

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ ABOU BEKR BELKAID
FACULTÉ DE TECHNOLOGIE
DÉPARTEMENT GÉNIE ÉLECTRIQUE ET ÉLECTRONIQUE

MÉMOIRE

pour obtenir le grade de
MAGISTER EN ÉLECTRONIQUE BIOMÉDICALE

présenté et soutenu publiquement
par

Settouti Nesma

le 28 Juin 2011

Titre:

Renforcement de l'Apprentissage Structural pour la Reconnaissance du Diabète

Jury

Président du jury. Pr. Bereksi Reguig Fethi,	UABB Tlemcen
Examineurs. Dr. Hadj Slimane Zine Eddine,	MCA UABB Tlemcen
. Dr. Bechar Hacène,	MAA UABB Tlemcen
Invités d'honneur. Pr. Alexandre Aussem,	UCB Lyon1
. Dr. Haytham Elghazel,	MC UCB Lyon1

Directeur de mémoire. Dr. Chikh Mohamed Amine, MCA UABB Tlemcen

Je dédie ce travail à :

Mes grands parents,

Mes parents,

Toute ma famille,

Mes amis,

Qu'ils trouvent ici l'expression de toute ma reconnaissance.

Remerciements

D'abord, j'adresse mes vifs remerciements à Mr CHIKH M.A, maitre de conférences à l'université de Tlemcen, en tant que Directeur de mémoire, qui s'est toujours montré à l'écoute et disponible tout au long de la réalisation de ce travail malgré ses charges administratives et pédagogiques, et surtout pour sa patience et ses judicieux conseils.

Ensuite, je désire adresser mes sincères remerciements à Mr BEREKSI-REGUIG F, Professeur à l'Université de Tlemcen, directeur du laboratoire Génie biomédical. Son savoir, sa méthodologie et son expérience m'ont beaucoup aidé. Et je tiens à lui témoigner toute ma gratitude pour la réussite de cette formation de Magister. Merci d'avoir accepté de présider ce jury.

Je tiens à remercier Mr HADJ SLIMANE Z.E, maitre de conférences à l'université de Tlemcen, pour l'intérêt porté à ce travail et sa présence parmi les membres du jury.

Mes remerciements à Mr BECHAR H, maitre de conférences à l'université de Tlemcen, qui a participé à l'examen de ce travail.

Je souhaite la bienvenue à nos convives : Professeur Alexandre AUSSEM, Dr Haythem ELGHAZEL, de l'Université Claude Bernard Lyon1, et les remercie de leur honorable présence.

Résumé

L'intelligibilité représente la force motrice la plus importante derrière la mise en œuvre des classifieurs à base flous pour les problèmes d'application médicale. L'expert devrait être capable de comprendre le classifieur et d'évaluer ses résultats. Les modèles à base de règles floues sont particulièrement adaptés, car ils sont constitués de simples règles linguistiques interprétables. Dans la littérature, la majorité des algorithmes basés sur un système d'inférence neuro-flous adaptatifs (ANFIS) ne fournissent pas suffisamment d'explications sur la façon d'obtenir les résultats d'inférence. Ce mémoire de Magister traite la possibilité d'augmenter l'interprétabilité du classifieur ANFIS avec l'apport de la méthode de clustering Fuzzy C-Means . Il montre comment un classifieur neuro-flou interprétable peut être obtenu par un processus d'apprentissage et comment les règles floues extraites peuvent améliorer son interprétation. Les résultats expérimentaux appliqués sur la base de données du diabète (UCI Machine Learning) montrent de fortes similitudes avec les règles appliquées par les experts. Les résultats sont comparés à d'autres travaux dans la littérature. L'approche proposée est simple et efficace pour clarifier la décision finale du classifieur, tout en préservant sa précision à un niveau satisfaisant.

Mots clés

Classification interprétable ; Règles floues ; FCM ; Neuro-Flou ; Base de données du dépôt UCI Machine Learning.

Abstract

Interpretability represents the most important driving force behind the implementation of fuzzy-based classifiers for medical application problems. The expert should be able to understand the classifier and to evaluate its results. Fuzzy rule based models are especially suitable, because they consist of simple linguistically interpretable rules. The majority of classifiers based on an adaptive neuro-fuzzy inference system (ANFIS) used in literature do not provide enough explanation of how their inference results have been obtained. This Magister thesis discusses the possibility to increase the interpretability of ANFIS classifier by the hybridation with the clustering method Fuzzy C means. It is shown how a readable neuro-fuzzy classifier can be obtained by a learning process and how fuzzy rules extracted can enhance its interpretability. Experimental results show that the proposed fuzzy classifier can achieve a good tradeoff between the accuracy and interpretability. Also the basic structure of the fuzzy rules which were automatically extracted from the UCI Machine learning database shows strong similarities to the rules applied by human experts. Results are compared to other approaches in the literature. The proposed approach gives more compact, interpretable and accurate classifier.

Keywords

Interpretable classification ; fuzzy rules ; FCM ; neuro-fuzzy ; UCI Machine Learning Database.

Table des matières

Résumé	ii
Abstract	iii
Table des matières	iv
Table des figures	vi
Liste des tableaux	viii
Liste d'abréviations	ix
Introduction	1
1 Problématique	3
1 Contexte	3
1.1 Définition du diabète	3
1.2 Tests pour le diagnostic du diabète	3
1.3 Les principaux types de diabète	4
1.4 Le diabète dans le Monde	5
1.5 Complications du diabète	7
1.6 Les facteurs agissant sur le diabète	10
2 Aide au diagnostic	16
3 Problématique	17
2 Etat de L'art	19
1 Introduction	19
2 Réseaux de neurones	19
2.1 Théorie	19
2.2 Reconnaissance du diabète par l'approche Neuronal dans la littérature	21
3 Adaptive Neuro-Fuzzy Inférence Système ANFIS	21
3.1 Théorie	21
3.2 Reconnaissance Neuro-floue du diabète dans la littérature	24
4 Fuzzy C-Means FCM	27
4.1 Théorie	27
4.2 Application de FCM dans la littérature	29
5 Hybridation Fuzzy C-Means & Neuronal Network (FCNN)	30
5.1 Principe	30
5.2 Applications de FCNN dans la littérature	30
6 Hybridation Fuzzy C-Means & Adaptive Neuro-Fuzzy Inference System	32
6.1 Principe	32
6.2 Applications de FCM-ANFIS dans la littérature	32

7	Synthèse et discussion des différentes techniques présentées pour la classification du diabète	36
8	Contribution	36
9	Conclusion	37
3	Propositions	38
1	Introduction	38
2	Base de données	38
3	Considération 1 : Classification par Fuzzy C-Means	41
3.1	Principe	41
3.2	Choix du nombre de centre C	41
3.3	Résultats de la classification avec FCM	46
3.4	Inconvénients de la technique	49
4	Considération 2 : Classification par l'hybridation Fuzzy C Means & Réseaux de neurones	50
4.1	Principe	50
4.2	Réseau de neurones mutlicouches RNMC	50
4.3	Experimentation	52
4.4	Conclusion	53
5	Considération 3 : Classification par l'hybridation Fuzzy C-Means & Adaptive Neuro-Fuzzy Inference System	54
5.1	Adaptive Neuro-Fuzzy Inference System	54
5.2	L'apprentissage du classifieur Neuro-flou	56
5.3	L'hybridation FCM-ANFIS	57
5.4	Résultats et discussion	61
5.5	Hybridation Soustractive Clustering (SC)-ANFIS	64
5.6	Conclusion	69
	Conclusion	71
	Bibliographie	73

Table des figures

1.1	Projection du nombre de personnes diabétiques dans différentes régions du monde. [Dia11]	6
1.2	Les 10 pays les plus touchés par le diabète en 2010 et leurs prévisions en 2025. [Alw11]	6
1.3	% de Mort relative au diabète (20-79 ans) dans différentes régions : NAC : Amérique du Nord ; SEA : Asie du sud-est ; MENA : Moyen Orient ; EUR : Europe ; WP : Pacifique ouest ; SACA : Amérique centrale et du Sud ; AFR : Afrique. [Alw11]	7
1.4	Les majeures complications du diabète. [Alw11]	8
1.5	Incidence des crises cardiaques pour les personnes diabétiques et non diabétiques durant une période de 7 ans [FCSFB01, Alw11]	9
1.6	Evolution des malades diabétiques souffrant d'insuffisance rénale signalés en Australie [HBC ⁺ 96, Alw11]	10
1.7	Prévision du diabète (20-79 ans) dans différentes régions : AFR : Afrique ; WP : Pacifique ouest ; SACA : Amérique centrale et du Sud ; EUR : Europe ; SEA : Asie du sud-est ; MENA : Moyen Orient ; NAC : Amérique du Nord. [Alw11]	11
1.8	A droite Age des personnes diabétiques par type de diabète [Par06], A gauche nombre de personnes diabétiques par Age [Alw11]	12
1.9	Prévalence du diabète Type 1(DID) et type 2 (DNID) dans des tranches d'âge de populations dans le monde [dedl94]	13
1.10	Pourcentage du diabète lié à la prise de poids par sous-régions définies par l'OMS. [OMS03]	14
1.11	Risque de transmission du diabète du type I dans une famille	15
2.1	Le neurone artificiel de McCulloch et Pitts (1943). w_i ($i = 1, 2, \dots, N$) est le poids de chacun des nœuds dans le réseau.	20
2.2	System Neuro-Flou (NFS)	21
2.3	Architecture ANFIS.	23
2.4	Architecture de Fuzzy C-Means & Adaptive Neuro-Fuzzy Inference System	33
3.1	Schéma représentatif de la procédure de classification du diabète.	38
3.2	Histogrammes des 9 paramètres de la base de données	39
3.3	Représentation de la base de données : distribution des diabétiques et non diabétiques dans les paramètres de la base de données.	40
3.4	Nombre de centres optimaux avec la technique x-means	43
3.5	Répartition des exemples sur 2 et 3 centres	44

3.6	Largeur maximale de la silhouette globale moyenne pour 2 et 3 centres	45
3.7	Schéma de convergence de la fonction objective	47
3.8	Répartition des 8 paramètres pour $c=2$	47
3.9	Architecture de FCNN [OCK06]	50
3.10	Structure générale d'un réseau de neurones multicouche RNMC .	51
3.11	L'architecture optimale du réseau de neurones obtenue	52
3.12	Architecture d'ANFIS	54
3.13	Modèle flou de Sugeno du Premier ordre	55
3.14	L'approche réalisée FCM-ANFIS	57
3.15	Schéma représentatif de la répartition des clusters en fonctions d'appartenance avec FCM	57
3.16	Les fonctions d'appartenance de la base de données Pima avec FCM	58
3.17	Schéma de principe du K-fold cross validation	59
3.18	Schéma représentatif de la répartition des clusters en des fonctions d'appartenance avec SC	65
3.19	Répartition des fonctions d'appartenance avec Soustractive clustering avec $r=0.8$	66
3.20	Répartition des fonctions d'appartenance avec Soustractive clustering avec $r=0.45$	66
3.21	Répartition des fonctions d'appartenance avec Soustractive clustering avec $r=0.5$	67
3.22	Répartition des fonctions d'appartenance avec Soustractive clustering avec $r=0.65$	67
3.23	Répartition des fonctions d'appartenance avec Soustractive clustering avec $r=1$	68

Liste des tableaux

1.1	Diagnostic du Test FPG [MBG06]	4
1.2	Diagnostic du Test HGPO [MBG06]	4
1.3	Exemple du diabète gestationnel avec le test aléatoire de glucose évaluant les valeurs supérieures à la normale [MBG06]	4
2.1	Quelques travaux avec MLPNN	22
2.3	Quelques travaux avec ANFIS	26
2.6	Quelques travaux de FCNN	32
2.9	Quelques travaux avec FCM-ANFIS	35
2.10	Caractéristiques des différentes techniques présentées pour la clas- sification du diabète	36
3.1	Les centroïdes obtenus pour chaque paramètre	46
3.2	Performances du classifieur FCM	48
3.3	Performances des classifieurs RNMC et FCNN	53
3.4	Les points modaux finaux avec FCM-ANFIS	59
3.5	Degrès de sollicitation des 2 règles de FCM-ANFIS	60
3.6	Performances du classifieur FCM-ANFIS avec différentes entrées .	62
3.7	Performances du classifieur SCM-ANFIS avec différents rayons . .	68

Liste d'abréviations

ACI : Analyse en Composantes Indépendantes
ACP : Analyse en composantes Principales
AG : Algorithme Génétique
ANFIS : Adaptive Neuro Fuzzy Inference System
BCD : Bladder Cancer Dataset
BMI : Body Mass Index
HCM : Hard C-Means
HGPO : Hyper Glycémie Provoquée par voie Orale
FCF : Fuzzy Classifier Function
FCM : Fuzzy C Means
FCNN : Fuzzy C Means Neuronal Network
FNN : Fuzzy Neuronal Network
FPG : The Fasting Plasma Glucose
IMC : Indice de masse corporelle
LERS : Learning from Examples based on Rough Sets
LR : Logistic Regression
LVQ : Learning Vector Quantization
PA : Pression artérielle
MFPCM : Modified Fuzzy Possibilistic C-Mean
ME : Mixture of Experts
MME : Modified Mixture of Experts
MLPNN : Multi-Layer Perceptron Neuronal Networks
MLPs : Multi-Layer Perceptron's
MSE : Mean Square Error
LM : Levenberg-Marquardt
NFGDM : Neuro Fuzzy Genetic Data Mining
NFS : Neuro Fuzzy System
NN : Neuronal Network
RNA : Réseau de Neurones Artificiel
RNMC : Réseau de Neurones Multi-Couches
SC : Subtractive Clustering
Se : Sensibilité
Sp : Spécificité
SVM : Support Vector Machine
TC : Taux de classification
TS : Takagi Sugeno
TSK : Takagi Sugeno Kang
UCI : University California Irvine

Introduction

L'une des motivations pour l'utilisation de systèmes à base de logique floue est due à leurs interprétations. L'interprétabilité permet de vérifier la plausibilité d'un système, conduisant à une maintenance aisée de ce dernier. Il peut également être utilisé pour acquérir des connaissances à partir d'un problème caractérisé par des exemples numériques [DS06]. Une amélioration de l'interprétabilité peut augmenter les performances de généralisation lorsque l'ensemble des données est petit.

L'interprétabilité d'une base de règles est généralement liée à la continuité, cohérence et l'exhaustivité [Gui01]. La continuité garantit que de légères variations dans l'entrée ne peut provoquer de grandes variations dans le résultat. La cohérence signifie que si deux ou plusieurs règles sont déclenchées à la fois, leurs conclusions seront cohérentes. L'exhaustivité signifie que pour tout vecteur d'entrée possible, au moins une règle est activée sans rupture d'inférence.

Lorsque les systèmes neuro-flous sont appliqués pour modéliser des fonctions non linéaires décrites par des ensembles d'apprentissage, le taux de performance peut être optimisé par la procédure d'apprentissage. Cependant, puisque l'apprentissage est orienté précision, il entraîne généralement une réduction de l'interprétabilité du système généré flou. La perte de l'interprétabilité peut être sous les formes suivantes [Jin03] :

1. Incomplétude des partitions floues : exemple deux voisins de sous-ensembles flous dans une partition floue qui ne chevauche pas.
2. Indiscernabilité des partitions floues : exemple les fonctions d'appartenance de deux sous-ensembles flous sont tellement semblables que la partition floue est indiscernable.
3. Inconsistance des règles floues : exemple Les fonctions d'appartenance perdent leurs sens prescrits physiques.
4. Trop de sous-ensembles flous.

Afin d'améliorer l'interprétabilité des systèmes neuro-flou, nous proposons dans ce mémoire de Magister d'automatiser l'apprentissage structurel par des méthodes de clustering dans le but d'augmenter la distinction des partitions

floues et de réduire le nombre de sous-ensembles flous. Cela nous permettra d'optimiser la connaissance avec un minimum de règles tout en gardant la performance à un niveau satisfaisant. L'évaluation de cette approche sera établie sur la base de données Diabète Indiennes PIMA du dépôt d'UCI Machine Learning Database.

Ce travail de Magister se situe dans le contexte général de l' « Aide au Diagnostic » médical, qui a pour but de réaliser un classifieur de diabète lisible et transparent. Le plan de mémoire est composé de :

- Chapitre 1 Problématique, il détaille le contexte, la problématique et les objectifs de ce travail de Magister.
- Chapitre 2 Etat de l'art, fait le tour des différents travaux existants dans la classification du diabète par les approches :
 - Réseaux de neurones,
 - Neuro flou,
 - Leurs hybridations avec FCM (méthode de regroupement)

Tout en présentant leurs propriétés qu'elles regroupent et en commentant leurs nombreux travaux dans la littérature scientifique. Il retrace aussi les questions de recherches traitées et les contributions mises en œuvre, en résumant les avantages et inconvénients de leur usage.

- Chapitre 3 Propositions, décrivant les implémentations faites de notre travail avec les résultats obtenus.
- En dernier lieu, une conclusion générale et des perspectives de ce travail de Magister seront présentées.

Chapitre 1

Problématique

Introduction

Le diabète est une maladie inopinée, sournoise et silencieuse qui peut survenir d'un moment à l'autre. Cette maladie est classée parmi les plus émergentes et se propage à une vitesse fulgurante. Sa gravité réside dans ses impacts néfastes sur plusieurs organes du corps. Le diagnostic de cette pathologie consiste à classer le patient suivant deux situations « diabétique ou sain » en analysant un certain nombre de paramètres qui la caractérise. Et vue le nombre important d'individus et la complexité d'interprétation des paramètres, il est utile et important de faire appel aux systèmes de classification des données. D'où la nécessité de conception d'un système d'aide au diagnostic pour seconder le médecin d'une façon complémentaire et à réduire au minimum les erreurs possibles qui peuvent survenir ; et cela en examinant des données médicales dans un temps plus court et d'une façon plus détaillée et plus précise.

1 Contexte

1.1 Définition du diabète

Le diabète est une condition chronique caractérisée par une insuffisance absolue ou relative de la sécrétion de l'insuline par le pancréas. L'insuline est une hormone qui régule la concentration de sucre dans le sang. L'hyperglycémie, ou concentration sanguine élevée de sucre, est un effet fréquent du diabète non contrôlé qui conduit avec le temps à des atteintes graves de nombreux systèmes organiques et plus particulièrement des nerfs et des vaisseaux sanguins.

1.2 Tests pour le diagnostic du diabète

Les tests suivants sont utilisés pour le diagnostic :

Test glycémie à jeûn (FPG) mesure la glycémie chez une personne qui n'a rien mangé pendant au moins 8 heures. Ce test est utilisé pour détecter le diabète et le pré diabète.

Résultats du glucose plasmatique (mg / dL)	Diagnostic
<= 99	Normal
100 – 125	Le pré-diabète (Anomalie de la glycémie à jeûn)
>= 126	Diabète

TABLE 1.1 – Diagnostic du Test FPG [MBG06]

Test oral de tolérance au glucose (HGPO) la glycémie est mesurée après le jeûne d’au moins 8 heures et ensuite 2 heures après consommation d’une boisson contenant du glucose. Ce test est appliqué pour diagnostiquer le diabète ainsi que son apparition (pré-diabète).

Résultats du glucose plasmatique après 2 heures (mg / dL)	Diagnostic
<= 139	Normal
140 – 199	Le pré-diabète (Tolérance au glucose)
>= 200	Diabète

TABLE 1.2 – Diagnostic du Test HGPO [MBG06]

Test aléatoire de glucose plasmatique appelé également test de plasma glucose occasionnel, comporte des mesures de glycémie prélevées aléatoirement après le dernier repas. Ce test ainsi que l’évaluation des symptômes, permettent de déceler le diabète, et non son apparition.

La prise	Résultat de glucose plasmatique (mg / dL)
A Jeûn	>= 95
A 1 heure	>= 180
A 2 heures	>= 155
A 3 heures	>= 140

TABLE 1.3 – Exemple du diabète gestationnel avec le test aléatoire de glucose évaluant les valeurs supérieures à la normale [MBG06]

1.3 Les principaux types de diabète

Le diabète de type 1

Précédemment connu sous le nom de diabète insulino-dépendant ou juvénile. Il est dû à une destruction, à plus de 90%, des cellules β (bêta) fabriquant l’insuline, hormone indispensable à la vie. Cette forme de diabète peut affecter l’individu dès son plus jeune âge et nécessite obligatoirement des injections pluri-quotidiennes d’insuline.

- Les symptômes sont les suivants : excrétion excessive d'urine (polyurie), sensation de soif (polydipsie), faim constante, perte de poids, altération de la vision et fatigue. Ces symptômes peuvent apparaître brutalement.

Le diabète de type 2

Précédemment appelé diabète non insulino-dépendant ou diabète de la maturité. C'est la forme de diabète la plus répandue. L'insuline produite n'est plus efficace. La sédentarité et une alimentation déséquilibrée et trop riche, en sont les principales causes.

- Ses symptômes peuvent être les mêmes que ceux du diabète de type 1 mais sont souvent moins marqués. De ce fait, la maladie peut être diagnostiquée plusieurs années après son apparition, et cela une fois les complications déjà présentes.
- Cette forme de diabète surgit généralement à l'âge adulte (après 40 ans), mais progresse actuellement chez les jeunes en raison d'un mauvais équilibre alimentaire et du manque d'activité physique dès l'enfance.

Le diabète gestationnel

Le diabète gestationnel est un trouble de la tolérance glucidique de gravité variable, survenant ou diagnostiqué pour la première fois pendant la grossesse, malgré le traitement appliqué et son évolution après l'accouchement.

Durant les deux derniers trimestres de la grossesse, il a été constaté que les besoins en insuline chez la femme enceinte sont deux à trois fois plus importants qu'en temps normal. Cela expliquerait l'augmentation progressive, de la production d'hormones « hyperglycémiantes » (par exemple, les hormones placentaires, le cortisol et les hormones de croissance), qui réduisent les effets de l'insuline sur l'organisme. Elles sont essentielles au bon déroulement d'une grossesse, donc à la santé du fœtus et de la mère. Les femmes qui ont une tendance au diabète sont celles dont le pancréas ne peut produire ce surplus d'insuline, lors de la grossesse. Chez elles, s'installe alors l'hyperglycémie [Bel11]. La moindre hyperglycémie est délétère pour le fœtus. Après l'accouchement, l'hyperglycémie disparaît, mais la femme reste à hauts risques de développer un diabète de type 2.

