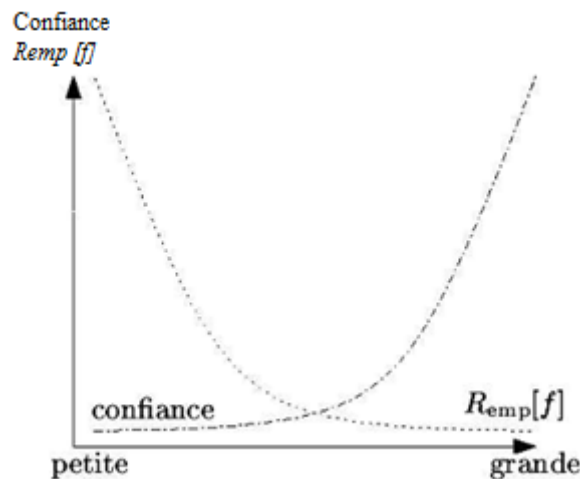


Figure II.3: Illustration de l'inégalité.

Les SVM fournissent une approche très intéressante de l'approximation statistique. Souvent, le nombre des exemples pour l'apprentissage est insuffisant pour que les estimateurs fournissent un modèle avec une bonne précision. D'un autre côté, l'acquisition d'un grand nombre d'exemples s'avère être souvent très coûteuse et peut même mener à des problèmes de sur-apprentissage dans le cas où la capacité du modèle est très complexe. Pour ces deux raisons, il faut arriver à un compromis entre la taille des échantillons et la précision recherchée. Dans ces cas spécifiques comme la reconnaissance de formes, il serait intéressant de trouver une mesure de la fiabilité de l'apprentissage, et d'avoir une mesure du taux d'erreur qui sera commis durant la phase de test. Ces nouvelles techniques unifient deux théories : Minimisation du risque empirique et Capacité d'apprentissage d'une famille de fonctions. C'est la Minimisation du Risque Structurel (MRS) [36].

Rappelons que dans tout problème par apprentissage, nous cherchons à classer correctement les futures observations grâce à la connaissance apprise.

On cherche à définir la loi de probabilité $P(x, y)$ qui, pour tout x (observations) donné, donne la probabilité qu'il appartienne à la classe d'étiquette y . Pour cela, il faut définir une fonction $f(x, \alpha)$ telle que $x \rightarrow f(x, \alpha)$. α est un ensemble de paramètres dépendant du problème et permettant une réponse optimale.

L'erreur de classification donnée par une telle machine est appelée le "risque attendu" ou simplement le "risque" et est définie par :

$$R(\alpha) = \frac{1}{2} \int |y - f(x, \alpha)| dP(x, \alpha) \quad (\text{II.5})$$

On remarque que si la densité de probabilité $P(x, \alpha)$ existe, alors la quantité $dP(x, \alpha)$ peut s'écrire $dP(x, \alpha) = P(x, \alpha) dx dy$.

On peut alors définir la quantité qui nous intéresse, appelée le "risque empirique" et qui est la moyenne des erreurs de mauvaise classification :

$$R_{emp}(\alpha) = \frac{1}{2p} \sum_{i=1}^n |y_i - f(x_i, \alpha)| \quad (\text{II.6})$$

Où n est le nombre d'échantillons. Nous remarquons qu'elle ne fait pas intervenir de distribution de probabilité et que $R_{emp}(\alpha)$ est un scalaire donné pour un choix particulier de α et pour un ensemble de couples $\{x_i, y_i\}$ donnés [7].

II.5. Principe de la technique SVM:

Le principe des séparateurs à vaste marge (SVM) consiste à projeter les données de l'espace d'entrée (appartenant à deux classes différentes) non-linéairement séparables dans un espace de plus grande dimension appelé espace de caractéristiques de façon à ce que les données deviennent linéairement séparables. Dans cet espace, on construit un hyperplan optimal séparant les classes tel que :

- Les vecteurs appartenant aux différentes classes se trouvent de différents côtés de l'hyperplan.
- La plus petite distance entre les vecteurs et l'hyperplan (la marge) soit maximale.

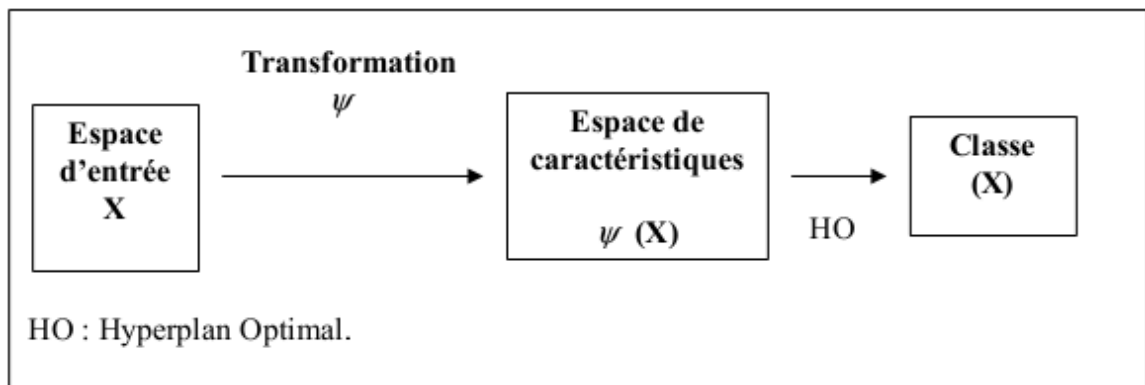


Figure II.4: principe de la technique SVM.

La méthode cherche alors l'hyperplan qui sépare les deux classes, en garantissant que la marge entre les classes soit maximale. Cela garantit une généralisation du principe car

de nouveaux exemples pourront ne pas être trop similaires à ceux utilisés pour trouver l'hyperplan mais être situés d'un côté ou l'autre de la frontière. L'intérêt de cette méthode est la sélection de vecteurs supports qui représentent les vecteurs discriminant grâce auxquels est déterminé l'hyperplan. Les exemples utilisés lors de la recherche de l'hyperplan ne sont alors plus utiles et seuls ces vecteurs supports sont utilisés pour classer un nouveau cas, ce qui peut être considéré comme un avantage pour cette méthode.

Exemple : Pour avoir une idée plus claire sur les SVM, voici un exemple inspiré du travail de B.Schölkopf [32] qui met en pratique le principe des SVM. Dans cet exemple, les données non-linéairement séparables dans \mathbb{R}^2 deviennent linéairement séparables dans \mathbb{R}^3 grâce à la transformation Ψ définie par :

$$\begin{aligned} \Psi : \mathbb{R}^2 &\rightarrow \mathbb{R}^3 \\ (x_1, x_2) &\rightarrow (x_1^2, \sqrt{2} x_1 x_2, x_2^2) \end{aligned} \quad (\text{II.7})$$

La Figure II.5 représente une simulation de cette transformation réalisée par J. Kharroubi [24] avec Matlab. Les données de la figure II.5.a ont été tirées aléatoirement dans \mathbb{R}^2 et la Figure II.5.b représente l'image de ces données dans \mathbb{R}^3 suivant la transformation Ψ .

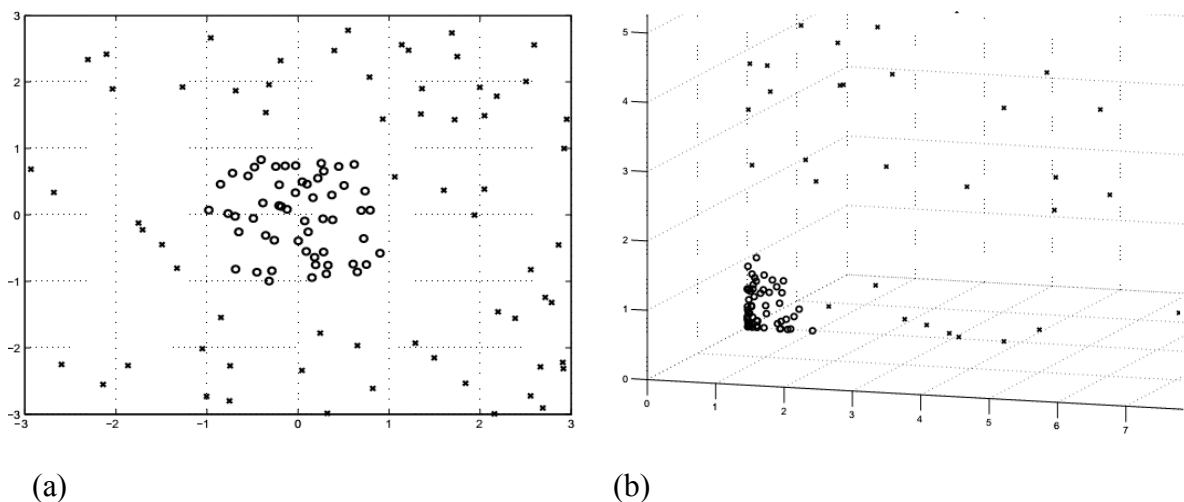


Figure II.5: Exemple montrant l'efficacité d'une transformation dans un espace de plus grande dimension pour faciliter le classement.

(a): exemples tirés aléatoirement dans \mathbb{R}^2 appartenant à deux classes non-linéairement séparables.

(b): l'image des exemples de la figure (a) dans \mathbb{R}^3 en utilisant la transformation Ψ .

II.6. Les SVM binaires:

II.6.1. Notion de base :

II.6.1.1. Hyperplan :

Plaçons-nous dans le cas d'une classification binaire (i.e. les exemples à classifier réparties en 2 classes). On appelle hyperplan séparateur un hyperplan qui sépare les deux classes figure II.6, en particulier il sépare leurs points d'apprentissage. Comme il n'est en générale pas possible d'en trouver un, on se contentera donc de chercher un hyperplan discriminant qui est une approximation au sens d'un critère à fixer (maximiser la distance entre ces deux classes) [26].

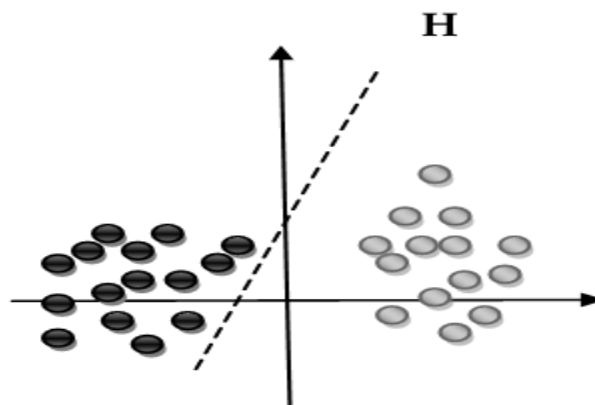


Figure II.6: l'hyperplan H qui sépare les deux ensembles de points.

II.6.1.2. Vecteur de support :

Une tâche de détermination de l'hyperplan séparable des SVM est d'utiliser seulement les points les plus proches (i.e. les points de la frontière entre les deux classes des données) parmi l'ensemble total d'apprentissage, ces points sont appelés vecteurs de support (figure II.7) [27].

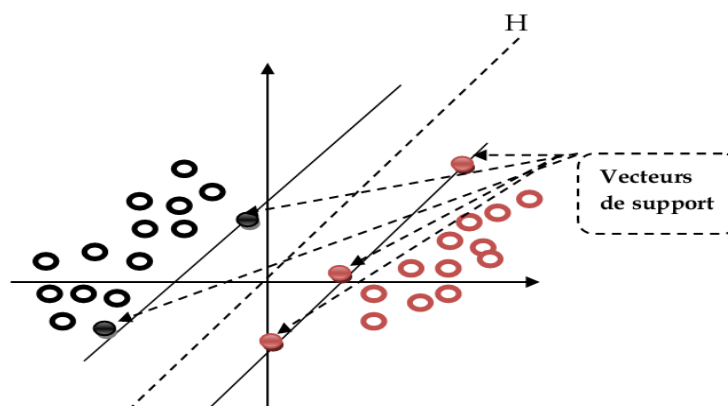


Figure II.7 : Les vecteurs de support.

II.6.1.3. 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 exemples d'apprentissage (i.e. la distance entre l'hyperplan et les vecteurs de support), cette distance est appelée la marge. (Figure II.8) [27].

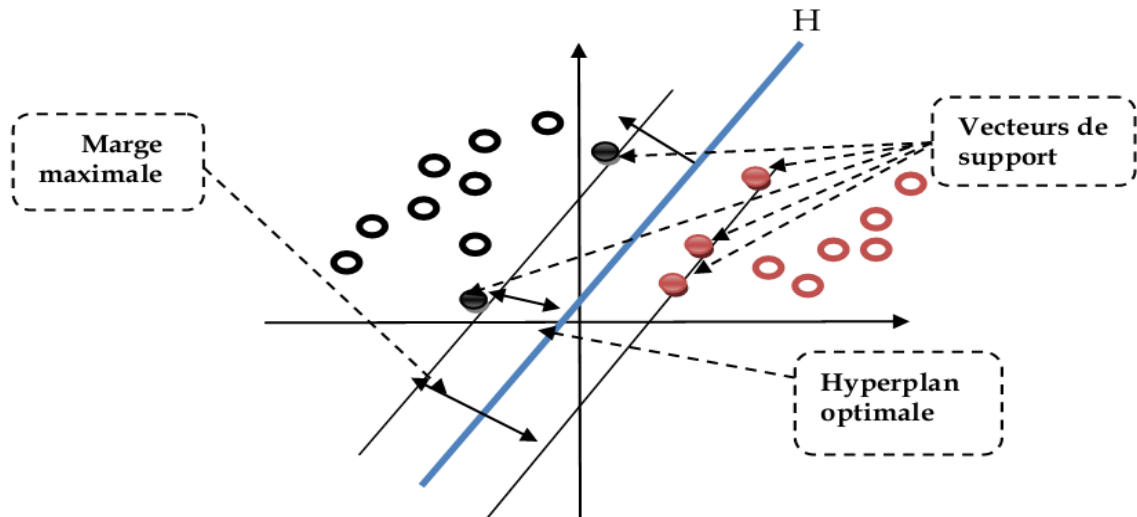


Figure II.8: hyperplan optimal, vecteurs de support et marge maximale.

II.6.2. Construction de l'hyperplan optimal:

Les machines à vecteurs de support forment une classe d'algorithmes d'apprentissage supervisés. Nous nous intéressons à une fonction notée f qui à toute entrée x fait correspondre une sortie $y = f(x)$. Le but est d'essayer d'apprendre f à partir d'un ensemble de couple $(x_i; y_i)$. Dans ce problème les machines à vecteurs support vont être utilisées pour classifier une nouvelle observation x en se limitant à deux classes $y \in \{-1, 1\}$.

II.6.2.1. Cas des données linéairement séparables:

On considère une base d'apprentissage B de m objets à deux classes linéairement séparables. Chaque objet est alors décrit par le couple $\{x_i, y_i\}$, $x_i \in \mathbb{R}^n$, $y_i \in \{-1, 1\}$ où x_i est le vecteur décrivant chaque objet dans l'espace de description de dimension n et y_i l'étiquette de la classe de cet objet [7].

Un classifieur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire. On peut exprimer une telle fonction par $w \in \mathbb{R}^n$ et $b \in \mathbb{R}$ tel que :

$$h(x) = \langle w, x \rangle + b = \sum_{i=1}^n w_i x_i + b \quad i \in [1 \dots m]. \quad (\text{II.8})$$

Dans cette partie, nous présentons la méthode générale de construction de l'Hyperplan Optimal (HO) qui sépare des données appartenant à deux classes différentes linéairement séparables. La figure II.9 donne une représentation visuelle de l'HO dans le cas des données linéairement séparables.

On suppose qu'il existe un hyperplan séparant les points de chaque classe, l'équation d'un tel hyperplan s'écrit alors : $\langle w, x \rangle + b = 0$ (II.9)

Où w est une normale à l'hyperplan, x est la donnée à classer, $b \in \mathbb{R}$.

Il existe une infinité d'hyperplans respectant cette équation, il faut trouver le meilleur. La notion de meilleur hyperplan est introduite en considérant deux hyperplans parallèles (H_1 et H_2) passant chacun par les points les plus proches de la frontière des classes qu'ils limitent et n'ont aucun point de la base d'apprentissage B entre eux (figure II.9). Leurs équations respectives sont :

$$H_1 : \langle w, x \rangle + b = 1 \quad \text{et} \quad H_2 : \langle w, x \rangle + b = -1 \quad (\text{II.10})$$

Pour garantir qu'il n'y a aucun point de B entre eux, tout point de B doit satisfaire l'inégalité suivante :

$$y_i \cdot (\langle w, x_i \rangle + b) \geq 1 \quad i \in [1 \dots m]. \quad (\text{II.11})$$

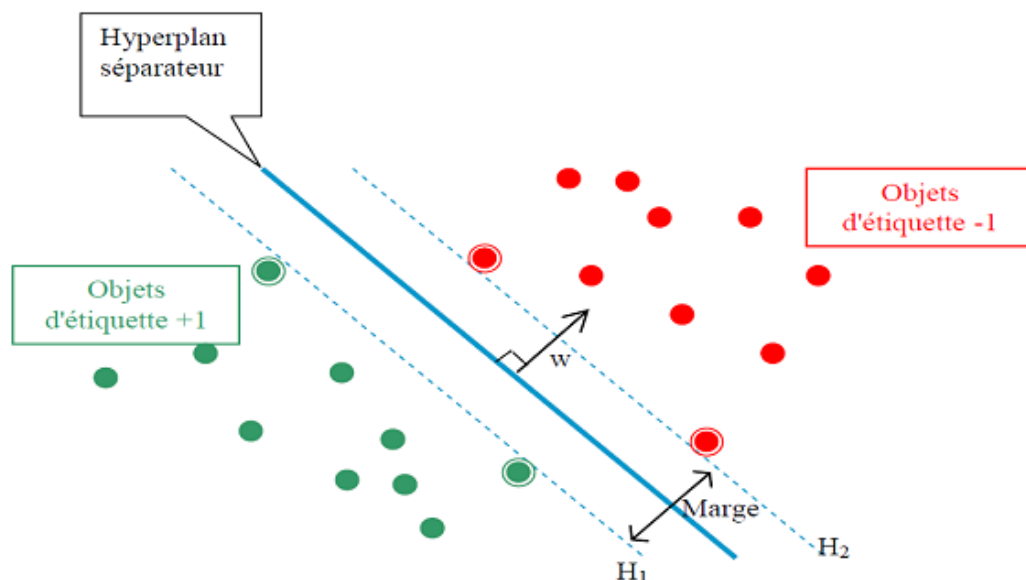


Figure II.9: les deux hyperplans qui séparent les deux classes des éléments.

Comme nous l'avons déjà mentionné, un HO est un hyperplan qui maximise la marge M

qui représente la plus petite distance entre les différentes données des deux classes et l'hyperplan. Maximiser la marge M est équivalent à maximiser la somme des distances des deux classes par rapport à l'hyperplan. Ainsi, la marge a l'expression mathématique suivante :

$$\begin{aligned}
 M &= \min_{x_i|y_i=1} \frac{\omega \cdot x + b}{\|\omega\|} - \max_{x_i|y_i=-1} \frac{\omega \cdot x + b}{\|\omega\|} \\
 &= \frac{1}{\|\omega\|} - \frac{-1}{\|\omega\|} = \frac{2}{\|\omega\|}
 \end{aligned} \tag{II.12}$$

Trouver l'hyperplan optimal revient donc à maximiser $\frac{2}{\|\omega\|}$. Ce qui est équivalent à minimiser $\|\omega\|$. Ceci est un problème de minimisation d'une fonction objective quadratique avec contraintes linéaires. Pour cela plusieurs méthodes ont été proposées :

Principe de Fermat (1638) : Les points qui minimisent où maximisent une fonction dérivable annule sa dérivée. Ils sont appelés points stationnaires.

Principe de Lagrange (1788) : Pour résoudre un problème d'optimisation sous contrainte, il suffit de rechercher un point stationnaire z_0 du lagrangien $L(z, \alpha)$ de la fonction g à optimiser et les fonctions C_i^g exprimant les contraintes.

$$L(z, \alpha) = g(z) + \sum_{i=1}^m \alpha_i C_i^g(z) \tag{II.13}$$

où les α_i sont des constantes appelées coefficients de Lagrange. $i \in [1 \dots m]$.

Principe de Kuhn-Tucker (1951) [24]: Avec des fonctions g et C convexes, il est toujours possible de trouver un point-selle (z_0, α^*) qui vérifie :

$$\min_z L(z, \alpha^*) = L(z_0, \alpha^*) = \max_{\alpha \geq 0} L(z_0, \alpha) \tag{II.14}$$

En appliquant le principe de Kuhn-Tucker, on est amené à rechercher un point-selle (w_0, b_0, α^0) . Le lagrangien correspondant à notre problème est :

$$L(w, b, \alpha) = \frac{1}{2} w \cdot w - \sum_{i=1}^m \alpha_i [y_i [(x_i \cdot w) + b] - 1] \tag{II.15}$$

Le lagrangien doit être minimal par rapport à w et b et maximal par rapport à $\alpha \geq 0$.

- $L(w, b, \alpha)$ est minimal par rapport à b :

$$\frac{\partial L(w, b, \alpha)}{\partial b} = 0 \Leftrightarrow \sum_{i=1}^m \alpha_i y_i = 0 \tag{II.16}$$

- $L(w, b, \alpha)$ est minimal par rapport à w :

$$\frac{\partial L(w, b, \alpha)}{\partial w} = 0 \Leftrightarrow w - \sum_{i=1}^m \alpha_i x_i y_i = 0 \tag{II.17}$$

On remplace (II.16) et (II.17) dans (II.15), on aura :

$$L(w, b, \alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \quad (\text{II.18})$$

$L(w, b, \alpha)$ est alors maximisé selon les $\alpha_i \geq 0$ en respectant les contraintes précédentes.

D'après le théorème de Kuhn-Tucker une condition nécessaire et suffisante pour que la solution α^0 soit optimal est: $\alpha_i^0 [y_i [(x_i \cdot w) + b] - 1] = 0$ pour $i = 1, \dots, m$ (II.19)

Ainsi on peut calculer w_0 et b_0 :

$$w_0 = \sum_{VS} \alpha_i^0 y_i x_i \quad (\text{II.20})$$

$$b_0 = \frac{1}{2} [(w_0 \cdot x^*(1))] + [(w_0 \cdot x^*(-1))] \quad (\text{II.21})$$

Tel que VS représente tout vecteur support x_i tel que $y_i [(x_i \cdot w) + b] = 1$

$$\text{Donc : } VS = \{x_i | \alpha_i > 0\} \quad (\text{II.22})$$

La fonction de classement **class** (x) est défini par:

$$\begin{cases} \text{class}(x) = \text{sign}[(w_0 \cdot x) + b_0] \\ = \text{sign}[\sum_{x \in VS} \alpha_i^0 y_i (x_i \cdot x) + b_0] \end{cases} \quad (\text{II.23})$$

Si **class**(x) est négative, x est de classe -1, si elle est positive, x est de classe 1 [24, 33].

II.6.2.2. Cas des données non-linéairement séparables:

Les machines à vecteurs supports présentées ci-dessus ne concernent que les modèles linéaires. Il serait complètement illusoire de se dire que l'on pourrait séparer efficacement n'importe quel jeu de données par un simple hyperplan. Si par exemple les données des deux classes se chevauchent sévèrement comme dans la figure II.10 de gauche, aucun hyperplan séparateur ne sera satisfaisant [7].

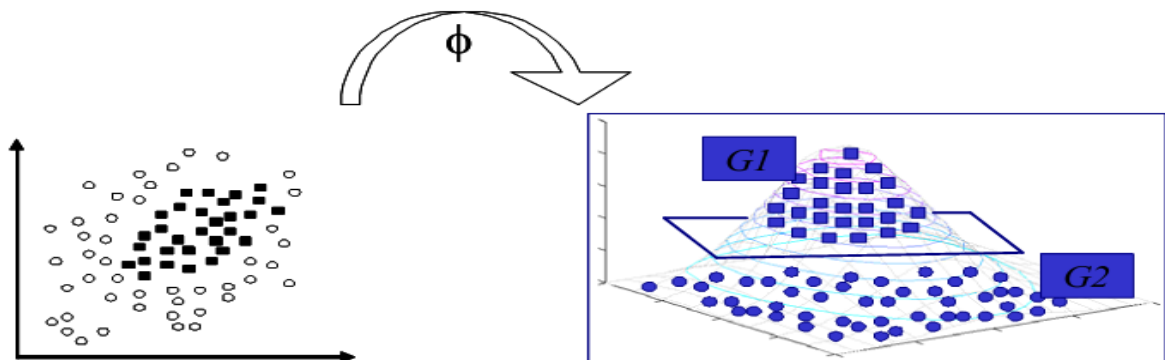


Figure II.10 : Exemple de plongement non-linéaire.

Dans ce cas où les données sont non-linéairement séparables (Figure II.10), l'hyperplan optimal est celui qui satisfait les conditions suivantes :

- La distance entre les vecteurs bien classés et l'hyperplan optimal doit être maximale.
- La distance entre les vecteurs mal classés et l'hyperplan optimal doit être minimale.

Pour formaliser tout cela, on introduit la notion de marge souple (soft margin), qui autorise un certain nombre d'éléments d'une classe à se trouver du mauvais côté de la frontière lors de l'apprentissage (Figure II.11). Ces éléments sont appelés des «outliers» [24, 36].

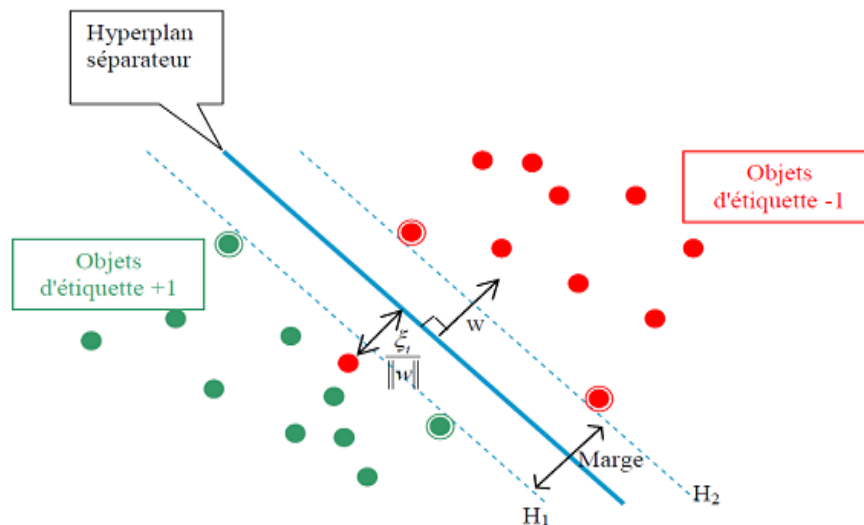


Figure II.11 : Principe de "l'outilier".

Certains éléments sont "autorisés" à se retrouver du mauvais côté de la frontière lors de la phase d'apprentissage.