1.4 Le diabète dans le Monde

- L'OMS a estimé le nombre de diabétiques dans le monde en 2000 à plus de 177 millions de cas. La projection en 2030 serait de 350 millions de cas, que seul un remède pourra freiner [Dia02]. (voir figures 1.1 et 1.2)

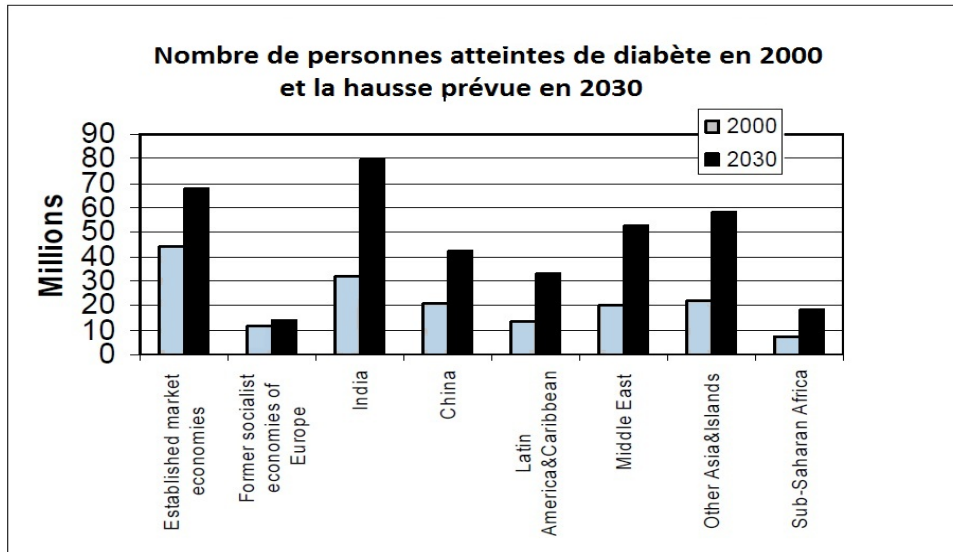


FIGURE 1.1 – Projection du nombre de personnes diabétiques dans différentes régions du monde. [Dia11]

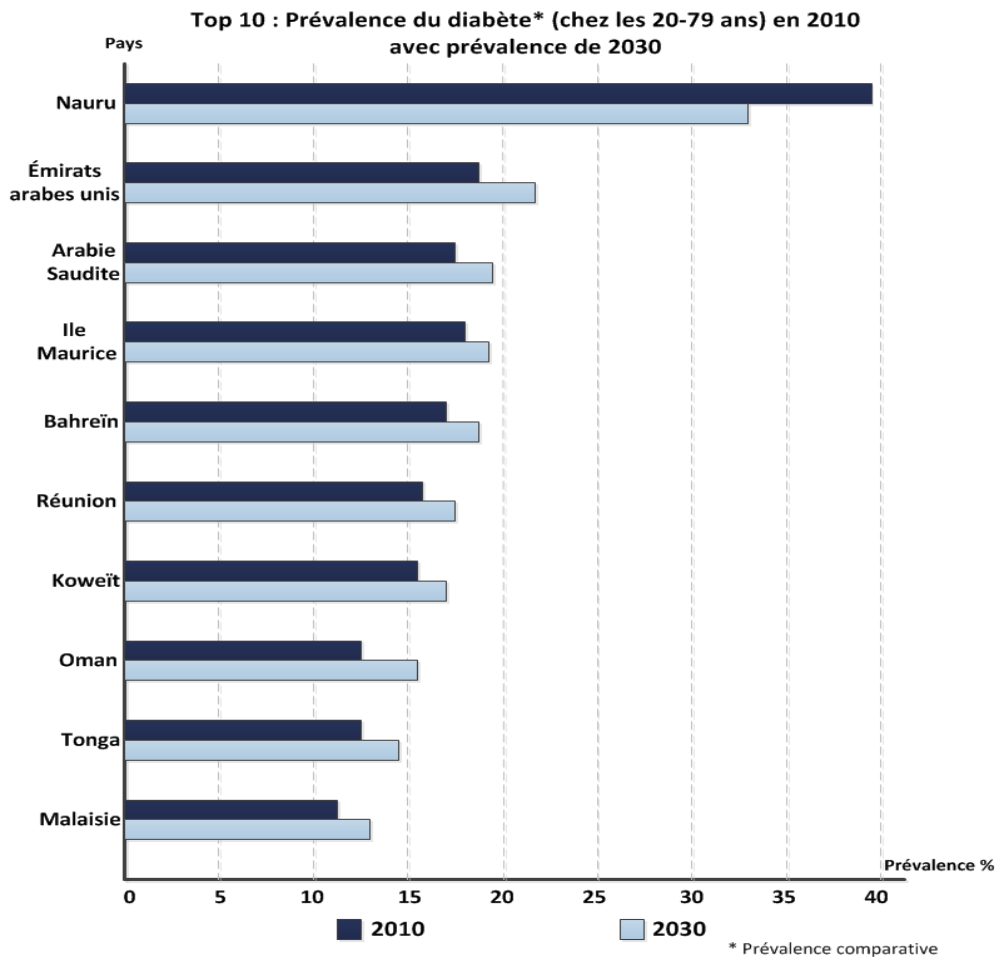


FIGURE 1.2 – Les 10 pays les plus touchés par le diabète en 2010 et leurs prévisions en 2030. [Alw11]

- En 2000, le diabète avait causé plus de 2,9 millions de décès dans le monde, ce qui le positionne au 5ème rang des principales causes de mortalité, selon les premières estimations de l’OMS [OMS07] ; (voir figure 1.3)
- 80% des décès par maladie chronique surgissent dans des pays à revenu faible ou intermédiaire. Son excès de mortalité attribué au diabète est semblable à celui du SIDA en l’an 2000 [OMS07] ;

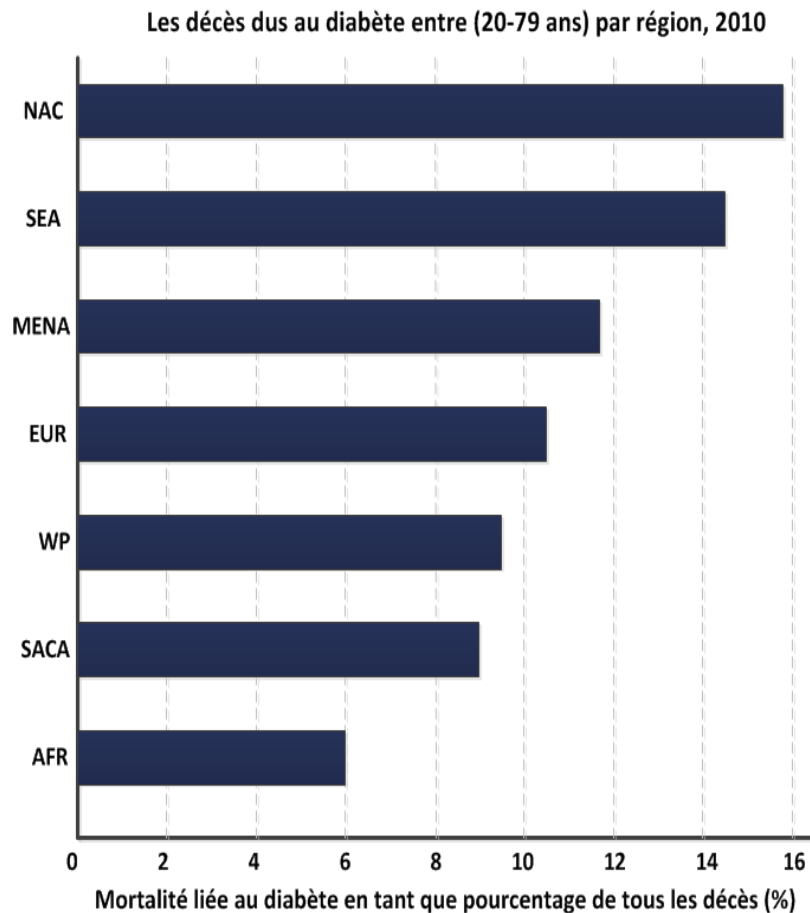


FIGURE 1.3 – % de Mort relative au diabète (20-79 ans) dans différentes régions : NAC : Amérique du Nord ; SEA : Asie du sud-est ; MENA : Moyen Orient ; EUR : Europe ; WP : Pacifique ouest ; SACA : Amérique centrale et du Sud ; AFR : Afrique. [Alw11]

- Dans le monde, la maladie rénale imputable au diabète (néphropathie diabétique) est la cause identifiable la plus répandue d’insuffisance rénale qui requiert une dialyse ou une greffe de rein. C’est aux Etats-Unis que l’incidence est la plus élevée : en effet, plus de 40% des personnes qui doivent suivre un traitement pour insuffisance rénale sont diabétiques [Dia03].

1.5 Complications du diabète

Avec le temps, le diabète peut endommager le coeur, les vaisseaux sanguins, les yeux, les reins et les nerfs (figure 1.4).

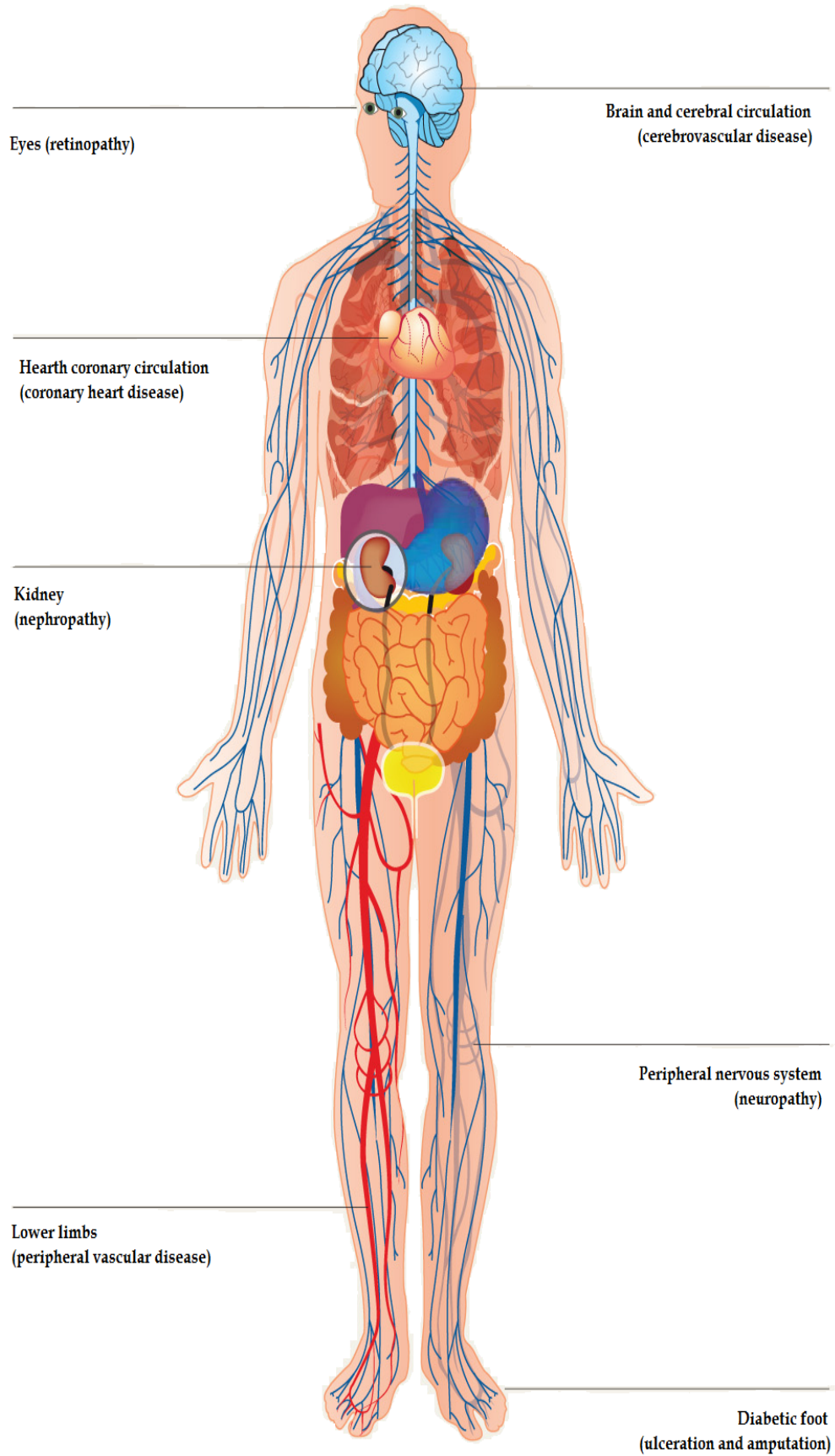


FIGURE 1.4 – Les majeures complications du diabète. [Alw11]

- Le diabète augmente le risque de cardiopathie et d'accident vasculaire cérébral. 50% des diabétiques meurent d'une maladie cardio-vasculaire (principalement cardiopathie et accident vasculaire cérébral) (figure 1.5).

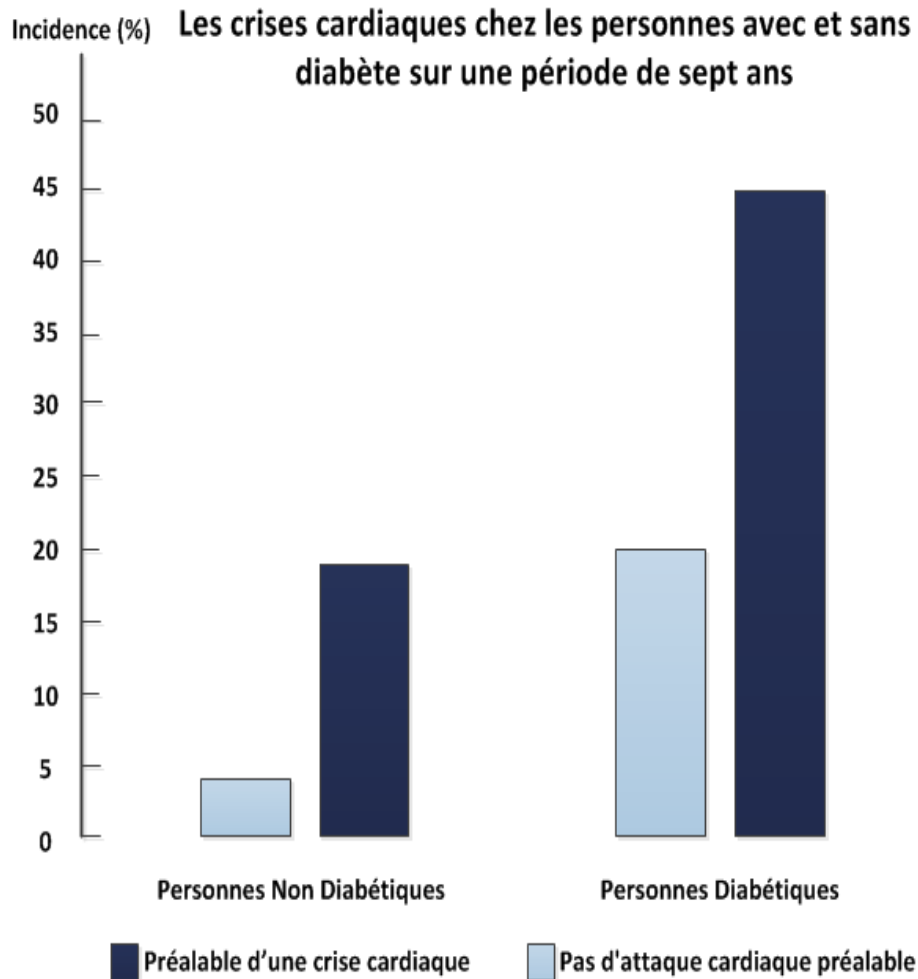


FIGURE 1.5 – Incidence des crises cardiaques pour les personnes diabétiques et non diabétiques durant une période de 7 ans [FCSFB01, Alw11]

- Associée à une diminution du débit sanguin, la neuropathie qui touche les pieds augmente la probabilité d'apparition d'ulcères des pieds et impliquant l'amputation des membres.
- La rétinopathie diabétique est une cause importante de cécité et survient par suite des lésions des petits vaisseaux sanguins de la rétine qui s'accumulent avec le temps. Au bout de 15 ans de diabète, près de 2% des sujets deviennent aveugles et environ 10% présentent des atteintes visuelles graves. [Dia11]
- Le diabète figure parmi les principales causes d'insuffisance rénale. 10 à 20% des diabétiques meurent d'une insuffisance rénale [Dia11](figure 1.6).

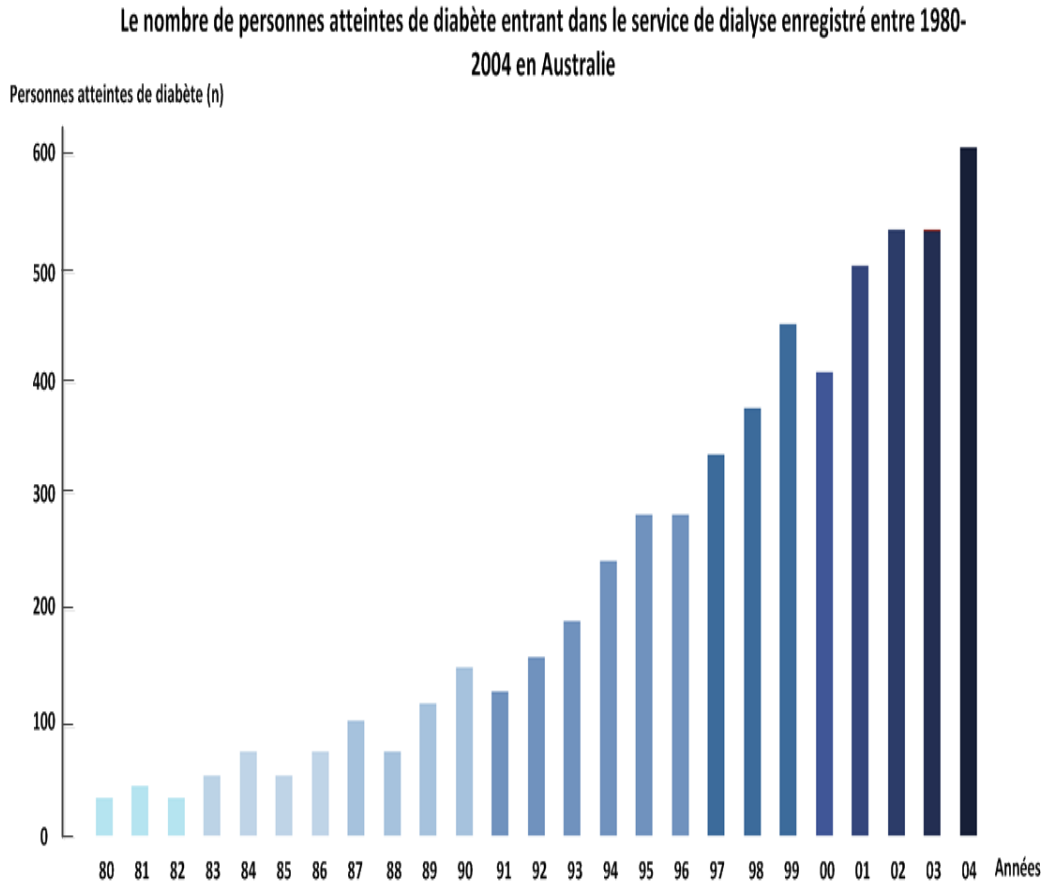


FIGURE 1.6 – Evolution des malades diabétiques souffrant d’insuffisance rénale signalés en Australie [HBC⁺96, Alw11]

- La neuropathie diabétique fait suite aux lésions nerveuses dues au diabète et touche jusqu’à 50% des diabétiques. Bien que de nombreux problèmes différents puissent résulter d’une neuropathie diabétique, les symptômes courants sont les suivants : fourmillement, douleur, engourdissement ou faiblesse au niveau des pieds et des mains.

1.6 Les facteurs agissant sur le diabète

Facteur d’âge

Selon la Fédération internationale du diabète IDF [Alw11], le diabète touche près de 200 millions de personnes de 20 à 79 ans dans le monde, soit 5% de la population mondiale dans cette tranche d’âge (7,8% de la population européenne des 20-79 ans) (voir figure 1.7). En 2025, 330 millions de personnes de 20 à 79 ans en seront atteintes, soit une prévalence de 6,3% (9,1% en Europe) : une augmentation en 20 ans de 74%.

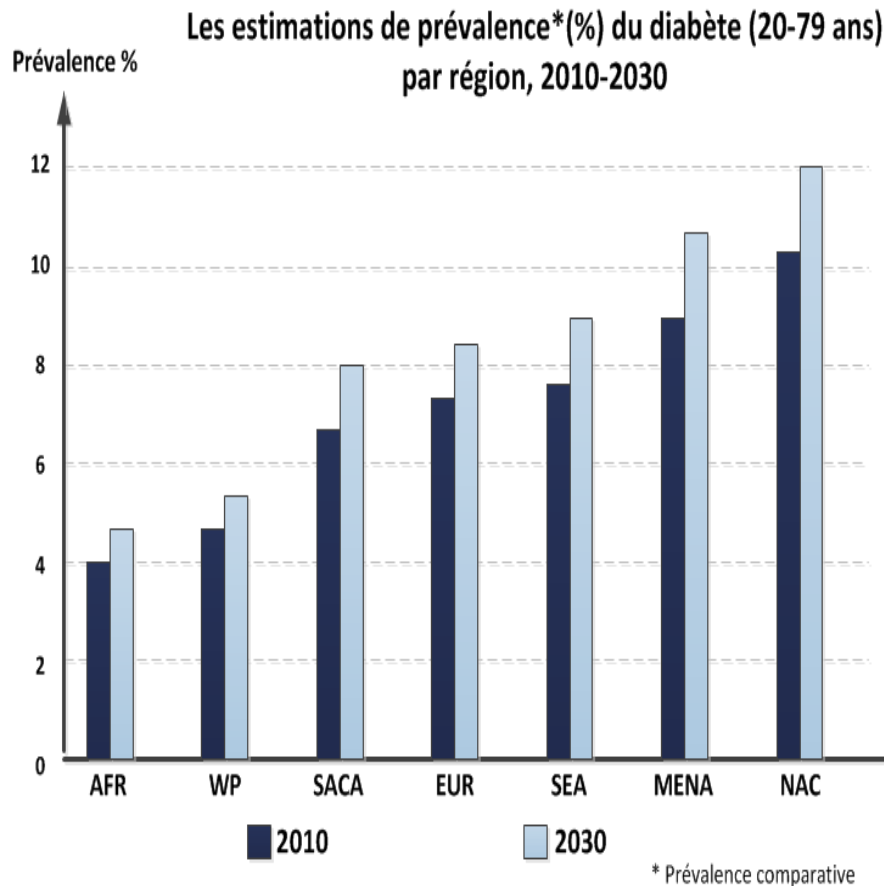


FIGURE 1.7 – Prévision du diabète (20-79 ans) dans différentes régions : AFR : Afrique ; WP : Pacifique ouest ; SACA : Amérique centrale et du Sud ; EUR : Europe ; SEA : Asie du sud-est ; MENA : Moyen Orient ; NAC : Amérique du Nord. [Alw11]

La prévalence du type 1 est très variable selon les pays : un enfant finlandais a 40 fois plus de risque d'être atteint qu'un enfant japonais (figure 1.9) Cela tend à accréditer l'influence de prédispositions génétiques, liées à des altérations multifformes de plusieurs gènes. Un deuxième point frappant est l'accroissement de l'incidence du type 1 (nombre de nouveaux cas par an) : 3 à 4% par an en moyenne selon l'étude européenne multicentrique EURODIAB [DLE04]. Plus inquiétant encore, cette augmentation est particulièrement visible chez les enfants de moins de 5 ans (voir figure 1.8).

Le diabète de type 2 augmente très vite en fréquence dans le monde, notamment dans les pays industrialisés. En seulement 5 ans [1993-1998], sa prévalence a augmenté de 35% aux Etats-Unis, où il touche 16 millions de personnes [Par06]. Un autre fait inquiétant, alors que le diabète de type 2 concernait les personnes de plus de 40 ans, il affecte de plus en plus d'enfants et d'adolescents, parallèlement à l'accroissement du nombre de jeunes obèses ; dans certains pays, comme au Japon, dans cette tranche d'âge, il dépasse même en fréquence le diabète de type 1. On remarquera aussi que les Indiens Pima (états unis d'Amérique) dont la tranche d'âge se situe entre 30-64 ont la plus grande prévalence >50% (figure 1.9).

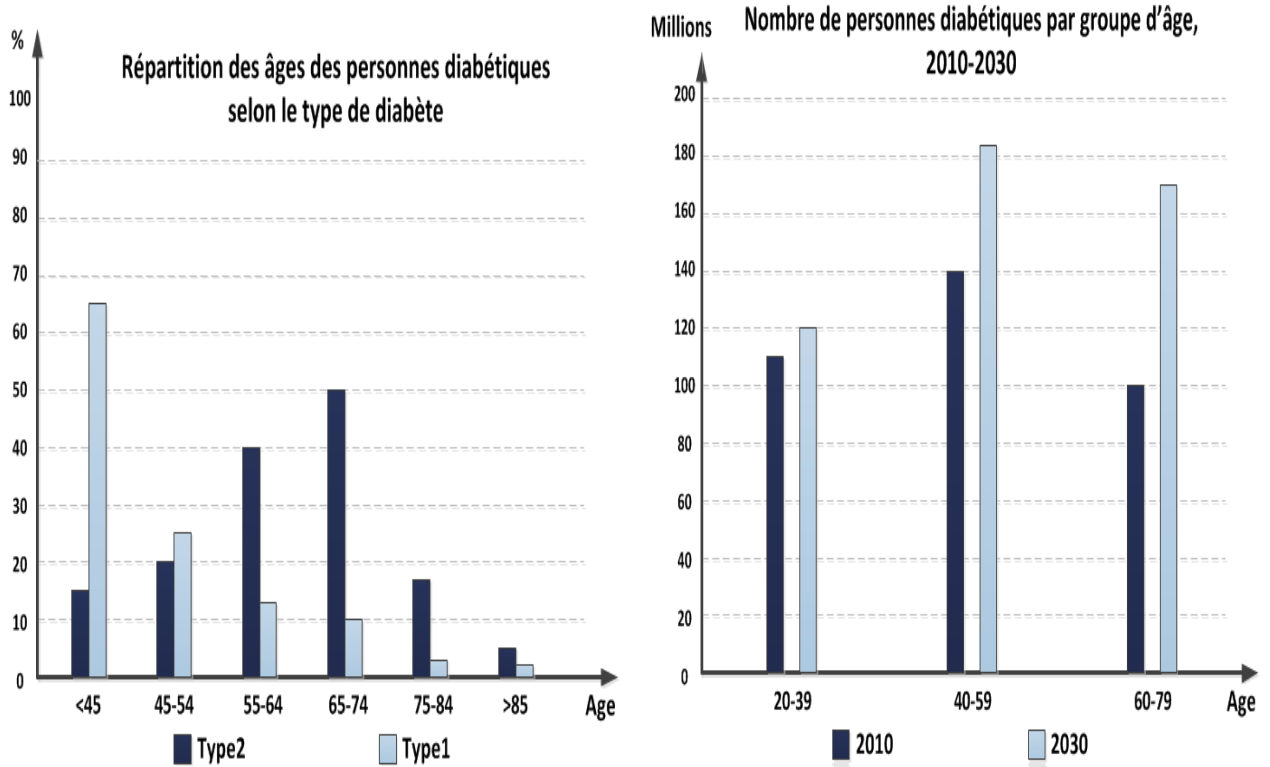


FIGURE 1.8 – A droite Age des personnes diabétiques par type de diabète [Par06], A gauche nombre de personnes diabétiques par Age [Alw11]

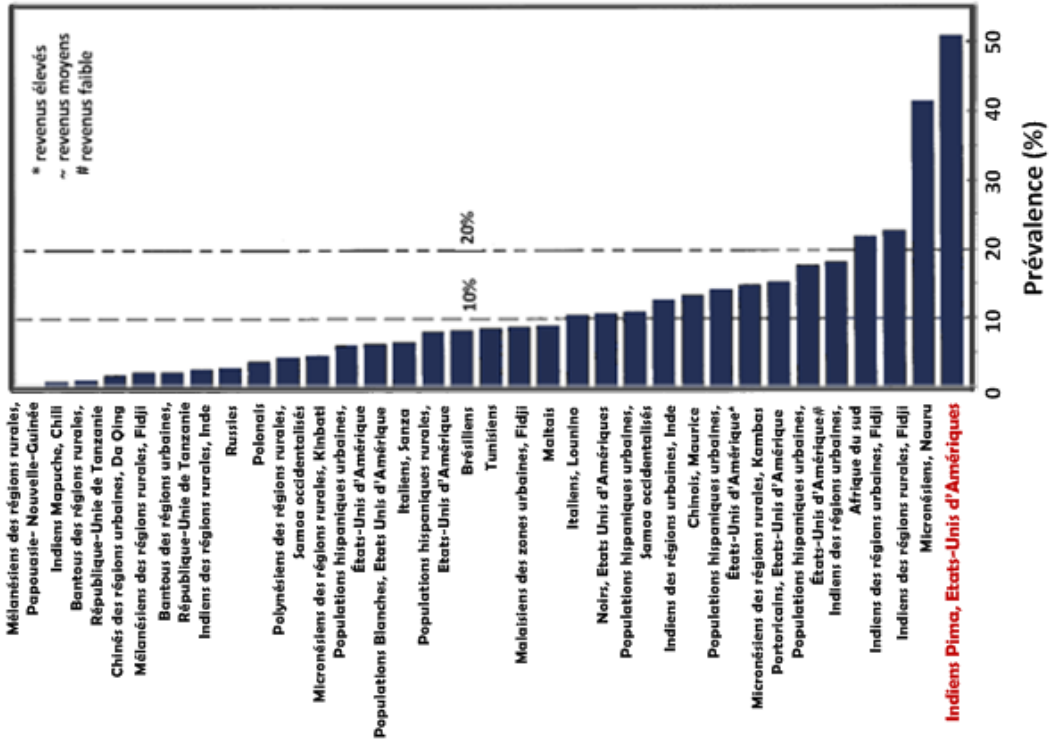
Facteur d'origine ethnique

Une étude réalisée au Colorado [WVB⁺00] prouve que les personnes d'origine hispanique ont 2 fois plus de risques de développer un diabète que les individus blancs non hispaniques après ajustement sur :

- L'âge,
- Le sexe,
- L'obésité,
- Les antécédents familiaux,
- Le niveau d'éducation,
- Les revenus.

Cela a permis une constatation selon les ethnies des différences de susceptibilité vis à vis du diabète. Ainsi, chez les indiens Mapuche au Chili, le diabète est presque inexistant dans la tranche d'âge 30-64 ans, alors que chez les Indiens Pima, on observe une prévalence supérieure à 50% pour la même tranche d'âge (figure 1.9). Aussi au Royaume Uni, la population d'âge mûr originaire d'Afrique noire ou des Caraïbes a une prééminence du diabète trois fois plus élevée que la population générale. Quand à la population originaire d'Asie du sud leurs prédispositions sont quatre fois plus élevées .

Prévalence du DNID chez les 30-64 ans dans quelques populations (correction faite de l'âge, les deux sexes confondus)



Incidence du DID chez les enfants de moins de 15 ans

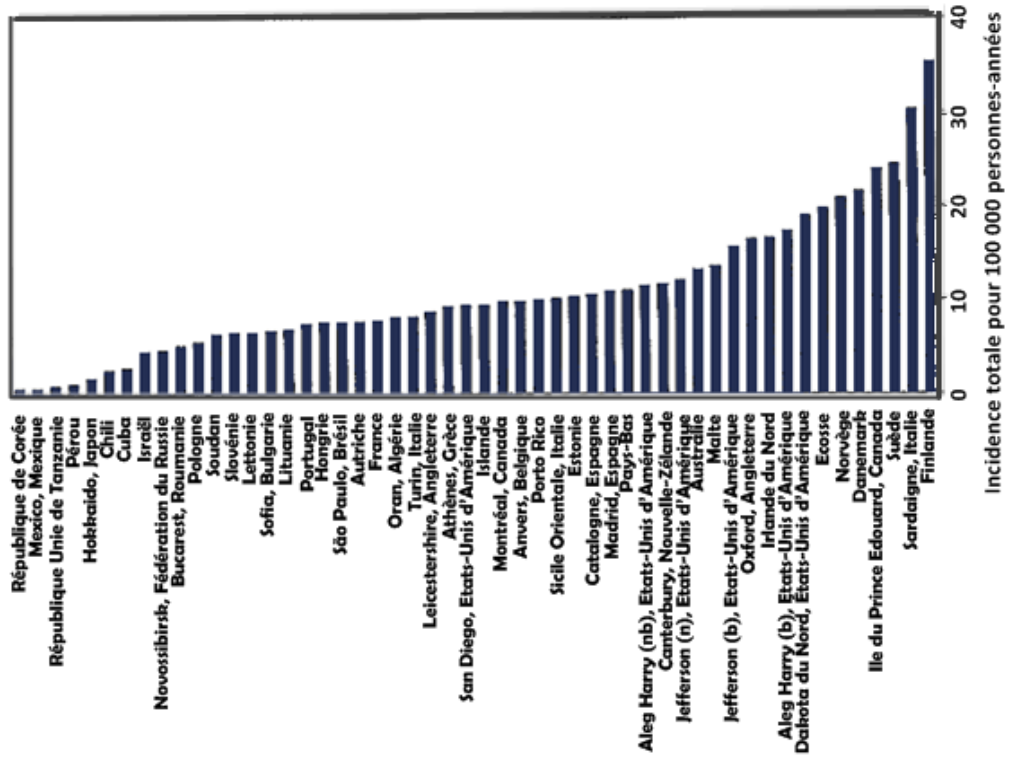


FIGURE 1.9 – Prévalence du diabète Type 1(DID) et type 2 (DNID) dans des tranches d'âge de populations dans le monde [ded194]

Facteur d'obésité (poids)

Le contrôle du poids corporel est d'une importance considérable pour la définition du risque du diabète car il affecte non seulement la résistance en insuline mais aussi les besoins en insuline et le contrôle ultérieur de la glycémie. Il est calculé par l'indice de masse corporelle (IMC) = le poids (kg) / taille (m).

Dans les pays occidentaux, environ 10 à 20% des patients atteints du diabète de type 2 ont un IMC normal. Donc ils ne nécessitent pas la modification de leurs apports caloriques car ils ont pu équilibrer la quantité de calories à prendre ainsi que la consommation d'énergie. L'attention sera retenue sur la teneur des aliments en glucides et leurs répartitions entre les principaux repas. Contrairement à ces patients plus de 80 à 90% des diabétiques de type 2 sont obèses (voir figure 1.10). L'excès de poids chez les enfants est particulièrement préoccupant. Une fois déclaré, le diabète de type 2 chez les enfants devient particulièrement alarmant. Chez les adultes, il a été clairement démontré qu'une perte de poids peut contrer la progression du diabète de type 2 et que des réductions de poids modestes pourraient diviser par deux le risque de diabète, voire le prévenir complètement.

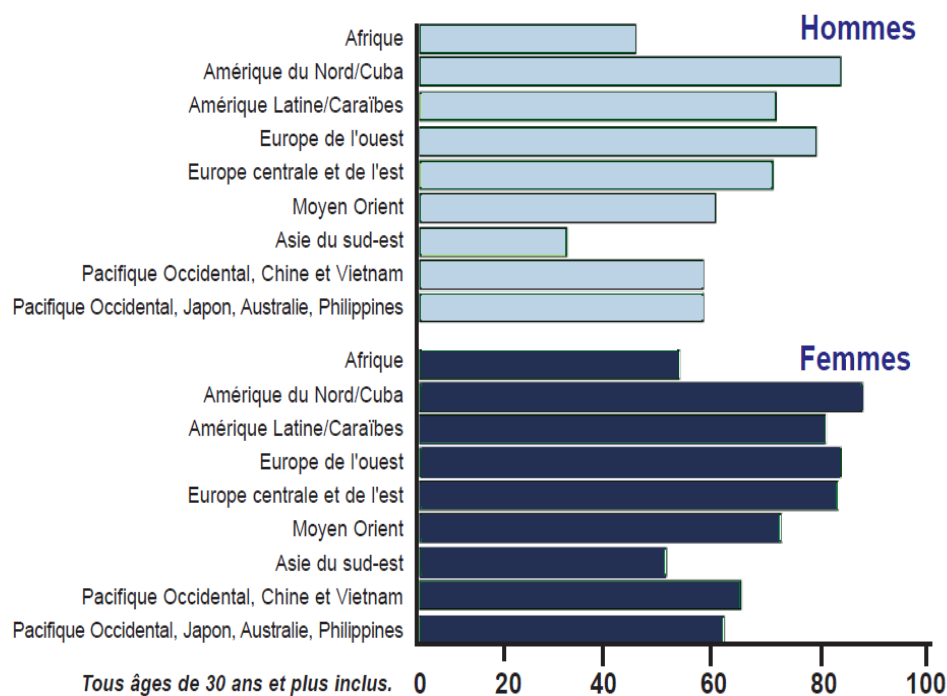


FIGURE 1.10 – Pourcentage du diabète lié à la prise de poids par sous-régions définies par l'OMS. [OMS03]

L'US National Institute of Diabetes et le Digestive and Kidney Diseases Diabetes Prevention Program [ftDPPRG02] ont fait une démonstration remarquable de l'effet d'une perte de poids à travers une alimentation saine et une plus grande activité physique. Ce programme a été particulièrement bénéfique pour les personnes de plus de 60 ans – près des trois quarts des nouveaux cas de diabète chez les plus de 60 ans ont pu être évités.

Cette étude illustrée par d'autres a apporté de l'espoir aux personnes en surcharge pondérale qui sont particulièrement prédisposées au diabète. Des changements de mode de vie ont entraîné une perte de poids de 5-7% permettant de réduire le diabète. En réduisant les graisses et l'apport en calories et en marchant une demi-heure supplémentaire par jour, l'incidence a pu être réduite de 58%. La plus grande réussite a été observée chez les personnes de plus de 60 ans avec une réduction du développement du diabète de 71%. Des données similaires sont ressorties d'études menées en Chine et dans plusieurs pays européens.

Facteur de risque génétique

La présence d'un diabétique dans une famille augmente le risque d'apparition du diabète chez les autres membres, cela favorise la participation génétique d'avoir le diabète du type 1. Dans la fratrie de sujets présentant un diabète de type 1, ce risque est d'environ 5%, alors que les enfants de parents diabétiques ont un risque de 2 à 3% si la mère est malade et de 5 à 6% s'il s'agit du père [Sky93] (figure 1.11).

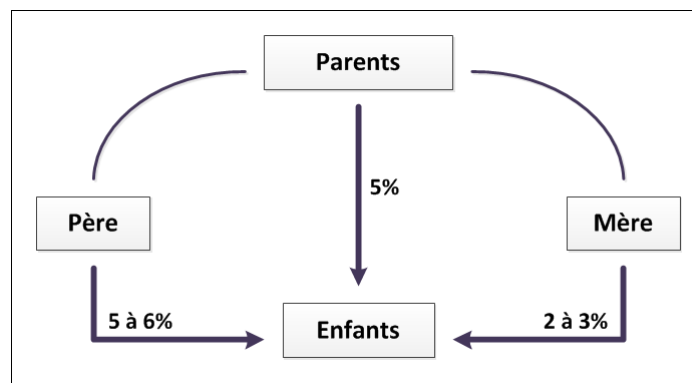


FIGURE 1.11 – Risque de transmission du diabète du type I dans une famille

Des études de concordance entre jumeaux ont démontré que si l'un d'entre eux est atteint de diabète, il y a une concordance plus importante chez les homozygotes (58% à 80% selon les études [Sky93]) que pour les hétérozygotes (17% à 40%). Ceci prouve un support génétique important au diabète, mais l'absence de concordance à 100% suggère également que cette prédisposition génétique est dépendante d'autres facteurs [NSK⁺87].

Facteur de la tension artérielle

L'étude UKPDS (United Kingdom Prospective Diabetes Study) [Lau08] a prouvé l'effet délétère de l'hypertension artérielle chez les malades atteints de diabète non-insulino dépendant. Cette expérience, d'une durée de 8 ans, avait pour but de déterminer les complications cardiovasculaires induites par l'hypertension artérielle. Deux groupes ont été analysés dans cette étude :

- Un groupe dont la pression artérielle (PA) est $< 150/85$ mmhg (n = 758), traité par IEC (Inhibiteur de conversion, captopril) et/ou bêta-bloquants (atenolol),
- Un groupe dont la PA est $< 180/105$ mmHg (n = 390), non traitée par des médicaments anti-hypertenseurs.

Les résultats sont les suivants : Le contrôle de la pression artérielle chez le diabétique non-insulino dépendant permet de limiter :

- L'atteinte MICROVASCULAIRE (atteinte du rein, de l'œil et des nerfs) (-37%)
- Les AVC (accidents vasculaires cérébraux) (-44%)
- Les décès liés au diabète (-32%)

S'il existe un diabète, il est très important de veiller à ce que la pression artérielle ne dépasse pas 130/80 mmHg. Au-delà de ce chiffre, le risque cardio-vasculaire devient alors très important et il est ainsi impératif d'avoir recours à un traitement anti-hypertenseur.

2 Aide au diagnostic

Les méthodes de classification ont pour but d'identifier les classes auxquelles appartiennent des objets à partir de certains traits descriptifs. Elles s'appliquent à un grand nombre d'activités humaines et conviennent en particulier au problème de la prise de décision automatisée [GT00]. Il s'agira, dans le cadre de notre travail, d'établir un diagnostic médical à partir de la description clinique d'un patient. Une première méthode possible pour résoudre ce type de problème est l'approche "systèmes experts". Dans ce contexte, la connaissance d'un expert (ou d'un groupe d'experts) est décrite sous forme de règles. Cet ensemble de règles forme un système expert qui est utilisé pour classer de nouveaux cas. Cette procédure, largement employée dans les années 80, dépend fortement de la capacité à extraire et à formaliser les connaissances de l'expert. Nous considérons ici un autre procédé pour lequel la classification sera tirée automatiquement à partir d'un ensemble d'exemples. Un exemple consiste en la description d'un cas avec la classification correspondante. Il s'agit donc d'induire une technique de classification générale à partir de données. Le problème est alors inductif, il est question en effet d'établir sur la base d'exemples une règle générale. La procédure générée devra classer correctement les paramètres du patient mais surtout avoir un bon pouvoir prédictif pour classer correctement de nouvelles descriptions.

Les méthodes utilisées par les systèmes d'apprentissage sont très nombreuses et sont issues de domaines scientifiques variés. Nous allons arbitrairement les séparer en deux groupes :

- Les méthodes de nature probabiliste ou statistique,
- Les méthodes issues de l'intelligence artificielle.

Les méthodes issues de l'intelligence artificielle sont des méthodes non paramétriques. On distingue les méthodes symboliques (la procédure de classification produite peut être écrite sous forme de règles) des méthodes non symboliques ou adaptatives (la procédure de classification produite est de type «boîte noire»). Les méthodes symboliques, les plus utilisées sont basées sur les systèmes de règles floues. Pour les méthodes adaptatives, nous citons en particulier les réseaux de neurones.

3 Problématique

La problématique de classification du diabète est un domaine qui a fait l'objet de plusieurs recherches. Certaines de ces recherches, telles quelles sont exposées dans le chapitre 2 état de l'art contribuent vers le renforcement de l'aide au diagnostic médical, par une caractérisation et une reconnaissance intelligente des données médicales.

Dans l'application des systèmes flous nous rencontrons toujours un compromis entre l'interprétabilité et la précision [BB97b]. Une solution interprétable peut être exprimée que par des termes linguistiques validée par l'expert humain. Elle n'emploie que des variables et règles floues tout en conservant les variables d'origine. Cela signifie, que pour l'obtention d'une solution interprétable, nous devons accepter un rendement sous-optimal. S'il s'avère que la précision n'est pas suffisante, nous devons réfléchir à une autre solution plus adéquate.

D'autre part, si nous nous intéressons à une solution très précise, alors nous risquons de perdre le sens linguistique définissant les modèles flous. Dans ce cas, nous choisissons d'appliquer une autre caractéristique des systèmes flous qui est : la combinaison des modèles locaux pour une solution globale. Pour sa réalisation, les modèles de type Sugeno [TS85] sont plus adaptées que les modèles de type Mamdani. Toutefois, nous devons examiner si un système flou est l'approche la plus appropriée pour la solution recherchée. Des techniques tels que les réseaux de fonction à base radiale, Kernel regression, réseaux B-spline, etc peuvent être utilisées. [BB97a, BH94].

Les classifieurs flous sont réalisés à partir de la connaissance basée sur des règles floues. Cependant, dans de nombreux cas ces connaissances sont soit inexistantes, ou partiellement disponibles. Dans ce cas, le classifieur peut être également créé à partir de l'apprentissage des données. Une possibilité est d'appliquer des techniques de clustering flou (apprentissage non supervisé) [BPKK99, HKKT99], et une autre qui consiste à utiliser des méthodes neuro-floues (apprentissage supervisé) [NKK97].

Si l'interprétabilité et la lisibilité sont des solutions importantes, les techniques neuro-floues peuvent fournir un moyen pratique pour l'obtention d'une classification à partir de données numériques. Les méthodes neuro-floues sont initialisées avec des fonctions d'appartenance. Et grâce à leurs algorithmes d'apprentissage, ils assurent aux ensembles flous le maintien de l'interprétabilité et la lisibi-

lité.

Le travail que nous vous présentons dans ce mémoire de Magister s'inscrit dans le contexte d'aide au diagnostic. Nous donnons un intérêt plus particulier à l'apprentissage automatique structurel et paramétrique, dans le but d'augmenter les performances du classifieur sur le plan précision et interprétabilité. Aussi nous mettons en évidence la connaissance extraite qui est sous exploitée dans les travaux de la littérature scientifique.

Chapitre 2

Etat de L'art

1 Introduction

Le diagnostic est parmi les activités les plus intéressantes dans la mise en œuvre de systèmes intelligents et robustes. Le principal objectif de ce chapitre est de développer un système automatisé d'aide à la décision pour la reconnaissance des diabétiques et les sujets présentant des facteurs de risque du diabète. Le but est de déterminer un schéma de classification optimal avec une précision de diagnostic élevée pour une transparence maximale. Les performances de plusieurs algorithmes de classification ont été testées et comparées pour la reconnaissance des diabétiques. Dans ce chapitre nous présentons les principales techniques utilisées dans la reconnaissance automatique du diabète. Nous citons en particulier les RNAs, FCM, ANFIS, FCNN, FCM-ANFIS. Enfin les résultats obtenus dans la littérature par ces différentes approches ont été présentés et comparés.

2 Réseaux de neurones

2.1 Théorie

Comme son nom l'indique, les Réseau de Neurones Artificiels (RNAs) sont des modèles mathématiques du système nerveux humain. En tant que tel, les RNA contiennent des «nœuds» interconnectés (les neurones) qui sont capables de traiter et de transmettre des informations d'un nœud à un autre [WBG97]. Les RNAs les plus élémentaires, décrits par McCulloch et Pitts 1943 [MP43](voir la figure 2.1), transmettent par anticipation les informations dans une seule direction. Les propriétés du réseau et la transmission de l'information au sein du réseau sont régies par l'architecture et la façon dont ces interconnexions sont modélisées [Gur97].

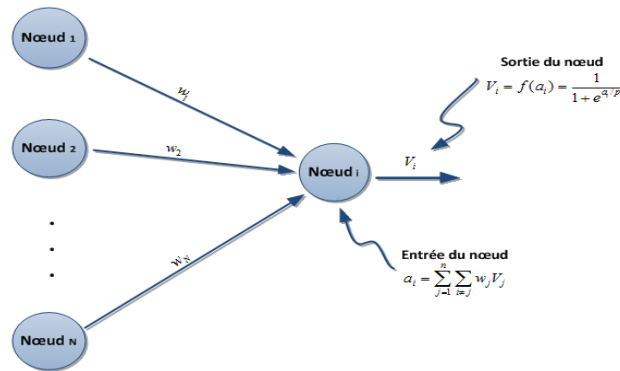


FIGURE 2.1 – Le neurone artificiel de McCulloch et Pitts (1943). w_i ($i = 1, 2, \dots, N$) est le poids de chacun des nœuds dans le réseau.

La rétro-propagation est une technique basée sur l'apprentissage supervisé. Elle est utilisée pour l'apprentissage des réseaux neuronaux artificiels. Cette technique a été décrite par P. Werbos en 1974 [Wer74], et développée plus tard par Rumelhart et al. en 1986 [RHW88]. Le principe de la rétropropagation est de traiter un ensemble d'échantillons de formation itérative, et de comparer les prévisions du réseau pour chaque échantillon avec l'étiquette des classes connues réelles. Afin de minimiser l'erreur quadratique moyenne entre la prédiction du réseau et la classe proprement dite, les poids sont modifiés pour chaque échantillon d'apprentissage [HK00].

L'algorithme de rétro propagation

La sortie de l'unité j est calculée en appliquant une fonction d'activation sigmoïde comme suit :

$$O_j = \frac{1}{1 + e^{-\sum_i w_{ij} O_i + \theta_j}} \quad (2.1)$$

Dans l'équation 2.1, w_{ij} est le poids de la connexion de l'unité i avec la couche précédente à l'unité j ; O_i est la sortie de l'unité i de la couche précédente, et θ_j est le pas de l'unité. La façon de mettre à jour les poids est la suivante :

$$w_{ij} = w_{ij} + l * Err_j O_j \quad (2.2)$$

Dans l'équation 2.2, l est le taux d'apprentissage. L'erreur est calculée pour une unité j dans la couche cachée, comme suit :

$$Err_j = O_j(1 - O_j) \sum_k Err_k w_{jk} \quad (2.3)$$

L'erreur est calculée pour une unité j dans la couche de sortie comme suit :

$$Err_j = O_j(1 - O_j)(T_j - O_j) \quad (2.4)$$

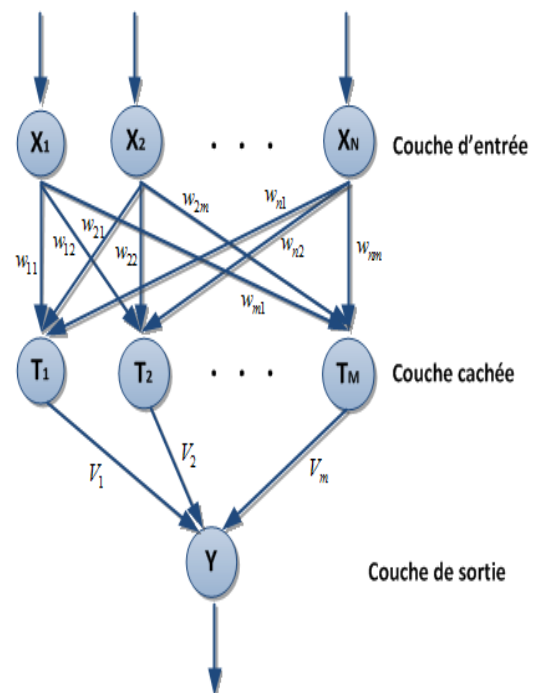


FIGURE - Réseau de neurones multicouches.

2.2 Reconnaissance du diabète par l'approche Neuronal dans la littérature

Largement employé dans la classification du diabète, le réseau de neurones multicouches a été utilisé avec l'algorithme rétropropagation pour reconnaître l'ensemble de données de la maladie du diabète Pima. Kayaer et Yıldırım [KY03], ont rapporté un taux de performance de 77.08%. En l'occurrence Temurtas et al. [TNT09] ont obtenu un taux de 82.37% suivant la même méthode. Aussi Vosoulipour et al. [VTM08] ont employé l'algorithme génétique pour la sélection des paramètres d'entrées et ont réalisé un taux de 77.60% pour la reconnaissance des diabétiques. Kahramanli et Allahverdi [KA08] ont eux aussi utilisé un réseau de neurones modifié, et ont obtenus une bonne reconnaissance du diabète avec un taux de 84.20%.

Le tableau 2.1 présente d'autres travaux utilisant l'approche neuronale pour l'identification de diabète.

3 Adaptive Neuro-Fuzzy Inférence Système ANFIS

3.1 Théorie

La modélisation Neuro-floue

Il existe plusieurs méthodes pour mettre en œuvre la technique neuro-floue de modélisation. En général, l'hybridation neuro-floue est faite de deux façons (Mitra et Hayashi [SM00]) :

- Un réseau de neurones équipé de la capacité de manutention du traitement de l'information floue, est appelé un réseau de neurones flou (FNN)
- Un système flou combiné avec les réseaux de neurones pour l'amélioration de certaines de ses caractéristiques, comme la flexibilité, la vitesse et la capacité d'adaptation, est appelé un système neuro flou (NFS).