On applique donc une marge d'erreur ξ_i à l'équation II.11 pour obtenir :

$$y_i \cdot (\langle w, x_i \rangle + b) \geq 1 - \xi_i \quad \text{pour } i \in [1 \dots m] \quad (\text{II.24})$$

Pour maximiser la marge, il ne suffit plus de minimiser $\|w\|$ mais la fonction (II.25) :

$$\Psi(\omega, \xi) = \frac{1}{2} \omega \cdot \omega + C \sum_{i=1}^m \xi_i \quad (\text{II.25})$$

Où C est un paramètre de régularisation. Elle permet de concéder moins d'importance aux erreurs. Cela mène à un problème dual légèrement différent de celui du cas des données linéairement séparables. Maximiser le lagrangien donné par l'équation (II.18) par rapport à α_i sous les contraintes suivantes :

$$\sum_{i=1}^m \alpha_i y_i = 0 \quad \text{avec } 0 \leq \alpha_i \leq C \quad \text{pour } i = 1 \dots m$$

Le calcul de la fonction de classification $f(x)$ reste exactement le même que pour le cas des données linéairement séparables.

La frontière définie est là aussi linéaire mais décalée en fonction du coût fixé. L'erreur due à ce type de séparation est souvent plus importante que l'erreur générée par une frontière non linéaire, c'est pourquoi une modélisation non linéaire a été intégrée, toujours en vue de minimiser le risque empirique. Cette méthode dite « astuce du noyau » (kernel trick) a été introduite par Aizerman en 1964 [41].

II.6.3. L'astuce du noyau (kernel trick):

Les équations de base font intervenir les données d'apprentissage (les échantillons et leur classe) sous la forme d'un produit scalaire. Il a été démontré qu'il est possible que les données initiales (de l'espace \mathbb{R}^d) soient linéairement séparables dans un espace euclidien $H = \mathbb{R}^e$ (avec $e > d$ fini ou non), dans lequel on projette l'espace de description \mathbb{R}^d à l'aide d'une fonction ϕ telle que :

$$\phi: \mathbb{R}^d \rightarrow \mathbb{R}^e, \phi(x) = \tilde{x}$$

Où x est un exemple estimé.

En effet, plus la dimension de l'espace de description est grande plus la probabilité de pouvoir trouver un hyperplan séparateur entre les classes est élevée. En transformant l'espace d'entrée en un espace de redescription H de très grande dimension, éventuellement infinie, il devient donc possible d'envisager d'utiliser la méthode de résolution quadratique précédente dans le cas de classes non linéairement séparables [25].

Si ces espaces sont munis d'un produit scalaire, on peut définir sur \mathbb{R}^d une fonction noyau telle que : $K: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^e \times \mathbb{R}^e, K(x, y) = \langle \phi(x), \phi(y) \rangle = \langle \tilde{x}, \tilde{y} \rangle$ (II.26)

Mais de telle sorte qu'on n'a jamais besoin de produire explicitement les $\phi(x)$.

L'algorithme des SVM est alors le même que dans le cas linéaire : il suffit de remplacer tous les produits scalaires par une fonction noyau non linéaire. Avec cette astuce, les SVM peuvent définir le meilleur hyperplan séparateur linéaire dans le nouvel espace \mathbb{R}^e en utilisant la fonction noyau appropriée.

En appliquant la technique au noyau à la fonction de Lagrange, on obtient la fonction suivante :

$$L(\omega, b, \alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (\text{II.27})$$

Et la fonction de décision devient donc :

$$y = \text{sign} \left(\sum_{i \in \text{SV}} \alpha_i y_i K(x_i, x_j) + b \right) \quad (\text{II.28})$$

Où SV est l'ensemble des indices des vecteurs de support issus de la phase d'apprentissage.

Toute fonction satisfaisant au critère de Mercer (qui se résumant essentiellement par être définie positive) peut être utilisée comme noyau [23]. Ce critère est le suivant :

Soit K une fonction symétrique, on définit alors la matrice $(K(x_i, x_j))_{i,j}$.

Si cette matrice est définie positive alors il existe une fonction ϕ tels que :

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (\text{II.29})$$

II.6.3.1. Exemples de noyaux :

Une famille de ces fonctions noyaux qui sont très appropriées aux besoins des SVM peut être définie, en voici les plus utilisés [23, 25]:

Noyau polynomial d'ordre d :

$$K(u, v) = (\langle u, v \rangle + 1)^d \quad (\text{II.30})$$

La dimension de l'espace transformé induit par un noyau polynomial est de l'ordre $\frac{(p+d)!}{p!d!}$ où p est la dimension de l'espace de départ.

Par exemple, il peut être montré que la fonction noyau polynomiale: $K(x, y) = \langle x, y \rangle^n$ réalise implicitement un produit scalaire dans l'espace des descripteurs correspondant à tous les produits d'exactly n dimensions. Ainsi pour $n = 2$ et $x, y \in \mathbb{R}^2$, on a :

$(x \cdot y)^2 = (x_1 y_1 + x_2 y_2)^2 = \langle \Phi(x), \Phi(y) \rangle$ qui correspond au changement de description par la fonction :

$$\Phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1 x_2)^T \in \mathbb{R}^3 \quad (\text{II.31})$$

Pour $n=3$ et $d=2$, un espace transformé de dimension 5 :

$$\Phi(x) = (x_1^3, \sqrt{3}x_1^2 x_2, x_1 x_2^2, \sqrt{3}x_1 x_2^2, x_2^3)^T \quad (\text{II.32})$$

On peut vérifier que : $(x \cdot y)^3 = \langle \Phi(x), \Phi(y) \rangle$

Pour $n=2$ et $d=2$, on peut aussi transformer autrement en dimension 3, ou en dimension 4 :

$$\Phi(\mathbf{x}) = ((x_1^2 - x_2^2), 2x_1x_2, (x_1^2 + x_2^2))^T$$

$$\Phi(x) = (x_1^2, x_1x_2, x_1x_2, x_2^2)^T$$

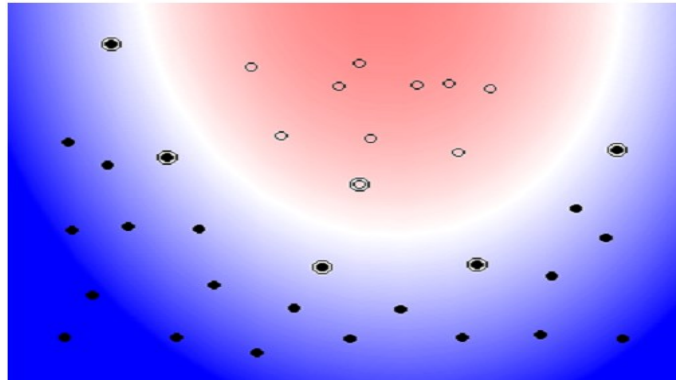


Figure II.12 : un modèle SVM avec fonction noyau polynomial d'ordre 2.

La fonction noyau sigmoïde : $K(u, v) = \tanh(u * v + c)$ (II.33)

La fonction noyau linéaire : $K(u, v) = u \cdot v$ (II.34)

La fonction noyau gaussien à Base Radiale RBF de largeur de bande σ :

$$K(u, v) = \exp\left(-\frac{\|u-v\|^2}{2\sigma^2}\right)$$
 (II.35)

Le paramètre σ permet de régler la largeur de la gaussienne. En prenant un σ grand, la similarité d'un exemple par rapport à ceux qui l'entourent sera assez élevée, alors qu'on prenant un σ tendant vers 0, l'exemple ne sera similaire à aucun autre.

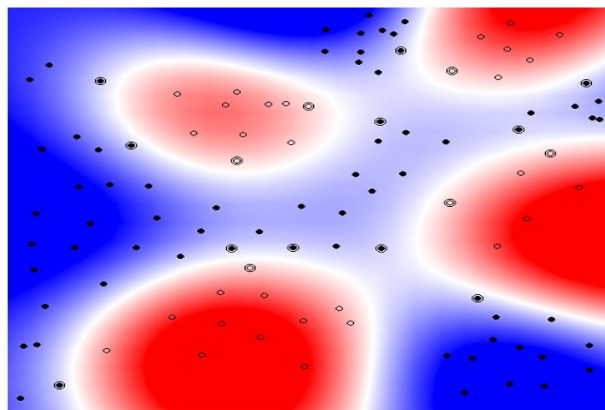


Figure II.13 : un modèle SVM avec la fonction noyau RBF.

II.7. SVM pour le cas multi-classes:

Nous abordons maintenant la question de l'adaptation pour un problème multi-classes, des Machines à Vecteurs de Support, originellement conçues sur un paradigme de

discrimination binaire. Nous emploierons par la suite le terme multi-classes pour désigner une classification impliquant plus de deux classes. Ce problème est antérieur à la création des SVM puisqu'il est légitimement posé par toute méthode discriminative, en particulier par la séparation par hyperplan linéaire. Ainsi, de nombreuses méthodes ont été proposées qui permettent de combiner les résultats de classifieurs binaires pour formuler une réponse multi-classes.

Les SVM multi-classes sont des modèles de l'apprentissage de conception relativement récente, dont l'étude est actuellement en plein essor. Cette situation nouvelle fait naître un besoin, celui de disposer d'une étude synthétique sur les SVM multi-classes, ou plus généralement l'utilisation de SVM pour la discrimination à catégories multiples.

II.7.1. Méthodes de décomposition:

Il existe deux stratégies pour faire la classification multi-classes :

II.7.1.1. La stratégie un contre tous:

L'architecture la plus ancienne, utilisée pour les machines à vecteurs supports multi-classes, est probablement la méthode un contre tous. Chaque SVM de la stratégie «un contre tous» étant entraîné à séparer les données d'une classe de celles de toutes les autres classes, chaque sortie est directement liée à la probabilité d'appartenance à l'une des classes [8]. Le problème consiste donc à calibrer les sorties de l'ensemble des SVM. Cette méthode construit m classifieurs binaires à vecteurs de supports où m est le nombre total des classes. L'apprentissage du $k^{\text{ème}}$ classifieur à vecteurs de supports s'effectue en considérant tous les exemples de la $k^{\text{ème}}$ classe dans la région positive et tous les autres exemples dans la région négative. Le $k^{\text{ème}}$ classifieur à vecteurs supports s'obtient en résolvant le problème

$$\begin{aligned} & \text{Minimiser } \frac{\|w_k\|^2}{2} + C \sum_{j=1}^l \xi_j^k \\ & \text{sous } \quad < w_k \cdot \phi(x_j) > + b_k \geq 1 - \xi_j^k \quad \text{si } y_i = k, \\ & \quad < w_k \cdot \phi(x_j) > + b_k \leq -1 + \xi_j^k \quad \text{à si } y_i \neq k, \\ & \quad \xi_j^k \geq 0 \quad j = 1, 2, \dots \dots \dots l. \end{aligned}$$

Où les $\phi(x_s)$ sont les transformées des x_s dans l'espace induit par la fonction ϕ et C est le paramètre de pénalité.

La résolution du problème (II.36) pour chaque valeur de $k \in \{1, 2, \dots, m\}$ donne lieu à m fonctions de décision : $f_k(x) = \langle w_k \cdot \varphi(x) \rangle + b_k$ (II.36)

Pratiquement, nous résolvons le problème dual correspondant au problème (II.33) ayant exactement l variables duales. En total, nous aurons à résoudre m problèmes quadratiques chacun à l variables. Ainsi, le temps d'apprentissage de cette méthode croît linéairement en fonction de m [8].

Une nouvelle observation x sera donc affectée à la classe k^* selon la règle de décision discrète : $k^* = \text{Arg} \max_{1 \leq k \leq m} h_k(x)$ (II.37)

Dans le cas multi-classes ($m > 2$), cette égalité peut être satisfaite pour plus que deux classes. Dans ce cas, l'observation x est dite non-classifiable. Toutes les observations x non-classifiables forment la région d'ambiguïté dite aussi région non-classifiable. Cette région est schématisée dans la Figure II.14.

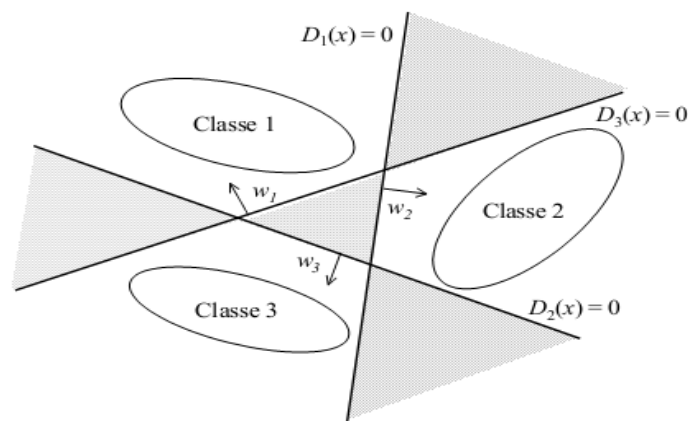


Figure II.14: L'espace hachuré représente la région d'ambiguïté pour l'approche « un contre tous » suite à la prise de décision discrète.

Afin de pouvoir classer une observation x tombant dans la région d'ambiguïté, la règle de décision continue a été utilisée. Cette règle est donnée par :

$$k^* = \text{Arg} \max_{1 \leq k \leq m} f_k(x) \quad (\text{II.38})$$

Géométriquement interprétée, tout nouvel exemple x est affecté à la classe qui correspond à l'hyperplan le plus éloigné. Ainsi, l'espace des variables explicatives X est subdivisé en m régions convexes, chacune correspondant à une classe. La figure II.14 donne un exemple de subdivision de l'espace X [8].

Cette approche est nommée « le gagnant emporte le tout ». L'inconvénient majeur de cette heuristique est qu'elle ne conserve pas les m frontières de séparation. La figure II.15

montre ce propos. Il est clair que cette heuristique a amélioré la règle de décision discrète, en revanche, elle perd totalement les capacités de généralisation des m hyperplans construits. Malheureusement, on ne dispose pas de borne pour l'erreur de généralisation de l'approche un-contre-tous [8].

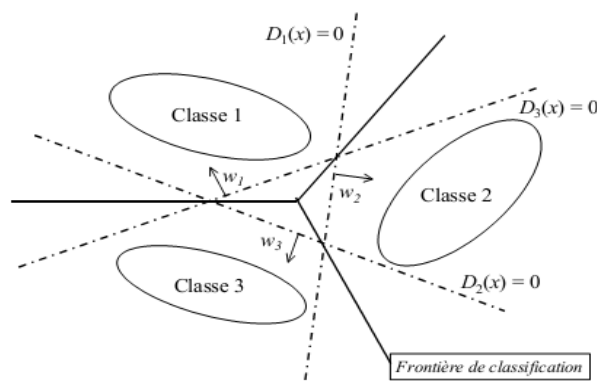


Figure II.15 : Règle de décision continue: les bissectrices des secteurs d'ambiguïté forment la nouvelle frontière de classification.

II.7.1.2. La stratégie un contre un:

Cette stratégie de décomposition a été adoptée par S. Knerret et al. [39] et utilisée pour la première fois dans le contexte des machines à vecteurs supports par Krebel [38]. Il consiste à construire $\frac{m(m-1)}{2}$ prédicteurs binaires chacun séparant uniquement deux classes tout en ignorant les autres. Chaque classifieur est donc entraîné à séparer les données d'une classe de celles d'une autre classe. Cette stratégie s'appelle « un contre un » ou en anglais « One Against One, (OAO) ». Notons par ailleurs, que cette stratégie peut aussi être rencontrée dans la littérature sous les noms de « pairwise coupling » [42, 43], « all pairs » [44] ou encore « round robin » [45].

L'hyperplan séparateur des classes k et s est la solution du problème d'optimisation suivant :

$$\text{Minimiser } \frac{\|w^{ks}\|^2}{2} + C \sum_{j=1}^l \xi_t^{ks}$$

$$\text{sous } \langle w^{ks}, \phi(x_t) \rangle + b^{ks} \geq 1 - \xi_t^{ks} \quad \text{si } y_t = k,$$

$$\langle w^{ks}, \phi(x_t) \rangle + b^{ks} \leq -1 + \xi_t^{ks} \quad \text{si } y_t = s,$$

$$\xi_t^{ks} \geq 0 \quad \forall t \in \{1, \dots, l_{ks}\}$$

Où l_{ks} est le nombre des observations issues des classes k et s .

Pratiquement, nous résolvons le problème dual correspondant au problème (II.39) ayant l_{ks} variables duales. Si chaque classe contient en moyenne $\frac{l}{m}$ exemples, nous aurons à

résoudre dans la phase d'apprentissage $\frac{m(m-1)}{2}$ problèmes quadratiques chacun dépendant à peu près de $\frac{2l}{m}$ variables [8].

L'approche un-contre-un consiste donc à construire un classifieur pour chaque paire de classes (k, s) définissant ainsi des fonctions de décision binaire $h_{ks} : X \subseteq \mathbb{R}^p \rightarrow \{-1, +1\}$

$$h_{ks}(x) = \text{sign}(f_{ks}(x)) = \begin{cases} +1 & \text{si } x \in \text{à la classe } k \\ -1 & \text{si } x \in \text{à la classe } s \end{cases} \quad (\text{II.39})$$

Pour des raisons de symétrie $h_{ks} \equiv -h_{sk}$ et on convient que $h_{kk} \equiv 0$ pour tout $k \in \{1, 2, \dots, m\}$. Sur la base des $\frac{m(m-1)}{2}$ fonctions de décision binaires h_{ks} , nous définissons m autres fonctions de décision de la façon suivante :

$$h_k(x) = \sum_{s=1}^m h_{ks}(x) \quad , k = 1, 2, \dots, m \quad (\text{II.40})$$

Et la règle de classification d'une nouvelle observation x est donnée par :

$$k^* = \text{Arg} \max_{1 \leq k \leq m} h_k(x) \quad (\text{II.41})$$

Cette règle connue sous le nom de vote majoritaire a été appliquée pour la première fois avec les SVM par Krebel [38].

Il peut arriver que cette règle soit satisfaite par plus qu'une classe, ainsi, une nouvelle observation x est dite non-classifiable et elle appartient à la région d'ambiguïté. Cette région est présentée par la figure II.16. Toute observation située dans la région d'ambiguïté est classée arbitrairement dans l'une des classes vérifiant cette dernière règle.

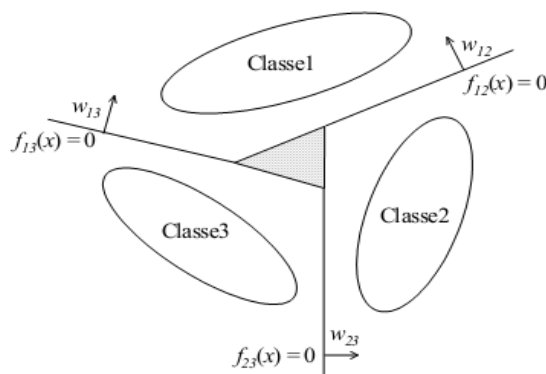


Figure II.16: La région d'ambiguïté hachurée est réduite pour l'approche un-contre-un.

Les avantages majeurs de cette combinaison sont : la conservation de bonnes parties des $\frac{m(m-1)}{2}$ hyperplans préalablement construits et la diminution de la région d'ambiguïté relativement à l'approche une-contre-reste. En revanche, son erreur de généralisation n'a pas

encore de majorant [8].

Plusieurs méthodes ont été proposées pour combiner les différents classifieurs issus de toutes les paires de classes. Chaque architecture vise à réduire le temps d'apprentissage et le temps de classification d'une nouvelle observation tout en améliorant les capacités de généralisation de la machine.

II.8. Avantages et inconvénients des SVM:

L'avantage de la création d'une fonction de décision avec l'algorithme des SVM est que la solution produite correspond à l'optimum d'une fonction convexe. Elle ne possède donc pas plusieurs optima locaux comme pour les réseaux de neurones (dans leur formulation classique), mais un optimum global. Cet optimum correspond à la minimisation du risque structurel et donc à la recherche d'une hypothèse possédant de bonnes capacités de généralisation à partir d'un espace d'hypothèses donné. L'espace d'hypothèses dépend du choix de la fonction noyau, ce qui est à la fois un avantage et un inconvénient. Un avantage car l'algorithme d'apprentissage reste identique. Le problème est que la fonction noyau a souvent des paramètres libres, comme la largeur de bande avec un noyau gaussien, et que la recherche de la valeur optimale de ces paramètres libres ne correspond plus à la recherche du minimum d'une fonction convexe [40].

La recherche des valeurs idéales de ces paramètres, appelée communément sélection de modèles, est critique. Illustrons cet état critique avec l'utilisation d'un noyau gaussien et une largeur de bande σ très faible. La fonction de décision produite, pour ces conditions, sera cohérente avec les données et elle aura autant de vecteurs de support que d'exemples. L'hyperplan sélectionné, bien que de marge géométrique maximum, a une marge faible. Les bornes supérieures vues précédemment sont très élevées et elles ne garantissent plus rien sur le pouvoir de généralisation. Donc la sélection d'une mauvaise fonction noyau (ou de ces paramètres) peut conduire à produire un effet de sur-apprentissage et c'est le principal désavantage des SVM [40].

Un autre désavantage des SVM est la durée importante de la phase d'entraînement, en particulier si une sélection de modèles doit être réalisée, bien que dans les dix dernières années plusieurs optimisations de l'algorithme initial aient été proposées.

Un dernier désavantage est la complexité de la fonction de décision produite lorsque la base d'apprentissage est de taille importante, même si, à la différence de la règle PPV,

l'ensemble des exemples n'est pas utilisé grâce à la notion de vecteur de support [40].

Conclusion :

Dans ce chapitre nous avons présenté les concepts de la méthode SVM, dont nous avons expliqué la démarche de construction de l'hyperplan optimal dans les cas des données linéairement séparables et non linéairement séparables ; en se basant sur les fondements mathématiques pour le cas binaire (à deux classes). Aussi nous avons défini quelques fonctions noyaux utilisés par les SVM non linéaires pour la transformation et présentation des vecteurs d'entrée dans l'espace des caractéristiques.

Nous avons ensuite décrit la démarche suivie pour généraliser cette technique pour le cas de multi-classes à partir du cas binaire en utilisant les trois approches : la discrimination binaire, Un-contre-Un ou Un-contre-Tous.

L'intérêt principal de la méthode SVM est sa facilité d'emploi, la mise en œuvre des algorithmes des SVM est peu coûteuse, mais le point noir est la durée des phases de tests qui sont assez longues lorsqu'il s'agit de régler les paramètres de la fonction noyau. Parmi d'autres avantages des SVM on cite : qu'ils possèdent des fondements théoriques solides, une optimisation très simple, et un temps de calcul court.

Les SVM peuvent également s'utiliser pour des tâches de prédiction des variables continues en fonction d'autres variables (température, évolution des marchés, etc.). Le champ d'application des SVM est donc large.

Chapitre III :

Application des SVM pour la classification de la sévérité de pollution



Introduction :

La transformée en ondelettes a l'avantage de décrire le signal dans le domaine temps / fréquence et donc de palier l'inconvénient de la transformée de Fourier qui a le désavantage de la perte de l'information temporelle. Le pavage du plan temps-fréquence induit par cette transformée a pour particularité de permettre une résolution temporelle fine aux hautes fréquences et une résolution fréquentielle fine aux basses fréquences. Cette propriété permet souvent une analyse intéressante du signal mais reste rigide. Il existe deux types de la transformée en ondelettes : la transformée en ondelettes continue (CWT pour Continuous Wavelet Transform) et la transformée en ondelettes discrète (DWT pour Discrete Wavelet Transform). Chaque transformation a ses avantages et ses inconvénients [50].

Dans la première partie du présent chapitre, nous abordons d'une manière explicite les deux types de la transformée en ondelettes (la transformée en ondelettes continue CWT et discrète DWT). Nous passons ensuite à l'analyse du signal du courant de fuite par la DWT en utilisant l'ondelette de Daubechies 4 comme ondelette analysante afin d'extraire les critères nécessaires à l'étude de l'état de surface des isolateurs pour les deux types de pollution : uniforme et non uniforme.

Dans la seconde partie, nous allons d'abord présenter les différentes classes associées à chaque concentration en kaolin. Ensuite, nous appliquerons les deux stratégies de la méthode SVM pour le cas multi-classe (one versus one et one versus all) afin de classer le degré de sévérité de pollution avec différentes fonctions noyaux pour les deux types de pollution en se basant sur les caractéristiques des signaux du courant de fuite.

III.1. Rappel sur la transformée de Fourier :

La transformée en ondelettes, comme la transformée de Fourier, est une transformation d'un espace d'entrée (espace-temps) vers un autre espace (espace fréquentiel) [51]. Dans le cas de la transformée de Fourier classique, la base de l'espace d'entrée est par définition $\{e(j\omega t), \omega \in R\}$, comme nous le montre la formule exprimant la transformée inverse de Fourier:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \exp(j\omega t) d\omega \quad (\text{III.1})$$

Dans le cas de la transformée de Fourier numérisée ou discrète (Fast Fourier Transform), les variables t et $\omega = 2\pi f$ prennent un nombre fini de valeurs espacées de $T = \Delta t$:

$$f[n] = \frac{1}{N} \sum_{k=0}^N \hat{f}[k] \exp\left(\frac{j2\pi kn}{N}\right) \quad (\text{III.2})$$

$$\text{La base devient discrète : } \left\{ e_k[n] = \exp\left(\frac{j2\pi kn}{N}\right), 0 \leq k \leq N \right\} \quad (\text{III.3})$$

Dans les deux cas, la base de l'espace fréquentiel est constituée de distributions de Dirac $\{\delta(\omega)\}$, la variable ω prenant des valeurs discrètes (FFT) ou variant de manière continue dans \mathbb{R} .

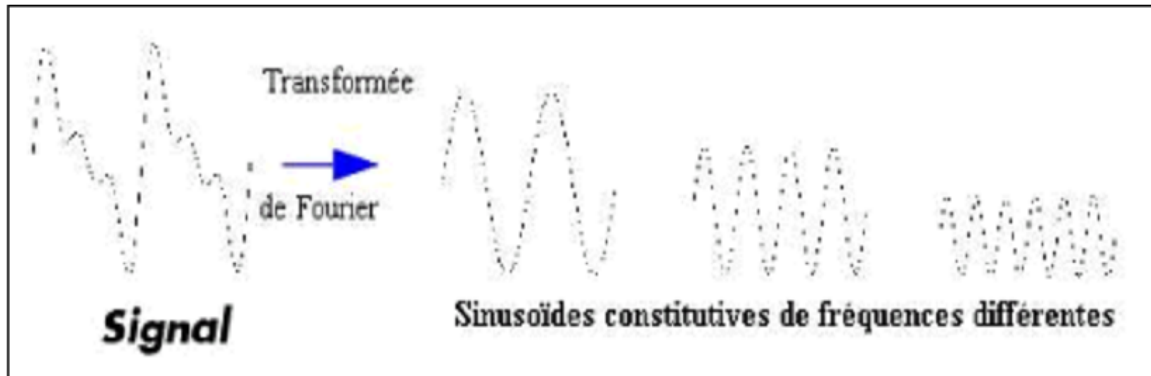


Figure III.1: Principe de la transformée de Fourier.

III.2. La transformée en ondelettes :

La théorie des ondelettes est apparue au début des années 1990. Elle touche de nombreux domaines des mathématiques, notamment le traitement du signal et des images. Dans la transformée en ondelettes, comme dans l'analyse de Fourier, on cherche à transformer un signal quelconque en une série de nombres que l'on pourra ensuite utiliser pour reconstruire au mieux le signal d'origine [52]. Cependant dans la transformée en ondelettes, on utilise plusieurs niveaux de résolution pour examiner le signal et faire ressortir les différentes variations [44].