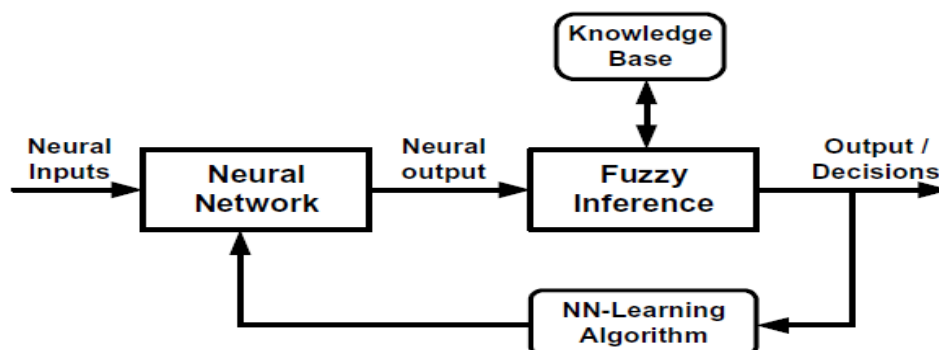


FIGURE 2.2 – System Neuro-Flou (NFS)

AUTEURS	ARTICLES / RAPPORTS	APPROCHE	EXPERIENCE	RÉSULTATS
Jonas Gu-zaitis, Antanas Verikas, Adas Gelziniš, and Marija Bacauskiene, 2009 [GVGB09]	A Framework for Designing a Fuzzy Rule-Based Classifier	Un cadre général pour la conception d'un système de classification fondé sur des règles floues est proposé. L'Algorithme Génétique développé partitionne l'espace de recherche et permet de fixer à la fois la structure et le calcul des paramètres du classifieur. Des simulations sur 4 bases de données (vote, cancer du sein, Diabète, inspection de surface) ont montré que les performances obtenues par le classifieur sont comparables, voire supérieures à la meilleure performance réalisée par l'utilisation des MLPNNs.	Dans la série d'expériences, la sélection des descripteurs a été activée et les caractéristiques propres de chaque patiente ont été trouvées par l'algorithme génétique (AG). La classification a été obtenue en utilisant les paramètres sélectionnés ainsi que la totalité des 8 paramètres pour des fins de comparaison.	Le taux de reconnaissance était de Tc=71.49%. Nous remarquons dans ce travail que la classification avec les paramètres sélectionnés par l'AG donnent de meilleurs résultats sauf pour la base de données du cancer.
Elif Derya Übeyli, 2008 [Ü08]	Modified Mixture of Experts for Diabetes Diagnosis	Le but de la présente étude était de vérifier l'exactitude des trois types de systèmes de diagnostic automatisés : les MLPNNs, les architectures modulaires du réseau de neurones : ME (mixture of experts), MMEs (modified mixture of experts), pour la détection des diabétiques. Ces classifieurs ont été entraînés avec les attributs (facteurs de risque du diabète) pour chaque enregistré dans la base diabète Pima.	Les expériences ont été réalisées sur 768 patients avec leurs 8 descripteurs, dont 268 sont diabétiques et le reste non-diabétiques.	<p>MME : Sp=99.43%Se=98.48% Tc=99.17%</p> <p>ME : Sp=98.01%Se=97.73% Tc=97.93%</p> <p>MLPNN : Sp=91.19%Se=92.42% Tc=91.53%</p>
Nesma Settouti, Tadj. E Fandi, M. Amine Chikh. 2009 [SCF09]	Reconnaissance par Neuro-Floue du Diabète	Les réseaux de neurones ont été appliqués pour la classification du diabète, en portant à leurs entrées les résultats réalisés par des méthodes de réduction des paramètres dans lesquelles deux procédures différentes sont adoptées : Analyse en Composantes Principales ACP et l'Analyse en composantes Indépendantes ACI. L'extraction des paramètres les plus pertinents à partir des ces méthodes de réduction sont présentés à l'entrée du réseau MLPNN.	Sur les 786 femmes Indiennes Pima seuls 392 ont été sélectionnées dont tous les paramètres sont complètes contrairement au reste à savoir 140, qui présentaient des données manquantes principalement en ce qui concerne la pression artérielle.	<p>8 paramètres : Tc=86.15%Se=87.91%Sp=82.05%</p> <p>ACP :3 composantes Tc=73.08%Se=98.90%Sp=12.82%</p> <p>Glu, Age, BMI : Tc=87.69%Se=91.21%Sp=79.49%</p> <p>ACI :3 composantes Tc=75.38%Se=86.81%Sp=48.72%</p> <p>Glu, Age, Insulin : Tc=89.23%Se=89.01%Sp=89.74%</p>

TABLE 2.1: Quelques travaux avec MLPNN

Dans le cadre de notre travail nous nous intéresserons au second modèle (Figure 2.2), le bloc de réseau de neurones entraîne le système d'inférence floue pour générer les décisions correspondantes.

Architecture d'ANFIS

Le modèle ANFIS est un réseau de neurones flou proposé par Jang [Jan93]. Ce réseau est composé de cinq couches représentées sur la figure 2.3 ci-dessous avec deux entrées X et Y. Pour simplifier la présentation de l'architecture d'ANFIS, deux règles floues « si alors » du modèle de Sugeno de premier ordre sont considérées :

Règle 1 : Si (x est A1) et (y est B1) alors ($f_1 = a_1x + b_1y + c_1$)

Règle 2 : Si (x est A2) et (y est B2) alors ($f_2 = a_2x + b_2y + c_2$)

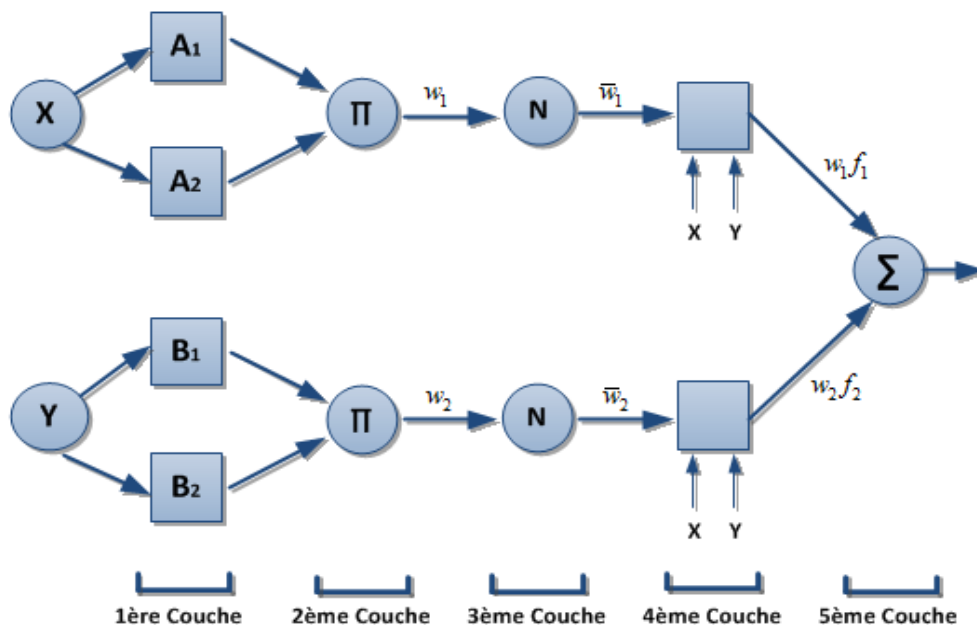


FIGURE 2.3 – Architecture ANFIS.

1ère couche : Fuzzification Cette couche permet de convertir les variables d'entrées dans l'espace flou à l'aide des fonctions d'appartenance.

$$O_i^1 = \mu_{A_i}(X) \quad \text{avec : } i = 1, 2 \quad (2.5)$$

$$O_i^1 = \mu_{B_{i-2}}(Y) \quad \text{avec : } i = 3, 4 \quad (2.6)$$

où $\mu_{A_i}(X)$ $\mu_{B_{i-2}}(Y)$ sont les degrés d'appartenance du ième terme de la partition floue des variables d'entrée. Différents types de fonctions d'appartenance peuvent être utilisés : triangle, trapèze ou gaussienne par exemple. Les paramètres définissant ces fonctions sont appelés prémisses.

2ème couche : Pondération des règles floues Chaque neurone est associé à une règle floue et permet de définir le poids de cette règle à l'aide d'une t-norme. La t-norme la plus employée est le produit :

$$O_i^2 = w_i = \mu_{A_i}(X)\mu_{B_{i-2}}(Y) \quad i = 1, 2 \quad (2.7)$$

3ème couche : Normalisation Cette couche a pour but de normaliser les différents poids.

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \quad (2.8)$$

4ème couche : Defuzzyfication Cette couche donne une sortie définie par une fonction du premier ordre des entrées pondérées (approche de Takagi-Sugeno).

$$O_i^4 = w_i f_i \quad \text{avec} \quad f = a_i \cdot X + b_i \cdot Y + c_i \quad \text{ou} \quad i = 1, 2 \quad (2.9)$$

La fonction f_i peut aussi être constante. Dans ce cas, tous les $a_i = b_i = 0$. Les paramètres a_i, b_i et c_i des fonctions f_i sont appelés paramètres conséquents.

5ème couche : Calcul de la sortie La sortie O_i^5 est définie par la somme des sorties de la couche précédente :

$$O_i^5 = \sum_{i=1}^2 w_i f_i = \frac{\sum_{i=1}^2 w_i f_i}{w_1 + w_2} \quad (2.10)$$

Phase d'Apprentissage :

L'ajustement des paramètres de l'ANFIS est réalisé lors de la phase d'apprentissage. Pour cela un ensemble de données associant les séquences d'entrées et de sorties est nécessaire. Pour la réalisation de cette phase, l'algorithme d'apprentissage hybride est utilisé.

L'algorithme d'apprentissage hybride est une association de deux méthodes celle de descente de gradient et la méthode d'estimation des moindres carrés. La première permet d'ajuster les prémisses en fixant les paramètres conséquents tandis que la seconde LSM (Least square Method) ajuste les paramètres conséquents en fixant les prémisses. [BAKF06]

3.2 Reconnaissance Neuro-floue du diabète dans la littérature

Elif Derya Ubeyli [Ü10] utilise le classifieur ANFIS pour diagnostiquer la maladie de diabète avec un dispositif de données caractérisé par 8 descripteurs. Dans la base de données de diabète d'Indiens Pima, il y a un total de 768 personnes, dont 268 sont des diabétiques et 500 non diabétiques. Les performances obtenues par cette approche sont : Spécificité 98.58% Sensibilité 96.97% Taux de classification 98.14%.

Le réseau neuro flou ANFIS a été utilisé précédemment par Çelikyılmaz et al. [CB07] pour la classification des diabétiques où le taux de reconnaissance a atteint 82%. Vosoulipour et al. [VTM08] a employé l'algorithme génétique pour la sélection des paramètres d'entrées donnant un taux de classification de 81.30%. Polat et Gunes [PG07] ont proposé l'analyse en composantes principales pour réduire l'ensemble de données du diabète de 8 à 4 variables d'entrée d'ANFIS atteignant un taux de classification de 89.47%, cependant ces résultats ont été discrédités par Temurtas et al. [TNT09] car ils ont obtenu avec la même approche un taux de reconnaissance de 66.78%. De leurs côtés, Wang et al [LWA⁺09] ont proposé dans leur étude, la classification par l'inférence floue en se basant sur des agents d'ontologie qui créent une structure des paramètres du diabète, réalisant des performances de 74.20% pour la reconnaissance de vrais positifs, 59.20% faux négatifs, et 71.50% de faux positifs.

Le tableau 2.3 présente quelques travaux de la reconnaissance du diabète avec l'approche Neuro floue.

AUTEURS	ARTICLES / RAPPORTS	APPROCHE	EXPÉRIENCE	RÉSULTATS
Robert Nowicki, 2009 [Now09]	Rough Neuro-Fuzzy Structures for Classification With Missing Data	<p>Ce papier réalise l'extension du système classique neuro-flou avec l'application de la théorie des ensembles approximatifs (Rough Neuro-Fuzzy) pour couvrir l'insuffisance des données incomplètes.</p> <ol style="list-style-type: none"> 1. Le Classifieur Rough Neuro-Fuzzy, règles obtenues par l'algorithme LEM-2 : détermine des intervalles de sous ensemble flou et non pas des paramètres. (LERS : Learning from Examples based on Rough Sets); 2. Le Classifieur Rough Neuro-Fuzzy, règles obtenues avec l'algorithme Wang and Mendel algorithm : génère la base de règles pour les ensembles flous définis à priori. 	<p>Les expériences ont été répétées 20 fois avec différentes répartitions de la série de données du diabète (768 patients) dans la phase d'apprentissage et de test. En utilisant 80 classifieurs la moyenne des résultats parvenues est de 15 667 200 décisions.</p>	<p>LERS : 8 paramètres : Tc=82.3%, Se=0%, Sp=17.7% 7 paramètres : Tc=41.7%, Se=54.7%, Sp=36% 6 paramètres : Tc=10.7%, Se=89%, Sp=0.3% Wang et Mendel : 8 paramètres : Tc=79.96%, Se=0%, Sp=20.1% 7 paramètres : Tc=0.2%, Se=99.8%, Sp=0% 6 paramètres : Tc=0%, Se=100%, Sp=0%</p>

AUTEURS	ARTICLES / RAPPORTS	APPROCHE	EXPERIENCE	RÉSULTATS
Narissara Eiamkarnitchat, Nipon Theerapong, Sansanee Auephanwiriyaikul 2010 [ETUA10]	A Novel Neuro-Fuzzy Method for Linguistic Feature Selection and Rule-Based Classification	La méthode proposée met en application un réseau neuro flou qui peut choisir les bonnes caractéristiques et les utiliser pour créer des règles de classification efficace. Les fonctions d'appartenance des variables linguistiques ont également été réalisées dans le processus d'apprentissage où un seul groupe de variables linguistiques va représenter chaque élément d'origine.	La méthode de validation croisée (10-fold cross validation) a été appliquée à l'ensemble de données diabète Pima (768). Où 5 des meilleurs paramètres informatifs de tous les ensembles de données ont été sélectionnés en fonction de leurs poids (plus le poids augmente, plus il est instructif) pour chaque classe.	Un taux de classification de 72.01% avec un nombre de règles réduit de 24 à 5 règles.
WANG Yuanzhen. ZHANG Zhibing. YI Bao-li, LI Huayang. 2005 [WZYL05]	Data Mining Based on Computational Intelligence	Cet article combine des outils de calculs intelligents comme le réseau de neurones, la logique floue et l'algorithme génétique pour développer une architecture de data mining (NFGDM). Les données d'entrée sont prétraitées par fuzzyfication, ensuite présentées au réseau probabiliste pour reconnaître les données selon les classes considérées. Une technique d'extraction de règles est ensuite appliquée afin d'extraire des connaissances explicites à partir des réseaux de neurones et la représente dans la forme de règles floues si-alors. Dans la phase finale, l'algorithme génétique est utilisé pour l'optimisation de ces dernières.	2 / 3 de la base de données ont été utilisés pour l'apprentissage et 1 / 3 pour les tests. La précision de la base de règles et la lisibilité ont été appliquées comme critères de départ pour l'évaluation de l'architecture.	Les résultats des tests de précision des bases de règles Originales : Nombre de règles : 21 TC=80.7% Extraites : Nombre de règles : 8 TC=83.1%
Nesma Settouti, Tadj. E Fandi, M. Amine Chikh. 2009 [SCF09]	Reconnaissance par Neuro-Floue du Diabète	Le modèle de Takagi-sugeno est adopté pour le réseau neuro-flou dans le cadre de la reconnaissance du diabète, avec l'application de différentes configurations en collaboration avec des experts. L'apprentissage est appliqué sur 262 personnes et a permis une optimisation des fonctions d'appartenance en modifiant ses points modaux.	Les performances du classifieur ont été testées sur les 130 patientes restantes ce qui englobent 91 non diabétiques et 39 diabétiques.	Age, Glu : Nbre règles=4 Tc= 86.92%, Age, Glu, BMI, PED : Nbre règles=24 Tc= 84.17% Age, Glu, BMI : Nbre règles=8 Tc= 82.17%

TABLE 2.3: Quelques travaux avec ANFIS

Le tableau 2.3 ne reflète que les exemples cités et ne prétend pas être exhaustif par rapport à toutes les recherches menées sur l'approche présentée. A partir des résultats obtenus à travers l'ensemble des travaux nous pouvons remarquer un certain nombre de points qui sont résumés comme suit :

- Premièrement, l'utilisation de la totalité de la base de données diabète Pima de 768 exemples sachant que 376 ont des données manquantes principalement dans le paramètre pression artérielle, ce qui nous mène à être sceptique par rapport aux résultats de la reconnaissance.
- Deuxièmement, la connaissance extraite par certains travaux sous forme de règles a été réalisée mais reste toute de même non exploitée, où le lien entre connaissances et performances reste absent.

Sachant que, l'usage des techniques intelligentes pour l'aide au diagnostic a comme objectif principal une adaptation personnalisée avec une transparence maximale.

4 Fuzzy C-Means FCM

4.1 Théorie

Principe du clustering

Le clustering (regroupement) est une méthode de description des données primaires dans l'exploration de données permettant de rassembler la plupart des données similaires. L'analyse de regroupement a joué un rôle important dans la résolution de nombreux problèmes comme la médecine, la psychologie, la biologie, la sociologie, la reconnaissance de formes et traitement d'images.

Le regroupement des objets de données dans des clusters est l'étape principale. Les algorithmes de clustering ont tenté d'évaluer l'interaction entre les modèles en organisant des schémas en groupes, de tel sorte que les modèles au sein d'un cluster sont plus semblables les uns aux autres. De grands nombres d'algorithmes de classification peuvent être trouvés dans la littérature [JD88, HKKT99].

Les données sont regroupées en fonction d'un certain nombre d'approches différentes. La classification avec un regroupement partitionnel ou hiérarchique sont deux des principales techniques. Le regroupement partitionnel commence souvent à partir d'une partition initiale et optimise (habituellement au niveau local) avec un critère de regroupement.

Le regroupement partitionnel porte essentiellement sur la tâche de partitionnement d'un ensemble d'entités dans un certain nombre d'homoclusters hétérogènes, par rapport à une mesure de similarité. En raison de la nature floue d'un certain nombre de problèmes pratiques, plusieurs méthodes floues de classification ont été élaborées à la suite de la théorie générale des ensembles flous esquissées par Zadeh [Zad65].

Les deux principales techniques du regroupement partitionnel sont : le regroupement rigide traditionnel, dans lequel chaque objet de données appartient à un cluster unique, et la classification floue, dans laquelle chaque objet est autorisé à appartenir à de nombreux groupements à des degrés divers d'appartenance.

L'algorithme Fuzzy C-Means FCM

L'algorithme Fuzzy c-Means (FCM) [Bez81] est l'un des algorithmes les plus largement utilisés dans les méthodes de regroupement flou. Il repose sur le concept de c-partition flou, présenté par Ruspini [Rus69].

L'algorithme FCM est basé sur la minimisation de la fonction objective suivant la formule suivante :

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty \quad (2.11)$$

Où :

- m est un nombre réel supérieur à 1,
- u_{ij} est le degré de l'adhésion de x_i dans le cluster j ,
- x_i est le i ème de d -dimension des données mesurées,
- c_j est le centre d -dimension de la grappe,
- $\|x_i - c_j\|^2$ est la distance Euclidienne entre les données mesurées et le centre.

Le partitionnement flou est réalisé à travers une optimisation itérative de la fonction objective indiquée ci-dessus, avec la mise à jour de l'adhésion u_{ij} et les centres de cluster c_j par :

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (2.12)$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m} \quad (2.13)$$

Cette itération s'arrête lorsque $\max_{ij} |u_{ij}^{(k+1)} - u_{ij}^{(k)}| < \epsilon$, où ϵ est un critère de terminaison entre 0 et 1, tandis que k est le nombre d'étapes d'itération. Cette procédure converge vers un minimum local ou un point-selle de J_m .

L'algorithme est composé des étapes suivantes :

Etape1 : Initialiser Matrice $U = [u_{ij}] = U(0)$

Etape2 : À k-étape : calculer les vecteurs de centres $C(K) = [c_j]$ Avec $U(K)$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m \cdot x_i}{\sum_{i=1}^N u_{ij}^m}$$

Etape3 : Mise à jour $U(K), U(K+1)$

$$u_{ij} = \frac{1}{\sum_{k=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}}$$

Etape4 : Si $\max_{ij} \|U(K+1) - U(K)\| < \epsilon$ alors STOP, sinon le retour à l'étape 2.

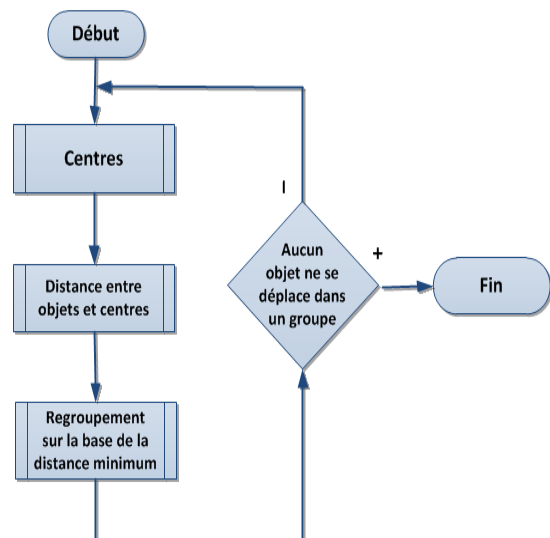


FIGURE - Organigramme de l'algorithme FCM.

L'utilisation de l'algorithme FCM exige la détermination de plusieurs paramètres, à savoir, c, m, ϵ . En outre, l'ensemble $U(0)$ de centres de cluster initial doit être défini. Même si aucune base théorique n'est disponible pour une bonne valeur de m , $1, 1 < m < 5$ est généralement reportée comme la plus utile des valeurs [RLCB86].

4.2 Application de FCM dans la littérature

L'algorithme fuzzy c-means (FCM) [Bez81], est un algorithme de groupement typique, qui a été employé dans une grande variété de l'ingénierie et disciplines scientifiques tel que la modélisation [ML00], la prise de décision [LS99], modèle de reconnaissance et la classification [CLF06], la segmentation [Kan08] ainsi que des domaines d'applications médicales [MR06]. Dans le domaine des applications médicales, FCM est utilisé avec succès pour modéliser et analyser les processus médicaux, la prise de décision [PKKS09], [LWA⁺09], la segmentation d'images [KSC06] [IKM10] [HSA09] et le classement des tumeurs [PSS⁺06] [SPGN06]. Des travaux récents ont démontré que FCM en comparaison avec HCM (Hard C-Means) [SK10] et MFPCM (Modified Fuzzy Possibilistic C-Mean) [RV10], peut fournir des solutions approximatives plus rapides, qui conviennent pour le traitement des questions liées à l'intelligibilité des modèles et des données incomplètes et bruitées.

5 Hybridation Fuzzy C-Means & Neuronal Network (FCNN)

5.1 Principe

Les RNs ont été appliquées à un large éventail de problèmes en ingénierie, en informatique et en mathématiques malgré le fait qu'ils présentent des lacunes dans l'apprentissage, cependant comme le montrent Hruschka et Natter [HN99], ces lacunes débutent plus avec des faiblesses dans la modélisation des RNs (leurs architectures spécifiques) plutôt que les défaillances des réseaux de neurones en général. En accord avec cette observation, nous proposons de modéliser le réseau de neurone avec une méthode de regroupement (clustering) Fuzzy c-means a son entrée, afin d'optimiser son architecture pour une meilleure classification.

5.2 Applications de FCNN dans la littérature

L'hybridation de la méthode fuzzy c-means FCM avec les réseaux de neurones a fait ses débuts dans le domaine médical et principalement dans la segmentation des images médicales avec FCM-réseau de Hopfield [CLM96], FCM-réseau de Kohonen [WMBOLS04] [JM08].

L'approche FCNN a aussi montré son intérêt dans les domaines de classification [HAJ09] [XH08] [LDZ00] [OCK06] optimisation [HEHM07] identification [LCT09] détection [CCLC99] et prédiction [Che07] [Sin08].

Le tableau 2.6 présente l'application de l'approche FCNN dans le domaine médical.

AUTEURS	ARTICLES	APPROCHE	EXPÉRIENCE	RÉSULTATS
Dipti Patra, Manab Kumar Das, Smita Pradhan. 2010 [PDP10]	Integration of FCM, PCA and Neural Networks For Classification of ECG Arrhythmias	Dans cet article, il est proposé d'intégrer FCM ACP et les réseaux de neurones (NN) pour la classification de l'ECG. Le regroupement FCM est employé pour la réduction de données. Le réseau de neurones (NN) comme classifieur. Une étude comparative a été réalisée entre quatre structures : FCNN, PCA-NN, FCM-ICA-NN, et la FCM-PCA-NN	A partir de la base de données MIT-BIH, six différents types de battements attribuant une arythmie cardiaque sont échantillonnés. Le but de l'utilisation de fuzzy c-means (FCM) est de diminuer le nombre de segments en regroupant des données similaires d'apprentissage. Les caractéristiques des modèles de formation obtenus en clusters sont extraites.	FCM-ICA-NN Training Error $3.99 \times 10^{-4}\%$ Test Error 0.45% FCM-PCA-NN Training Error $3.99 \times 10^{-4}\%$ Test Error 0.13% FCNN Training Error $3.99 \times 10^{-4}\%$ Test Error 0.25% PCA-NN Training Error $3.99 \times 10^{-4}\%$ Test Error 0.21%

AUTEURS	ARTICLES	APPROCHE	EXPÉRIENCE	RÉSULTATS
Sang-Hyun Hwang, Dongwon Kim, Tae-Koo Kang, and Gwi-Tae Park. 2009 [HKK ⁺ 07]	Medical Diagnosis System of Breast Cancer Using FCM Based Parallel Neural Networks	<p>Une méthode pour le diagnostic médical du cancer du sein à base de classification floue (fuzzy c means) et des réseaux de neurones parallèles est proposée. La première étape est de classer les données du cancer du sein en utilisant Fuzzy C-means (FCM). La seconde est la conception des réseaux de neurones en utilisant les clusters réalisés par FCM. Les expérimentations ont été faites en 3 étapes.</p> <ul style="list-style-type: none"> - Etape 1 : la base de données WBCD (Wisconsin Breast Cancer Data) est classée avec Fuzzy C-means. - Etape 2 : créer une table d'analyseur de décision qui sélectionne un des réseaux de neurones optimaux. - Etape 3 : tester avec des réseaux de neurones parallèles. 	L'expérimentation est réalisée avec 683 individus dans la base de données WBCD. Des expérimentations ont été simulées avec 400 données d'apprentissage et 283 données de test. Les résultats sont obtenus en variant le nombre de centre.	<p>No cluster Tc= 95.5842%</p> <p>2 cluster Tc= 96.0342%</p> <p>3 cluster Tc= 98.1284%</p> <p>4 cluster Tc= 98.2462%</p> <p>5 cluster Tc= 99.0459%</p> <p>6 cluster Tc= 99.2933%</p> <p>7 cluster Tc= 99.5289%</p>
Hassan H. Haseena & Paul K. Joseph & Abraham T. Mathew, 2010 [HJM10]	Classification of Arrhythmia Using Hybrid Networks	Détection des arythmies basée sur le traitement des signaux numériques de l'ECG. L'approche de cet article se concentre sur un Fuzzy C-Means (FCM), regroupé avec un réseau de neurone probabiliste (PNN) et un réseau de neurone multicouche (MLPNN) pour la discrimination de huit types de battements ECG. Des paramètres tels que : AR, SE sont extraits de chaque battement de l'ECG et la réduction de fonctionnalité a été réalisée à l'aide de FCM.	Les expérimentations ont été testées sur la base de données MIT-BIH.	<p>PNN-FCM : Tc=99.05%</p> <p>MLPNN-FCM : Tc= 97.14%</p>

AUTEURS	ARTICLES	APPROCHE	EXPÉRIENCE	RÉSULTATS
Rahime Ceylan, Yuksel Ozbay. 2007 [CO07]	Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using artificial neural network	Dans cet article, quatre structures différentes, FCNN, PCA-NN, FCM-PCA-NN et WT-NN sont présentées pour la classification de L'ECG. Les signaux ECG pris de la base de données MIT-BIH, sont appliqués dans la reconnaissance de 10 arythmies différentes : le rythme sinusal normal, bradycardie sinusale, tachycardie ventriculaire, arythmie sinusale, contraction prématurée auriculaire, rythme rapide, bloc de branche droit, bloc de branche gauche, fibrillation auriculaire et palpitation auriculaire.	Avant les tests, les structures proposées sont entraînées par l'algorithme de rétro propagation. Ensuite elles sont testées en utilisant des enregistrements d'ECG de 92 patients (40 hommes et 52 femmes, avec un âge moyen de $39,75 \pm 19,06$)	NN : Training Error 0.28% Test Error 0.22% FCNN : Training Error 9.92×10^{-10} % Test Error 0.19% PCA-NN : Training Error 7.07×10^{-9} % Test Error 0.95% FCM-PCA-NN : Training Error 7.76×10^{-9} % Test Error 0.91% WT-NN : Training Error 4.97×10^{-9} % Test Error 0.4%

TABLE 2.6: Quelques travaux de FCNN

6 Hybridation Fuzzy C-Means & Adaptive Neuro-Fuzzy Inference System

6.1 Principe

Lorsque le nombre de paramètres à l'entrée d'ANFIS augmente, la partition des fonctions d'appartenance et le nombre de règles floues deviennent un problème croissant. Dans ce mémoire, nous traitons ce problème en adoptant la stratégie de regroupement de FCM (Figure 2.4). Le regroupement des données permet d'identifier les paramètres initiaux et la structure basée sur la partition de dispersion. Par conséquent, la technique peut être utilisée pour réduire la dimension des modèles ainsi que le temps d'apprentissage, puisque le nombre de règles floues est égal au nombre de fonctions d'appartenance qu'elle que soit la dimension des entrants. Cette approche ANFIS modifiée est nommée FCM-ANFIS [KLK⁺01].

6.2 Applications de FCM-ANFIS dans la littérature

La méthode FCM-ANFIS a déjà été appliquée dans plusieurs domaines : classification [Elr06], prédiction [SBAS09], reconnaissance [KLK⁺01], détection [ZJW08], l'optimisation [FA10a], dans le domaine médical avec dernièrement la classifica-

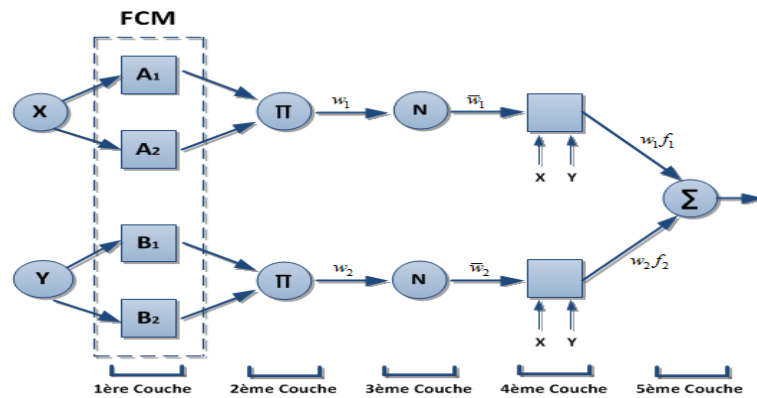


FIGURE 2.4 – Architecture de Fuzzy C-Means & Adaptive Neuro-Fuzzy Inference System

tion d’une base de données du diabète [OY07] où les données sont recueillies par la Faculté des sciences informatiques et d’information, Université de Technologie de Malaisie, où les résultats obtenus ont atteint les 72,66%.

Le tableau 2.7 présente l’application de l’approche FCM-ANFIS dans la littérature.

AUTEURS	ARTICLES	APPROCHE	EXPÉRIENCE	RÉSULTATS
M.N. SYED-AHMAD, A. Ben-senouci, S. A. Al-ghamdi, A. M. A. Ghany. 2000 [SABAG00]	Short-term load forecasting Using adaptive neuro-fuzzy inference system (anfis) Application to aleppo load demand	Ce papier présente les résultats préliminaires de l’application de l’inférence adaptative Neuro-Fuzzy System (ANFIS) avec FCM pour la prévision à court terme de la consommation d’énergie électrique.	Collecte de données à partir d’Alep Electric Company, à Alep, en Syrie, où la consommation horaire pour les trois dernières années (1997 - 1999) a été obtenue.	Il a été constaté que, les prévisions d’ANFIS correspondent au mieux à la demande réelle. L’erreur quadratique moyenne (MSE) a été estimée à environ 0,5%.
Deok Hee Nam and Edward Asikele. 2009 [NA09]	Data Mining of Image Segments Data with Reduced Neurofuzzy System	Ce travail traite de la reconnaissance des formes à l’échelle de l’image en fonction des caractéristiques de ces dernières, tout en réduisant la dimension originale des images.	Pour évaluer la performance du système, les systèmes neuro-flous sont appliqués avec une analyse multi variée, y compris l’analyse factorielle, l’analyse en composantes principales, et Fuzzy C-means.	Pour un nombre limité de données d’apprentissage d’entrée-sortie, l’algorithme proposé offre de meilleurs résultats en comparaison avec les techniques de la segmentation d’images.

AUTEURS	ARTICLES	APPROCHE	EXPÉRIENCE	RÉSULTATS
Seung-Heon Lee, M. Viswanathan, and Young-Kyu Yang. 2006 [LVY06]	Link Speed Estimation and Incident Detection Using Clustering and Neuro-fuzzy Methods	Ce papier rentre dans le contexte de développement des systèmes avancés d'information aux voyageurs (ATIS) au sein des systèmes de transport intelligents (ITS) pour l'estimation optimale du temps de voyage (autoroute) et de détection des incidents avec une précision raisonnable.	L'objectif de la recherche est une analyse exhaustive des informations recueillies auprès de sources GPS en utilisant l'algorithme fuzzy c-means (FCM), qui fournit l'estimation de la vitesse de liaison. Le FCM est employé pour extraire des motifs à partir des données du trafic routier collectées dans un réseau très animé des rues du centre-ville. L'estimation de vitesse de liaison est effectuée à l'aide des techniques de lissage. Enfin, l'algorithme neuro-flou est appliqué pour la tâche de détection d'incident de la circulation routière.	Cette recherche a amélioré la précision dans l'estimation de la vitesse du lien et le travail de détection des incidents. Cette procédure est réalisée avec la fusion des données provenant de multiples sources et informations sur la vitesse des principales artères et autoroutes.
Hai-Tao He, Xiao-Nan Luo, and Bao-Lu Liu. 2005 [HLL05]	Detecting Anomalous Network Traffic with Combined Fuzzy-Based Approaches	Ce document présente les approches combinées FCM-ANFIS afin de déceler les réseaux de trafic anormaux tels que DoS/DDoS ou les attaques probables.	L'idée de base de l'algorithme est : utiliser ANFIS à espace multidimensionnel caractéristique (MD) de connexions réseau pour le transformer en un espace compact caractéristique unidimensionnelle (1-D). D'autre part le regroupement FCM est appliqué pour classer l'espace 1-D dans la fonction anormale et la fonction normale. ACP est également utilisé pour la réduction dimensionnelle de l'espace des caractéristiques originales au cours de l'extraction de ces dernières.	Un ensemble de données publiquement disponibles DRAPA/KDD99 est utilisé pour démontrer les approches. Les résultats montrent leur précision dans la détection d'anomalies des connexions réseaux.

AUTEURS	ARTICLES	APPROCHE	EXPÉRIENCE	RÉSULTATS
Shrishail Hire-math. Sarat Kumar Patra, 2010 [HP10]	Transmission Rate Prediction for Cognitive Radio Using Adaptive Neural Fuzzy Inference System	L'article présente la cinquième génération de radio intelligente qui est la radio cognitive (CR) système qui fonctionne sur le taux de données de prévision. Il propose la méthode FCM-ANFIS comme régime d'apprentissage pour introduire l'intelligence.	La performance de ce travail est comparée au régime de base du réseau de neurones avec une complexité réduite.	ANFIS : Nbres règles=32 Taux appr=91% Taux valid=89% ANFIS-FCM : Nbres règles=10 Taux appr=86% Taux valid =84% Nbres règles=15 Taux appr=91% Taux valid=83% Nbres règles=20 Taux appr=98% Taux valid= 91% RNs : Taux appr=83% Taux valid=81%
Wei Ji, Raouf N.G. Naguib, John. Macall, Dobrila Petrovic, Elena Gaura, Mohamed Ghoneim. 2003 [JNM+03]	Prognostic prediction of Bilharziasis-related bladder cancer by neuro-fuzzy classifier	Un classifieur hybride FCM-neuro-flou est appliqué pour déterminer les résultats à long terme des patients atteints de cancer de la vessie bilharziose. Le même ensemble de données est également analysé par un réseau (MLPNN) et la régression logistique, qui sont tous deux largement utilisés dans le domaine d'aide à la décision médicale.	Afin de mieux évaluer la valeur de ce classifieur neuro-flou, des données du domaine d'oncologie : The Wisconsin Breast Cancer Data (WBCD) et bladder cancer dataset (BCD), sont examinées par les trois méthodes. Les ensembles de données WBCD et BCD sont divisés sur la base de 50% pour l'apprentissage et 50% pour les tests	Régression logistique : WBCD Tc= 96.5% BCD Tc = 73.1% MLPNN : WBCD Tc= 97.4% BCD Tc = 78.2% FCM-ANFIS : WBCD Tc= 97.1% BCD Tc = 84.9%

TABLE 2.9: Quelques travaux avec FCM-ANFIS

7 Synthèse et discussion des différentes techniques présentées pour la classification du diabète

Techniques et approches	Réseaux de neurones NN	Neuro-flou ANFIS	Fuzzy C-means FCM	FC NN	FCM-ANFIS
Paramètres	Poids synaptiques	Degrés d'appartenance	Degrés d'appartenance	Poids synaptiques	Degrés d'appartenance
Structure	- Nombre de neurones - Nombre de couches cachées	- 5 couches - Nombres de règles = Nombres de fonctions d'appartenance	Nombre de centre	FCM fixe les neurones d'entrées et l'architecture du réseau	- 5 couches - FCM réalise le partitionnement des fonctions d'appartenance
Connaissance	Absente	Base de Règles floues	Degrés d'appartenance	Absente	Base de règles floues réduites

TABLE 2.10 – Caractéristiques des différentes techniques présentées pour la classification du diabète

Le tableau 2.10 résume les discussions faites dans cette section à propos des différentes techniques. Certaines entrées seront une généralisation ; nous nous référons donc aux discussions précédentes pour une comparaison plus approfondie. Nous remarquons un certain nombre de points : Les approches ANFIS, FCNN, FCM-ANFIS offrent une grande puissance d'apprentissage grâce à l'apport des réseaux de neurones. Cependant, concernant la connaissance et l'interprétabilité, NN s'avère très limité (boite noire) à l'instar des techniques floues ANFIS, Fuzzy C-Means. Un compromis doit être réalisé entre les différentes caractéristiques pour le choix du classifieur adapté à la tâche de reconnaissance.

8 Contribution

Les données Pima ont été exploitées bien souvent dans la classification avec des méthodes très différentes. Une première étude de ces données a été réalisée par Smith et al. [SED⁺88] qui les a illustrées en utilisant une variété de méthodes. Par la suite différents travaux ont contribué à l'amélioration des systèmes d'aide au diagnostic du diabète. Les meilleurs résultats à notre connaissance, sont ceux d'Ubeyli [Ü10] pour la classification d'ensemble du diabète Indiens Pima avec la méthode Adaptive Neuro Fuzzy Inference System ANFIS. Cette dernière a fait ressortir un taux de reconnaissance de 98.14%, Spécificité 98.58%, Sensibilité 96.97%. Des résultats très prometteurs mais aucune information sur l'architecture du réseau ainsi que la répartition des fonctions d'appartenance. La connaissance

réalisée a été mise à l'écart, ce qui limite la technique ANFIS juste sur les performances de classification sans montrer la capacité d'interprétation des résultats.

En parcourant la littérature scientifique impliquant les différentes méthodes plus particulièrement (Réseau de neurones, adaptive neuro fuzzy inférence système ANFIS) relatives au diabète et autres, nous constatons en premier lieu que le pré-traitement des données avec l'Analyse en Composantes Principales (ACP) comme exemple n'a pas amélioré les résultats, mais en appliquant l'algorithme Fuzzy C-means, la précision de classification a été un peu plus élevée. En Second lieu ce que nous remarquons que toutes ces approches se focalisent sur les performances des méthodes avec l'absence de l'intrétabilité, rares sont les cas où la connaissance réalisée est abordée mais demeure inexploitée. Ce qui implique la non transparence des systèmes sachant que l'objectif principal des systèmes d'aide au diagnostic médical est la lisibilité.

Dans ce mémoire de magister notre travail s'oriente vers le renforcement de l'aide au diagnostic médical par une caractérisation et une reconnaissance interprétable des données du diabète. Cela consiste à étudier plus particulièrement l'intérêt d'automatiser l'apprentissage structurel et paramétrique afin d'augmenter les performances du classifieur sur le plan précision et interprétabilité en renforçant la connaissance extraite.

9 Conclusion

Nous remarquons que les méthodes dites intelligentes comme les réseaux de neurones, la logique floue, les algorithmes génétiques et leurs hybridations ont été largement utilisés dans la littérature et en particulier dans le domaine médical. Après exploration de l'état de l'art nous déduisons que l'apport de Fuzzy C-Means (FCM) peut être d'une grande importance dans l'optimisation de la connaissance.

Chapitre 3

Propositions

1 Introduction

Dans ce chapitre nous élaborons nos différentes contributions proposées en parcourant l'état de l'art (voir chapitre 2), Pour cela nous divisons ce chapitre en 3 parties où chacune d'elles traite des résultats d'une considération appliquée. La procédure de classification du diabète est représentée dans le schéma suivant :

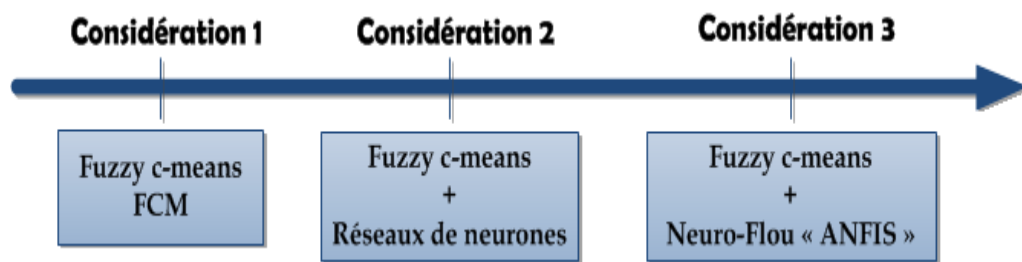


FIGURE 3.1 – Schéma représentatif de la procédure de classification du diabète.

Dans la considération 1 la classification du diabète est mise en œuvre avec l'algorithme non supervisé Fuzzy C-Means. La considération 2 présente l'application de Fuzzy C-means en amont du réseau de neurones multicouches afin d'augmenter son efficacité. La considération 3 porte son intérêt sur l'apport de la méthode FCM pour améliorer l'interprétabilité du classifieur Neuro Flou ANFIS. Pour l'implémentation de ces approches, nous avons fait recours aux logiciels : R 2.12.0 (open source) et Matlab 2010a.

2 Base de données

Dans ce mémoire de Magister nous utilisons la base de données médicale réelle (Indians Diabetes Pima). L'ensemble de données a été choisi du dépôt d'UCI [FA10b] qui réalise une étude sur 786 femmes Indiennes Pima, Ces mêmes femmes, qui ont stoppé leurs migrations en Arizona, États-Unis, adoptant un mode de vie occidentalisé, développent un diabète dans presque 50% des cas (record mondial, figure 1.9 chapitre1).

Parmi ces 782 femmes nous utilisons uniquement 392, qui ont des paramètres complets contrairement au reste à savoir les 140, celles-ci présentaient des données manquantes principalement sur la pression artérielle.

Le diagnostic est une valeur binaire variable «classe» qui permet de savoir si le patient montre des signes de diabète selon les critères de l'Organisation Mondiale de la Santé. Les huit descripteurs cliniques sont :

1. Npreg : nombre de grossesses,
2. Glu : concentration du glucose plasmatique,
3. BP : tension artérielle diastolique,
4. SKIN : épaisseur de pli de peau du triceps,
5. Insuline : dose d'insuline,
6. BMI : index de masse corporelle,
7. PED : fonction de pedigree de diabète (l'hérédité).
8. Age : âge

Les figures 3.2 et 3.3 montrent respectivement les histogrammes et la répartition des diabétiques et non diabétiques des paramètres de la base de données :

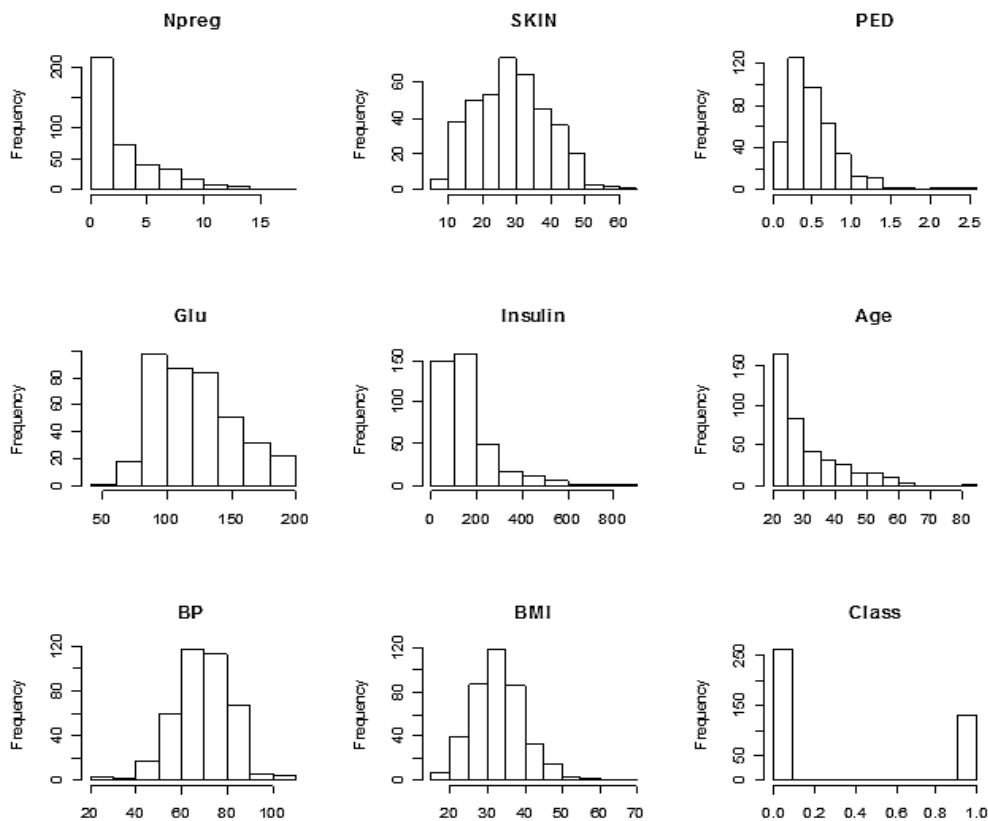


FIGURE 3.2 – Histogrammes des 9 paramètres de la base de données

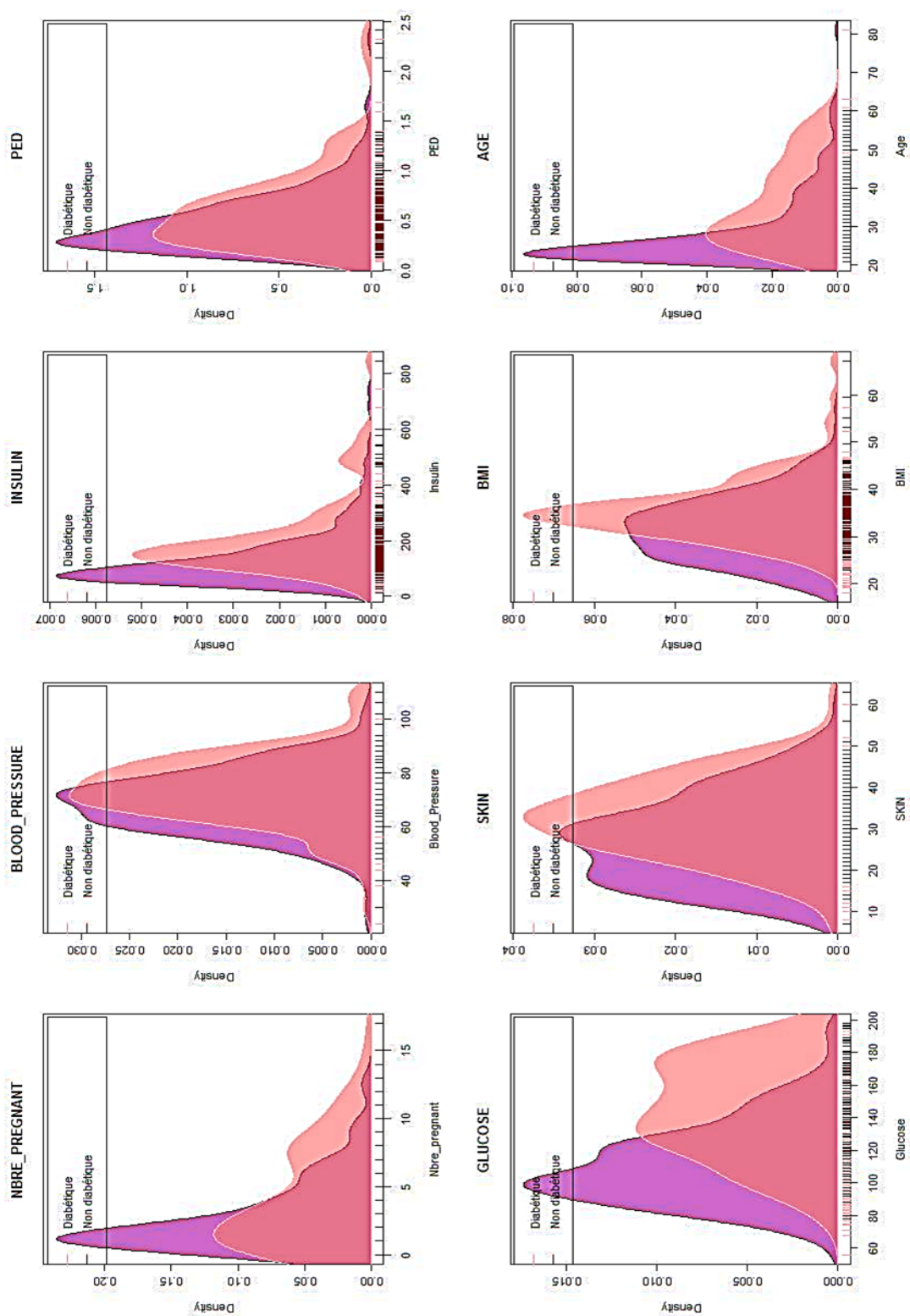


FIGURE 3.3 – Représentation de la base de données : distribution des diabétiques et non diabétiques dans les paramètres de la base de données.