Le traitement du signal par ondelettes permet d'élargir les intervalles de temps lorsque l'on recherche plus de précision sur les composantes de basses fréquences, elle permet également de rétrécir les zones pour l'extraction d'informations sur les hautes fréquences cela démontre la capacité des ondelettes à garder une bonne résolution temporelle et fréquentielle. L'analyse par la méthode des ondelettes nous permet aussi de partitionner l'énergie du signal en différentes bandes de fréquences pour séparer toutes ses composantes [48].

La transformée en ondelette permet d'appliquer une analyse multi-résolution (AMR) sur le signal étudié. Cette analyse qu'il conviendrait d'appeler temps-échelle, utilise une gamme très étendue d'échelles pour analyser le signal. L'analyse multi-résolution de la transformée

en ondelettes est équivalente à une décomposition atomique temps-échelle. Chacun des atomes peut s'interpréter comme étant une projection locale du signal analysé et il est obtenu à partir d'une ondelette $\Psi(t)$ unique par une translation en temps et une dilatation.

Partant d'une fonction bien localisée; dans le plan temps-échelle, on lui associe la famille d'ondelettes $\Psi_{(b,a)}(t)$ générées par des translations et des dilatations de $\Psi(t)$:

$$\Psi_{(b,a)}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \quad , b \in \mathbb{R}, a > 0 \quad (\text{III.4})$$

Les ondelettes sont de formes d'ondes standards, proportionnelles au paramètre de dilatation a (variable d'échelle). La transformée en ondelettes est aussi interprétée comme étant un processus de filtrage du signal analysé par un filtre passe-bande de bande passante variable. C'est le paramètre a qui fixe la valeur de cette bande [53].

III.2.1. L'ondelette d'analyse:

Un grand nombre d'ondelettes analysantes a été construit. Plusieurs types de ces ondelettes ont été rapportés dans la littérature dont certaines portent le nom de leurs créateurs. Les plus connues sont représentées dans la figure suivante : Daubechies, Gaussienne, Haar, Mexican Hat, Meyer et Morlet [54].

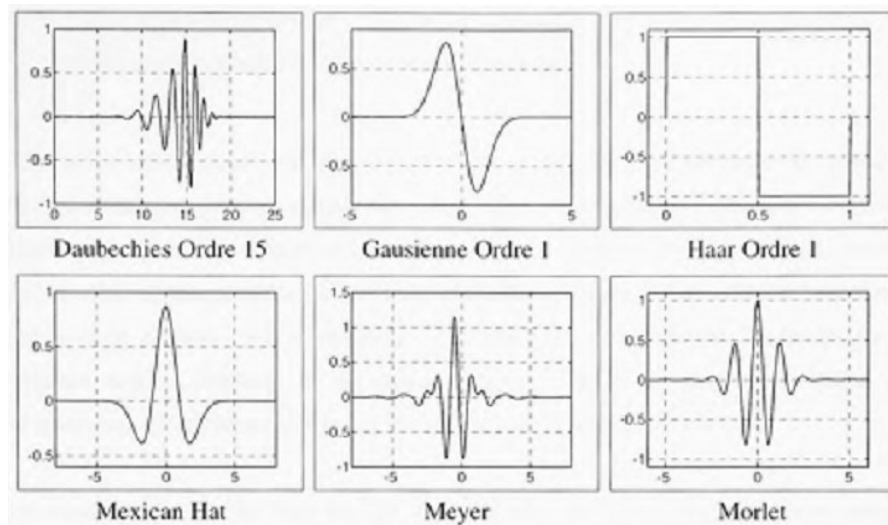


Figure III.2: Forme d'onde des ondelettes couramment utilisées.

Le choix de l'ondelette analysante n'est pas arbitraire. Pour être admissible, l'ondelette $\Psi(t)$ doit respecter les conditions suivantes :

$-\Psi(t)$ ainsi que son carré doivent être intégrables :

$$\int |\Psi(t)| dt < \infty \quad \text{et} \quad \int |\Psi(t)|^2 dt < \infty \quad (\text{III.5})$$

Cette condition implique que l'énergie de l'ondelette doit être finie.

$$- \text{si } \hat{\Psi}(t) \text{ représente la TF de } \Psi(t), \text{ alors : } \int \frac{|\hat{\Psi}(\omega)|^2}{\omega} d\omega < \infty \quad (\text{III.6})$$

Cette condition implique que la moyenne de $\Psi(t)$ soit nulle ($\hat{\Psi}(0) = 0$ où intégrale de $\Psi(t) dt=0$).

III.2.2. La transformée en ondelettes continues (CWT) :

La transformée en ondelette continue (CWT) utilise des ondelettes dont l'enveloppe rétrécit avec l'échelle en gardant le même nombre d'oscillations. Fondamentalement, la transformée en ondelette se calcule de la même manière que la transformée de Fourier fenêtrée, autrement dit on regarde la corrélation au sens du produit scalaire entre le signal à analyser et la famille de fonction $\Psi_{a,b}(t)$ [62, 64, 65].

Une ondelette est une onde qui oscille à fréquence constante et qui s'annule au bout d'un instant donné. La fonction analysante peut être choisie en toute liberté contrairement à la transformée de Fourier qui se restreint uniquement aux fonctions sinusoïdales.

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{b}} \Psi\left(\frac{t-a}{b}\right) \quad (u \in R_+, S) \quad (\text{III.7})$$

Ψ : l'ondelette mère

a : l'échelle de l'ondelette

b : la translation de l'ondelette

III.2.2.2. Scalogramme :

Le scalogramme d'un signal est défini par : $P_w = (b, \xi) = |W_f(a, b)|^2 = |W_f(a, \frac{n}{\xi})|^2$ (III.8)

Le scalogramme normalisé vaut : $\frac{n}{\xi} \cdot P_w f(b, \xi)$ (III.9)

La normalisation revient à multiplier l'expression par un facteur égal à $\frac{\xi}{n}$. n est le centre de fréquence de l'ondelette élémentaire qui a une énergie concentrée dans un intervalle positif centré en n/s dont la taille est donnée par $1/s$. Dans le domaine temps-fréquence de l'analyse par ondelettes, un atome d'ondelette (Figure II.12) est représenté par le rectangle de Heisenberg de largeur temporelle $s * \sigma_t$ et de largeur fréquentielle $\frac{1}{s * \sigma_w}$; σ_t et σ_w correspondent aux dimensions du rectangle de Heisenberg pour l'ondelette mère. La

largeur temporelle et fréquentielle restent constantes pour toute la transformée car :

$$S_{\text{rect}} = \sigma_w * \sigma_t.$$

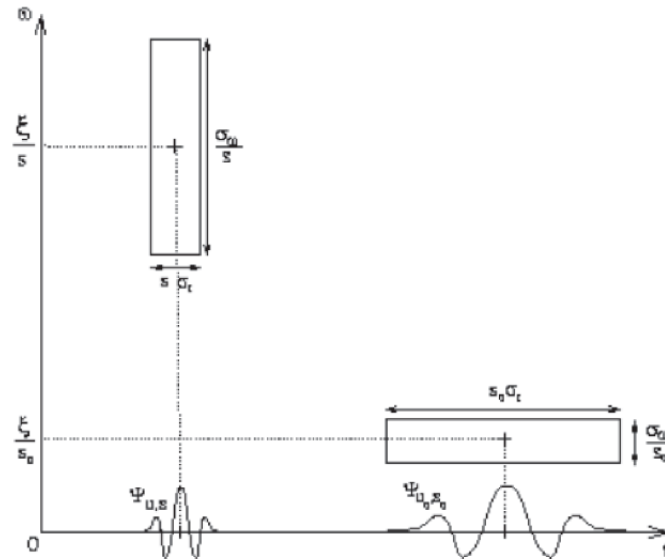


Figure III.3: Atome de la transformée en ondelettes.

III.2.2.3. Calcul des coefficients de la CWT :

La procédure de calcul des coefficients s'effectue en multipliant le signal par l'ondelette analysante puis en calculant l'intégrale du produit. Ce processus est accompli en quatre étapes [48]:

- Dès le départ, on compare l'ondelette à une portion du signal.
- On calcule le coefficient C qui étudie le degré de similitude entre la portion du signal et l'ondelette analysante. Ces derniers sont similaires lorsque la valeur numérique du coefficient C est élevée. Plus précisément, lorsque l'énergie du signal et celle de l'ondelette sont relativement proches, le coefficient C peut être interprété comme un facteur de corrélation.

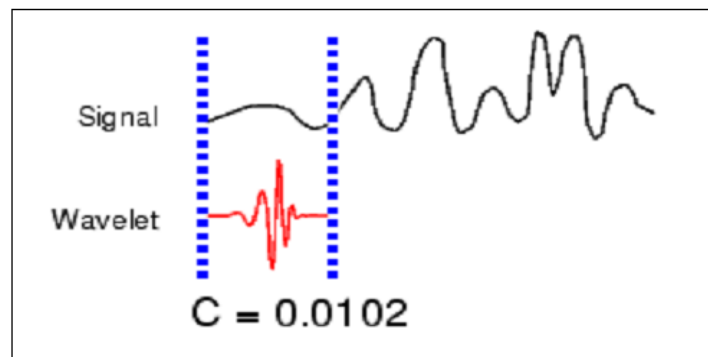


Figure III.4: Comparaison de l'ondelette à une portion du signal

- On translate l'ondelette vers la droite jusqu'au balayage total du signal (Figure III.5).

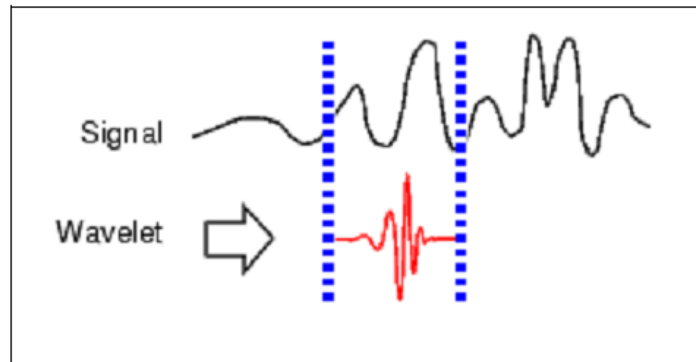


Figure III.5: Balayage du signal par l'ondelette analysante.

- On passe ensuite à une échelle plus élevée qui induit une dilatation temporelle de l'ondelette analysante. Cela a pour effet d'augmenter la portion de comparaison (Figure III.6).

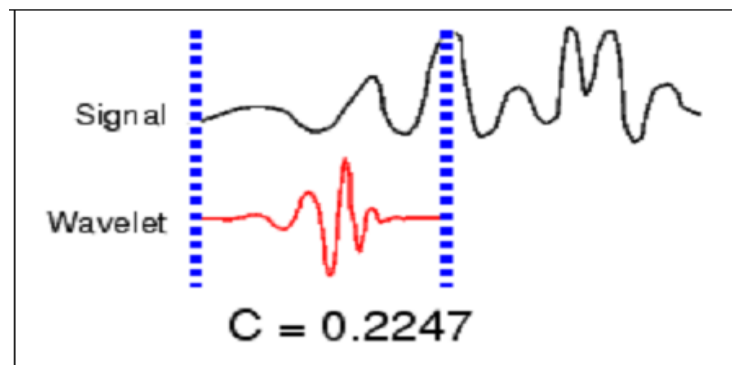


Figure III.6: Dilatation de l'ondelette et balayage du signal.

Cette opération est répétée pour chaque échelle jusqu'au calcul des coefficients pour toutes les échelles et pour différentes portions du signal. Les coefficients constituent le résultat d'une régression effectuée sur le signal original.

III.2.3. L'analyse multi-résolution (AMR):

L'analyse multi résolution consiste à décomposer successivement le signal suivant un terme d'approximation a_i et un terme de détail d_i aux différentes échelles. Cette décomposition est obtenue en appliquant un banc de filtres orthogonaux construit de façon à permettre une reconstruction parfaite du signal initial. Ces filtres correspondent en fait à une paire d'ondelettes analysantes dites fonction d'échelle et fonction ondelette. L'implémentation de l'analyse multi résolution est facilitée par l'emploi de filtres grâce à un schéma de type pyramidal. Le signal s peut alors s'exprimer de la manière suivante : [63]

$$s \cong d_1 + d_2 + d_3 \dots + d_i + a_i$$

III.2.4. La transformée en ondelettes discrètes (DWT) :

La transformée en ondelette discrète (Discrete Wavelet Transform ou DWT) est considérée comme un processus de décomposition du signal. Le signal d'origine $s(t)$ traverse deux filtres complémentaires : passe-haut et passe-bas, et émerge en tant que deux signaux: respectivement le signal d'approximations A et le signal de détails D comme le montre la Figure suivante [47].

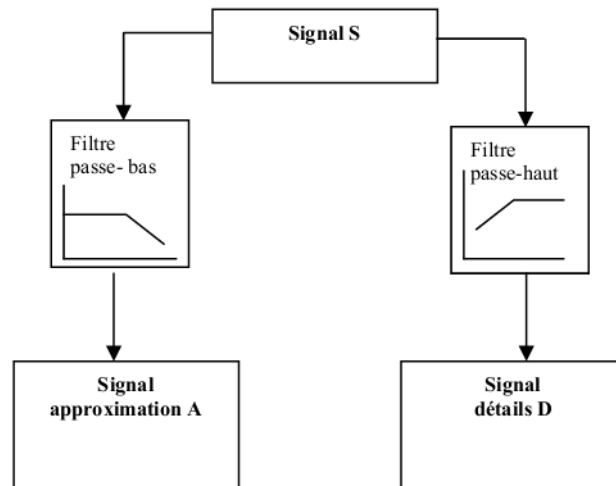


Figure III.7: décomposition du signal s en approximations et détails.

Pour beaucoup de signaux, le contenu de basse fréquence est la pièce la plus importante. Il donne au signal son identité. Le contenu à haute fréquence, d'autre part, nous donne la saveur ou la nuance. Dans la transformée en ondelette, on dit souvent l'approximation et le détail [47].

III.2.4.1. Niveau maximal de décomposition :

Avant de décomposer un signal en approximations et détails, il y a lieu de considérer un niveau maximal de décomposition. Pour la DWT, le nombre maximal pour lequel un signal est décomposé est obtenu par la relation [60]:

$$J_{ful} = \text{fix}(\log_2(N)) \quad (\text{III.10})$$

N est la longueur du signal échantillonné et la fonction fix donne le nombre entier le plus proche de la valeur obtenue après calcul du logarithme à base 2 de N .

III.2.4.2. Choix de l'ondelette analysante :

Le choix de l'ondelette analysante joue un rôle primordial pour la compréhension des phénomènes transitoires comme les impulsions des décharges électriques observées sur les signaux du courant de fuite pour un état de pollution d'isolateurs sous des conditions

humides. Il est cependant connu que les ondelettes Daubechies 4 ‘db4’ et Daubechies 6 ‘db6’ sont plus appropriées pour la détection des perturbations de courtes durées ; alors que les ondelettes db8 et db10 sont adéquates pour l’étude des perturbations de relatives longues durées [17,43].

L’expression mathématique de l’ondelette mère complexe de Daubechies 4 est :

$$\Psi(t) = e^{-\frac{t^2}{2}} * (e^{-j\omega t} - e^{-\frac{\omega^2}{2}}) \quad (\text{III.11})$$

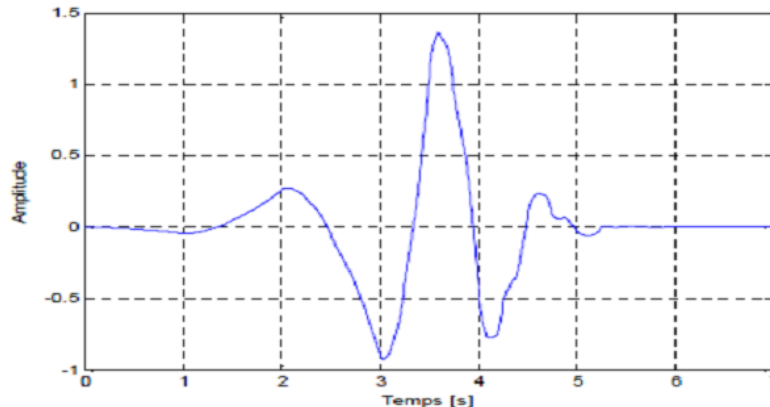


Figure III.8: Forme d’onde de l’ondelette de Daubechies 4.

L’ondelette de Daubechies 4 est choisie pour appliquer la DWT sur le signal du courant de fuite qui affiche des pics de hautes fréquences à partir d’un certain niveau de tension. Cette ondelette semble adéquate pour la détermination des bandes fréquentielles qui caractérisent les perturbations de hautes fréquences via l’analyse multi-résolution. A cet effet, la détermination du niveau de décomposition est importante pour récupérer de bonnes informations sans trop de calculs.

III.3. Décomposition du signal du courant de fuite par DWT :

Dans le présent travail, nous étudions la corrélation qui pourrait exister entre le signal courant de fuite (en termes de forme et d’amplitude) et la sévérité de la pollution caractérisant l’état de surface de l’isolateur. Pour cela nous allons appliquer la DWT sur les différents signaux du courant de fuite. On peut décomposer ces signaux par la DWT en utilisant différentes ondelettes mère (Haar, Daubechies 4, Coiflet, Daubechies 8). L’ondelette la plus adéquate pour notre étude est l’ondelette de Daubechies 4.

Le choix du niveau de décomposition est lié à la taille de notre signal, il est calculé à l’aide de l’équation (III.10). Pour notre étude, la taille des signaux analysés est de 10000

points donc la décomposition par DWT doit être réalisée jusqu'au niveau 10 afin de récupérer toutes les informations fréquentielles et temporelles contenues dans les signaux du courant de fuite. La figure suivante représente un signal du courant de fuite et ses 10 détails.

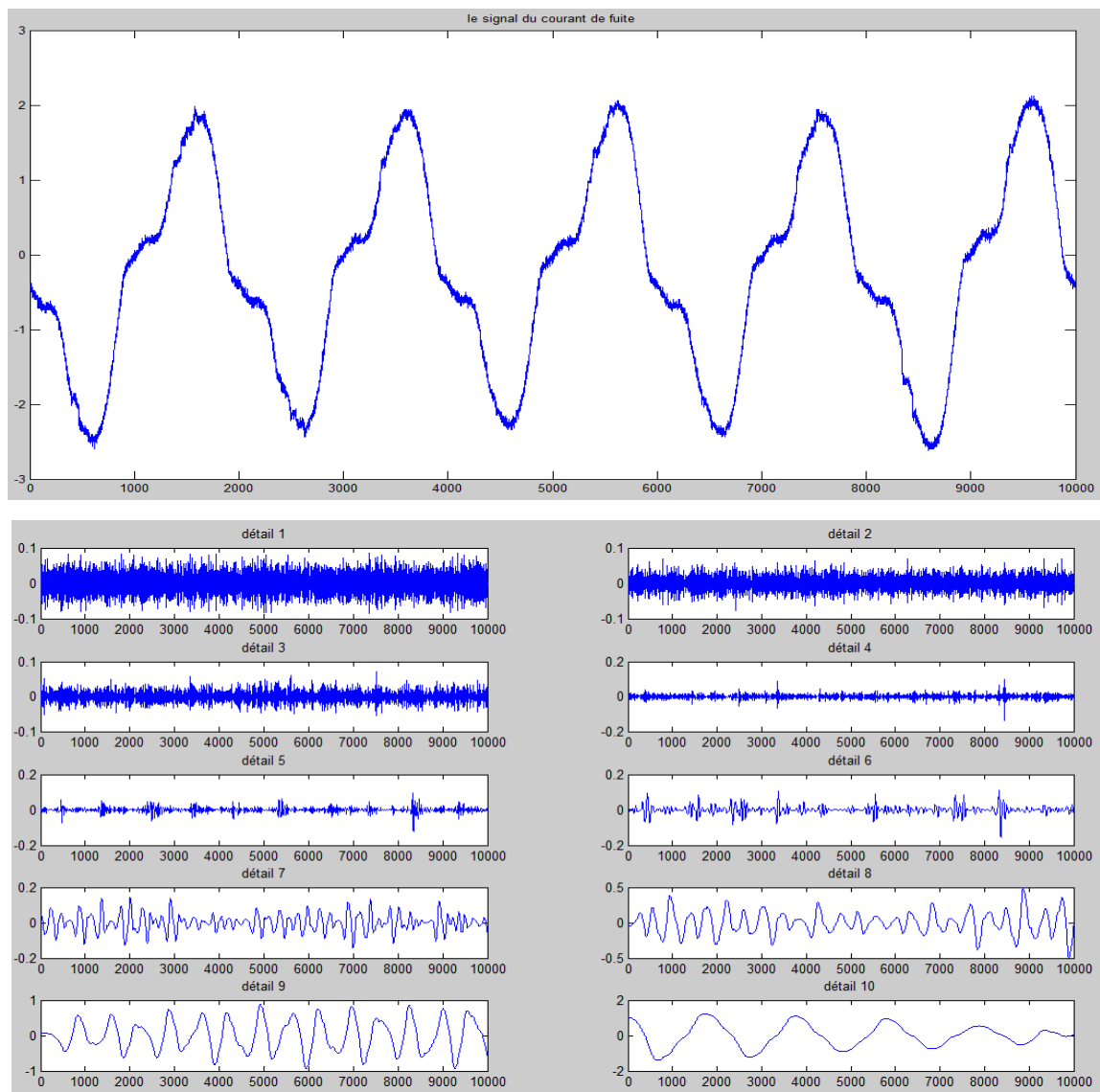


Figure III.9: Représentation de la forme d'onde du signal du courant de fuite et ses dix détails.

La base de données utilisée dans notre étude a été réalisée par Mr B.MOULA [58], les essais ont été effectués au laboratoire de haute tension de l'Ecole Nationale Polytechnique. Pour Cinq conductivités volumiques étudiées sous pollution uniforme à savoir (1.2; 3.5; 6; 10 et 20mS/cm) et pour plusieurs paliers de tension appliquée, il prélève cinq fois le signal du courant de fuite via un oscilloscope numérique TEKTRONIX.

La forme d'onde du courant de fuite est très sensible à l'état de surface de l'isolateur pollué. Par ailleurs, l'amplitude du courant de fuite augmente avec l'augmentation de la

tension appliquée et/ou de la conductivité volumique de la solution polluante.

III.4. Calcul des caractéristiques des coefficients de détails d'ondelette :

Après la décomposition du signal du courant de fuite, nous avons calculé pour chaque détail les valeurs des maxima, minima, moyennes ainsi que la standard déviation (*écart type*). Le choix de ces quatre paramètres semble être suffisant pour bien identifier notre signal.

Le maximum et le minimum nous donnent une information sur les valeurs crêtes des détails à chaque niveau de décomposition. La standard déviation (*écart type*) est calculée pour les détails afin d'identifier les transitions d'énergie présentes dans le signal du courant de fuite [49]. Elle est donnée par l'expression :

$$\delta = \sqrt{\frac{\sum_{i=1}^N (d_n - d_m)^2}{N-1}} \quad (\text{III.12})$$

δ : la standard déviation (*écart type*) du signal d .

d_n : correspond à un échantillon du détail d .

d_m : la moyenne du signal d .

N : correspond à la taille du signal d .

Par la suite nous allons étudier la variation de ces paramètres en fonction de la conductivité volumique pour différents paliers de tension et pour les deux types de pollution.

III.4.1. Etude de la standard déviation des détails du signal du courant de fuite pour une pollution avec 100 g/l de Kaolin :

Afin de réaliser cette étude, nous avons calculé le standard déviation (*écart type*) pour les dix détails en utilisant la formule (III-8). Après avoir obtenu les valeurs de la standard déviation (*écart type*), Nous avons établi une représentation graphique de ces dernières en fonction de la conductivité volumique de l'agent polluant pour différents paliers de tension appliquée et pour les deux types de pollution comme le montre les figures suivantes.

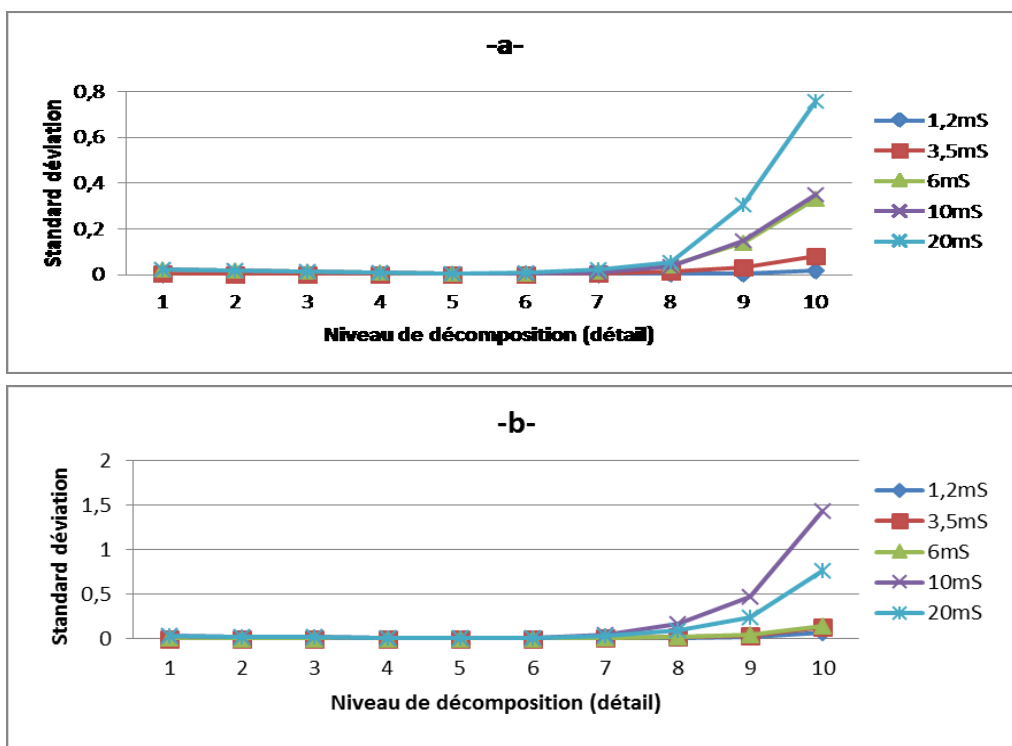


Figure III.10: Evolution de la standard déviation (écart type) des détails du signal du courant de fuite pour une tension appliquée de 5 kV_{eff}. a) pollution uniforme. b) pollution

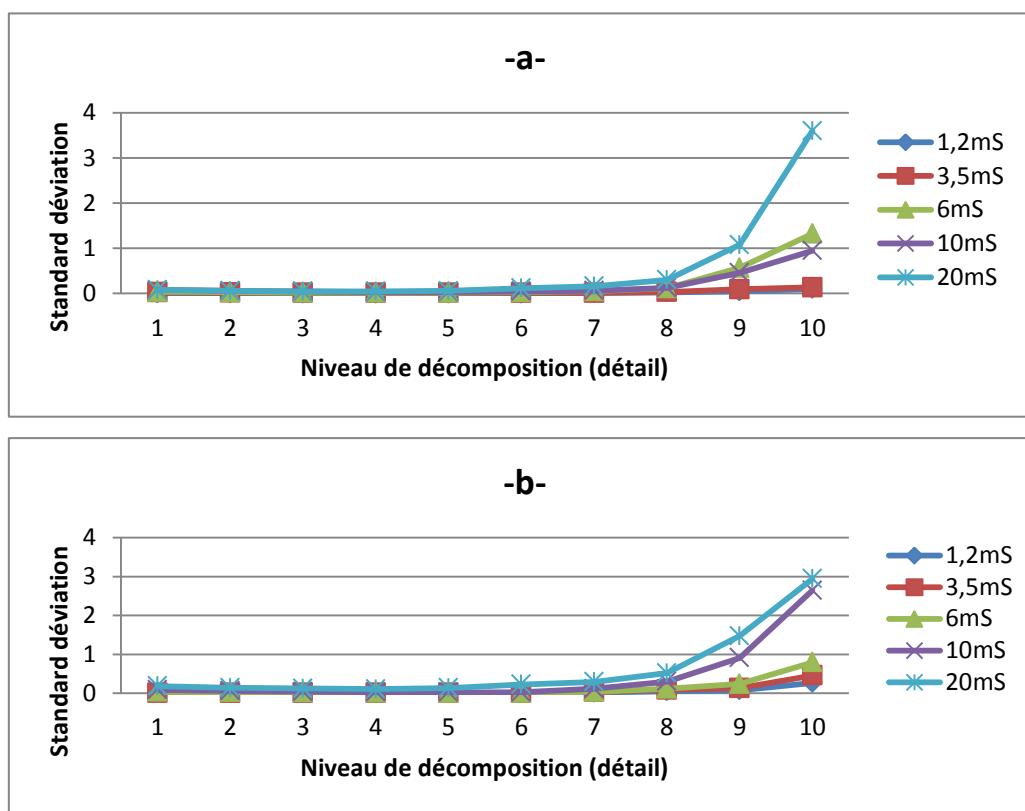


Figure III.11: Evolution de la standard déviation (écart type) des détails du signal du courant de fuite pour une tension appliquée de 15 kV_{eff}. a) pollution uniforme. b) pollution

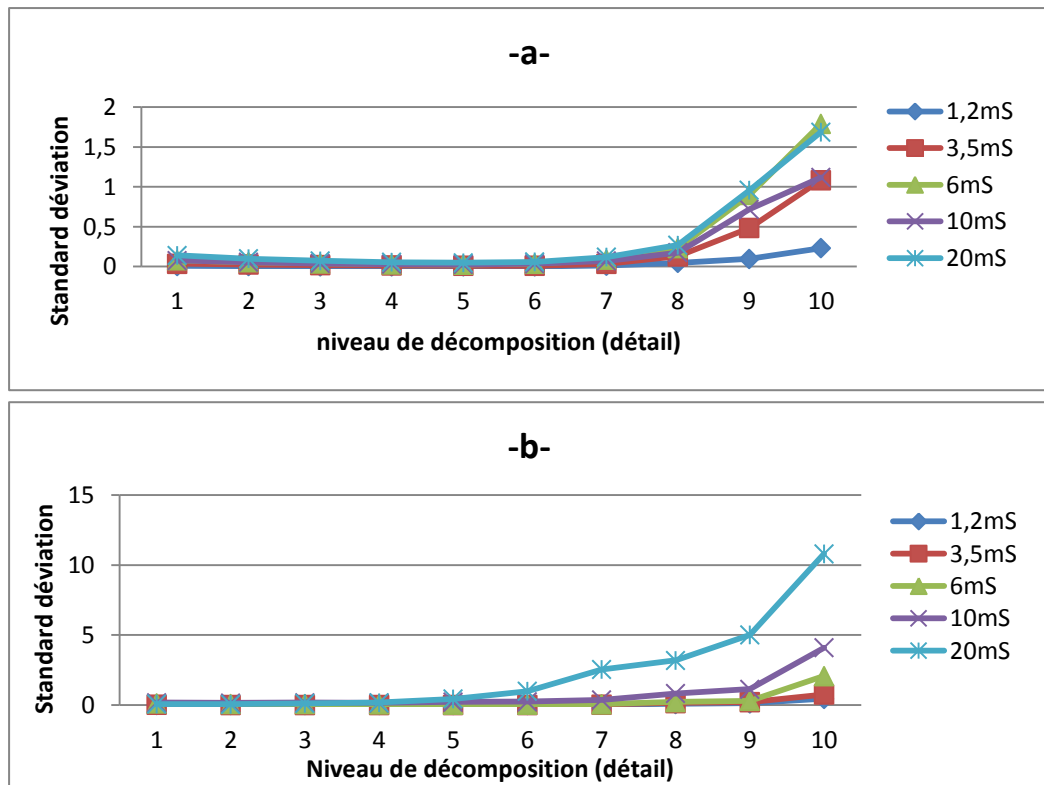


Figure III.11: Evolution de la standard déviation (écart type) des détails du signal du courant de fuite pour une tension appliquée de $25 \text{ kV}_{\text{eff}}$. a) pollution uniforme. b) pollution non uniforme.

Interprétation :

Nous constatons que les valeurs de la standard déviation augmentent avec le niveau de décomposition. Cette augmentation est vérifiée pour tous les niveaux de tension appliqués ainsi pour les deux types de pollution. La standard déviation est d'autant plus importante que la conductivité volumique de la solution polluante est élevée. Nous observons que l'augmentation de la standard déviation est remarquable à partir du huitième niveau de décomposition pour le cas de pollution uniforme et à partir du septième niveau de décomposition pour le cas de pollution non uniforme.

Suite à cette étude nous constatons que les valeurs de la standard déviation peuvent être considérées comme information importante sur le degré de sévérité de pollution car leurs valeurs sont liées à la conductivité volumique de l'agent polluant.

III.4.2. Etude des maximums des détails du signal du courant de fuite pour une pollution avec 100 g/l de Kaolin :

Afin de réaliser cette étude, nous avons calculé les valeurs des maximums pour les dix détails. Une représentation graphique de ces dernières en fonction de la conductivité

volumique de la solution polluante a été établie pour différents paliers de tension appliquée comme le montre les figures suivantes.

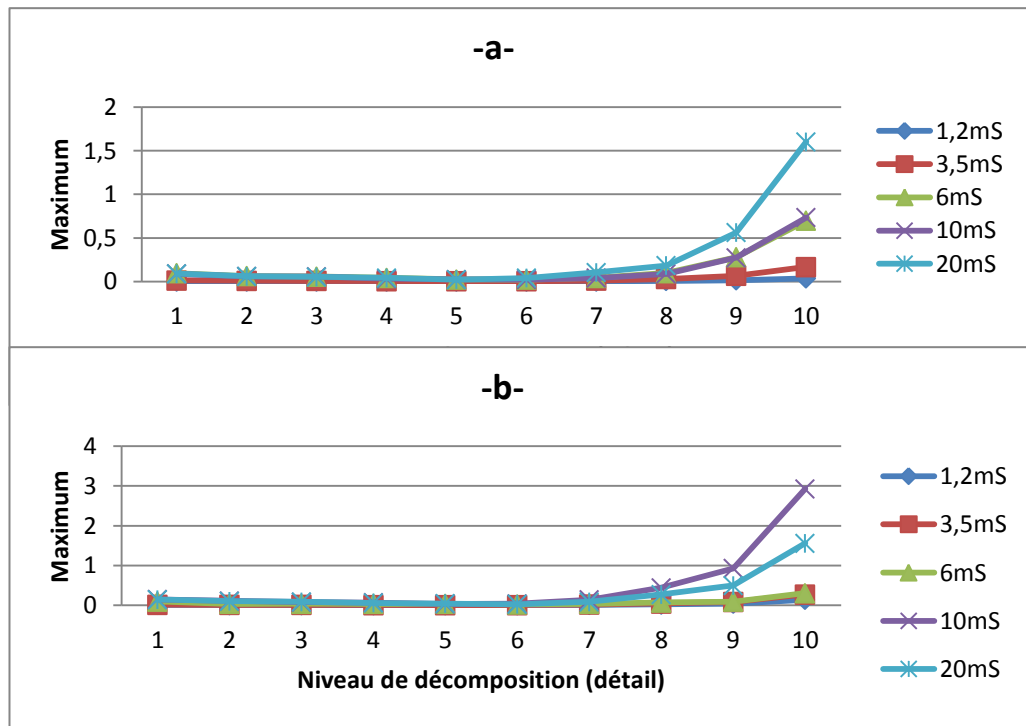


Figure III.13: Evolution des valeurs des maxima des détails du signal du courant de fuite pour une tension appliquée de 5 kV_{eff} . a) pollution uniforme. b) pollution non uniforme.

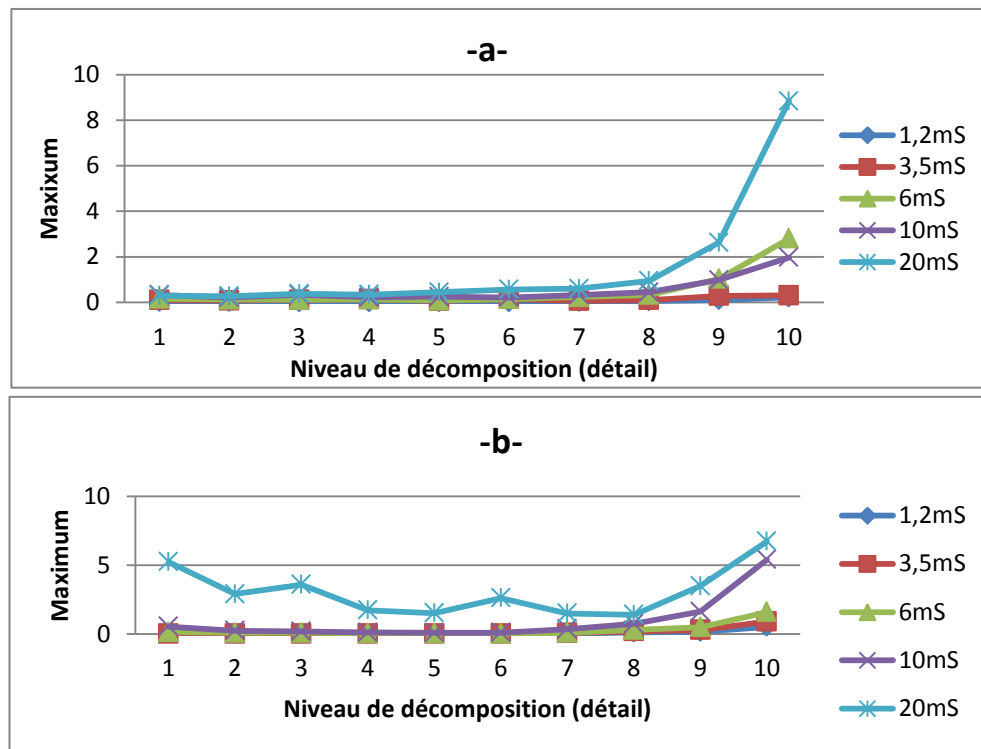


Figure III.14: Evolution des valeurs des maxima des détails du signal du courant de fuite pour une tension appliquée de 15 kV_{eff} . a) pollution uniforme. b) pollution non uniforme.

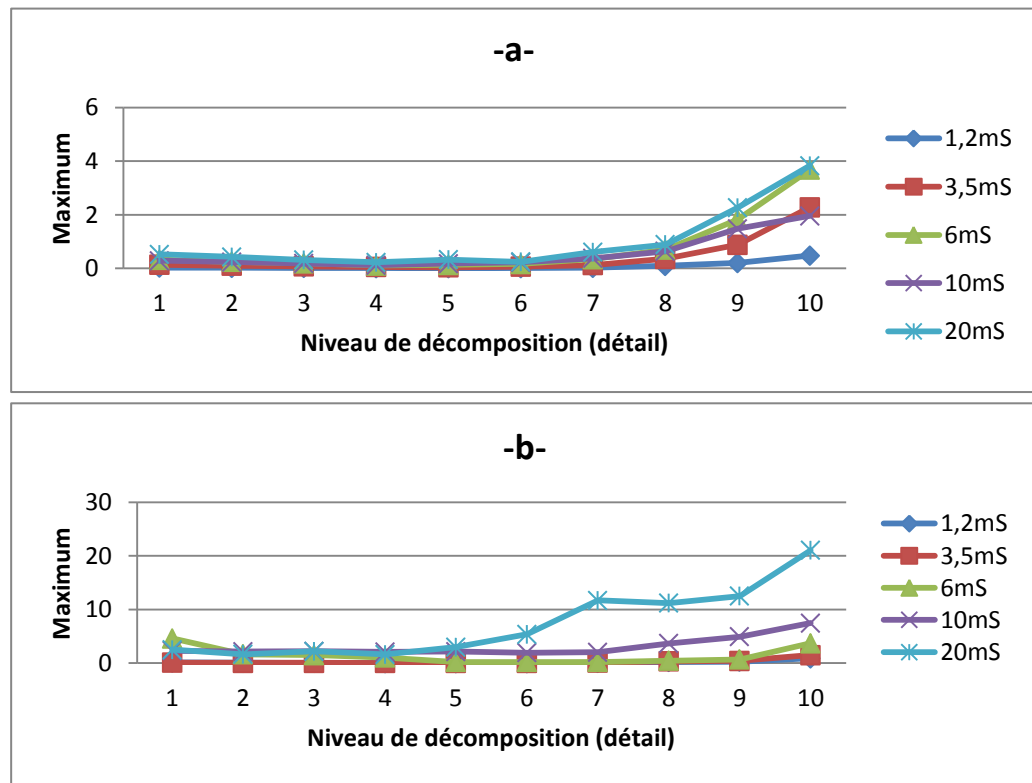


Figure III.15: Evolution des valeurs des maximums des détails du signal du courant de fuite pour une tension appliquée de 25 kV_{eff}. a) pollution uniforme. b) pollution non uniforme.

Interprétation :

Quelques soit les niveaux de tension appliquée, les valeurs des maximums en générale augmentent avec le niveau de décomposition. Cette augmentation est remarquable à partir du huitième niveau de décomposition pour le cas de pollution uniforme et à partir du septième niveau pour le cas de pollution non uniforme. Les valeurs des maximums sont d'autant plus importantes que la conductivité volumique de l'agent polluant est élevée.

Cette étude nous permet de dire que les valeurs des maximums peuvent être considérées comme une deuxième information importante pour la prédiction de l'état de surface des isolateurs car leurs valeurs sont aussi proportionnelles à la conductivité.

III.4.3. Etude des minimums des détails du signal du courant de fuite pour une pollution avec 100 g/l de Kaolin :

Afin de réaliser cette étude, nous avons calculé les valeurs des minimums pour les dix détails du signal du courant de fuite. Nous avons établi une représentation graphique de ces derniers en fonction de la conductivité volumique de la solution polluante pour trois paliers de tension comme le montre les figures suivantes.

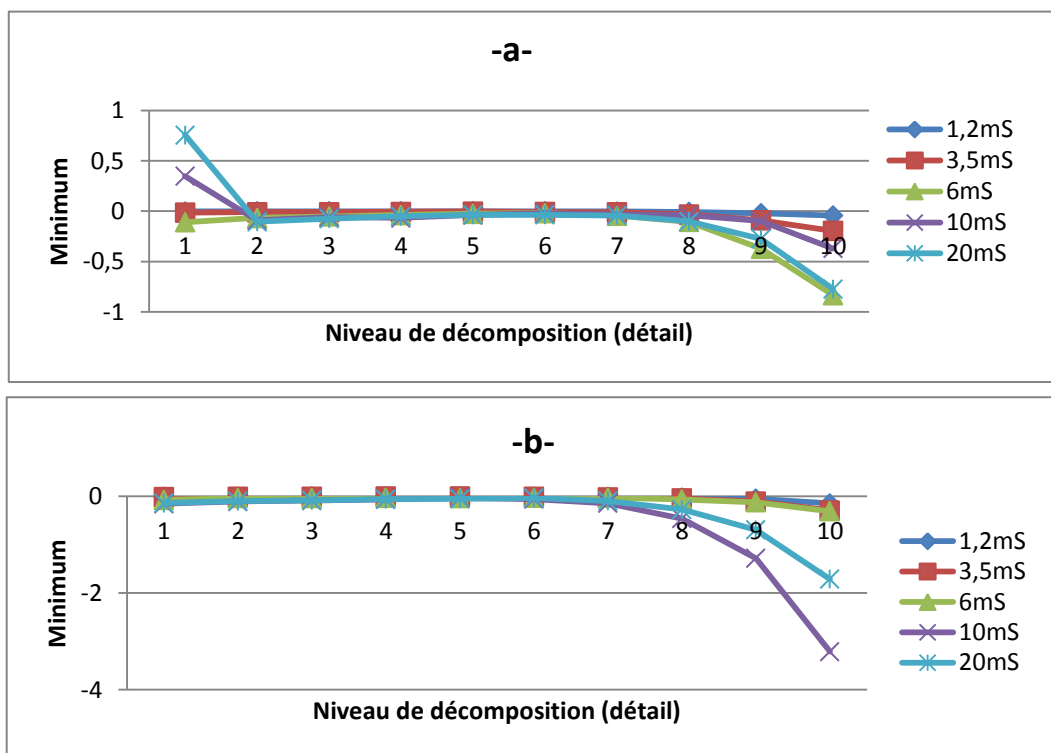


Figure III.16: Evolution des valeurs des minimums des détails du signal du courant de fuite pour une tension appliquée de 5 kV_{eff}. a) pollution uniforme. b) pollution non uniforme.

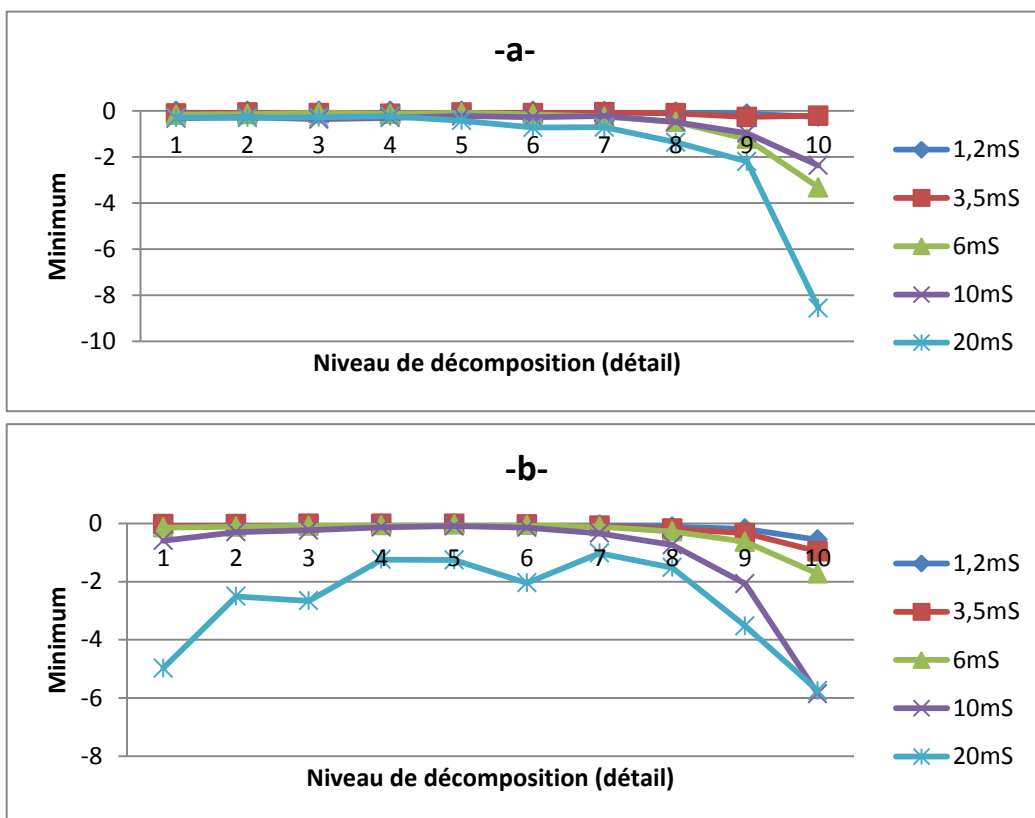


Figure III.17: Evolution des valeurs des minimums des détails du signal du courant de fuite pour une tension appliquée de 15 kV_{eff}. a) pollution uniforme. b) pollution non uniforme.

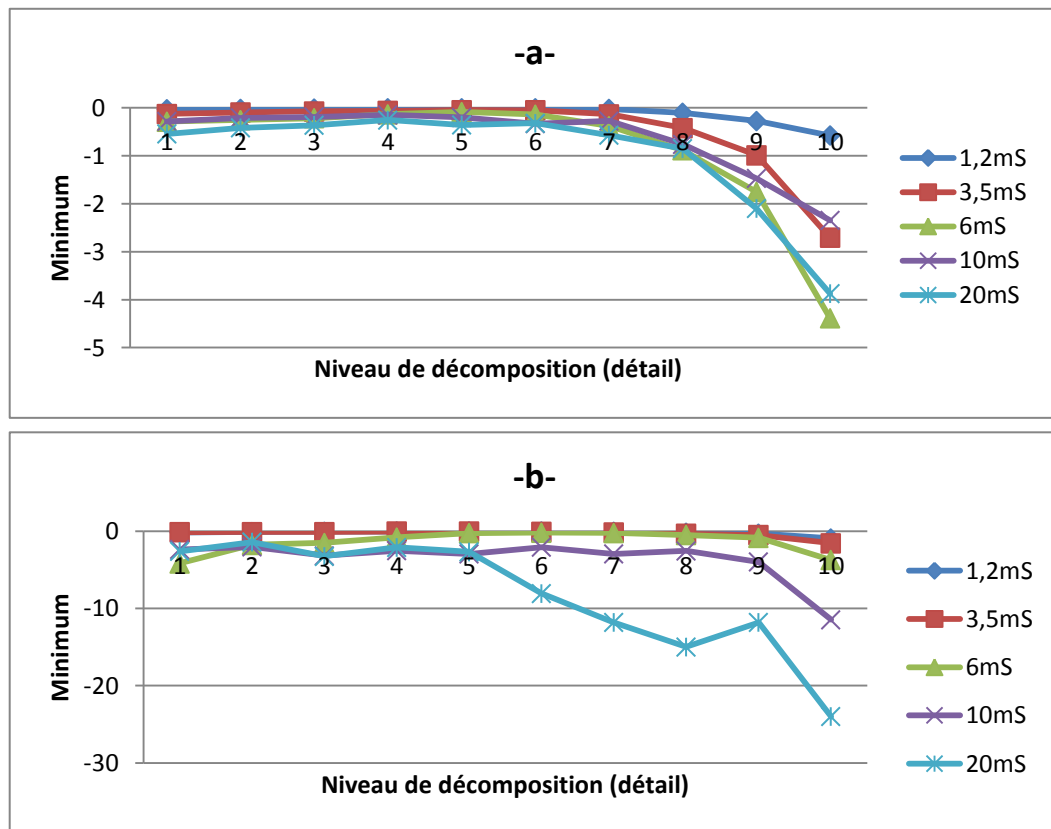


Figure III.18: Evolution des valeurs des minimums des détails du signal du courant de fuite pour une tension appliquée de 25 kV_{eff} . a) pollution uniforme. b) pollution non uniforme.

Interprétation :

Les valeurs des minimums diminuent avec le niveau de décomposition pour les différents niveaux de tensions et ils sont d'autant plus faibles que la conductivité de l'agent polluant est élevée.

L'étude des valeurs des minimums des détails des signaux du courant de fuite nous permet de constater que les valeurs des minimums peuvent être considérées comme une troisième information importante pour la prédiction de l'état de surface des isolateurs.

III.4.4. Etude des moyennes des détails du signal du courant de fuite pour une pollution avec 100 g/l de Kaolin :

Afin de réaliser cette étude, nous avons calculé les valeurs moyennes pour les dix détails. Nous avons établi une représentation graphique de ces derniers en fonction de la conductivité volumique de la solution polluante pour trois paliers de tension appliquée comme le montre les figures suivantes.

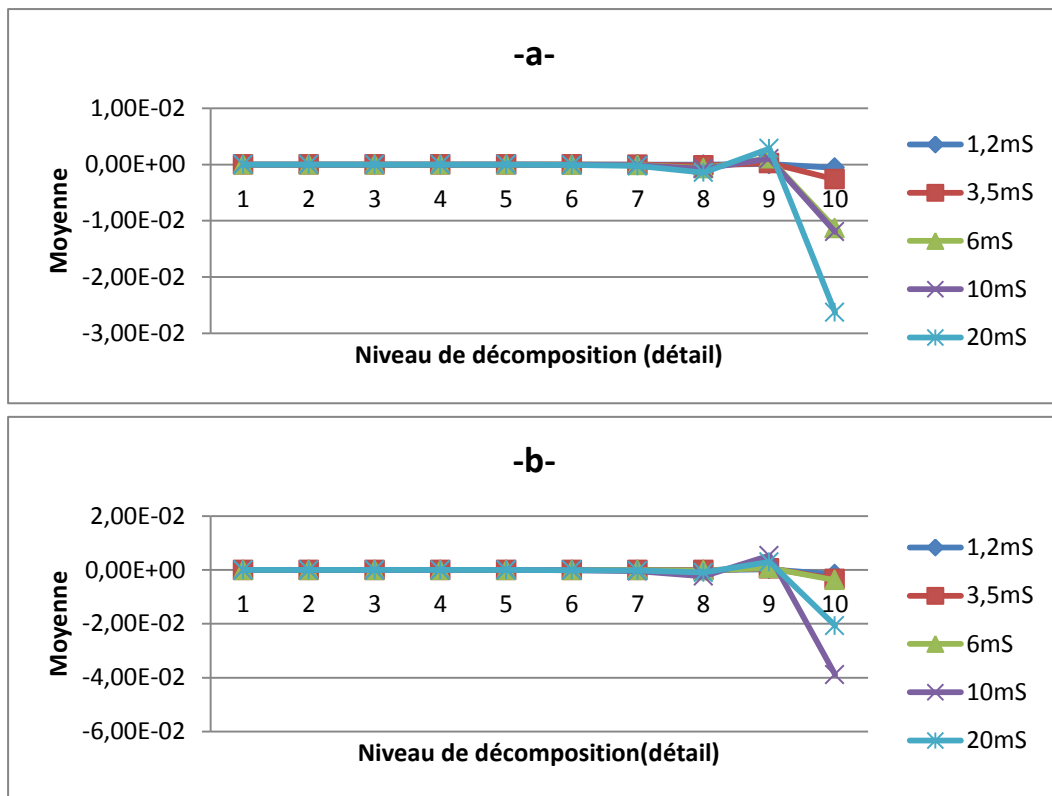


Figure III.19: Evolution des valeurs des moyennes des détails du signal du courant de fuite pour une tension appliquée de 5 kVeff. a) pollution uniforme. b) pollution non uniforme.