3 Considération 1 : Classification par Fuzzy C-Means

3.1 Principe

Fuzzy C-Means (FCM) est un algorithme de classification non-supervisée floue. Issu de l'algorithme des C-Moyens (C-means), il introduit la notion d'ensemble flou dans la définition des classes : chaque point dans l'ensemble des données appartient à chaque cluster avec un certain degré, et tous les clusters sont caractérisés par leur centre de gravité. Comme les autres algorithmes de classification non supervisée, il utilise un critère de minimisation des distances intra-classe et de maximisation des distances inter-classe, mais en donnant un certain degré d'appartenance à chaque classe pour chaque point. Cet algorithme nécessite la connaissance préalable du nombre de clusters et génère les classes par un processus itératif en minimisant une fonction objective.

La classification floue des objets est décrite par une matrice floue à n lignes et c colonnes dans laquelle n est le nombre d'objets de données et c est le nombre de grappes. μ_{ij} , l'élément de la i -ème ligne et la colonne j . Dans ce chapitre nous élaborons les différentes contributions proposées en parcourant l'état de l'art (voir chapitre 2), Pour cela nous divisons ce chapitre en 3 parties distinctes. Chaque partie traite des résultats d'une considération appliquée. La procédure de classification du diabète est représentée dans le schéma suivant : dans μ , représente le degré de fonction d'appartenance de l' i -ème objet avec le pôle j .

Les étapes de l'algorithme FCM (Dunn [Dun74] ; Bezdek [Bez81]) sont présentées ci-dessous :

Algorithm 1 Pseudo code de Fuzzy C-Means

- 1: **for** $t = 1, 2, \dots$ **do**
 - 2: **Step1** Calcule des centres de clusters $c_i^{(t)} = \frac{\sum_{j=1}^N (\mu_{ij}^{(t-1)})^m x_j}{\sum_{j=1}^N (\mu_{ij}^{(t-1)})^m}$
 - 3: **Step2** Calcule des distances D_{ijA}^2 avec :

$$D_{ijA}^2 = (x_j \checkmark c_i)^T A (x_j \checkmark c_i), \quad 1 \leq i \leq n_c, 1 \leq j \leq N.$$
 - 4: **Step3** Mise à jour de la matrice des partitions floues :

$$\mu_{ij}^{(t-1)} = \frac{1}{\sum_{k=1}^{n_c} (D_{ijA} / D_{kjA})^{2/(m-1)}}$$
 - 5: **end for**
 - 6: **return** $\|U^{(t)} - U^{(t-1)}\| < \epsilon$
-

Le FCM est un algorithme très puissant, mais son inconvénient réside dans l'initialisation des centres. Pour cette raison, il nous a paru intéressant de développer un procédé dans lequel nous n'aurions pas à fixer le nombre de classes mais qui déterminerait automatiquement un nombre de classe optimal. C'est ce qu'on appellera indice de validité des clusters (cluster validity index).

3.2 Choix du nombre de centre C

Plusieurs indices de validité de clusters sont proposés dans la littérature. Bezdek a défini deux indices : le Coefficient de partition (V_{PC}) [Bez74] et l'Entropie de

Partition (V_{PE}) [Bez81]. Ils sont sensibles au bruit et à la variation de l'exposant m . D'autres indices V_{FS} et V_{XB} sont proposés respectivement par Fukayama et Sugeno [FS89] et Xie-Beni [XB91]; V_{FS} est sensible aux valeurs élevées et basses de m , V_{XB} donne de bonnes réponses sur un large choix pour $c = 2, \dots, 10$ et $1 \leq m \leq 7$. Cependant, il décroît rapidement avec l'augmentation du nombre de clusters. Wu et al. [WY05] ont apporté une amélioration à cet indice. Halkidi et al. [HBV02] ont défini $V_{S_{D_{bw}}}$ basé sur les propriétés de compacité et de séparation de l'ensemble des données. Cet indice donne de bons résultats en cas de classes compactes et bien séparées, notamment quand il n'y a pas de chevauchement.

Tous ces indices restent performant mais cependant, ils sont incapables de déterminer le nombre réel de clusters dans le cas où il y a un réel chevauchement.

En considérant tous ces indices et leurs limites, nous nous sommes intéressés à une nouvelle méthode x-means basée sur l'algorithme K-means pour optimiser notre choix du nombre de centre de manière claire et rapide. Afin de consolider les résultats de cette technique nous avons appliqué l'indice de validité de cluster Silhouette pour juger la qualité de toutes solutions de clustering.

x-means

Bien que K-means est principalement employé comme méthode de clustering, elle possède quelques limites :

- Le nombre de clusters K doit être fourni par l'utilisateur,
- La recherche est sujette à des minima locaux.

Pelleg et Moore [PM00] ont proposé des solutions pour la première limite, et une autre partielle pour la seconde. Ils introduisent un nouvel algorithme de manière efficace, pour les recherches de l'espace des lieux de clusters et le nombre de clusters. Cela donne lieu à un algorithme, statistiquement fondé qui génère à la fois le nombre de classes et leurs paramètres. Les expériences montrent que cette technique révèle le véritable nombre de classes (clusters) dans la distribution sous-jacente, et qu'il est plus rapide que K-means pour différentes valeurs de K [PM00], ainsi que tous les indices de validité de clusters utilisés dans la littérature. Cette méthode proposée est appelée x-means. L'algorithme se compose de deux opérations qui se répètent jusqu'à la fin.

Algorithm 2 The x-means algorithm

- 1: **Improve – Params**, consists of running conventional K-means to convergence
 - 2: **Improve – Structure**, finds out **If** and **Where** new centroids should appear.
 - 3: **If** $K > K_{max}$ stop and report the best scoring model found during the search.
 - 4: **else**, goto 1.
-

- Apprentissage paramétrique : l'opération the Improve-Params : applique le conventionnel K-means pour la convergence.

- Apprentissage structurel : l'opération The Improve-Structure : trouve S_i et O_i de nouveaux centres de gravité devront apparaître.

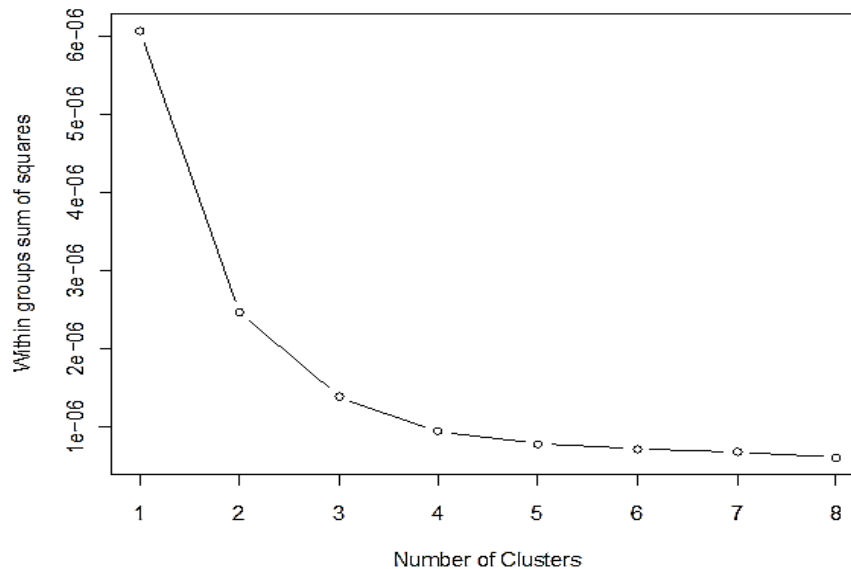


FIGURE 3.4 – Nombre de centres optimaux avec la technique x-means

La figure 3.4 montre les résultats d'application de l'algorithme x-means pour notre base de données qui nous a permis de déterminer un nombre de cluster optimal de 1 à 5 pôles. Le meilleur choix du nombre de clusters selon la figure 3.4 correspond à $c = 2$, où la somme des carrés au sein des groupes est de $2e^{-8}$ (valeur relativement faible).

Nous appliquons l'indice Silhouette pour confirmer et renforcer le choix de cluster ;

La technique de validation Silhouette

Cette technique [KR90] calcule :

- La largeur de silhouette pour chaque échantillon,
- La largeur moyenne de silhouette pour chaque centre,
- La largeur moyenne hors silhouette de l'ensemble des données.

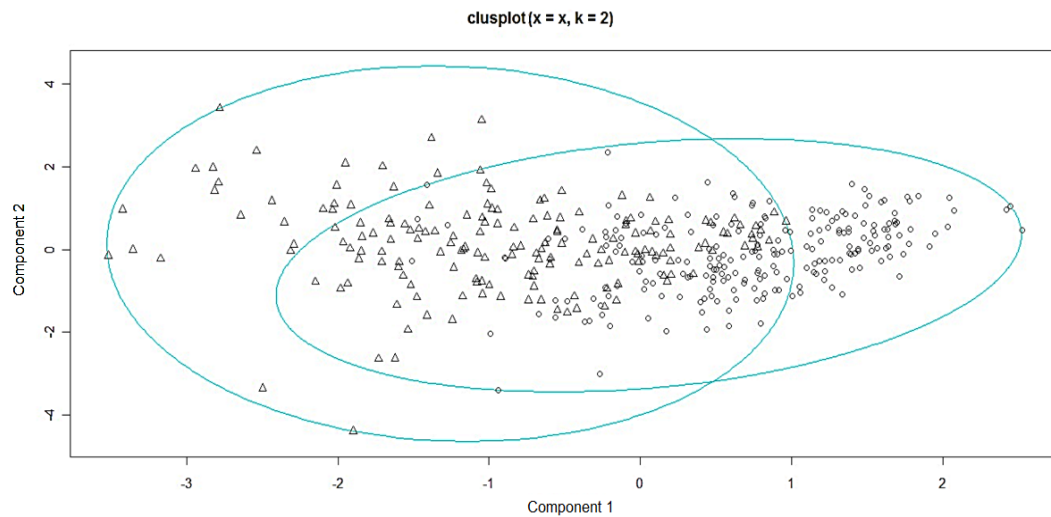
En utilisant cette technique, chaque groupe pourrait être représenté par une silhouette, elle est basée sur la comparaison de son étanchéité et de la séparation. La largeur moyenne de la silhouette pourrait être appliquée pour l'évaluation de la validité de clustering et aussi pour décider sur le bon choix du nombre de centres sélectionnés.

Supposons que a représente la distance moyenne d'un point par rapport aux autres points du centre auquel ils sont attribués, et b représente le minimum des distances moyennes du point à partir des points des autres centres. Le calcul de la largeur silhouette peut être défini comme suit :

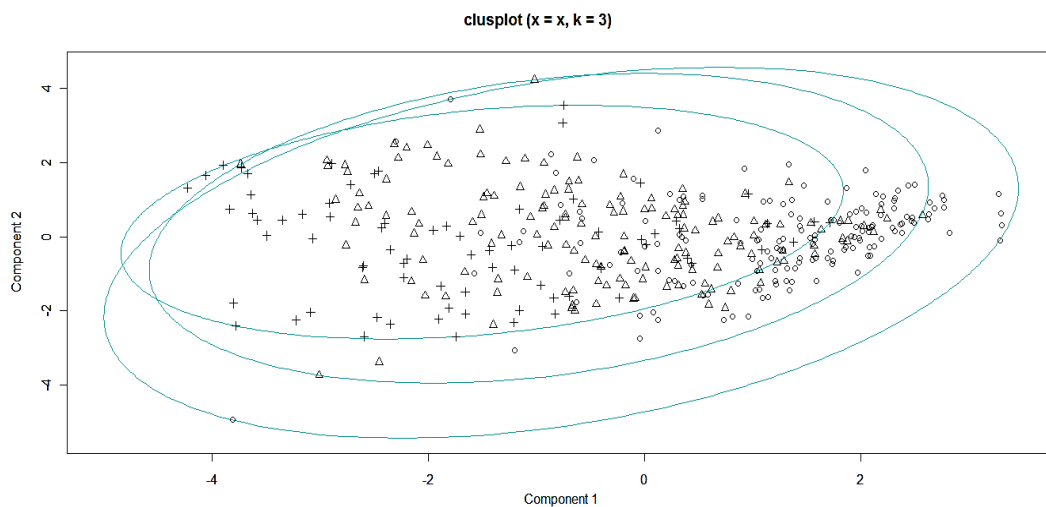
$$S(i) = \frac{(b(i) - a(i))}{\max[a(i), b(i)]}$$

Sachant que $-1 \leq S(i) \leq 1$. Si la valeur de la silhouette est proche de 1, cela signifie que l'échantillon est «bien-classé» et il a été affecté à un groupe tout à fait approprié. Si la valeur de la silhouette est proche de zéro, cela signifie que cet échantillon pourrait être attribué à un autre plus proche cluster. Si la valeur de la silhouette est proche de -1, cela signifie que l'échantillon est «mal classé» (il se trouve quelque part entre les clusters). Par conséquent, le nombre de clusters avec une largeur maximale de la silhouette globale moyenne est considéré comme le nombre optimal de centres.

L'application de l'indice Silhouette sur la base de données Pima avec un nombre de centre égal à 2 et 3 nous donne les résultats suivants (figure 3.5) :



Ces deux composantes expliquent 79.14% de la variabilité



Ces deux composantes expliquent 49.94% de la variabilité

FIGURE 3.5 – Répartition des exemples sur 2 et 3 centres

Avec un nombre de centre égal à 2 les deux composantes obtenues comportent plus de 79% de la variabilité des points (voir figure 3.5), la répartition des exemples est clairement dispersée sur les 2 centres représentés par des cercles. Avec $c=3$ les 3 composantes n'expliquent que 49.94% de la variabilité des points. La figure 3.5 montre que les 3 centres ont un fort degré de chevauchement. Aussi on remarque clairement qu'avec $c=3$ la distinction est médiocre par rapport à 2 clusters. La solution à deux clusters est la plus satisfaisante, car elle correspond à une partition intuitive, bien qu'elle n'isole pas l'exception.

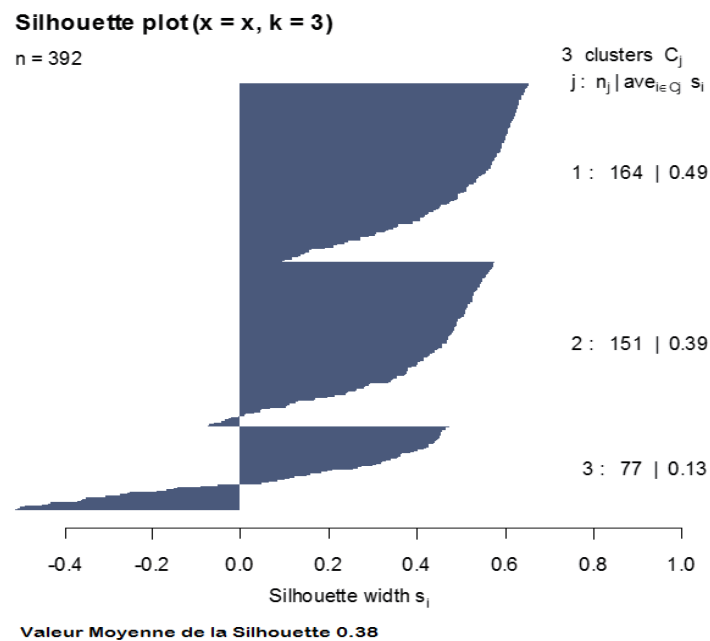
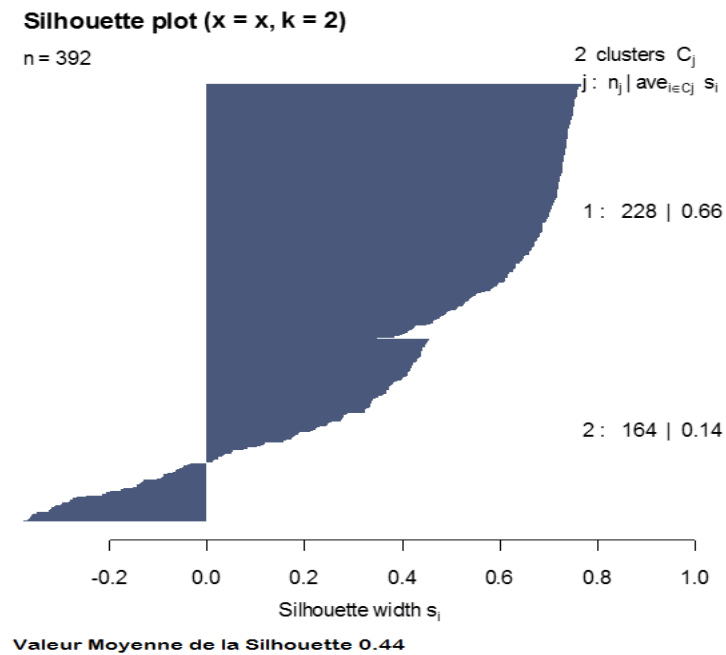


FIGURE 3.6 – Largeur maximale de la silhouette globale moyenne pour 2 et 3 centres

Cela est confirmé par l'index de silhouette où la largeur maximale de la silhouette globale moyenne pour les 2 centres est meilleure. Effectivement la définition de l'index silhouette du nombre optimal de cluster stipule que si la valeur de la silhouette est proche de 1, cela signifie que l'échantillon est «bien-classé» et il a été affecté à un groupe tout à fait approprié. Dans notre cas le plus optimal étant de 0.44 avec 2 clusters (figure 3.6), ce nombre de centres peut être considéré comme bien classé en observant les chevauchements dans les exemples de la base (figure 3.3).

3.3 Résultats de la classification avec FCM

Le clustering est souvent confondu avec la classification, mais il y a une certaine différence entre les deux. Dans la classification les objets sont affectés à des catégories définies a priori, ce qui n'est pas le cas du regroupement, alors nous pourrions dire que le clustering est une classification non supervisée : pas de classes prédéfinies.

Nous résumons les principales étapes réalisées par l'application de l'algorithme Fuzzy C-Means sur la base de données Pima comme suit :

1. La fixation arbitraire d'une matrice d'appartenance.
2. Le calcul des centroïdes des classes. Les résultats obtenus sont présentés dans le tableau 3.1

<i>Paramètres</i>	<i>Centre1</i>	<i>Centre2</i>
Npreg	3.866298	3.170495
Glu	153.6923	114.8574
BP	69.88424	72.73502
Skin	31.73031	28.35179
Insulin	361.9835	109.7955
BMI	35.44282	32.35977
PED	0.5701096	0.5036238
Age	33.84151	29.93615

TABLE 3.1 – Les centroïdes obtenus pour chaque paramètre

3. Le réajustement de la matrice d'appartenance suivant la position des centroïdes.
4. Calcul du critère de minimisation et retour à l'étape 2 s'il y a non convergence de critère.

La figure 3.7 montre la convergence obtenue de l'algorithme FCM à la 31ème itérations, avec une erreur égale à $5.14e^{-13}$, et $m=2$. Le regroupement réalisé sur les exemples pour chaque paramètre de la base de données est représenté dans la figure 3.8.

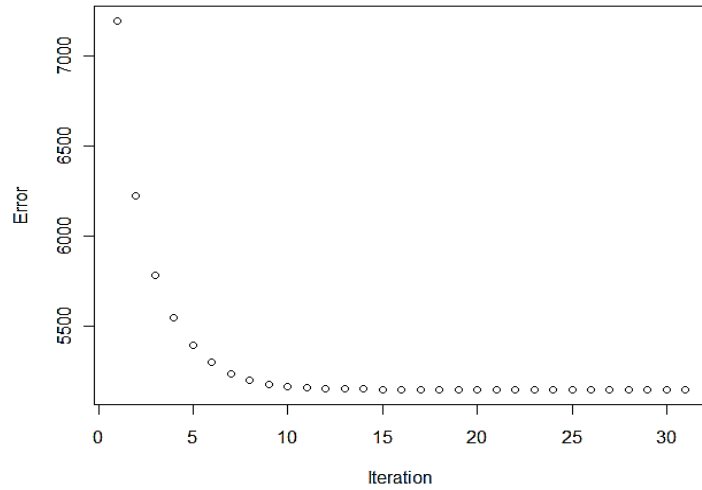


FIGURE 3.7 – Schéma de convergence de la fonction objective

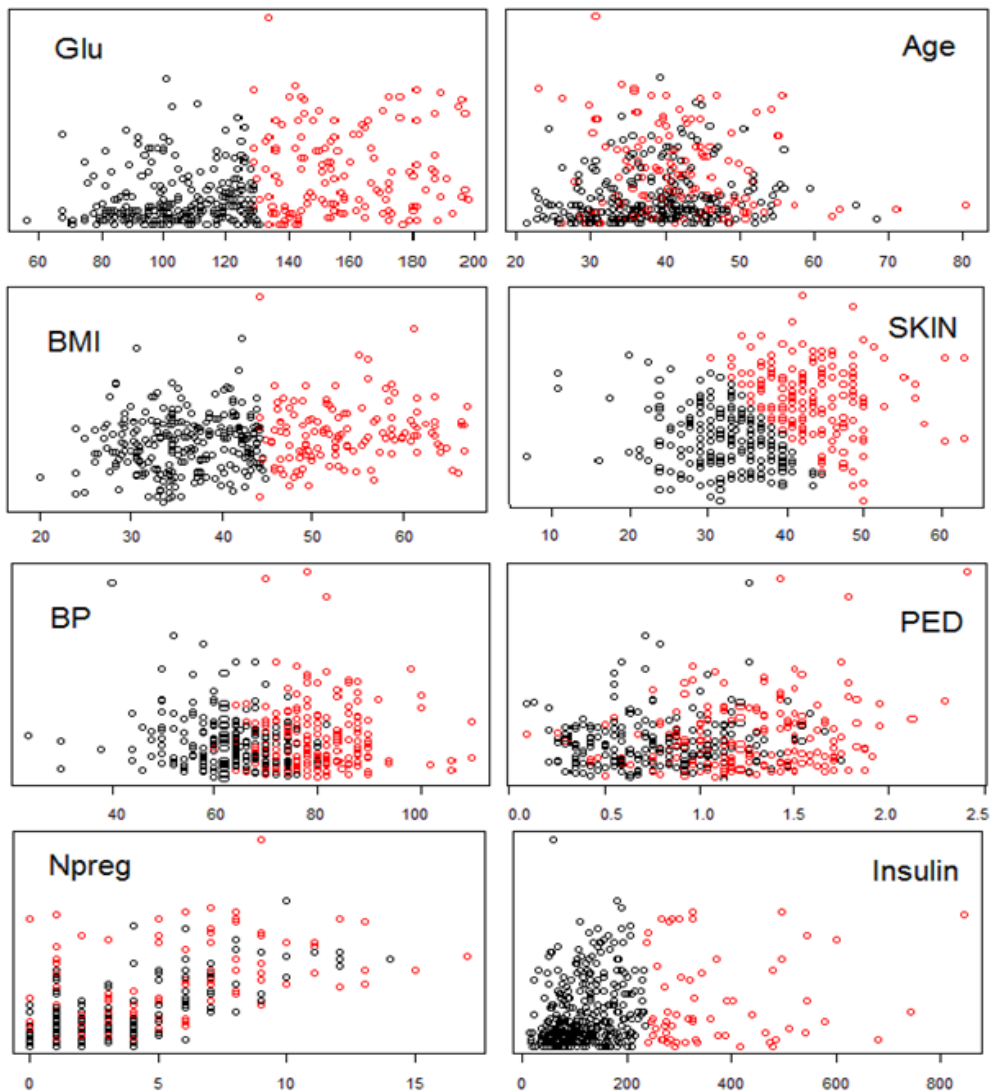


FIGURE 3.8 – Répartition des 8 paramètres pour $c=2$

Les performances du classifieur FCM implémenté ont été évaluées par le calcul du pourcentage de sensibilité (SE), la spécificité (SP) et taux de classification (TC), les définitions de ces derniers sont respectivement comme suit :

- Sensibilité (Se%) : $[Se = 100 * TP / (VP + FN)]$ on appelle sensibilité (Se) du test sa capacité de donner un résultat positif quand la maladie est présente. Représente ceux qui sont correctement détectés parmi tous les événements réels.
- Spécificité (Sp %) : $[Sp = 100 TN * / (VN + FP)]$ on appelle spécificité du test cette capacité de donner un résultat négatif quand la maladie est absente. Elle est représentée pour détecter les patients non diabétiques.
- Taux de classification (CC %) : $[CC = 100 * (TP + TN) / (TN + TP + FN + FP)]$ est le taux de reconnaissance.
 - VP : diabétique classé diabétique ;
 - FP : non diabétique classé diabétique ;
 - VN : non diabétique classé non diabétique ;
 - FN : diabétique classé non diabétique.

Les résultats de l'évaluation du classifieur en termes de taux de classification, sensibilité et spécificité sont résumés dans le tableau 3.2, sachant que les données contiennent 130 diabétiques et 262 non-diabétiques.