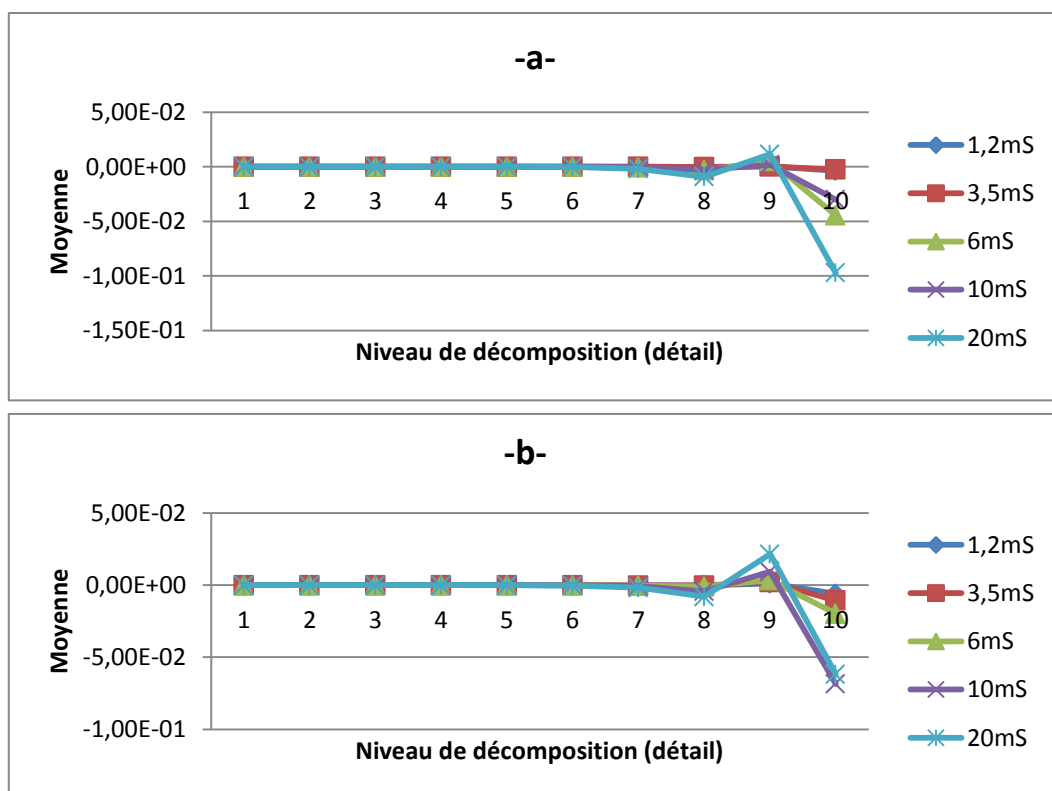


Figure III.19: Evolution des valeurs des moyennes des détails du signal du courant de fuite pour une tension appliquée de 15 kVeff. a) pollution uniforme. b) pollution non uniforme.

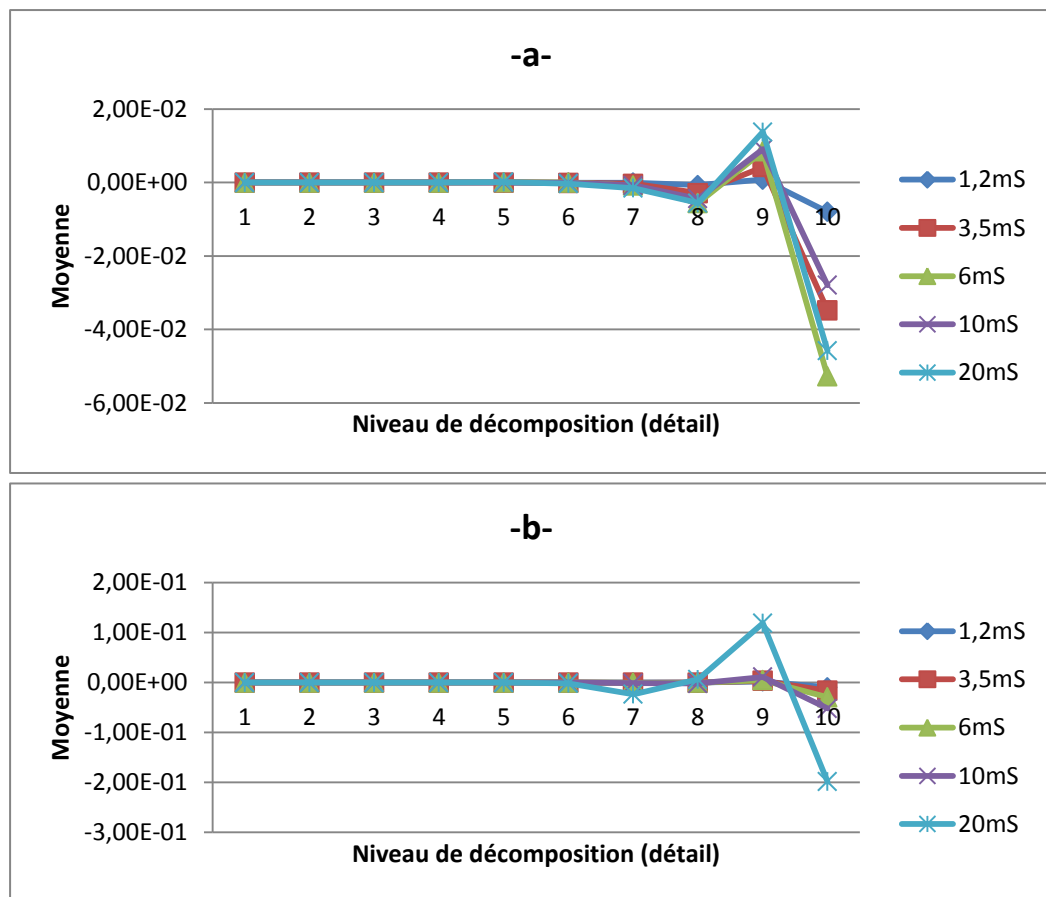


Figure III.19: Evolution des valeurs des moyennes des détails du signal du courant de fuite pour une tension appliquée de 25 kVeff. a) pollution uniforme. b) pollution non uniforme

Interprétation :

Pour les trois niveaux de tension appliquée, nous constatons une diminution légère des valeurs des moyennes entre le premier et le huitième niveau de décomposition. Par ailleurs, nous observant un pic au niveau du neuvième détail. Les valeurs des moyennes diminuent de nouveau dans le dixième détail.

Les résultats de l'étude des valeurs des moyennes des détails des signaux du courant de fuite nous pousse à l'utiliser comme une quatrième information pour la prédiction de l'état de surface des isolateurs.

III.5. Les Classes de pollution :

Dans notre étude, la classification de la sévérité de pollution est un point important. Dans la littérature, il est noté que cette sévérité est généralement caractérisée par l'intensité de précipitation ainsi que par la conductivité des dépôts polluants. La sévérité considérée est alors exprimée en termes de salinité équivalente. En effet, on peut faire correspondre à chaque

site, soumis indifféremment à une pollution naturelle ou industrielle, une solution saline équivalente. Pour notre cas, nous distinguons quatre classes de sévérité [59], dont on va utiliser trois d'entre elles, à savoir :

Classe1 : pollution faible d'origine naturelle (Sauf marine). Elle concerne généralement des zones non industrielles ayant une faible densité d'habitation, situées loin de la mer. Un isolement normal convient facilement pour ces zones. La salinité équivalente maximale retenue pour cette classe correspond à une pulvérisation d'une solution de chlorure de sodium de concentration 2.5g/l [59].

Classe2 : pollution moyenne d'origine naturelle (Sauf marine) ou faiblement industrielle. Elle correspond à des zones à moyenne densité d'habitation et situées loin de la mer. Ces zones nécessitent un isolement renforcé, la concentration de la salinité équivalente maximale retenue est de 10g/l [59].

Classe3 : pollution forte d'origine marine et industrielle. Les zones considérées possèdent une forte densité d'habitation et située près de la mer. Elles sont exposées à des embruns marins relativement forts. Ces sites nécessitent un isolement important, la concentration de la salinité équivalente maximale est de 80g/l [59].

En tenant compte de la base de données de B. Moula [58] et des normes définies ci-dessus, nous proposons un tableau (Tableau IV-1) récapitulatif des classes assignées aux différentes conductivités utilisées expérimentalement.

Concentration en Kaolin [g/l]	Conductivité volumique [mS/cm]	Concentration de sel utilisé [g/l]	Classe
40	1,2 et 3,5	Entre 0.7 et 2.5	1
	6	Entre 3 et 10	2
	10 et 20	Inférieur à 60	3
100	1,2 et 3,5	Entre 0.7 et 2.5	1
	6	Entre 3 et 10	2
	10 et 20	Inférieur à 60	3
200	1,2 et 3,5	Entre 0.7 et 2.5	1
	6	Entre 3 et 10	2
	10 et 20	Inférieur à 60	3

Tableau III.22: Tableau récapitulatif des différentes classes assignées aux différentes conductivités utilisées.

III.6. Application des SVM pour la classification de sévérité de pollution des isolateurs :

III.6.1. Composition du vecteur d'état :

Afin de juger l'état de surface d'un isolateur, nous avons besoin d'un certain nombre de critères le concernant sur lesquels nous allons nous baser. Dans le paragraphe précédent, nous avons cité les différents critères de classification et nous avons établi des liens entre ces derniers et la sévérité de pollution.

Le vecteur d'état qu'on a utilisé pour la formation du modèle est V (1×40), ce dernier est composé des coefficients extraits des coefficients liés aux détails d'ondelettes. On a extrait dix détails pour chaque signal, par la suite on a tiré les standards déviations, les minimums, les maximums et les moyennes de ces dix détails pour construire le vecteur d'état suivant :

Std1 _dstd10 _d	Min1 _dmin10 _d	Max1 _dmax10 _d	Moy1 _dmoy10 _d
---	---	---	---

III.6.2. Entraînement du modèle :

III.6.2.1. Base de données pour le cas de la pollution uniforme :

Après les mesures expérimentales sur les isolateurs avec la pollution uniforme, chaque signal du courant de fuite, est traité par la technique de transformée en ondelettes discrète (DWT) pour extraire différents critères de jugement détaillés précédemment et pour chaque signal du courant de fuite nous faisons correspondre un vecteur V (1×40).

Nous répétons ce processus pour 106 signaux expérimentaux -base de données de B. Moula [58] pour obtenir une matrice de (106×40), dont 95 échantillons servent de base de données d'entraînements pour le modèle à l'aide du logiciel MATLAB®, les 11 échantillons restants servent de base de données de validation du modèle.

III.6.2.2. Base de données pour le cas de la pollution non uniforme :

Après les mesures expérimentales sur les isolateurs avec la pollution non uniforme, chaque signal du courant de fuite est traité de la même manière que pour les signaux du courant de fuite en cas de pollution uniforme.

Parmi les 75 signaux de la base de données, 66 sont utilisés pour l'entraînement du modèle SVM correspondant au type de pollution non uniforme ; et les 9 qui restent sont utilisés pour la validation.

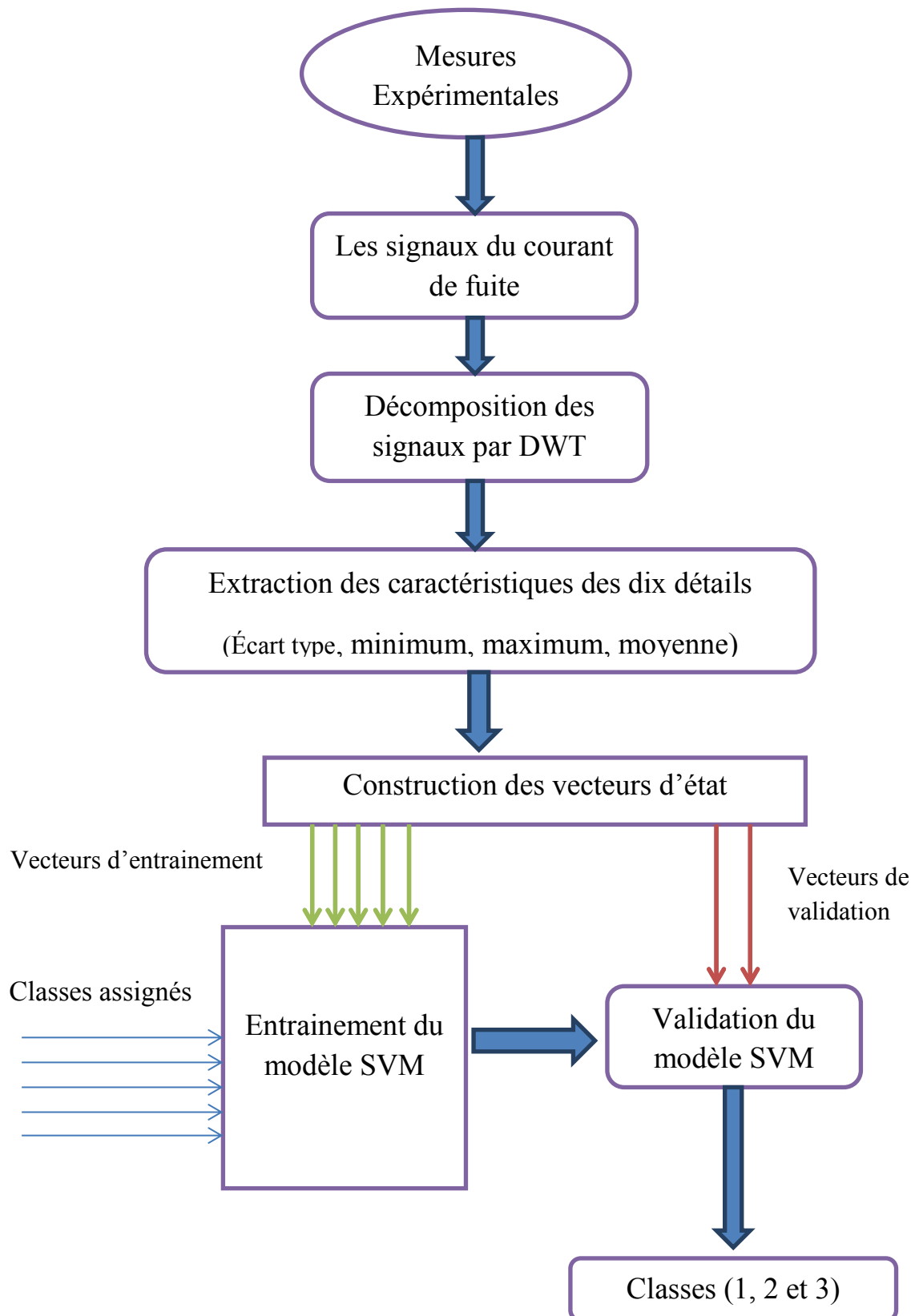


Figure III.23: La procédure utilisée pour la classification.

Remarque :

La base de données dispose de plusieurs paliers de tension appliquée, dans notre travail nous avons choisi 20, 21, 25 et 28 kV comme tension des signaux servant dans la construction et la validation du modèle SVM.

La base de données dispose de deux types de signaux, ceux du courant de fuite et ceux de la tension. Dans notre modèle SVM nous avons utilisé les signaux du courant de fuite afin d'avoir de bons résultats.

III.6.3. Les différentes fonctions noyaux appliquées :

$$\text{La fonction noyau linéaire : } K(u, v) = u \cdot v \quad (\text{III.13})$$

$$\text{Noyau polynomial d'ordre } d : K(u, v) = (\langle u, v \rangle + 1)^d, \text{ dans notre cas on a pris deux valeurs : } d=2 \text{ donc l'expression devient : } K(u, v) = (\langle u, v \rangle + 1)^2 \quad (\text{III.14})$$

$$\text{Et } d=3 \text{ donc : } K(u, v) = (\langle u, v \rangle + 1)^3 \quad (\text{III.15})$$

$$\text{La fonction à Base Radiale RBF : } K(u, v) = \exp\left(-\frac{\|u-v\|^2}{2\sigma^2}\right) \quad (\text{III.16})$$

Dans notre étude on a pris $\sigma = 1$.

$$\text{La fonction noyau sigmoïde : } K(u, v) = \tanh(u, v + c), \text{ dans notre étude on a pris } c = -1 \text{ et } c = -2 \text{ dont les fonctions noyau sont : } K(u, v) = \tanh(u * v - 1) \quad (\text{III.17})$$

$$\text{Et } K(u, v) = \tanh(u * v - 2) \quad (\text{III.18})$$

III.7. Validation et résultat de classification :

Les résultats de classification sont représentés dans les tableaux suivants :

III.7.1. Cas de pollution uniforme :

Echantillons (n°)	Conductivité (mS/cm)	Classe assignée	Résultats obtenus avec différentes les fonctions noyaux					
			Linéaire	RBF	Polynomial		Sigmoïde	
					d=2	d=3	c=-1	c=-2
1	1.2	1	1	3*	1	1	2*	1
2	1.2	1	1	1	1	1	1	3*
3	3.5	1	1	3*	2*	2*	3*	1
4	3.5	1	3*	3*	3*	1	1	2*
5	6	2	3*	2	2	2	3*	3*

6	6	2	2	3*	2	2	2	3*
7	6	2	2	3*	3*	2	2	3*
8	10	3	1*	3	2*	2*	2*	3
9	10	3	2*	3	2*	3	1*	2*
10	20	3	3	3	2*	3	1*	3
11	20	3	2*	3	3	3	2*	1*
Exactitude de classification (%)			54.54	54.54	45.45	81.81	36.36	36.36

Tableau III.2: Classification SVM multi-classes en utilisant la stratégie « un contre tous ».

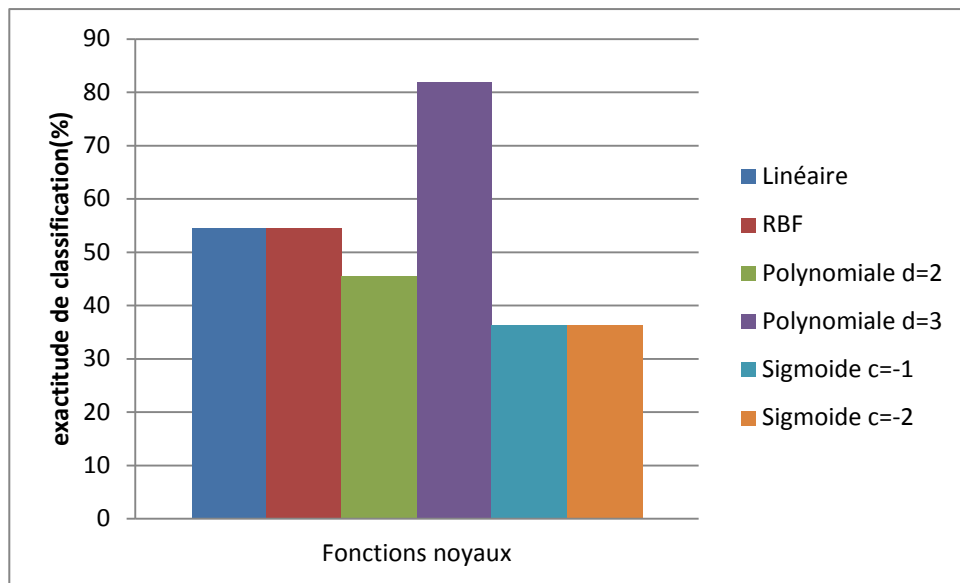


Figure III.24 : représentation de l'exactitude de classification des différentes fonctions noyaux avec la stratégie **un contre tous** pour le cas de **pollution uniforme**.

Echantillons (n°)	Conductivité (mS/cm)	Classe assignée	Résultats obtenus avec les différentes fonctions noyaux					
			Linéaire	RBF	Polynomial		Sigmoïde	
					d=2	d=3	c=-1	c=-2
1	1.2	1	1	3*	2*	1	1	3*
2	1.2	1	1	1	1	1	1	1
3	3.5	1	1	3*	2*	2*	1	1
4	3.5	1	3*	3*	1	1	3*	1
5	6	2	2	2	2	2	1*	3*
6	6	2	3*	3*	2	3*	2	2

7	6	2	3*	3*	2	3*	2	2
8	10	3	2*	3	3	3	1*	3
9	10	3	2*	3	3	3	2*	2*
10	20	3	2*	3	2*	3	2*	3
11	20	3	2*	3	3	3	3	2*
Exactitude de classification (%)			36.36	54.54	72.72	72.72	54.54	63.63

Tableau III.3: Classification SVM multi-classes en utilisant la stratégie « un contre un ».

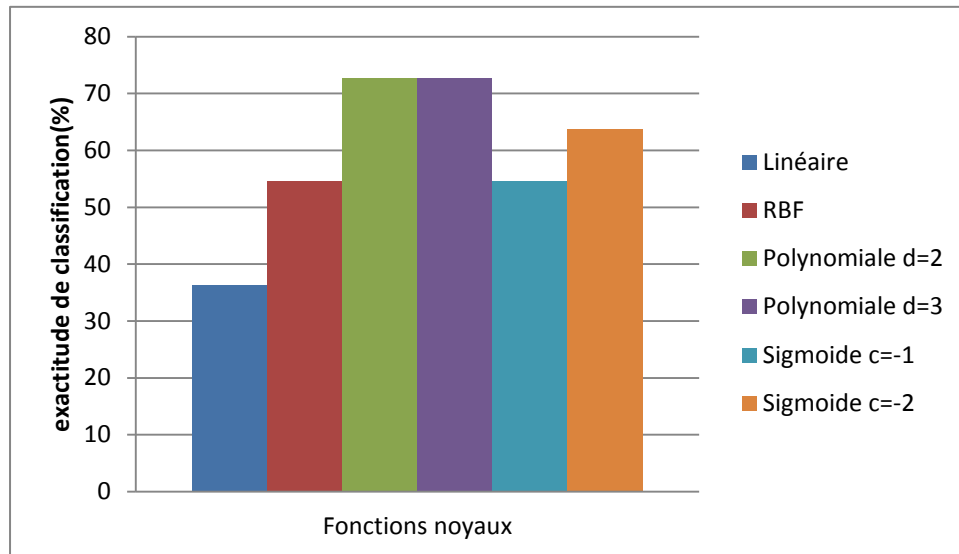


Figure III.25 : représentation de l'exactitude de classification des différentes fonctions noyaux avec la stratégie **un contre un** pour le cas de **pollution uniforme**.

III.7.2.Cas de pollution non uniforme :

Echantillons (n°)	Conductivité (mS/cm)	Classe assignée	Résultats obtenus avec les différentes fonctions noyaux					
			Linéaire	RBF	Polynomial		Sigmoïde	
					d=2	d=3	c=-1	c=-2
1	1.2	1	1	1	1	1	1	1
2	3.5	1	1	1	1	1	1	1
3	3.5	1	1	1	1	1	1	1
4	1.2	1	1	1	1	1	1	1
5	6	2	1*	1*	2	2	1*	1*
6	6	2	1*	3*	3*	3*	1*	1*
7	10	3	3	3	2*	1*	3	3

8	20	3	2*	3	3	3	3	3
9	20	3	2*	3	3	3	2*	3
Exactitude de classification (%)			55.55	77.78	77.78	77.78	66.66	77.78

Tableau III.4: Classification SVM multi-classes en utilisant la stratégie « un contre tous ».

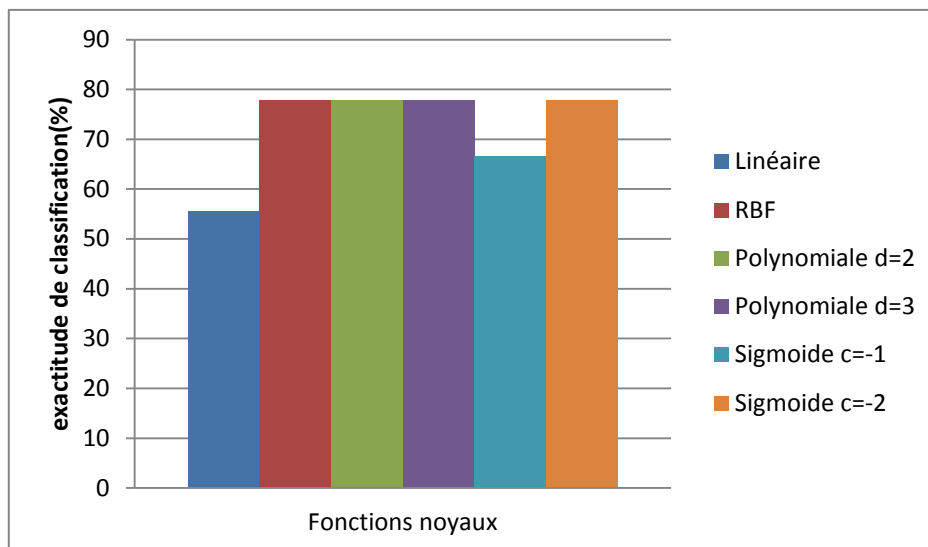


Figure III.26 : représentation de l'exactitude de classification des différentes fonctions noyaux avec la stratégie **un contre tous** pour le cas de **pollution non uniforme**.

Echantillon (n°)	Conductivité (mS/cm)	Classe assignée	Résultats obtenus avec les différentes fonctions noyaux					
			Linéaire	RBF	Polynomial		Sigmoïde	
					d=2	d=3	c=-1	c=-2
1	1.2	1	1	1	1	1	1	1
2	3.5	1	1	1	1	1	1	1
3	3.5	1	1	1	1	1	1	1
4	1.2	1	1	1	1	1	1	1
5	6	2	2	2	2	2	1*	1*
6	6	2	2	3*	3*	3*	2	2
7	10	3	3	3	2*	1*	2*	2*
8	20	3	2*	3	3	3	2*	2*
9	20	3	3	3	3	1*	2*	2*
Exactitude de classification (%)			88.89	88.89	77.78	66.66	55.55	55.55

Tableau III.5: Classification SVM multi classe en utilisant la stratégie « un contre un ».

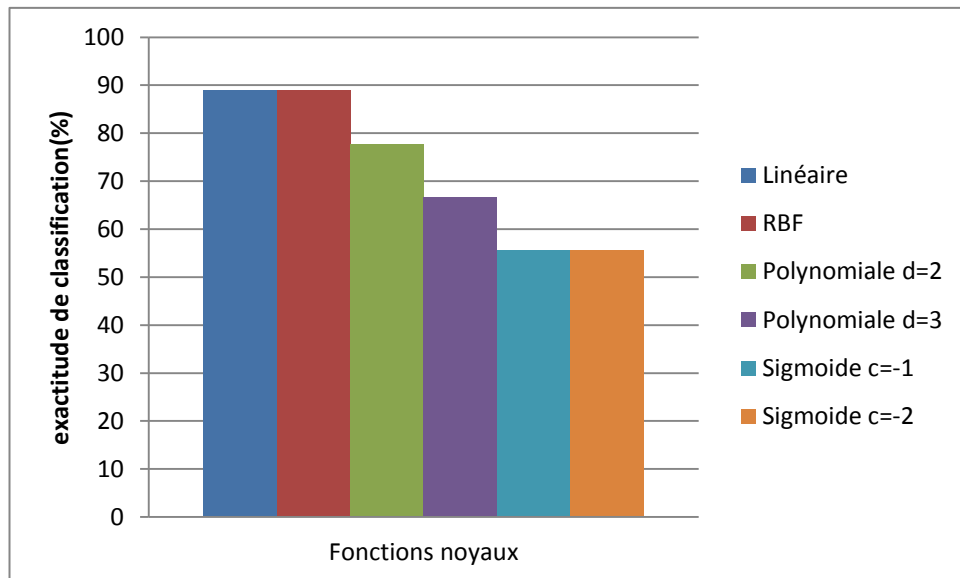


Figure III.27 : représentation de l'exactitude de classification des différentes fonctions noyaux avec la stratégie **un contre un** pour le cas de **pollution non uniforme**.

III.8. Interprétation des résultats :

III.8.1. Pollution uniforme :

Stratégie un contre tous :

Dans le cas de la stratégie « un contre tous », l'utilisation du noyau polynomial d'ordre $d=3$ a donné le meilleur résultat avec neuf bonnes classifications sur onze donc un pourcentage de 81.82%. Les deux noyaux RBF Gaussien et linéaire donnent des résultats similaires avec six bonnes classifications sur onze ce qui correspond à un pourcentage de 54.54%. Tandis que, le noyau sigmoïde ne donne pas de bons résultats pour cette stratégie « un contre tous ». Nous notons un pourcentage d'exactitude de 36.36% pour les deux cas $c=-1$ et $c=-2$.

Stratégie un contre un :

Dans le cas de la stratégie « un contre un », l'utilisation du noyau polynomial d'ordre $d=2$ et $d=3$ donne les meilleurs résultats avec huit bonnes prédictions sur onze donc un pourcentage de 72.72% pour les deux cas. Le noyau sigmoïde pour le paramètre $c = -2$ donne aussi un bon résultat avec un pourcentage de 63.63%. Le noyau sigmoïde avec $c = -1$ et le noyau RBF Gaussien donnent dans ce cas les mêmes résultats avec six classifications adéquates sur onze donc une exactitude de classification de 54.54%. Cependant le noyau linéaire reste le noyau le moins précis pour cette stratégie un contre un avec seulement quatre

classifications justes sur onze ce qui donne un pourcentage de 36.36%.

III.8.2. Pollution non uniforme :

Stratégie un contre tous :

Dans le cas de la stratégie « un contre tous », les quatre noyaux RBF, polynomial d'ordre $d=2$, polynomial d'ordre $d=3$ et sigmoïde avec le paramètre $c = -2$ donnent des résultats performants avec sept classifications justes sur neuf donc un pourcentage de 77.77%. Le noyau sigmoïde de paramètre $c = -1$ donne des résultats moins performants par rapport aux autres noyaux avec six bonnes classifications sur neuf ce qui veut dire 66.66%. Enfin pour le noyau linéaire, les résultats de classifications sont moyens avec 55.55% donc cinq classifications justes sur neuf.

Stratégie un contre un :

Dans le cas de la stratégie « un contre un », les 2 noyaux RBF Gaussien et linéaire donnent de bons résultats avec huit classifications justes sur neuf donc un pourcentage de 88.88%. Le noyau polynomial donne des résultats moins performants que les deux premiers noyaux avec sept bonnes classifications sur neuf pour $d=2$ et six sur neuf pour $d=3$. Pour le noyau sigmoïde avec $c = -1$ et $c = -2$, les résultats de classifications sont intermédiaires avec cinq classifications justes sur neuf ce qui donne un pourcentage de 55.55%.

Conclusion :

Dans la première partie de ce chapitre, nous avons étudié la corrélation qui pourrait exister entre le signal du courant de fuite et la sévérité de la pollution caractérisant l'état de surface de l'isolateur. La technique de décomposition par la DWT est une technique très intéressante qui nous a permis d'étudier les détails en tirant les valeurs de la standard déviation, des maximums, des moyennes et des minimums pour les dix niveaux de décomposition. Les critères obtenus à l'issue de cette étude ont servis comme base de données pour l'entraînement des modèles SVM qu'on a construit par la suite.

Dans la seconde partie de ce chapitre, nous avons présenté la procédure que nous avons utilisée pour la construction des modèles SVM ainsi que les résultats de classification pour les deux stratégies avec les différentes fonctions noyaux.

Nous concluons pour le cas de notre base de données que la stratégie « un contre tous » avec le noyau polynomial d'ordre $d=3$ pour le cas de pollution uniforme est le meilleur modèle à utiliser. Pour le cas de pollution non uniforme, le meilleur modèle a utilisé est celui

de la stratégie « un contre un » avec la fonction noyau RBF Gaussien.

En dernier lieu, il faut mentionner que la méthode SVM est une méthode d'apprentissage supervisé performante et fiable pour la classification.

Chapitre IV :

Application de la LDA pour la classification de la sévérité de pollution



Introduction :

L'objectif de la classification supervisée est d'affecter des individus à des groupes définis a priori à partir des mesures effectuées sur des variables. Dans ce contexte, les analyses discriminantes linéaire et quadratique sont parmi les méthodes les plus populaires. Elles sont fondées sur des hypothèses de multinormalité [66].

L'analyse discriminante classique suppose que l'échantillon d'apprentissage et l'échantillon test qui contient les individus à classer, sont issus d'une même population. Depuis les travaux de Fisher (1936) qui a introduit une règle de discrimination linéaire entre deux groupes, de nombreuses évolutions ont été proposées [67].

Dans le présent chapitre, nous allons présenter la théorie concernant l'analyse discriminante ainsi que les principales méthodes génératives dont la plus connue est l'Analyse Discriminante Linéaire (LDA). Par la suite, on va étudier la classification de la sévérité de pollution des isolateurs de haute tension en utilisant cette fois ci l'Analyse Discriminante Linéaire (LDA).

IV.1. Analyse discriminante :

L'analyse discriminante est le nom donné à la classification dans le cadre supervisé. Elle se distingue de la classification non supervisée par le fait que des observations dont on connaît l'appartenance aux classes sont disponibles pour apprendre la règle de décision (on parlera aussi parfois de classifieur). Ces observations, dites d'apprentissage, supervisent la construction du classifieur [68].

Le problème de l'analyse discriminante est de prédire l'appartenance d'une observation x décrit par p variables explicatives, à une classe parmi k classes $C_1 \dots C_k$ définies a priori. Afin de prédire l'appartenance de l'observation x à une des k classes, nous disposons d'un ensemble d'apprentissage :

$$A = \{(x_1, C_1), \dots, (x_n, C_n), x_i \in \mathbb{R}^p, C_i \in \{1, \dots, k\}\} \quad (\text{IV.1})$$

Où le vecteur x_i est la $i^{\text{ème}}$ observation et C_i indique le numéro de la classe à laquelle x_i appartient. Nous allons donc utiliser l'échantillon A pour construire une règle de décision δ qui associe tout vecteur x de \mathbb{R}^p à la classe correspondante.

$$\delta: \mathbb{R}^p \rightarrow \{1, \dots, k\} \quad (\text{IV.2})$$

$$x \rightarrow C$$

On distingue classiquement deux objectifs principaux en Analyse Discriminante:

-Descriptif : l'aspect descriptif vise à trouver une représentation qui permet l'interprétation des groupes grâce aux variables explicatives.

-Décisionnel: dans ce cas, on cherche à définir la bonne affectation d'un nouveau individu dont on ne connaît que les valeurs des variables explicatives. Cet aspect est particulièrement apprécié dans des domaines où l'aspect diagnostic est essentiel [69].

Dans notre travail, nous nous intéresserons plus particulièrement à l'aspect décisionnel qui est le plus important et souvent le plus délicat. Nous allons tout d'abord rappeler le modèle probabiliste de la discrimination avant de décrire les principales méthodes paramétriques d'Analyse Discriminante.

IV.2. La règle de décision de Bayes :

Les méthodes paramétriques d'Analyse Discriminante font une hypothèse de normalité des classes [69]. Les densités de probabilité des variables explicatives conditionnellement aux classes $p(x|C_i), \forall i = 1, \dots, k$ sont supposées celles de lois normales $N(\mu_i; S_i)$ de moyennes μ_i et de matrice de variance S_i :

$$p(x|C_i) = \frac{1}{(2\pi)^{\frac{p}{2}} (\det S_i)^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (x - \mu_i)^t S_i^{-1} (x - \mu_i)\right) \quad (\text{IV.3})$$

La règle de décision optimale δ^* , dite règle de Bayes, affecte le point x à la classe qui a la probabilité a posteriori maximum. La règle de décision δ^* prend la forme suivante [69]:

$$x \in C_{i^*}, \text{ si } i^* = \operatorname{argmax}_{i=1, \dots, k} \{p(C_i|x)\} \quad (\text{IV.4})$$

La formule de Bayes permet d'obtenir la probabilité que l'individu x provienne de la classe C_i :

$$p(C_i|x) = \frac{p(C_i|x)\pi_i}{p(x)} \quad (\text{IV.5})$$

$$\text{Où } \pi_i \text{ est la probabilité a priori de la classe } C_i \text{ et } p(x) = \sum_{i=1}^k p(x|C_i) \quad (\text{IV.6})$$

Par conséquent et comme les dénominateurs de l'équation (IV.5) sont communs pour chacune des k classes, la règle de Bayes consiste donc à affecter x à la classe C_{i^*} si:

$$i^* = \operatorname{argmax}_{i=1, \dots, k} \{\pi_i p(x|C_i)\} \quad (\text{IV.7})$$

Afin de faciliter l'écriture des règles de décision par la suite, nous allons définir la fonction de coût K_i conditionnellement à la classe $C_i, \forall i = 1, \dots, k$ de la façon suivante:

$$K_i : \mathbb{R}^p \rightarrow \mathbb{R}$$

$$x \rightarrow -2 \log(\pi_i p(x|C_i)) \quad (\text{IV.8})$$

Avec cette notation, la règle de Bayes consiste donc à affecter x à la classe C_{i^*} si [69]:

$$i^* = \operatorname{argmin}_{i=1,\dots,k} \{K_i(x)\} \quad (\text{IV.9})$$

IV.3. Les principales méthodes d'Analyse Discriminante :

Dans cette section, nous allons décrire brièvement les principales méthodes paramétriques de l'Analyse Discriminante.

IV.3.1. Analyse Discriminante Quadratique (QDA) :

Si l'on suppose que les densités conditionnelles des classes $p(x|C_i)$ sont gaussiennes, on obtient alors l'Analyse Discriminante Quadratique (QDA) qui doit son nom au fait qu'elle réalise des séparations quadratiques entre les classes [68]. La fonction de coût K_i prend dans ce cas la forme suivante :

$$K_i(x) = (x - \mu_i)^t S_i^{-1} (x - \mu_i) + \log(\det S_i) - 2\log(\pi_i) + C^{te} \quad (\text{IV.10})$$

Où la constante C^{te} représente une quantité commune à toutes les classes et n'intervient donc pas dans la règle de décision.

La figure IV.1 (gauche) montre la frontière de décision de QDA pour un jeu de données artificielles de dimension 2.

IV.3.2. Analyse Discriminante Linéaire (LDA) :

L'Analyse Discriminante Linéaire (LDA) est basée sur le critère de la maximisation de l'écart entre les classes et la minimisation de la dispersion autour de la moyenne de chaque classe [68]. Cette analyse combine linéairement les données de manière à maximiser la discrimination entre les classes. Le nombre d'axes générés par la LDA est inférieur ou égal au nombre d'étiquettes différentes.

L'analyse discriminante maximise la répartition des N observations dans leurs classes respectives, contrairement à l'analyse factorielle qu'on va voir par la suite qui maximise la variance des N observations en fonction des variables [68].

La LDA est généralement considérée comme une méthode d'analyse discriminante à part entière plutôt que comme une méthode de régularisation de QDA, car elle fait l'hypothèse supplémentaire d'égalité des matrices de variances, i.e. $\forall i = 1, \dots, k, S_i = S$.

L'analyse discriminante linéaire doit également son nom au fait qu'elle réalise des séparations linéaires entre les classes [68].

Les termes $\log(\det S)$ et $x^t S^{-1} x$ ne dépendent plus des classes comme dans le cas de QDA, mais ils peuvent être utilisés pour le calcul de la règle de décision. La fonction de coût K_i prend la forme suivante :

$$K_i(x) = \mu_i^t S^{-1} x - 2 \mu_i^t S^{-1} x - 2 \log(\pi_i) + C^{te} \quad (\text{IV.11})$$

Les estimateurs des proportions et des moyennes sont les mêmes que ceux de QDA et l'estimateur du maximum de vraisemblance de la matrice S est la matrice de covariance intra-classe empirique W :

$$\hat{S} = W = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^n s_{ij} (x_j - \hat{\mu}_i) (x_j - \hat{\mu}_i)^t \quad (\text{IV.12})$$

La figure IV.1 (droite) montre la frontière de décision de LDA pour des données artificielles de dimension 2. En pratique, l'analyse discriminante linéaire est fréquemment utilisée car elle offre un bon compromis entre pertinence et complexité. D'autre part, elle fournit des résultats robustes aux fluctuations sur les hypothèses de normalité des classes et d'égalité des matrices de variance [68].

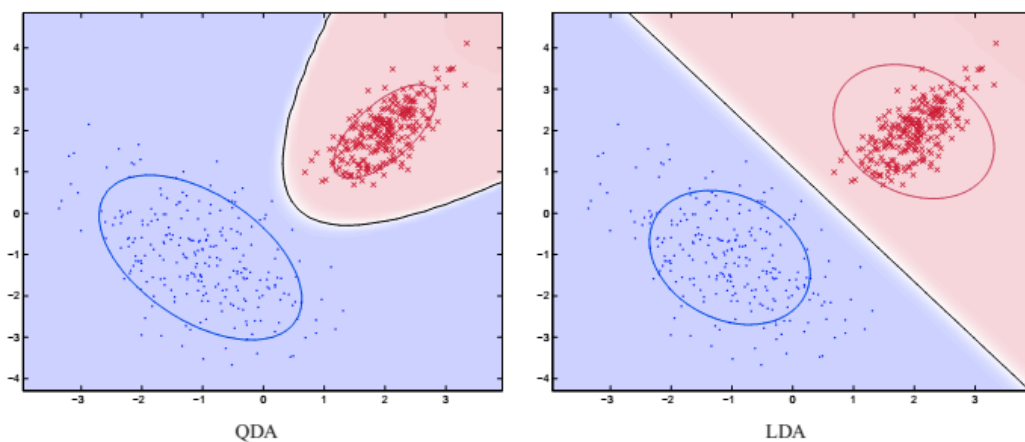


Figure IV.1: Frontières de décision de l'Analyse Discriminante Quadratique (QDA) et de l'Analyse Discriminante Linéaire (LDA) sur un même jeu de données en dimension 2.

IV.3.3. Analyse Factorielle Discriminante (FDA) :

L'Analyse Factorielle Discriminante combine une étape qui relève de la réduction de dimension et une étape de discrimination [69]. En effet, effectuer une Analyse Factorielle Discriminante consiste à projeter les données de \mathbb{R}^p sur les $d = (k - 1)$ axes discriminants qui maximisent le rapport de la variance interclasse et de la variance intra-classe, puis

d'apprendre la règle de décision δ^* sur les données projetées. Nous avons donc besoin de définir les matrices de variance inter et intra-classe. La matrice de variance interclasse est définie par:

$$B = \sum_{i=1}^k \pi_i (\mu_i - \mu) (\mu_i - \mu)^t \quad (\text{IV.13})$$

$$\text{Où } \mu = \sum_{i=1}^k \pi_i \mu_i \quad (\text{IV.14})$$

D'autre part, la matrice de variance intra-classe est définie par:

$$W = \sum_{i=1}^k \pi_i \Sigma_i \quad (\text{IV.15})$$

Notons également que le théorème de Huyghens nous permet d'obtenir la relation suivante qui lie les matrices de variance inter et intra-classe à la matrice de variance totale:

$$S_{totale} = B + W \quad (\text{IV.16})$$

Nous souhaitons trouver une représentation des données qui permet de discriminer les groupes le mieux possible. Pour ce faire, il faut que les projections des k centres de gravité soient le plus séparée possible, tandis que les données de chaque classe doivent se projeter de façon groupée autour du centre de gravité de leur classe. Nous recherchons donc une représentation des données qui maximise la variance interclasse et qui minimise la variance intra-classe [69]. Avec les notations et résultats précédents, les axes de la projection recherchés satisfont le problème d'optimisation suivant: $\max_u \frac{u' B u}{u' S_{totale} u}$ (IV.17)

On sait que ce maximum est atteint pour u vecteur propre $S_{totale}^{-1} B$ associé à sa plus grande valeur propre. La figure IV.2 illustre le choix d'un axe de projection permettant de discriminer au mieux les classes [69].

Une fois la projection déterminée, on peut alors effectuer une Analyse Discriminante Linéaire (ou Quadratique). Cette stratégie qui combine réduction de dimension et discrimination est souvent profitable car les données de chaque classe n'occupent en général pas la totalité de l'espace et cela permet de réduire le nombre de paramètres à estimer. Cette méthode se révèle relativement efficace sur des données de grande dimension.

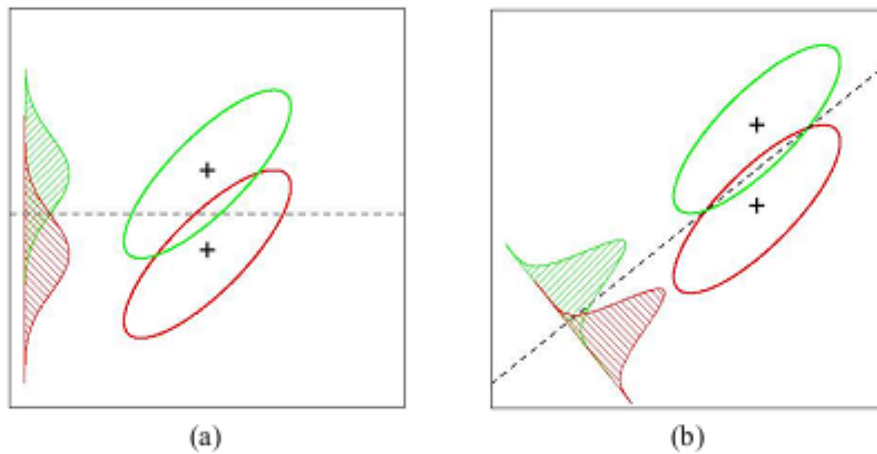


Figure IV.2: l'axe principal de la figure(a) ne permet pas de discriminer efficacement les deux groupes alors que celui de la figure(b) possède un bon pouvoir discriminant.

IV.4. La classification avec l'analyse discriminante linéaire :

En se basant sur la théorie de l'analyse discriminante linéaire, le principe de classification est le suivant :

- Les vecteurs à classifier x_i , $i = 1 \dots n$.
- Les classes correspondantes C_i , $i=1, \dots, m$. Tel que m est le nombre des classes.
- Le centre de gravité des vecteurs x_i : $\mu = \frac{1}{n} \sum_{k=1}^n x_k$ (IV.18)
- le centre de gravité des classes c_i : $\mu_i = \frac{1}{n_i} \sum_{x_k \in C_k} x_k$ (IV.19)

La LDA nous permet de calculer les mesures suivantes :

$$\text{La variance de la } i^{\text{ème}} \text{ classe : } S_i = \sum_{x_k \in C_k} (x_k - \mu_i)(x_k - \mu_i)^t \quad (\text{IV.20})$$

$$\text{La matrice de variance intra-classes : } W = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} (x_k - \mu_i)(x_k - \mu_i)^t \quad (\text{IV.21})$$

$$\text{La matrice de variance interclasse : } B = \frac{1}{n} \sum_{i=1}^m (\mu_k - \mu)(\mu_k - \mu)^t \quad (\text{IV.22})$$

$$\text{La matrice totale de variance : } S_{\text{totale}} = W + B \quad (\text{IV.23})$$

Fisher définit la séparation entre un certain nombre de classes en maximisant le rapport, tel que : $S = \frac{B^2}{W^2}$ (IV.24)

Le numérateur de l'expression IV.24 traduit la séparation des classes, alors que le dénominateur reflète la compacité des classes. Plus le rapport est grand, plus les classes sont séparées. Ainsi, d'une manière générale, pour m classes, et pour un paramètre α

donné, l'expression précédente devient :

$$S = \sum_{i=1}^m \sum_{j=1}^{m-1} \frac{\mu_i(\alpha) - \mu_j(\alpha)}{n_i \sigma_i^2(\alpha) - n_j \sigma_j^2(\alpha)} \quad (\text{IV.25})$$

σ_i^2 est la $i^{\text{ème}}$ variance des vecteurs de la classe C_i , tel que :

$$\sigma_i^2(\alpha) = \frac{1}{n_i} \sum_{j=1}^{n_i} [x_{ij}(\alpha) - \mu_i(\alpha)]^2 \quad (\text{IV.26})$$

De ce fait, le critère de Fisher a pour but de maximiser la variance interclasse, et de minimiser la variance intra-classe. Ainsi, dans le cas des deux classes et pour un paramètre, le critère de Fisher s'exprime de la manière suivante :

$$S(\alpha) = \frac{\mu_1(\alpha) - \mu_2(\alpha)}{n_1 \sigma_1^2(\alpha) - n_2 \sigma_2^2(\alpha)} \quad (\text{IV.36})$$

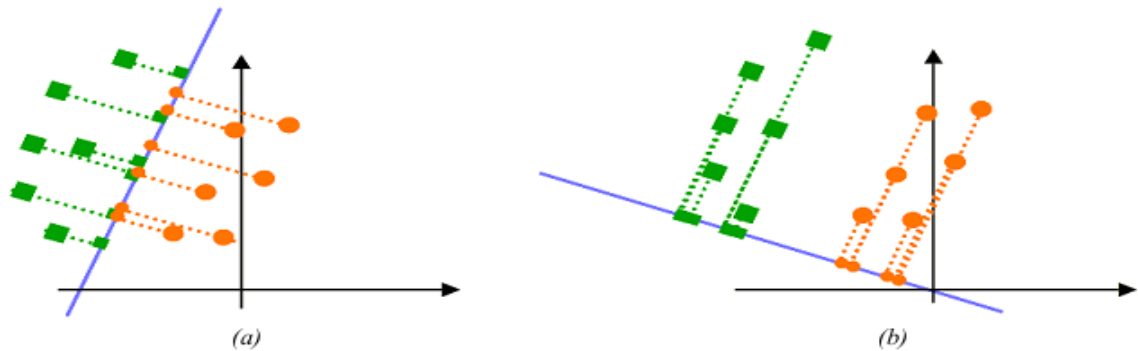


Figure IV.3: Exemple d'une séparation de deux classes avec la droite discriminante de Fisher. (a) mauvaise séparation. (b) bonne séparation.

IV.5. Application de LDA pour la classification de la sévérité de pollution des isolateurs :

IV.5.1. Composition de la matrice d'apprentissage :

Les vecteurs d'apprentissage que nous allons utiliser pour la classification avec LDA sont les mêmes que nous avons utilisé précédemment avec les SVM. Nous rappelons que les éléments de ces vecteurs sont extraits à partir des détails des signaux du courant de fuite en calculons leurs standard déviation, minimum, maximum et leurs moyennes.

La base de donnée qu'on a utilisé pour l'apprentissage et la validation est la même qu'on a utilisé avec la classification par les SVM.

IV.5.2. Validation et résultat de classification :

Les résultats obtenus par la classification en utilisant l'Analyse Discriminante Linéaire

sont données dans les deux tableaux suivants:

Echantillon n°	Conductivité (mS/cm)	Classe assignée	Résultat obtenu
1	1.2	1	1
2	1.2	1	1
3	3.5	1	1
4	3.5	1	3*
5	6	2	1*
6	6	2	3*
7	6	2	2
8	10	3	2*
9	10	3	3
10	20	3	3
11	20	3	2*
Exactitude de classification (%)			54.54

Tableau IV.1: Résultats de la classification en utilisant la LDA dans le cas de pollution uniforme.

Echantillon n°	Conductivité (mS/cm)	Classe assignée	Résultat obtenu
1	1.2	1	1
2	3.5	1	1
3	3.5	1	2*
4	1.2	1	1

5	6	2	2
6	6	2	1*
7	10	3	2*
8	20	3	1*
9	20	3	2*
Exactitude de classification (%)			44.44

I *Tableau IV.2: Résultats de la classification en utilisant la LDA dans le cas de pollution*

La classification avec l'analyse discriminante linéaire dans le cas de la pollution uniforme a donné six bonnes classifications sur un ensemble de onze échantillons ce qui correspond à 54.54%.

En ce qui concerne la pollution non uniforme, la classification a donné quatre bonnes classifications sur un ensemble de neuf échantillons donc un pourcentage d'exactitude de classification de 45.44%.

IV.6. Comparaison entre SVM et LDA :

Dans le but de comparer les deux méthodes de classification SVM et LDA, on va choisir d'abord la meilleur stratégie et la meilleur fonction noyau pour la méthode SVM pour les deux cas de pollution, par la suite on va la comparer avec LDA.

Les résultats obtenus dans le précédent chapitre montrent que la fonction noyau la plus performante pour le cas de pollution uniforme est la fonction noyau polynomial d'ordre $d=3$ pour la stratégie un contre tous avec un pourcentage d'exactitude de 81.82%.

Pour le cas de pollution non uniforme, le meilleur résultat est celui qu'on a obtenu avec la fonction noyau RBF avec la stratégie un contre un dont le pourcentage de classification est de 88.89%.

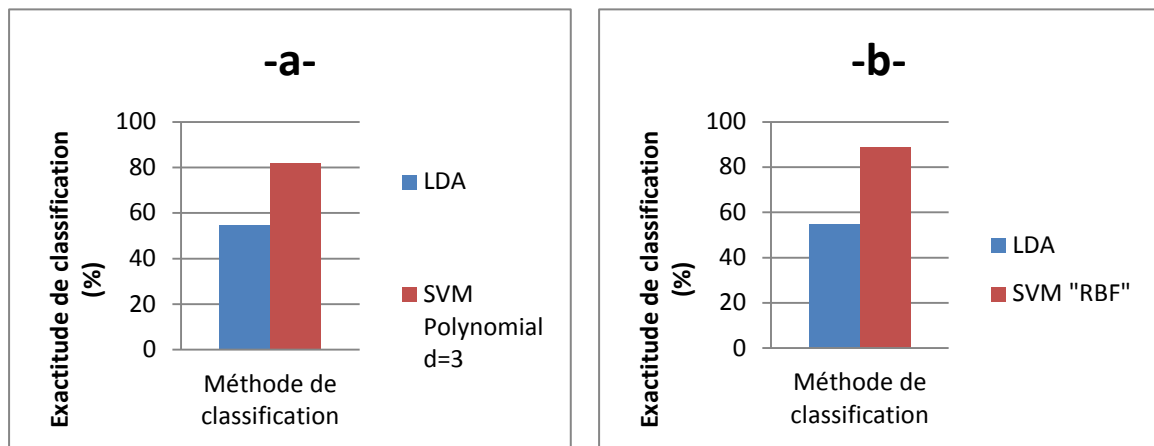


Figure IV.4 : Comparaison entre les deux méthodes de classification SVM et LDA.

(a) pollution uniforme (b) pollution non uniforme.

Pour les deux cas de pollution, les résultats obtenus montrent clairement que la méthode SVM surpasse la méthode LDA. Ce résultat peut être expliqué par le fait que la méthode SVM cherche l'hyperplan qui sépare les exemples positifs des exemples négatifs, en garantissant que la marge entre le plus proche des positifs et des négatifs soit maximale, alors intuitivement, cela garantit un bon niveau de généralisation car de nouveaux exemples pourront ne pas être trop similaires à ceux utilisés pour trouver l'hyperplan mais être tout de même situés franchement d'un côté ou de l'autre de la frontière [71].

Conclusion :

Dans ce quatrième chapitre nous avons parlé sur l'analyse discriminante et ces trois méthodes de classification QDA, LDA et l'analyse discriminante factorielle. Ensuite, nous avons utilisé la LDA pour faire une classification du degré de sévérité de pollution des isolateurs en utilisant la même base de données qu'on a utilisée avec la méthode SVM.

En deuxième lieu, on a pris les meilleurs résultats de classification par la méthode SVM afin de les comparer avec la méthode LDA. Nous concluons d'après les résultats trouvés que les SVM donnent des résultats plus performants par rapport à la LDA pour les deux types de pollution avec des pourcentages plus élevés.

Finalement, on peut dire aussi que l'analyse discriminante linéaire (LDA) est une méthode d'apprentissage supervisé qui peut être utilisée pour la classification sauf qu'elle est généralement moins performante et fiable par rapport à la méthode SVM.