	TC%	Se%	Sp%	VP	VN	FP	FN
8 paramètres							
2 clusters	76.02	66.15	80.92	86	212	50	44

TABLE 3.2 – Performances du classifieur FCM

L'objectif de FCM est la partition de données numériques dans des clusters. Grâce à l'apport du flou, l'appartenance d'un point de données à un cluster spécifique est donnée par la valeur d'appartenance du point de données à ce cluster. Cela rend la technique efficace dans la classification d'où des résultats assez bons avec un taux de classification atteignant les 76% ou la reconnaissance des non diabétiques est de 212/262 et les diabétiques 86/130. Le nombre des cas mal classés (Faux Positif=50 et Faux Négatif=44), peut être expliqué par les chevauchements constatés dans certains paramètres tel que Age, PED (figure 3.8).

Ce comportement est dû à deux particularités de FCM. D'une part, il ne prend pas directement en compte la séparabilité des groupes : la fonction objective pénalise le manque de compacité (i.e. une forte variance) des groupes constitués, et la non ressemblance entre les données de groupes différents. De plus, FCM tend à produire des groupes de même taille, qui équilibrent les influences des données sur les positions des prototypes. Aussi, il a des difficultés à représenter des groupes de petites tailles, et donc ressortir les exceptions. Cela n'empêche pas l'algorithme FCM de démontrer sa puissance dans le regroupement des données de manière non supervisée.

3.4 Inconvénients de la technique

En plus de la nécessité de la connaissance au préalable du nombre de clusters dans l'application de l'algorithme FCM, les résultats de FCM ne semblent pas très stables et cela à cause de la sélection aléatoire des centres qui rend le processus itératif à converger facilement dans la solution optimale locale. Plusieurs travaux ont présenté quelques variantes de FCM pour contrer ses effets. Certaines variantes visent à réduire l'influence des points d'appartenance faible aux clusters, afin qu'ils n'interviennent pas dans la définition des prototypes correspondants : elles autorisent une appartenance binaire pour des données distinctes des centres des clusters [RTK95], pénalisent les degrés proches de 0.5 [HK01] ou utilisent une métrique plus robuste que la métrique euclidienne [WY02]. On pourrait attendre de ces approches qu'elles permettent de stabiliser un prototype sur des données isolées, sans qu'il soit attiré vers des clusters de tailles supérieures. Cependant, l'effet obtenu est que les prototypes ne sont pas attirés vers les données nettement différentes, et représentent mieux les groupes de tailles moyennes. Ces méthodes répondent donc à un objectif de robustesse [DK97], mais n'excluent pas les exceptions.

4 Considération 2 : Classification par l'hybridation Fuzzy C Means & Réseaux de neurones

Introduction

Dans cette considération, nous réduisons l'architecture d'un classifieur Neuronal classique pour la reconnaissance des diabétiques. Pour cela nous utilisons l'approche FCM comme outil de réduction. Le vecteur d'entrée du classifieur est représenté par des degrés d'appartenance aux centres déjà définis par FCM. Cette hybridation améliorera considérablement le temps d'apprentissage.

4.1 Principe

Différentes solutions pour la reconnaissance du diabète ont été présentées dans la littérature, comme les techniques de réseaux de neurones mutlicouches RNMC, carte auto-organisatrice et la méthode LVQ. Dans cette partie nous proposons la combinaison de l'algorithme de Fuzzy c-means et les réseaux de neurones mutlicouches RNMC connectés en cascade, nommée Fuzzy Clustering Neural Network (FCNN) [OCK06]. La structure d'un tel réseau est représentée dans la figure 3.9.

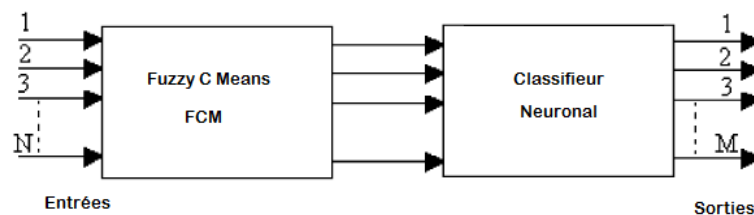


FIGURE 3.9 – Architecture de FCNN [OCK06]

L'algorithme de fuzzy C-means concerne le regroupement des données d'entrée. Toutefois, il utilise un regroupement flou, dans lequel le vecteur d'entrée x est pré-classifié avec des valeurs d'appartenance différentes. La pénétration de l'espace de données est meilleure et la localisation du vecteur d'entrée x dans l'espace des données est plus précise. Les sorties de l'algorithme FCM forment le vecteur d'entrée au classifieur neuronal. Ce dernier sera utilisé pour la reconnaissance automatique des diabétiques.

4.2 Réseau de neurones mutlicouches RNMC

Dans cette partie, un réseau de neurones mutlicouches RNMC est appliqué. Sa phase d'apprentissage est réalisée par l'algorithme de rétro propagation des erreurs. Les paramètres d'entrées de RNMC sont formés par des centres de cluster qui sont obtenus par le regroupement FCM. La figure 3.10 montre une structure générale de RNMC. L'apprentissage par rétro propagation fait appel à un algorithme de descente de gradient itératif visant à minimiser les erreurs quadratiques moyennes entre la sortie réelle d'un RNMC et la sortie désirée [Hai98, KoTA03].

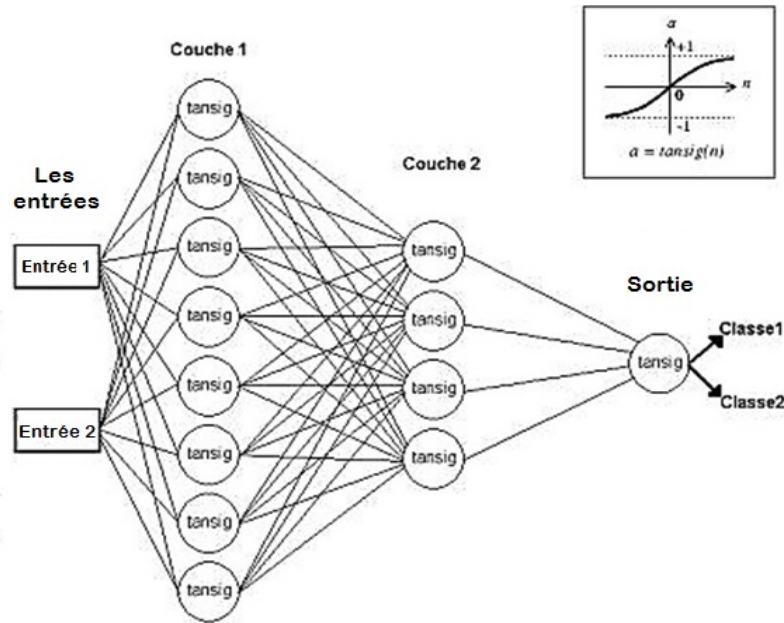


FIGURE 3.10 – Structure générale d’un réseau de neurones mutlicouches RNMC

L’algorithme de rétro propagation

Algorithm 3 Principe de l’algorithme de rétro propagation

- 1: **Initialisation** : Affecter à tous les poids des valeurs aléatoires réelles.
- 2: **Présentation des entrees et la sortie desirée** :
Présenter le vecteur d’entrée $x(1), x(2), \dots, x(N)$ et leurs correspondantes sorties désirées $d(1), d(2), \dots, d(N)$, une paire à la fois, où N est le nombre d’exemple d’apprentissage.
- 3: **Calcul des sorties reelles** : calcule les sorties y_1, y_2, \dots, y_{NM}

$$y_i = \phi \left(\sum_{j=1}^{N_{M-1}} w_{ij}^{(M-1)} x_j^{(M-1)} + b_i^{(M-1)} \right), \quad i = 1, \dots, N_{M-1}.$$

- 4: **Adaptation des poids (w_{ij}) et les biais (b_i)** :

$$\begin{aligned} \Delta w_{ij}^{(l-1)}(n) &= \mu x_j(n) \delta_i^{(l-1)}(n) \\ \Delta b_i^{(l-1)}(n) &= \mu \delta_i^{(l-1)}(n) \end{aligned}$$

Où :

$$\delta_i^{(l-1)}(n) = \begin{cases} \varphi'(net_i^{(l-1)})[d_i - y_i(n)], & l = M \\ \varphi'(net_i^{(l-1)}) \sum_k w_{ki} \delta_k^{(l)}(n) & 1 \leq l \leq M \end{cases}$$

Avec :

- $x_j(n)$ représente la sortie du noeud j à l’itération n ,
- l est la couche,

- K est le nombre de noeuds de sortie du réseau de neurones,
- M est la couche de sortie,
- φ est la fonction d'activation.
- Le pas d'apprentissage est représenté par μ .

Nous remarquons, que la valeur élevée du pas d'apprentissage peut aboutir à une convergence plus rapide, mais peut aussi entraîner des oscillations. Afin d'atteindre une convergence plus rapide avec une oscillation minimale, un terme dynamique peut être ajouté à l'équation de mise à jour des poids. Dans ce travail, nous choisissons $\mu=1$ via expérimentation. Après avoir terminé la procédure d'apprentissage du réseau de neurones, les poids du RNMC sont enregistrés et prêts à être appliqués dans la phase de test.

4.3 Experimentation

Dans cette étude, une approche fondée sur le regroupement est adoptée sur la base de données diabète. Les degrés d'appartenance aux centres constituent les paramètres d'entrée du RNMC.

Implémentation du classifieur Neuronal

Les 392 exemples d'apprentissage ont été regroupés à l'aide de l'algorithme FCM. Donc chaque exemple sera représenté par 2 degrés d'apprentissage ($c=2$), la matrice de taille 392×2 . La couche d'entrée du classifieur sera fixée à 2 neurones. Le nombre optimal de nœuds cachés a été déterminé au moyen d'expérimentations. Les résultats expérimentaux montrent que le nombre optimal de nœuds cachés est de 4 avec la plus haute précision de 83.97% pour 100 epochs. L'architecture optimale du réseau de neurones est représentée dans la figure 3.11 avec 2 :4 :1.

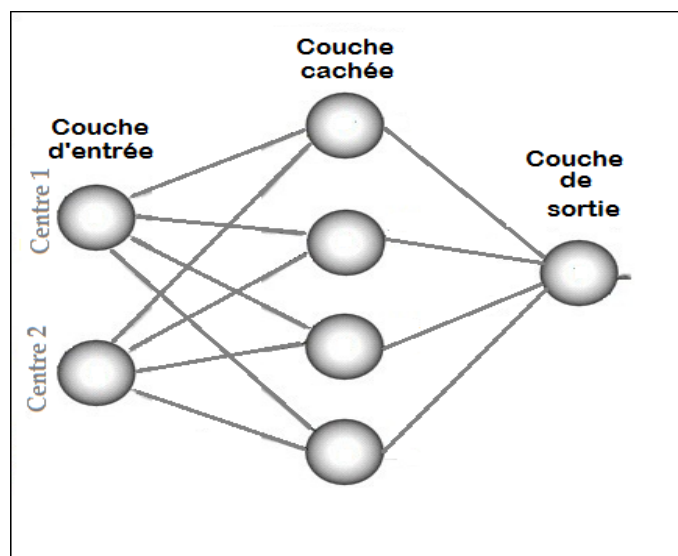


FIGURE 3.11 – L'architecture optimale du réseau de neurones obtenue

Résultats des tests

La qualité de l'apprentissage augmente avec la taille de l'ensemble d'apprentissage. Il existe peu de résultats théoriques sur les tailles d'échantillons nécessaires à l'ensemble d'apprentissage et de test. Nous ne disposons que de résultats empiriques afférant au problème qui englobent plusieurs centaines d'exemples. La répartition de l'échantillon entre les deux ensembles se fait en général sur la base des proportions 1/2, 1/2 pour chacun des deux ensembles ou 2/3 pour l'ensemble d'apprentissage et 1/3 pour l'ensemble de test.

L'apprentissage est réalisé avec 262 exemples (171 non diabétiques et 91 diabétiques). Les 130 échantillons restants sont employés pour la phase de test (91 non diabétiques et 39 diabétiques).

L'évaluation des performances de cette approche est estimée par le taux de classification, la sensibilité, la spécificité. Les résultats obtenus sont résumés dans le tableau 3.3.

Méthodes	Tc%	Se%	Sp%	Architecture	epochs
RNMC	85.50	90.22	74.36	8 :4 :1	800
FCNN	83.97	91.30	66.67	2 :4 :1	100

TABLE 3.3 – Performances des classifieurs RNMC et FCNN

Le perceptron multicouche est une structure très couramment utilisée vue sa simplicité. On détermine, selon les besoins : le nombre de couches cachées, le nombre de neurones par couches cachées et la fonction linéaire ou non linéaire de chaque couche. Ce qui donne d'une part une bonne flexibilité au réseau, et d'autre part le choix du vecteur d'entrée est devenu très important pour une bonne classification. En effet la qualité de l'apprentissage dépend énormément de la pertinence des paramètres d'entrées. Le regroupement des données réalisé par l'algorithme FCM a permis de maintenir une bonne classification avec un temps d'apprentissage rapide et une structure réduite.

4.4 Conclusion

Dans cette étude, le nouveau FCNN a été élaboré et présenté afin de classer les diabétiques. Une évaluation comparative des performances de FCNN avec RNMC montrent des résultats satisfaisants pour la classification du diabète. Les réseaux de neurones RNMC sont toujours capables de réaliser une bonne reconnaissance, toutefois ils prennent plus de temps pour l'apprentissage. L'objectif dans le développement de FCNN est de parvenir à des résultats plus optimaux avec des caractéristiques réduites de la base de données. Il a été démontré que le temps d'apprentissage de FCNN était de 60% moins que celui requis par le RNMC. En outre, cette technique, qui intègre la méthode de regroupement flou, et l'apprentissage des réseaux de neurones par rétro propagation a permis une réduction des paramètres d'entrées tout en ayant un taux de classification acceptable. Nous espérons que les performances de la méthode seront améliorées avec une base de données comportant un nombre d'exemples et de paramètres plus important.

5 Considération 3 : Classification par l'hybridation Fuzzy C-Means & Adaptive Neuro-Fuzzy Inference System

5.1 Adaptive Neuro-Fuzzy Inference System

Les systèmes flous ont été conçus pour exploiter les informations linguistiques des connaissances d'experts pour l'étude d'un système [JS95, PY98]. Toutefois, au cours des années, le concept s'est élargi et les systèmes adaptatifs flous ont pris de l'importance pour l'extraction automatique des connaissances (en l'absence d'un expert humain) à partir d'un ensemble de données. Ils ont été mis en œuvre pour construire des modèles efficaces avec des algorithmes d'apprentissage sophistiqués. Bon nombre de ces systèmes adaptatifs sont développés par hybridation avec d'autres méthodes, par exemple réseaux de neurones, algorithmes génétiques, programmation évolutive, des méthodes probabilistes, etc [Bon97]. Jusqu'à présent, le plus populaire de ces systèmes hybrides, est celui du neuro-flou (les réseaux de neurones flous) qui exploite la force d'apprentissage des réseaux de neurones ainsi que la facilité de compréhension linguistique du système à base de règles floues. L'apprentissage est utilisé de manière adaptative, afin d'ajuster les règles dans la base de connaissances, et de produire ou d'optimiser les fonctions d'appartenance du système flou.

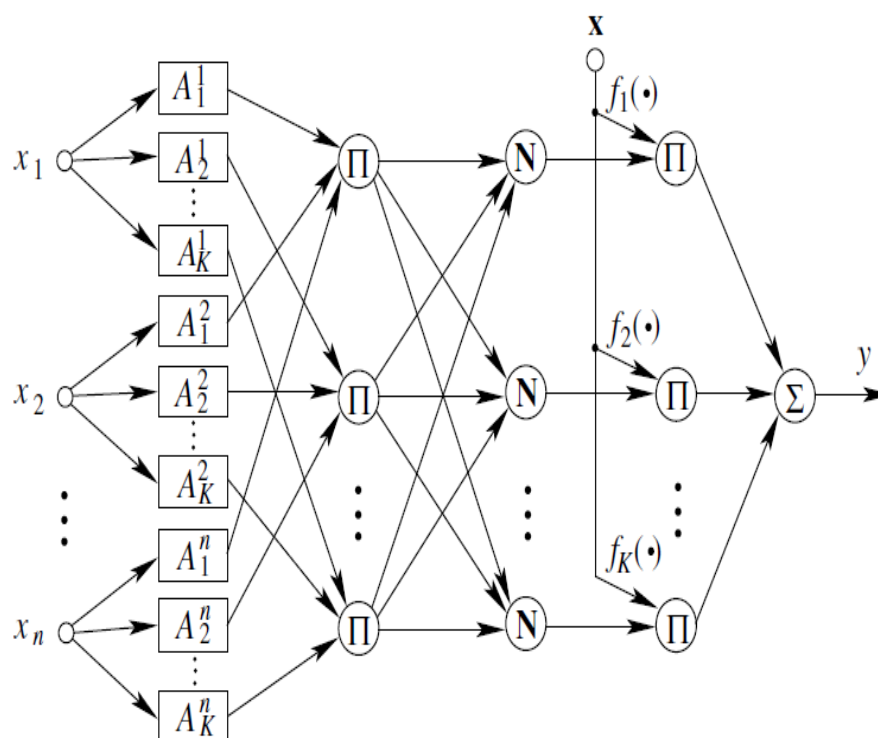


FIGURE 3.12 – Architecture d'ANFIS

Jang dans [Jan93] a mis au point ANFIS (Adaptive Neuro-Fuzzy Inference System) qui est un système neuro-flou. Avec une structure de réseau de neurones,

où chaque couche est un composant du système neuro-flou (Figure 3.12). Il applique le modèle de règle de TSK (Takagi-Sugeno-Kang) flou [SK88] dans lequel les règles ont un caractère flou juste dans la partie « SI », tandis que dans la partie « Alors », il y a des dépendances fonctionnelles.

$$R^{(r)}: \text{SI } x \text{ est } A^r \\ \text{ALORS } y_r = f^{(r)}(x_1, x_2, \dots, x_n).$$

Le mécanisme de raisonnement flou de TSK est résumé dans la figure 3.13 [JM96]. Les moyennes pondérées sont utilisées afin d'éviter les calculs complexes dans le processus de défuzzification.

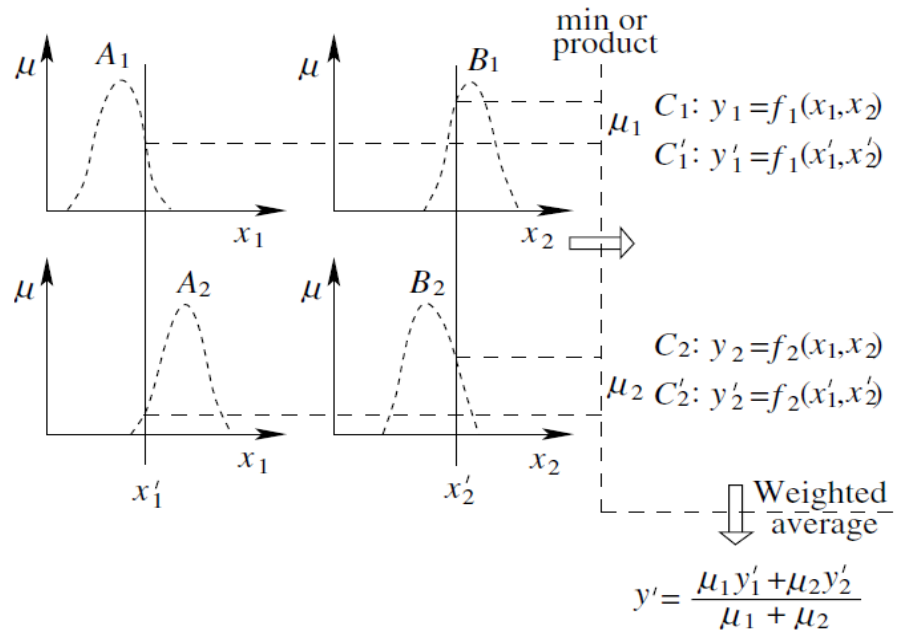


FIGURE 3.13 – Modèle flou de Sugeno du Premier ordre

Algorithm 4 Principe de l'architecture d'ANFIS a deux entrées

1: **1^{re} couche** : Fuzzyfication

$$O_i^1 = \mu_{A_i}(X) \quad \text{avec } : i = 1, 2$$

$$O_i^1 = \mu_{B_{i-2}}(Y) \quad \text{avec } : i = 3, 4$$

2: **2^{me} couche** : Pondération des règles floues

$$O_i^2 = w_i = \mu_{A_i}(X)\mu_{B_{i-2}}(Y) \quad i = 1, 2$$

3: **3^{me} couche** : Normalisation

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1, 2$$

4: **4^{me} couche** : Défuzzyfication

$$O_i^4 = w_i f_i \quad \text{avec } f = a_i \cdot X + b_i \cdot Y + c_i \quad o \quad i = 1, 2$$

5: **5^{me} couche** : Calcul de la sortie

$$O_i^5 = \sum_{i=1}^2 w_i f_i = \frac{\sum_{i=1}^2 w_i f_i}{w_1 + w_2}$$

Toutefois, dans la conception des systèmes neuro-flous, on relève plusieurs difficultés :

- Détermination du nombre de fonctions d'appartenance : il est lié à l'identification de la structure [Sun94],
- Choix des centres et de la largeur des fonctions d'appartenance,
- Besoin d'une base a large exemples pour tester les modèles,
- Perte de la capacité de généralisation pour une base de données importante,
- Explosion combinatoire du nombre de règles floues (temps d'apprentissage),
- Fixation des paramètres de fonctions d'appartenance.

Une partition optimale de l'espace d'entrée peut diminuer le nombre de règles et augmenter la vitesse d'apprentissage. Par conséquent, choisir le nombre de fonctions d'appartenance et trouver les paramètres optimaux de ces dernières deviennent des questions ouvertes.

Pour remédier aux problèmes suscités, nous exploitons les méthodes de regroupement pour la répartition des données et permettre ainsi la régularisation des contraintes qui s'appliquent sur les paramètres des fonctions d'appartenance floues. Cela en éliminant la similarité par la fusion des partitions floues afin d'améliorer leurs distinctions et de diminuer le nombre de sous-ensembles flous [SBKvN98]. De ce fait, automatiser et optimiser la structure et les paramètres des fonctions d'appartenance.

Pour résoudre ce problème, nous proposons dans cette partie l'implémentation de l'algorithme Fuzzy C-means et à titre comparatif l'algorithme Soustractive clustering

5.2 L'apprentissage du classifieur Neuro-flou

Dans ce travail, deux phases d'apprentissage sont nécessaires : Premièrement, une phase d'apprentissage structurel est appliquée pour déterminer le bon partitionnement de l'espace d'entrée (nombre de fonctions d'appartenance pour chaque entrée). Deuxièmement, une phase d'apprentissage paramétrique est utilisée pour affiner les fonctions d'appartenance et les paramètres en conséquence. Il y a plusieurs façons de combiner l'apprentissage structurel et paramétrique dans un classifieur neuro-flou. Elles peuvent être réalisées de manière séquentielle : apprentissage de la structure en premier lieu pour trouver la structure appropriée d'un système neuro-flou et l'apprentissage paramétrique est ensuite réalisé pour réajuster les paramètres. Dans d'autres cas, seul l'apprentissage paramétrique ou structurel est nécessaire lorsque la structure (règles floues) ou les paramètres (fonctions d'appartenance) sont donnés par des experts. Par ailleurs la structure de certains modèles neuro-flous est fixée à priori.

ANFIS ajuste les paramètres de fonction d'appartenance en adoptant soit un algorithme de rétro propagation seul ou en combinaison avec une estimation des moindres carrés, pour réduire une certaine mesure d'erreur définie par la somme d'écart quadratique entre les résultats réels et désirés.

5.3 L'hybridation FCM-ANFIS

Cette approche se concentre sur l'optimisation du classifieur ANFIS pour la reconnaissance du diabète. De ce fait, nous proposons la stratégie de la FCM clustering. Le regroupement des données permet d'identifier la structure basée sur la partition de dispersion (figure 3.14). Ainsi, cette approche peut être adoptée pour réduire la dimension du classifieur ainsi que le temps d'apprentissage, puisque le nombre de règles floues est égal au nombre de fonctions d'appartenance quelle que soit la dimension des entrées [DJZ06].

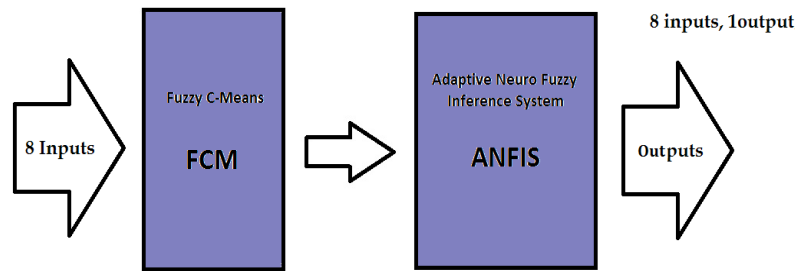


FIGURE 3.14 – L'approche réalisée FCM-ANFIS

FCM tente de partitionner les données numériques dans des clusters. L'appartenance d'un point de données à un cluster spécifique est exprimée par la valeur d'appartenance de ce point à ce cluster. La valeur d'appartenance est calculée par la minimisation d'une fonction objective de FCM, qui recherche l'appartenance ressortant le moins d'erreur.