Chapitre V :

Caractérisation de l'état de surface des isolateurs par la technique des ondelettes

Introduction :

Dans ce chapitre, nous allons appliquer la technique des ondelettes pour la caractérisation de l'état de surface des isolateurs pollués. En premier lieu, nous exploitons les signaux de la tension appliquée pour tracer les scalogrammes et les courbes des coefficients d'ondelettes pour différentes conductivités (3.5, 10 et 20 mS/cm).

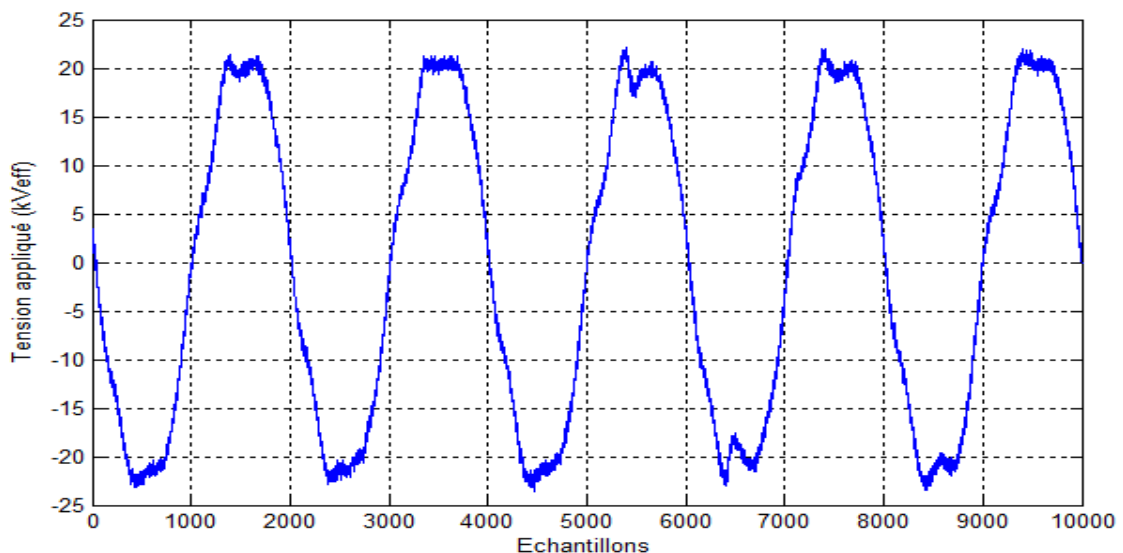
En deuxième lieu, nous traçons les scalogrammes et les courbes des coefficients d'ondelettes pour différentes concentrations en kaolin (40, 100 et 200 g/l) en gardant la conductivité volumique et la tension appliquée constantes.

V.1. Pollution uniforme :

Le traitement des signaux de la tension appliquée est effectué en utilisant la technique de la CWT.

V.1.1 Influence de la conductivité volumique:

Pour étudier l'influence de la conductivité volumique on a tracé les scalogrammes des signaux du courant de fuite pour trois conductivités différentes 3.5, 10 et 20 mS, en fixant la concentration du kaolin (100 g/l) et la tension appliquée (15 kV_{eff}).

**(a)**

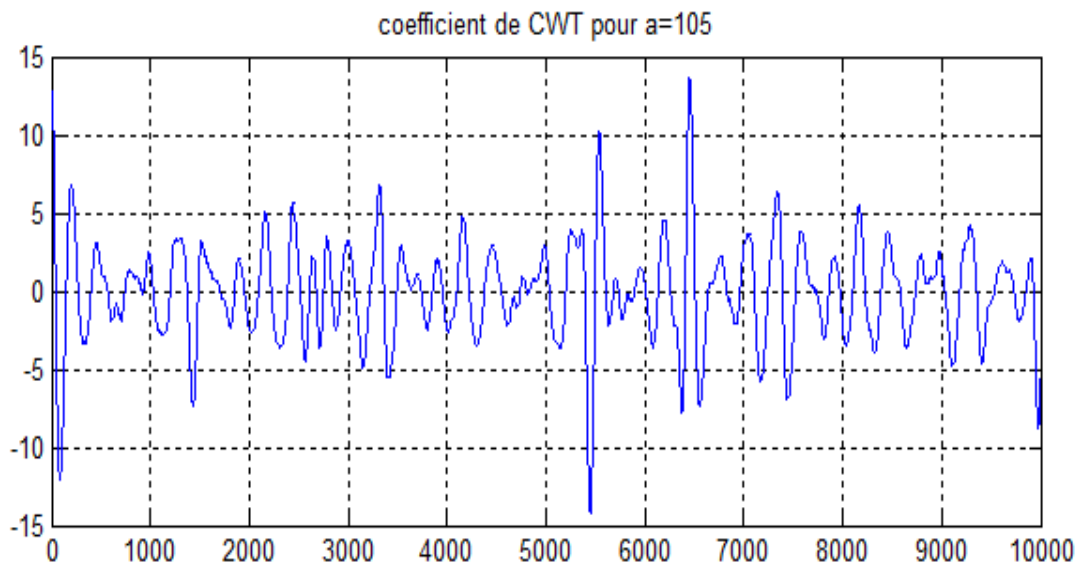
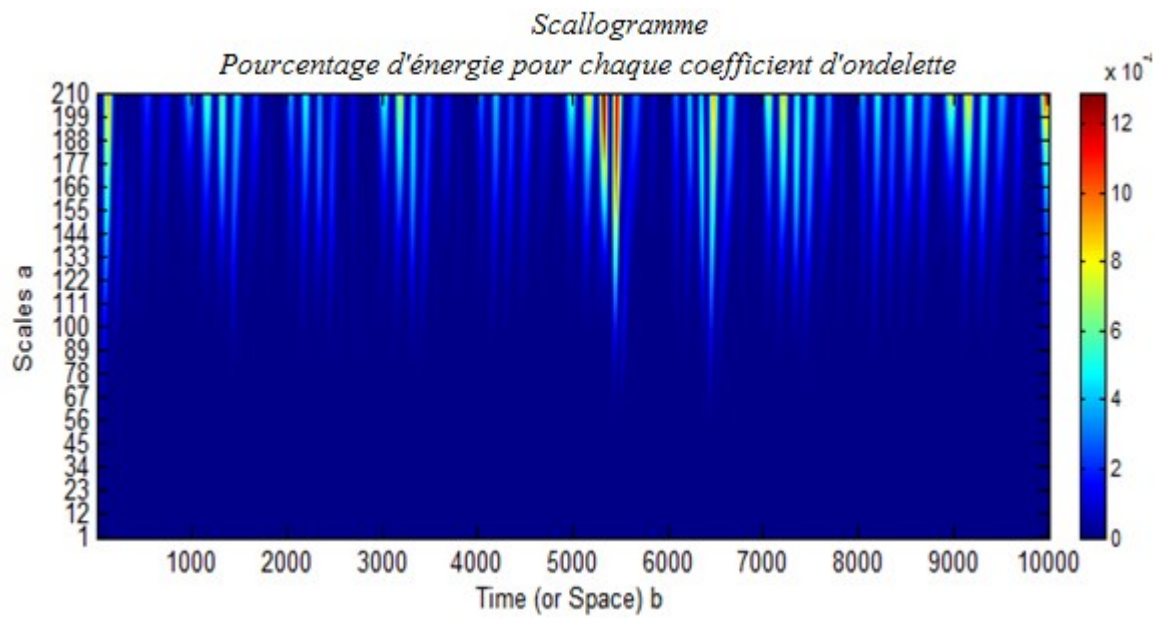


Figure V.1 : (a) Forme d'onde de la tension appliquée (b) Scalogramme correspondant (c) Coefficients de la CWT pour $a = 105$ Conductivité 3.5 mS/cm.

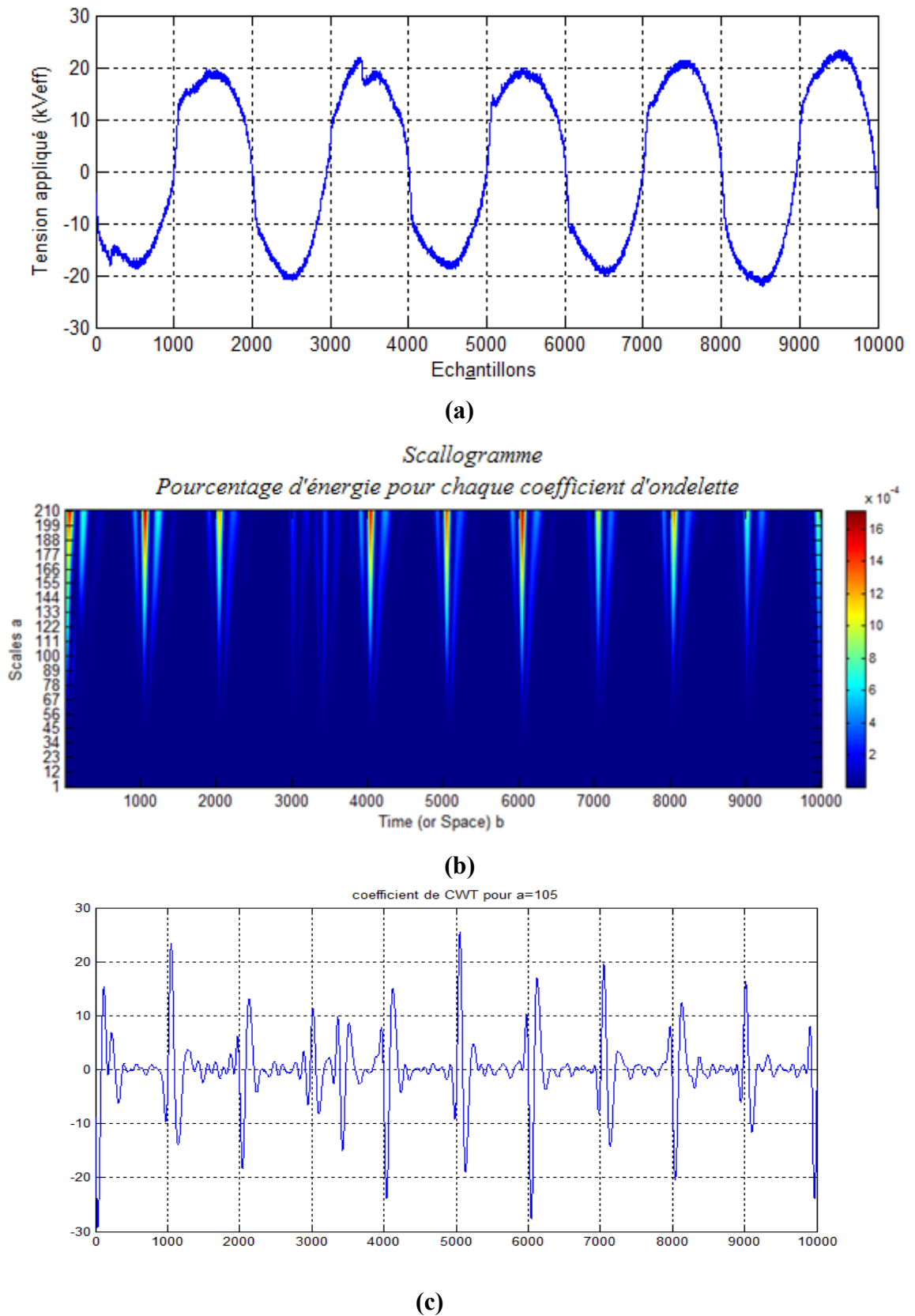
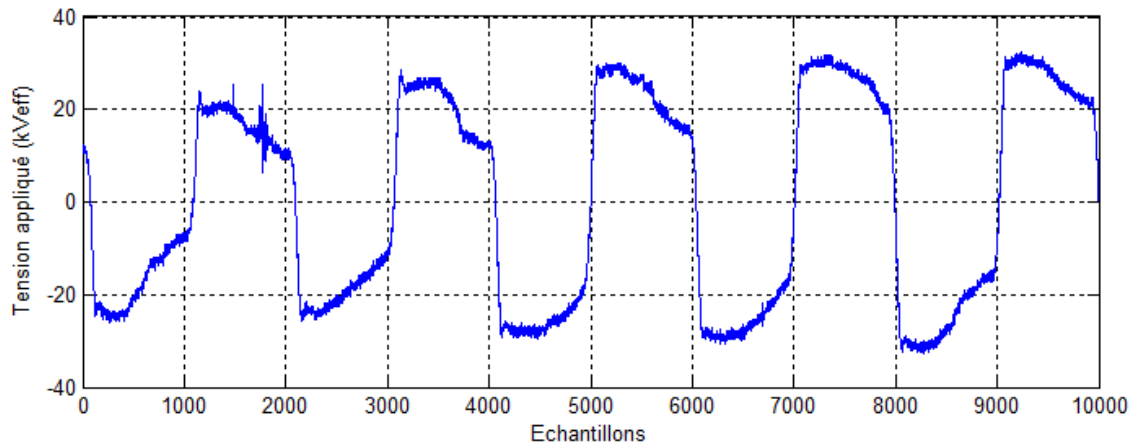
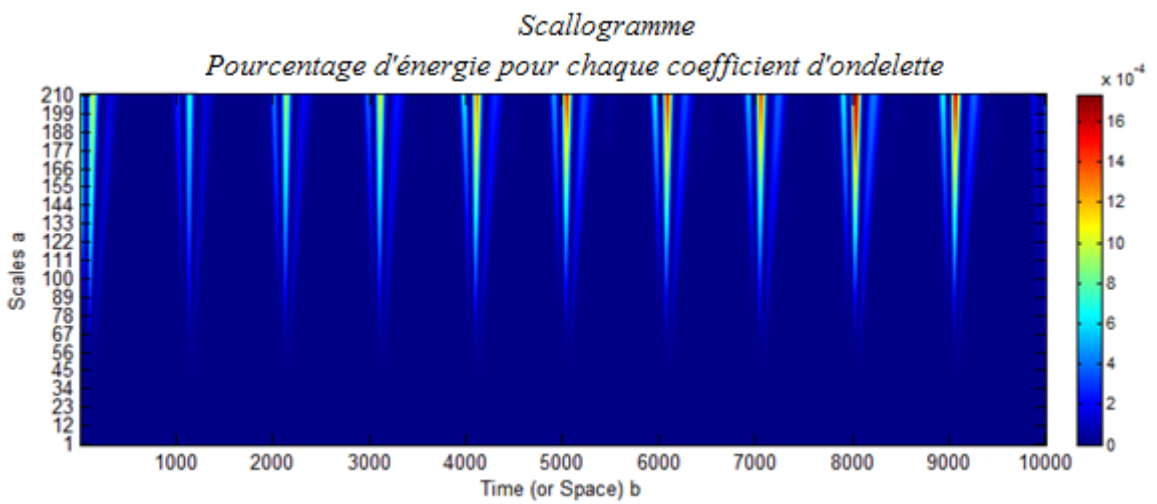


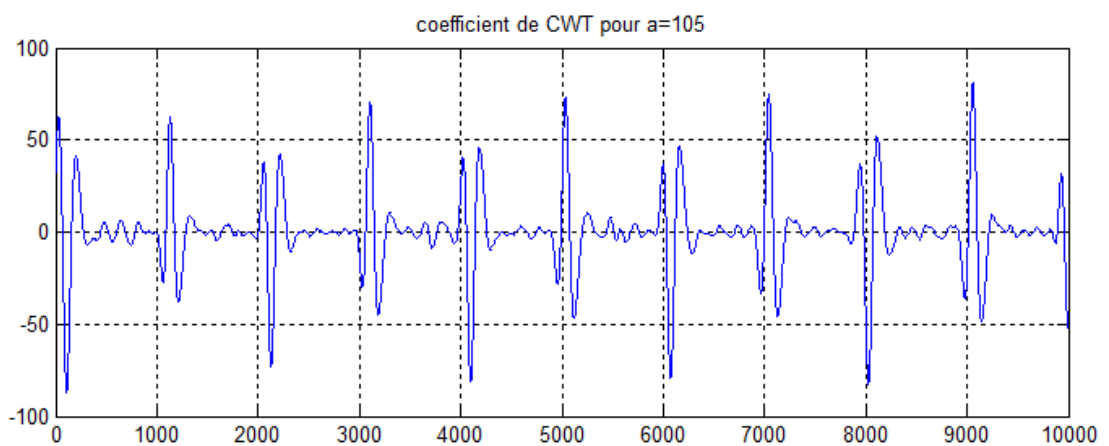
Figure V.2 : (a) Forme d'onde de la tension appliquée (b) Scalogramme correspondant (c) Coefficients de la CWT pour $a = 105$. Conductivité 10 mS/cm



(a)



(b)



(c)

Figure V.3: (a) Forme d'onde de la tension appliquée (b) Scalogramme correspondant (c) Exemple des Coefficients de la CWT pour $a = 105$. Conductivité 20 mS/cm

Discussion des résultats :**Conductivité volumique 3.5mS/cm :**

Pour la conductivité de 3.5 mS/cm, nous remarquons que le signal de la tension appliquée affiche des distorsions au niveau de la crête et au voisinage de l'axe des temps. Son scalogramme montre que les coefficients ont des valeurs maximales (couleur rouge intense). Pour des échelles moyennes et grandes (de 133 à 210) c-à-d des perturbations de basses et moyennes fréquences apparaissent. Cela signifie que ces perturbations oscillent à des fréquences faibles et moyennes indiquant que l'activité des décharges parallèles à la surface est moindre.

Les coefficients de la CWT pour l'échelle $a = 105$ montrent qu'une petite corrélation entre l'oscillation de l'ondelette de Daubechies 4 et celle de la transformée en ondelette de la tension. Cette corrélation est plus en plus évidente dans les deux points qui apparaissent au milieu du signal de la CWT et aux extrémités.

Conductivité volumique 10 mS/cm :

Pour la conductivité de 10 mS/cm, nous remarquons des distorsions dans tout le signal de la tension appliquée plus précisément dans les crêtes. Son scalogramme montre que les coefficients ont une valeur maximale dans plusieurs régions du signal pour des échelles plus faibles (de 122 à 210). Cela est expliqué par l'apparition des fréquences un peu plus élevées indiquant l'augmentation des décharges partielles au niveau de la surface du modèle. L'apparition de ces décharges partielles, engendre l'assèchement du modèle.

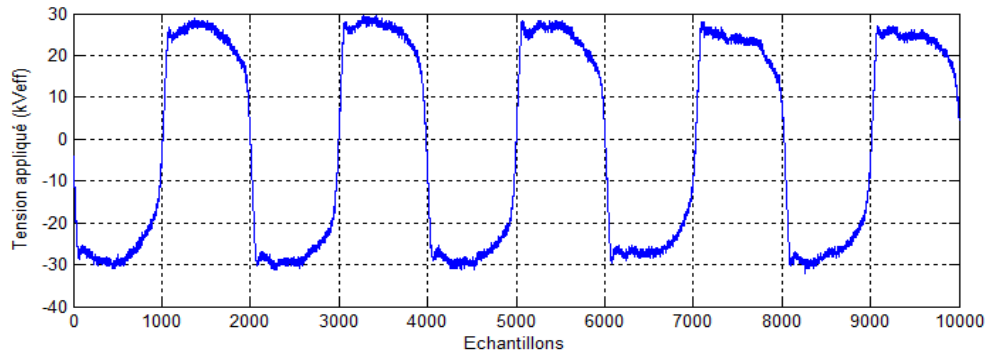
Les coefficients de la CWT pour une échelle de $a=105$ montre une similitude entre l'ondelette analysante Daubechies 4 et la transformée en ondelette continue du signal étudié.

Conductivité volumique 20 mS/cm :

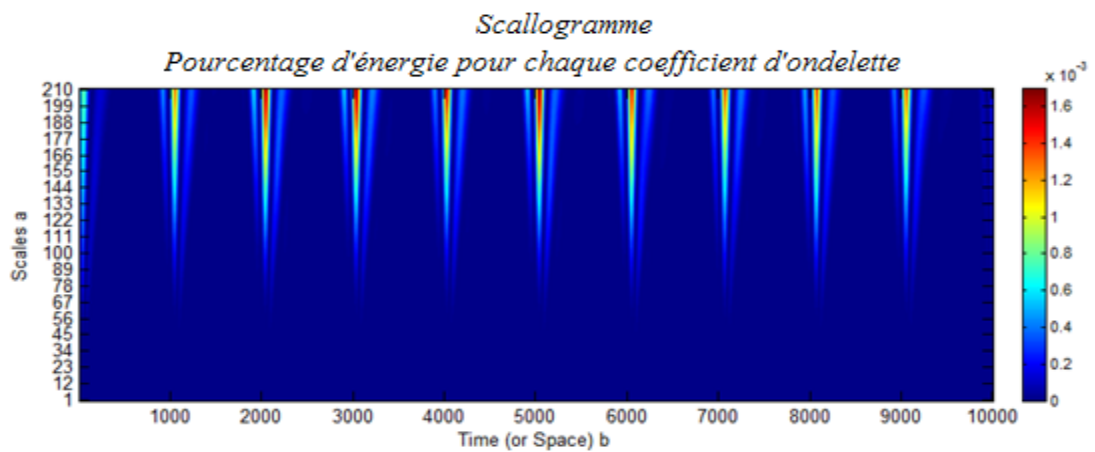
Pour cette conductivité, nous constatons que les perturbations sur l'onde de la tension appliquée s'intensifient avec des chutes de tension plus remarquables au niveau de la crête. Le scalogramme montre que ces chutes de tension sont causées à des instants espacés de la même manière ou on peut remarquer des coefficients très élevés de la CWT pour des grandes échelles allons de 144 à 210. Cela est expliqué par l'apparition des basses fréquences et donc des perturbations et des décharges moindre à la surface.

V.1.2. Influence de la concentration en kaolin :

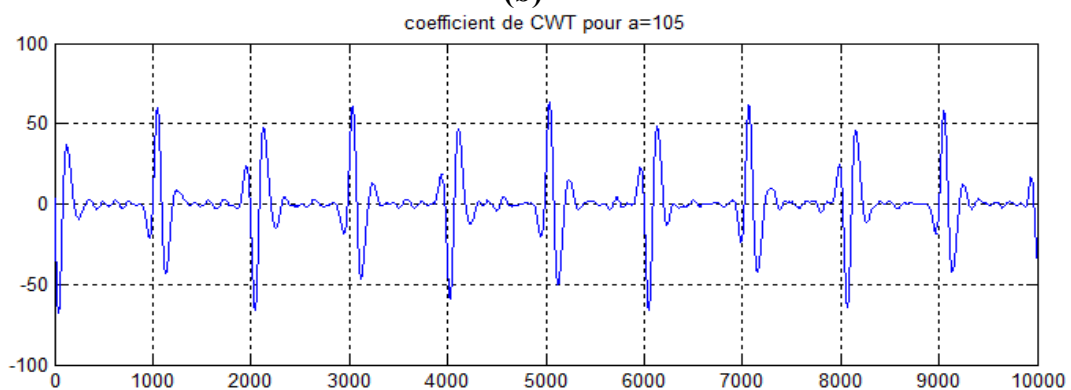
Pour étudier l'influence de la concentration en kaolin nous, avons tracé les scalogrammes des signaux de la tension appliquée pour trois concentrations différentes 40, 100 et 200 g/l, en fixant la conductivité volumique à 20 mS/cm.



(a)

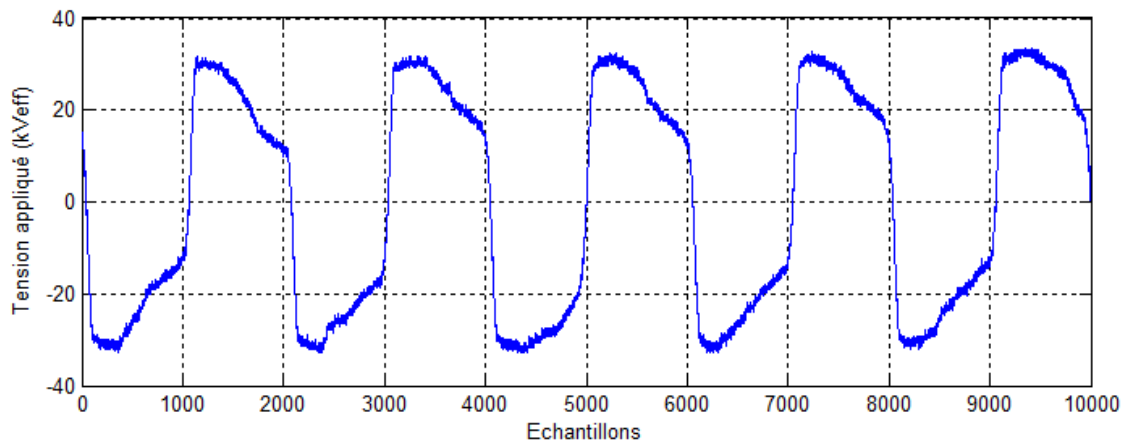


(b)

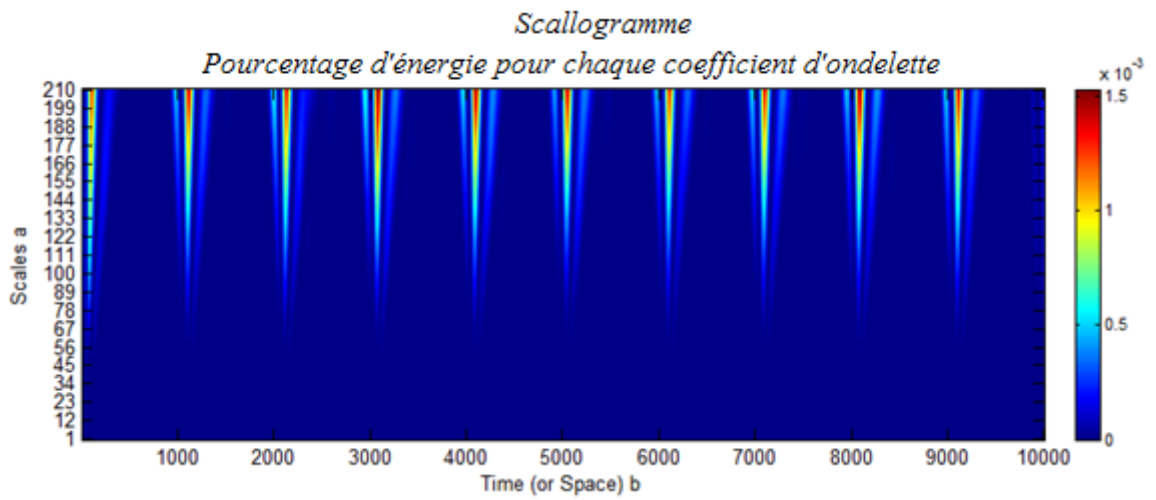


(c)

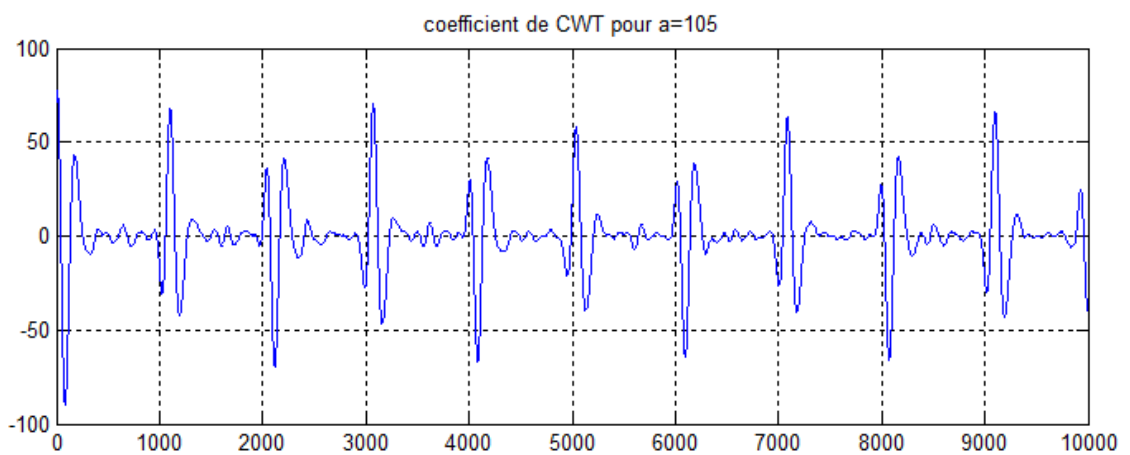
Figure V.4 : Concentration en kaolin de 40g/l : (a) Forme d'onde de la tension appliquée 25 kV_{eff} (b) Scalogramme correspondant (c) Exemple des Coefficients de la CWT pour a = 105



(a)

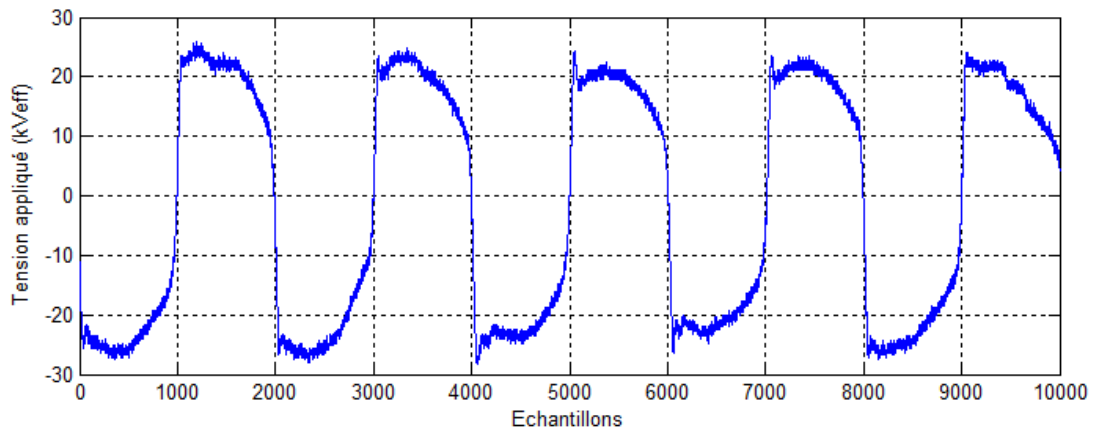


(b)

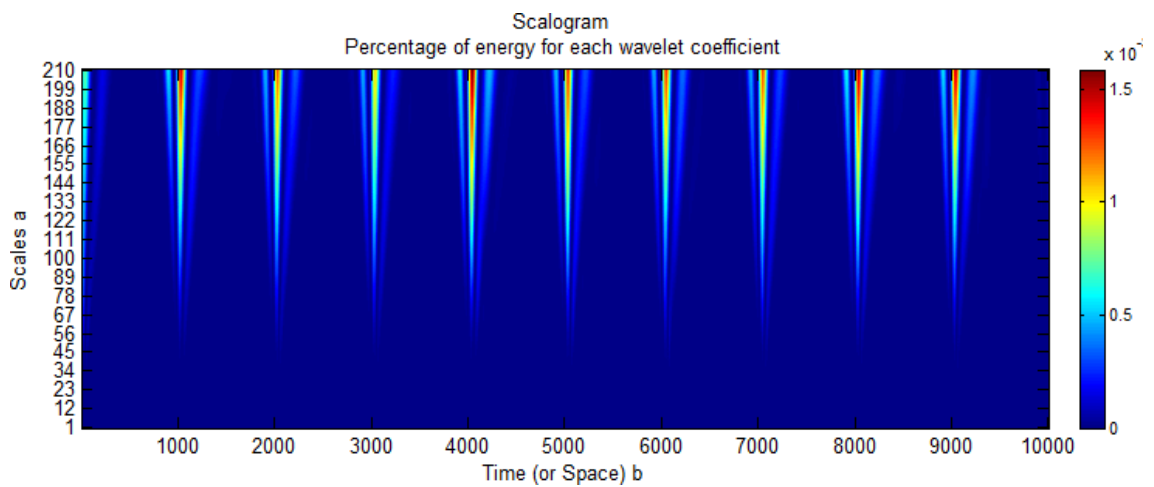


(c)

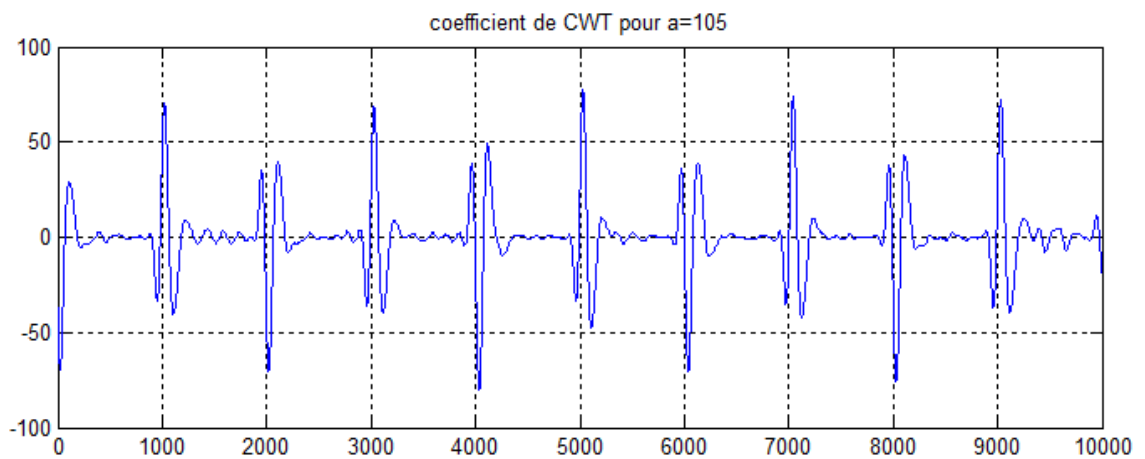
Figure V.6 : Concentration en kaolin de 100g/l : (a) Forme d'onde de la tension appliquée 20 kV_{eff} (b) Scalogramme correspondant (c) Coefficients de la CWT pour $a = 105$



(a)



(b)



(c)

Figure V.5 : Concentration en kaolin de 200 g/l : (a) Forme d'onde de la tension appliquée 25 kV_{eff} (b) Scalogramme correspondant (c) Coefficients de la CWT pour $a = 105$

Discussion des résultats :**Concentration en kaolin de 40g/l :**

Pour une conductivité de 20mS/cm et une concentration en kaolin de 40g/l, les distorsions occupent pratiquement toutes les régions du signal de la tension appliquée. Le scalogramme indique que les coefficients de la CWT sont très intenses pour des échelles plus élevés à partir de 166 à 210, ce qui justifie l'apparition des décharges partielles de basses fréquences.

Les coefficients de la CWT pour une échelle de $a=105$ montrent une similitude entre l'ondelette analysante Daubechies 4 et la transformée continue du signal appliqué au niveau des crêtes du signal de la tension appliquée ou les distorsions sont plus visibles.

Concentration de kaolin 100g/l :

Pour cette concentration en kaolin de 100g/l et avec une conductivité élevée, les perturbations sont remarquables dans toutes les régions du signal de la tension appliquée. Le scalogramme indique que les coefficients de la CWT sont très intenses pour des échelles un peu élevés entre 155 et 210, ce qui est expliqué par l'apparition des décharges partielles de basses fréquences.

Les coefficients de la CWT pour une échelle de $a=105$ montrent une corrélation entre l'ondelette analysante Daubechies 4 et la CWT de la tension appliquée.

Concentration de kaolin 200g/l :

Pour une concentration en kaolin de 200g/l, les distorsions occupent pratiquement toutes les régions du signal de la tension appliquée. Le scalogramme indique que les coefficients de la CWT sont très intenses pour des échelles plus réduites à partir de 144. Ce qui peut être traduit par l'apparition des décharges partielles de basses et moyennes fréquences.

Les coefficients de la CWT pour une échelle $a=105$ montre une bonne corrélation entre l'ondelette analysante Daubechies 4 et la transformée continue du signal au niveau des crêtes positives et négatives du signal de la tension appliquée.

L'apparition des fréquences un peu plus élevées indiquent l'augmentation des décharges partielles au niveau de la surface de l'isolateur pollué.

Conclusion :

Le scalogramme représente une projection des coefficients calculés pour des échelles données sur un plan orthogonal. D'après l'étude faite dans ce chapitre, l'application de la CWT sur les signaux de la tension appliquée et le scalogramme montre leur efficacité pour le diagnostic de la sévérité de pollution des isolateurs.

Concernant l'analyse par le scalogramme, nous concluons que l'augmentation de la conductivité peut être observée par la couleur des coefficients qui spécifient leurs valeurs selon qu'elles soient faibles ou élevées. Un coefficient élevé pour une grande échelle indique une oscillation du signal à basse fréquence et inversement pour une échelle de plus petite valeur. Dans notre cas, la fréquence des perturbations occasionnées sur le signal de la tension augmente avec la sévérité de pollution.

Le diagnostic de l'état de surface des isolateurs peut être fait à travers une échelle bien déterminée (105 dans notre cas) qui nous renseigne sur la pollution du modèle grâce aux coefficients de cette même échelle. Lorsque la corrélation entre la CWT du signal et l'ondelette analysante est quasi inexistante avec une faible amplitude, la conductivité de la surface est faible. Cette corrélation est autant plus apparente avec l'augmentation de la conductivité volumique qui fait augmenter l'amplitude des coefficients de la CWT ainsi que le nombre d'oscillations qui ressemblent beaucoup à celles de l'ondelette choisie.

La concentration du kaolin n'as pas une grande influence sur l'amplitude des coefficients de la CWT. Pour les trois cas de concentration de kaolin étudiés, nous concluons aussi qu'il y a une corrélation entre l'ondelette analysante (Daubechies 4) et les coefficients de la CWT.

Conclusion

générale



Ce travail nous a permis d'appliquer de nouvelles méthodes de traitement de signal pour la classification de la sévérité de pollution des isolateurs HT, en vue de diagnostiquer l'état de surface de l'isolateur tout en se référant à la norme CEI 71 pour classer le degré de sévérité de pollution à trois classes différentes : faible pollution, moyenne pollution et forte pollution.

Nous avons effectué une synthèse bibliographique sur les plus récents travaux concernant la classification en utilisant les supports vecteur machine (SVM) dans plusieurs domaines. Cette étude nous a donné une idée sur l'utilisation de cette méthode (SVM) et ces techniques pour faire la classification et la prévision.

Par la suite, nous avons approfondi nos connaissances sur les types d'apprentissage, la théorie des SVM comme une des méthodes d'apprentissage supervisé et son utilisation pour la classification dans le cas binaire et le cas multi-classes afin de les utiliser pour la classification de sévérité de pollution des isolateurs de haute tension par la suite.

En premier lieu, nous avons décomposé les signaux du courant de fuite par la DWT pour avoir les caractéristiques des dix premiers détails. Ensuite, l'étude de la variation de la standard déviation, de maximum, de minimum et de la moyenne en fonction du niveau de décomposition et en fonction de la conductivité de l'agent polluant (1.2, 3.5, 6, 10 et 20 mS/cm) nous renseigne sur la sévérité de pollution. Ces dernières caractéristiques sont utilisées comme entrées pour le modèle SVM que nous avons construit. Les résultats de cette classification montrent que la meilleure stratégie dans le cas de pollution uniforme est la stratégie «un contre tous» avec la fonction noyau polynomiale d'ordre $d=3$ avec un pourcentage de 81.82 %. Pour le cas de pollution non uniforme la meilleure stratégie est «un contre un» avec la fonction noyau RBF et un pourcentage de 88.89%.

En deuxième lieu, nous avons approfondie nos connaissances sur l'analyse discriminante linéaire (LDA) et son utilisation pour la classification. Ensuite, on a effectué la classification de la sévérité de pollution par la LDA sur les mêmes échantillons utilisés dans le troisième chapitre avec les SVM afin de comparer les deux méthodes. Pour le cas de pollution uniforme la LDA a donné un pourcentage de 54.54% et seulement 44.45% dans le cas de pollution non uniforme. La comparaison des résultats de classifications des deux méthodes ont montrés que la classification par les SVM est plus performante et plus exacte par rapport à

la LDA. Finalement, on peut dire aussi que l'analyse discriminante linéaire (LDA) est une méthode d'apprentissage supervisé qui peut être utilisée pour la classification sauf qu'elle est moins performante que la méthode SVM.

Finalement, Le diagnostic de l'état de surface de l'isolateur par la technique des ondelettes nous a également permis de caractériser la sévérité de la pollution grâce au signal de la tension appliquée (pollution uniforme) via la CWT par l'intermédiaire du scalogramme qui est une projection des coefficients calculés pour des échelles données sur un plan orthogonal et pendant l'activité des décharges parallèles. En effet, plus la conductivité volumique de la solution polluante et la concentration en kaolin augmentent, plus les distorsions de la tension tendent vers les hautes fréquences. Mais l'influence de la conductivité volumique est plus remarquable que celle de la concentration de kaolin.

Références bibliographiques

- [1] M. Tegar, « Modélisations d'isolateurs pollués soumis à divers paramètres électro géométriques », Thèse de Doctorat d'Etat, Département d'électrotechnique, Ecole Nationale Polytechnique, Alger, 2003.
- [2] M. A. B. Elkosairg, E. E. I. Sharaine, M. M Awad, H. E. M Zoroura, M. M. Khalifa et A.Nossier, « Performance de chaînes d'isolateurs haute tension soumises à la pollution du désertique », rapport 33-05, France, 1970.
- [3] M.T.Gencoglu, M.Cebeci, « The pollution flashover on high voltage insulators », Electric Power Systems Research 78, 2008, pp. 1914–1921.
- [4] W.Heise, G.F.luxa, G.Revrery, M. P. Verma, « Estimation de la méthode d'essais sous pollution artificielle par couche solide », CIGRE, rapport 33-09, 1982.
- [5] A. Cimador et S. Vilet, « La Pollution des Isolateurs », EDF-Epure, Rapport N° 27, France, Juillet 1990.
- [6] T. Suda, « Frequency Characteristics of Leakage Current Waveforms of Artificially Polluted Suspension insulators », IEEE Trans. Dielectr. Electr. Insul, Vol. 8, pp.705 -709, 2005.
- [7] X.Yan Shi, C.Sun and Z. Zhang, « Evaluating the Safety Condition of Porcelain Insulators by the Time and Frequency Characteristics of LC Based on Artificial Pollution Tests », IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 17, N°. 2; April 2010, pp1-9.
- [8] F. Amidi et k. Ouerdane, « Les effets de la pollution sur les ouvrage électrique à 30kV en région littorales », Projet de fin d'études, Université des sciences et technologies Houari Boumediene, Alger, Juin 1998.
- [9] Xingliang Jiang, Yunfeng Xia, Qin Hu, Zhenyu Li and Shihua Zhao, « Evaluation of Insulator's Pollution Condition Based on Leakage Current », Chongqing University, Chongqing 400044, China, 14 September 2011.

- [10] Gubanski, S.M, Dernfalk, A., Andersson, J., Hillborg, H., « Diagnostic methods for outdoor polymeric insulators », IEEE Trans. Dielectr. Electr. Insul. 14, (5), 2007, pp.1065-1080.
- [11] T. Suda, « Frequency Characteristics of leakage current waveforms of artificially polluted suspension insulators », IEEE Trans. Dielectr. Electr. Insul., 8,(4), 2005, pp.705-709.
- [12] S. Banerjee, A. Lahiri, K. Bhattacharya, « Optimization of Support Insulators Used in HV Systems Using Support Vector Machine », IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 14, N^o. 2; April 2007, pp.360-367.
- [13] L. Hao, P.L. Lewin, « Partial Discharge Source Discrimination using a Support Vector Machine », IEEE Transactions on Dielectrics and Electrical Insulation Vol. 17, N^o.1, February 2010.
- [14] Jae H. Min and Young-Chan Lee, « Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters », ELSEVIER Expert Systems with Applications, China, 2005.
- [15] Z.Huang, H.Chen, Hsu, C.J.Chen et S .Wu, « Credit rating analysis with support vector machine and neural networks: A market comparative study », Decision Support Systems, 37, 543–558, China, 2004.
- [16] B. Hudgins, P. Parker, and R. Scott, « A new strategy for multifunction myoelectric control », IEEE Trans. Biomed. Eng., vol.40, N^o. 1, pp. 82–94, Jan 1993.
- [17] C.Chang and C.J. Lin, « LIBSVM: A library for support vector machines », <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [18] Y. Huang, K. Englehart, B. Hudgins, and A. D. C. Chan, « A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses », IEEE Trans. Biomed. Eng. Vol. 52, no. 11, pp. 1801–1811, Novembre 2005.

- [19] Vincent Rapp, Thibaud Senechal, Kevin Bailly et Lionel Prevost, « Machine à Vecteurs Supports Multi-Noyau pour la détection de points caractéristiques du visage », Rapport de recherche, Université Pierre et Marie Curie, Paris, France.
- [20] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, « SimpleMKL », Journal of Machine Learning Research, vol. 9, p.p 2491, 2008.
- [21] T. Kanade, Y. Tian, and J. Cohn, “Comprehensive database for facial expression analysis », Proc. IEEE Conf. Face and Gesture Recognition (FG’00), pp. 46, 2000.
- [22] T. Sim, S. Baker, and M. Bsat, « The CMU Pose, Illumination, and Expression Database, Pattern Analysis and Machine Intelligence », IEEE Trans., vol. 25, N°. 1, p. 1615, 2003.
- [23] A. Martinez and R. Benavente, « The AR face database », Tech. report, CVC Technical report, 1998.
- [24] J. Kharroubi, « Étude de Techniques de Classement, Machines à Vecteurs Supports pour la Vérification Automatique du Locuteur », Thèse de Doctorat, Ecole nationale supérieure des télécommunications, France, 2002.
- [25] Pascal Vincent, « Modèles à noyaux à structure locale », Thèse présentée à la Faculté des études supérieures en vue de l’obtention du grade de Philosophiæ Doctor (Ph.D.) en informatique, Université de Montréal, Canada, Octobre 2003.
- [26] Mohamadally Hasan et Fomani Boris, « SVM : Machines à Vecteurs de Support ou Séparateurs a Vastes Marges », Thèse de Doctorat, Versailles St Quentin, France, Janvier 2006.
- [27] Senoussaoui Mohammed, « application des modèles de markov cachés & les machines à vecteurs de support pour la reconnaissance des caractères isolés d’écriture en ligne », SIMPA, 2007.
- [28] Emmanuel Jakobowicz, « La discrimination à plus de deux classes Comparaison de plusieurs approches issues des Support Vector Machines (SVM) », Ecole Polytechnique Fédérale de Lausanne, Suisse.

- [29] Ludovic Mercier, « Les machines à vecteurs support pour la classification en imagerie hyperspectrale : implémentation et mise en œuvre », centre d'enseignement de Grenoble, France, février 2010.
- [30] S. Parfait, « Classification de spectres et recherche de biomarqueurs en spectroscopie par résonance magnétique nucléaire du proton dans les tumeurs prostatiques », Thèse de Doctorat, Université de Bourgogne, Dijon, Décembre 2010.
- [31] Anis Ben Ishak, « sélection des variables par les Machines à Vecteurs de Support pour la discrimination binaire et multi-classe en grande dimension », Thèse de Doctorat, Faculté des Sciences de Luminy, Université de la méditerranée (Aix-Marseille II), Septembre 2007.
- [32] B. Scholköpf, « Support vector learning », R. Oldenbourg verlag, Munich, 1997.
- [33] C.M. Bishop, « Pattern Recognition and Machine Learning », General library system, Information Science and Statistics, Springer, New York, USA, 2006.
- [34] Nicolas Turenne, « Apprentissage automatique (Machine Learning) », INRA, 2006.
- [35] V. Vapnik, « The nature of statistical learning theory », Springer, New York, 1995.
- [36] Hanifi Majdoulayne, « Extraction de caractéristiques de texture pour la classification d'images satellites », Thèse de Doctorat, Université Toulouse III–Paul Sabatier, France, Novembre 2009.
- [37] M. Pontiland et A. Verri, « Properties of support vector machines », Technical Report AI Memo1612, MIT, 1998.
- [38] U. H. G. Krebel, « Pairwise classification and support vector machines », The MIT Press, Cambridge, 1999, pp. 255-268.
- [39] S. Knerr, L. Personnaz, and G. Dreyfus, « Single-layer learning revisited: a stepwise procedure for building and training a neural network », NATO ASI series, Vol.F68, Ecole superieur de physique et de chimie industrielles de la vie de Paris, Laboratoire d'électronique, Paris, France, 1990.

- [40] Gilles Lebrun, « Sélection de modèles pour la classification supervisée avec des SVM (Séparateurs à Vaste Marge).Application en traitement et analyse d'images », Université de Caen/Basse-Normandie, France, Novembre 2006.
- [41] A. Aizerman, E.M. Braverman and L.I. Rozoner, « Theoretical foundations of the potential function method in pattern recognition learning », Automation and Remote Control, 28(1964), pp. 821-837.
- [42] T. Hastie et R. Tibshirani, « Classification by pairwise coupling », The annals of statistics, 26(1) (1998), pp. 451-471.
- [43] M. Moreira, E. Mayoraz, « Improved pairwise coupling classification with correcting classifiers », European conference on machine learning, Germany, pp. 160-171, 21-24, April 1998.
- [44] P. Savicky et J. Fürnkranz, « Combining pairwise classifiers with stacking », International symposium on intelligent data analysis, Berlin, Germany, pp. 219-229, 2003.
- [45] J. Fürnkranz, « Round robin classification », Journal of Machine Learning Research, 2, pp. 721-747, 2002.
- [46] R. Sarathi, S. Chandrasekar et N. Yoshimura, « Investigations into the Surface Condition of the Silicone Rubber Insulation Material using Multiresolution Signal Decomposition », IEEE Trans. Power Delivey, Vol. 21, pp.243 -252, 2006
- [47] Amirouche Annouche et Mohamed Said Berkani, « Traitement et analyse du signal du courant de fuite pour la caractérisation des surfaces isolantes polluées par apprentissage supervisé », Projet de fin d'études, Ecole nationale polytechnique, Alger, Juin 2011.
- [48] Harzelli Imadeddine et Maanani Yacine, « Mesure du courant de fuite et de la charge électrique d'un isolateur pollué de haute tension. Caractérisation de son état de surface en utilisant la transformée en ondelette », Projet de fin d'études, Ecole nationale polytechnique, Alger, Juin 2011.

- [49] A.Cavallini et G.C. Montanari, « Investigations on Leakage Current and Phase Angle Characteristics of Porcelain and Polymeric Insulator under contaminated Conditions », IEEE Transactions on Dielectrics and Electrical Insulation Vol. 16, No. 2; April 2009, pp1 -10.
- [50] O. Bennouna, O. Robin, H. Chafouk et J.P. Roux, « Diagnostic et détection de défauts des systèmes embarqués dans l'automobile », Rapport scientifique, Saint Etienne du Rouvray, France.
- [51] Michel Vacher et Dan Istrate, « Notes de lecture sur la Transformée en Ondelettes : Communication Langagière et Interaction Personne-Système », Grenoble, France, Septembre 2004.
- [52] S.M. Debbal et F. Bereksi-Reguig, « Analyse spectro-temporelle des bruits cardiaques par les transformées discrete et continue d'ondelettes », Sciences & Technologie B – N°23, Juin 2005, pp. 5-15.
- [53] Landrin Patrick Flandrin, « Temps-fréquence », Edition Hes, Collection traitement du signal, 1998.
- [54] Lee-Pierre Belley, « Séparation de sources aveugle utilisant la transformée en ondelettes continues », Thèse de Doctorat, Montréal, Janvier 2008.
- [56] N. Smail, « Application de la Transformée en Ondelettes pour la recherche et l'analyse de textes ». Thèse de Doctorat, université Paris-Est, 27 Janvier 2009.
- [57] Y.Meyer, et M.N owal, « La surprenante ascension des Ondelettes », Société d'éditions scientifiques, Paris, France, Février 2005, pp. 56-59. 0029 -5671.
- [58] B. Moula, A. Mekhaldi et M. Tegar, « Comparaison des Performances d'un Isolateur Haute Tension Soumis aux Conditions de Pollution Artificielle », CNHT'2011– 8ème Conférence Nationale sur la Haute Tension, Tiaret, 09-11 Mai 2011, pp 33-37.
- [59] « Artificial pollution tests on high-voltage insulators to be used on AC systems », IEC 85-1, 1991.

- [60] H.V. Luong, « Étude de la méthode de la transformation en ondelette et l'application à la compression des images », Rapport de stage, 15 juillet 2005.
- [61] K.Dinesh, S.Ashwin, V.Harish, E.VarathKanth, C.Venkataseshaiah et R. Sarathi, « Understanding the surface condition of Insulation Material Using Wavelets », CBIP – International Conference in Maintenance Technologies of EHV Transmission Lines, New Delhi, India, April 2002.
- [62] M.A.Douar et M.C.Bouzidi, « Caractérisation de l'état de surface d'un isolateur pollué à l'aide des signaux du courant de fuite et de la tension appliquée », Projet de fin d'étude, Département de Génie Electrique, Ecole Nationale Polytechnique d'Alger, 2009.
- [63] Fabrice Chane-Ming, Franck Molinaro et Jean Leveau, « Applications de l'analyse en ondelettes sur un signal géophysique », Laboratoire de Physique de l'Atmosphère Université de la Réunion, France, Septembre 1997.
- [64] M.G.E. Schneiders, « Wavelets in Control Engineering », Master's thesis, Eindhoven University of Technology, August 2001.
- [65] A.Bouzida, « Diagnostic de défauts de la machine asynchrone à cage par la technique d'ondelette », Mémoire de Magister, Ecole nationale polytechnique d'Alger 2008.
- [66] Stéphane Verdun, Véronique Cariou et El Mostafa Qannari, « Discrimination et classification supervisée en référence à des prototypes », Rapport de recherche, Ecole Nationale Vétérinaire Agroalimentaire et de l'Alimentation Nantes Atlantique –ONIRIS, 2010.
- [67] Julien Jacques et Christophe Biernacki, « Analyse discriminante sur données binaires lorsque les populations d'apprentissage et de test sont différentes », Laboratoire de Statistiques et Analyse des Données, Université Pierre Mendès, France.
- [68] Charles Bouveyron, « Modélisation et classification des données de grande dimension application à l'analyse d'image », Thèse de Doctorat, Université JOSEPH FOURIER – Grenoble1, Septembre 2006.

[69] Charles Bouveyron, Stéphane Girard et Cordelia Schmid, « Analyse Discriminante de Haute Dimension », Rapport scientifique, Institut nationale de recherche en informatique et en automatique, France, Janvier 2005.

[70] Guillaume Bouchard, « Les modèles génératifs en classification supervisée et applications à la catégorisation d'images et à la fiabilité industrielle », Thèse de Doctorat, Institut National Recherche en Informatique et Automatique, France, 2005.