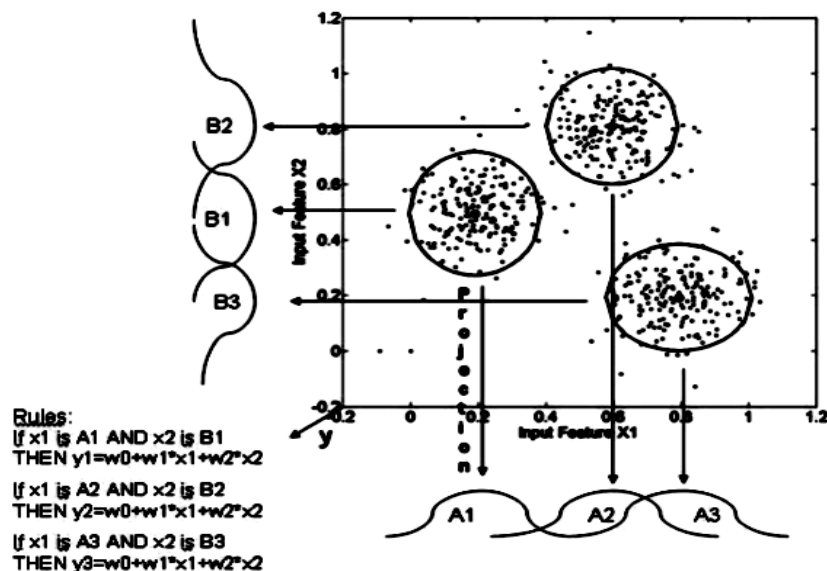


FIGURE 3.15 – Schéma représentatif de la répartition des clusters en fonctions d'appartenance avec FCM

On distingue dans la figure 3.16, la répartition obtenue sur deux centres (deux fonctions d'appartenance) pour les variables de la base de données Indiens Pima.

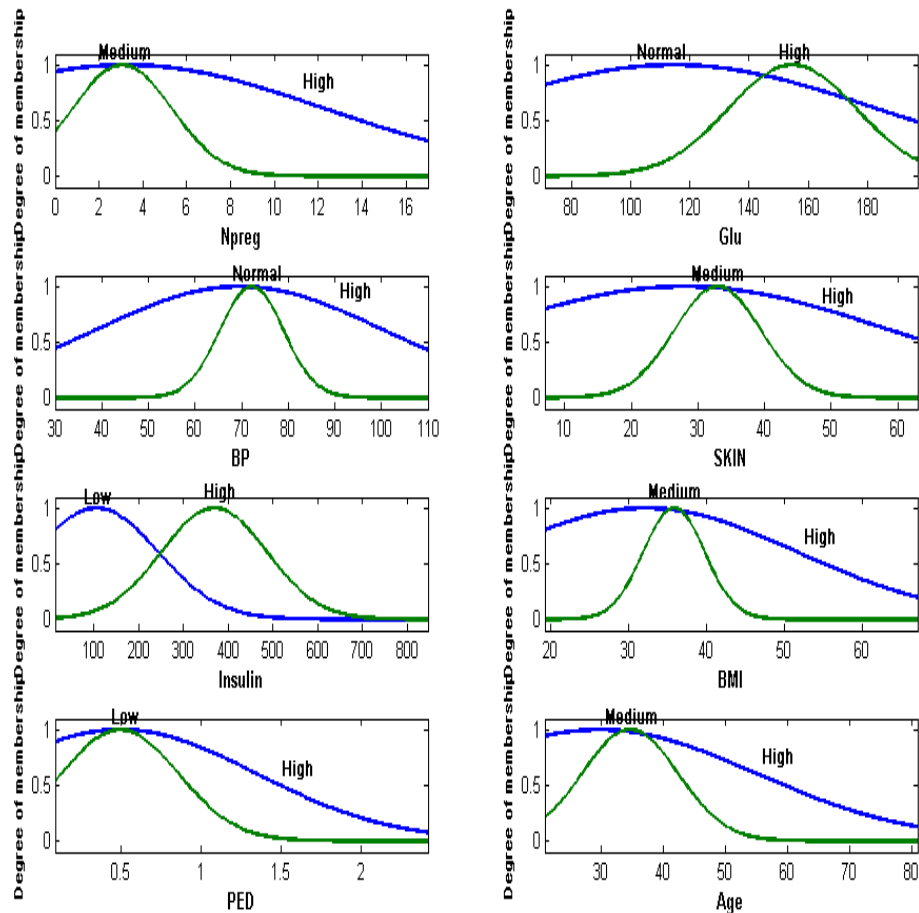


FIGURE 3.16 – Les fonctions d’appartenance de la base de données Pima avec FCM

Pendant l’apprentissage d’ANFIS, l’algorithme d’optimisation hybride est employé pour déterminer les valeurs optimales des paramètres du SIF (Système d’Inférence Floue) de type TS. Il utilise une combinaison de la méthode des moindres carrés et de la descente de gradient par rétro-propagation pour l’apprentissage paramétrique des fonctions d’appartenance du SIF [JM96, JYAD96].

Dans la 1^{ère} phase (forward) les paramètres de la prémisse sont supposés fixes et les valeurs optimales des paramètres en conséquence sont estimées en appliquant la méthode des moindres carrés. Ensuite, la sortie du système est calculée et l’erreur observée est utilisée pour ajuster les paramètres de la prémisse par le biais de l’algorithme de rétro propagation standard. Dans le passage vers l’arrière (backward), les paramètres résultants sont supposés fixes et l’algorithme de descente de gradient est utilisé pour trouver les valeurs optimales des paramètres de la prémisse.

Afin d’optimiser les valeurs des paramètres du modèle neuro-flou, l’apprentissage et la validation croisée des sous-ensembles sont adoptés. Pour le test des performances la méthode K-cross validation est appliquée. Son principe est de

diviser la base de données en k parties d'échantillon d'apprentissage/échantillon de test (figure 3.17). Le modèle est bâti sur l'échantillon d'apprentissage et validé sur l'échantillon de test. L'erreur est estimée en calculant l'erreur quadratique moyenne (MSE).

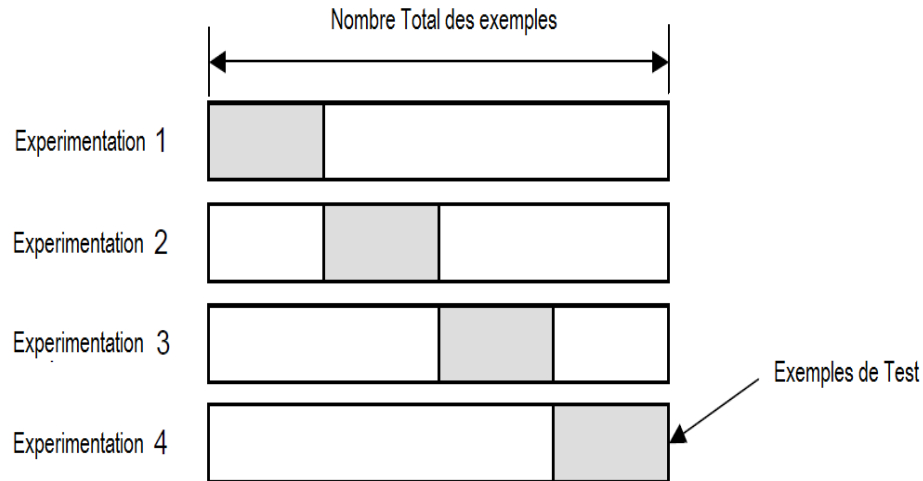


FIGURE 3.17 – Schéma de principe du K-fold cross validation

Après 3 cross validation le meilleur résultat est obtenu en 100 epochs (MSE apprentissage=0,1012, MSE validation = 0.1023), Cela a permis de réaliser un bon apprentissage du classifieur FCM-ANFIS. Le tableau 3.4 présente les points modaux finaux des fonctions d'appartenance.

Npreg	Medium 1 - 3 - 9	High 0 - 5 - 17
Glu	Normal 71.6 - 101.3 - 180	High 85.1 - 155.8 - 200
BP	Low 55 - 71 - 92	High 30 - 70 - 110
Skin	Medium 16 - 33 - 51	High 9 - 30 - 70
Insulin	Low 14 - 100 - 460	High 80 - 380 - 690
BMI	Normal 18.2 - 31.83 - 51	High 19 - 34.79 - 59
Ped	Low 0.085 - 05- 1.45	High 0.85 - 0.6 – 2.42
Age	Medium 21 - 27.94 -56.1	High 21 - 35.09- 61.8

TABLE 3.4 – Les points modaux finaux avec FCM-ANFIS

La connaissance extraite du modèle FCM-ANFIS

Le classifieur FCM-ANFIS génère une base de connaissances de 2 règles de classification.

Règle 1 : *If (Npreg is High) and (Glu is High) and (BP is High) and (Skin is Medium) and (Insulin is High) and (BMI is High) and (PED is High) and (Age is High) then (Class is Diabetic)*

Règle 2 : *If (Npreg is Medium) and (Glu is Normal) and (BP is Low) and (Skin is High) and (Insulin is High) and (BMI is Medium) and (PED is Low) and (Age is Medium) then (Class is Non Diabetic)*

Le degré de sollicitation et les prototypes pour chaque règle sont indiqués dans le tableau 3.5. Leurs définitions respectives dans notre travail sont les suivantes :

- Degré de sollicitation = $\frac{\text{Nombre de personnes qui activent cette règle à plus de 50\%}}{\text{Le nombre total des personnes de chaque classe (diabétiques et non diabétiques)}}$
- Prototype des règles, exprime le nombre de personnes activant cette règle à plus de 90%.

	Personnes Non Diabétiques	Personnes Diabétiques
Degrés de sollicitation Pour la Règle Diabétique R_1	20%	34.61%
Degrés de sollicitation Pour la Règle Non diabétique R_2	75.38%	12.30%
Prototype dans la phase d'apprentissage	46	17
Prototype dans la phase de test	46	15

TABLE 3.5 – Degrès de sollicitation des 2 règles de FCM-ANFIS

Dans cette partie, chaque attribut d'entrée a deux fonctions d'appartenance, en toute logique 256 (2^8) règles floues seront générées, augmentant sensiblement la complexité globale de la base de connaissances du classifieur flou. On a remarqué que l'application de fuzzy c-means avec ANFIS a réduit considérablement cette complexité. En appliquant cette technique, le nombre de règles a été minimisé de 256 à 2, ce qui diminue la complexité de la base de connaissances de manière significative. Comme le montre le tableau 3.5, 75,38% d'exemples non-diabétiques ont sollicité la règle floue 2 (avec un degré de sollicitation de plus de 50%). En revanche seulement 34.61% des diabétiques ont sollicité la règle floue 1 (avec un degré de sollicitation de plus de 50%).

Sachant que la règle 2 est sollicitée à 75.38%, elle prédit qu'une personne d'âge moyen avec un glucose normal et une surcharge pondérale est classée non diabétique. La règle 1 compatible avec le raisonnement des experts dans le domaine médical pour le diagnostic du diabète. Elle stipule qu'un patient avec une hyperglycémie, un âge grand et une obésité, elle est sollicitée à 34.61%. La règle diabétique R_1 à été sollicitée par un degré de 20% par les personnes non diabétiques (FP), quant à la règle non diabétique R_2 elle a été sollicitée par les diabétique (FN) (voir tableau 3.5).

5.4 Résultats et discussion

Les performances du classifieur implémenté ont été évaluées par le calcul du pourcentage de sensibilité (SE), la spécificité (SP) et la précision de classification (CC), leurs définitions respectives sont les suivantes :

La précision de classification (Classification accuracy)

$$Precision = \frac{\sum_{i=1}^{|T|} assess(t_i)}{|T|}$$

$$assess(t) = \begin{cases} 0 & \text{if } classif(t) = t.c \\ 1 & \text{otherwise} \end{cases}$$

Où

- T : l'ensemble de test, $t \in T$,
- c : la classe de l'objet t .
- $classif(t)$: renvoie la classification de t par l'algorithme.

Sensibilité

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

Spécificité

$$Specificite = \frac{TN}{TN + FP}$$

Où TP, TN, FP et FN sont notés respectivement :

- TP : diabétique classé diabétique ;
- FP : non diabétique classé diabétique ;
- TN : non diabétique classé non diabétique ;
- FN : diabétique classé non diabétique.

La base "Indiens Pima" du diabète avec 392 exemples (262 non diabétiques et 130 diabétiques) a été utilisée pour évaluer le classifieur en termes de précision, sensibilité et spécificité, en appliquant pour la phase de test l'approche 3 cross validation (voir table 3.6)

Nombre de centre	8 paramètres	C=2	C=3	C=4
Précision %	73.08	83.85	83.08	81.54
Se %	66.67	82.05	69.23	82.05
Sp %	75.82	84.62	89.01	81.32
FP	62	42	31	45
TN	200	220	231	217
TP	96	109	97	105
FN	34	21	33	25
Nombre de règles	256	2	3	4

TABLE 3.6 – Performances du classifieur FCM-ANFIS avec différentes entrées

Avec un minimum de clusters (c=2) et seulement deux règles floues, l'approche de la FCM-ANFIS a donné les meilleurs résultats avec une précision= 83,85%, = Se 82,05% et Sp = 84,62% comparés aux autres cas (voir le tableau 3.6), tout en donnant un nombre réduit de cas mal classés (FP=42 et FN=21).

Patient numéro 107 personne diabétique correctement classée : cas d'un TP

Le tableau 1 des caractéristiques montre que le patient diabétique N°107 a été correctement classé comme diabétique avec une apparente hérédité (PED) et une valeur d'insuline et de glucose élevée.

Paramètres	Patient N° 107
Npreg	3.0000
Glu	173.0000
BP	78.0000
Skin	39.0000
insulin	185.0000
BMI	33.8000
PEd	0.97
Age	31.0000
Class	1

TABLE.1 Paramètres du patient 107.

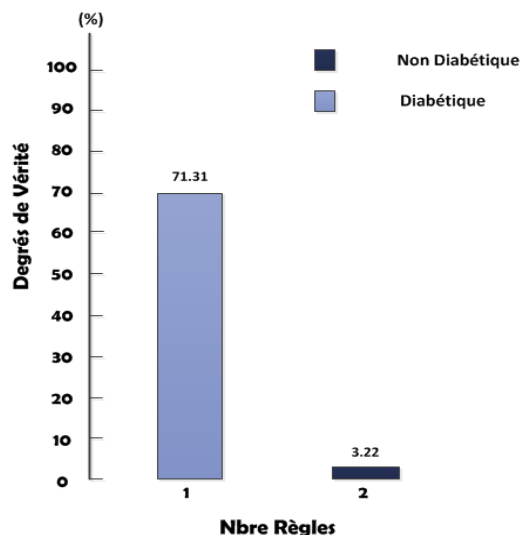


FIGURE.1 Les règles floues activées pour le patient N°107 correctement classé diabétique cas (TP)

FCM-ANFIS a permis d'avoir une connaissance plus ciblée avec la règle 1 diabétique qui a été activée à plus de 70%.

Patient numéro 28 personne non diabétique correctement classée : cas d'un TN

220 personnes non diabétiques ont été correctement classées ; le tableau 2 des paramètres du patient N °28 est un exemple de cas correctement reconnu comme non diabétique.

Paramètres	Patient N° 28
Npreg	1.0000
Glu	87.0000
BP	68.0000
Skin	34.0000
insulin	77.0000
BMI	37.8000
PEd	0.401
Age	24.0000
Class	0

TABLE.2 Paramètres du patient 28.

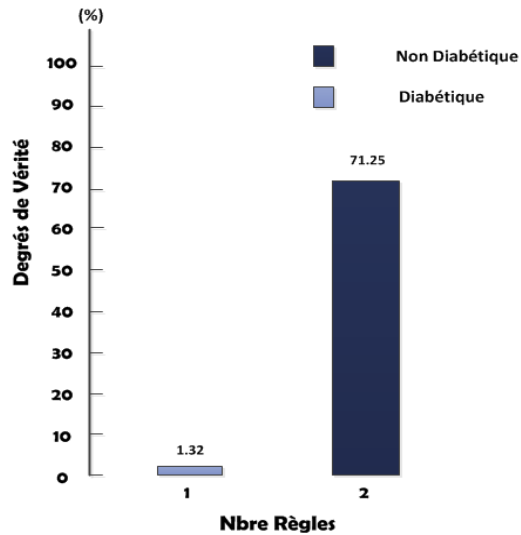


FIGURE.2 Les règles floues activées pour le patient N°28 correctement classé non diabétique cas (TN)

La règle 2 non diabétique a été activée avec un degré de vérité de 71.25%.

Patient numéro 102 (diabétique) reconnu comme non diabétique : cas d'un FN mal classé

Les caractéristiques du patient N°102 diabétique sont présentées ci-dessous (voir tableau 3), cet exemple est mal classé, c'est à dire diabétique reconnu comme non diabétique. Il présente une valeur normale de glucose et une quantité d'insuline faible. La règle floue R_2 est activée avec un degré de vérité d'environ 74%.

Paramètres	Patient N° 102
Npreg	2.0000
Glu	93.0000
BP	78.0000
Skin	64.0000
insulin	32.0000
BMI	38.8000
PEd	0.674
Age	23.0000
Class	1

TABLE.3 Paramètres du patient 102.

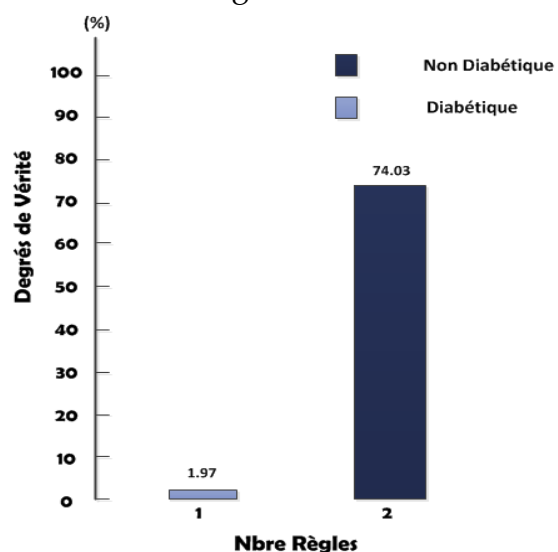


FIGURE.3 Les règles floues activées pour le patient N°102 (diabétique) classé non diabétique (FN)

Patient numéro 68 (non diabétique) reconnu comme diabétique : cas d'un FP mal classé

Le tableau 4 décrit les caractéristiques d'un cas non-diabétique, ce patient N°68 est mal classé, c'est à dire non diabétique reconnu comme diabétique (fausse alarme), il présente une hyperglycémie. La règle floue R_1 est activée avec un degré de vérité égale à 60,71%.

Paramètres	Patient N° 68
Npreg	2.0000
Glu	157.0000
BP	74.0000
Skin	35.0000
insulin	440.0000
BMI	39.4000
PEd	0.134
Age	30.0000
Class	0

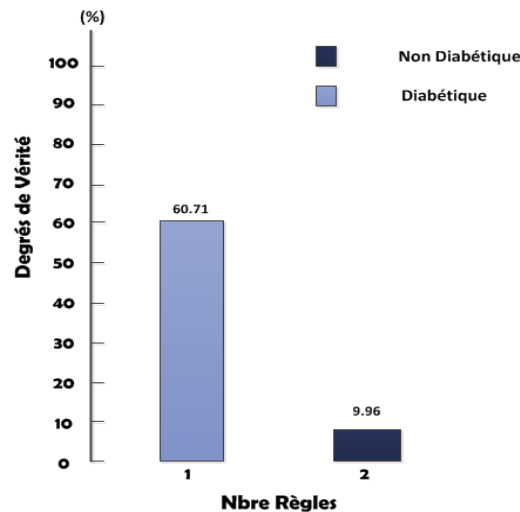


TABLE.4 Paramètres du patient 68.

FIGURE.4 Les règles floues activées pour le patient N° 68 (non diabétique) classé diabétique (FP)

A partir des différents exemples, nous constatons que les règles floues sont sollicitées en fonction des paramètres du vecteur d'entrée. Ensuite le moteur d'inférence floue évalue ses entrées en utilisant la base de connaissances pour diagnostiquer la classe du patient. Ces règles floues activées se rapprochent du raisonnement de l'expert médical. Le classifieur neuro-flou justifie ses résultats par les règles floues les plus sollicitées. Ainsi, l'expert peut facilement vérifier la plausibilité du classifieur flou, et voir les raisons d'une bonne ou mauvaise classification d'un patient à l'aide des paramètres le définissant ainsi que le degré de vérité des règles floues activées.

Les résultats expérimentaux ont prouvé que l'approche proposée avec Fuzzy C-Means (FCM) est simple et efficace pour montrer l'interprétabilité du classifieur neuro-flou en réduisant les 256 règles à 2 règles efficaces tout en préservant la précision du modèle à un niveau satisfaisant.

5.5 Hybridation Soustractive Clustering (SC)-ANFIS

Pour des fins de comparaison nous avons eu recours à un autre algorithme de clustering Soustractive Clustering (SC) pour voir si la répartition des entrées avec un autre principe de regroupement pourra améliorer les performances d'ANFIS.

Principe de Soustractive Clustering (SC)

Cette technique est appliquée lorsqu'il n'y a pas une idée claire sur le nombre de centres pour la répartition de données. La méthode subclustering est une extension de la méthode de classification proposée par Yager [Yag08]. Il suppose que chaque point de données est un centre potentiel de l'amas et calcule le potentiel de chaque point de données par la mesure de la densité des points de données l'entourant. L'algorithme sélectionne d'abord le point de données avec le plus grand potentiel en tant que centre de l'amas, puis évince les points de données à proximité du centre du premier groupe (déterminer par un rayon), pour calculer le centre prochain et ainsi de suite. Son algorithme est résumé comme suit :

Algorithm 5 Procédure de l'algorithme Soustractive clustering

1: **Etape1**

Choisir le point de donnée avec le plus grand potentiel d'être le centre du premier groupe

2: **Etape2**

Évincer tous les points dans le voisinage du 1^{er} centre de l'amas (déterminé par le rayon)

3: **Etape3**

Déterminer le prochain centre

4: **Etape4**

Itérer le processus jusqu'à ce que toutes les données soient dans un rayon du centre d'un amas.

Nous pourrions dire que l'algorithme SC réduit la complexité de calcul et donne une répartition des centres de cluster en fonction de la mesure de densité ainsi que le rayon (voir figure 3.18), sachant que chaque centre est un point des données avec le plus grand potentiel.

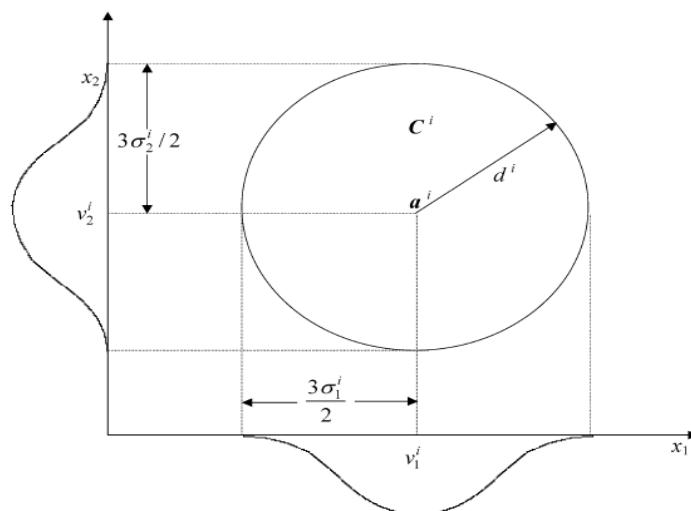


FIGURE 3.18 – Schéma représentatif de la répartition des clusters en des fonctions d'appartenance avec SC

L'application de l'algorithme soustractive clustering sur la base de données Pima a donné la répartition suivante :

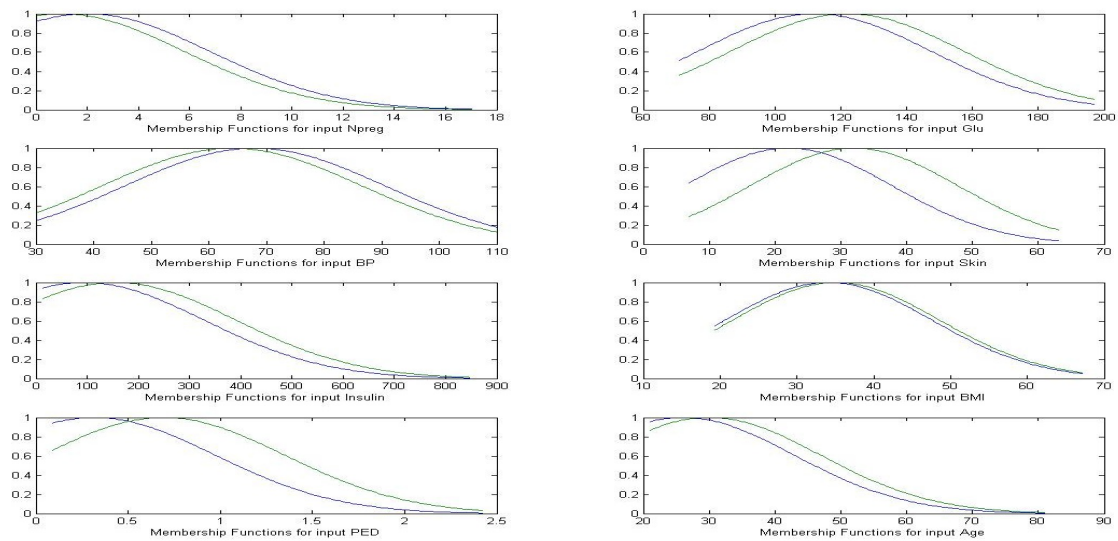


FIGURE 3.19 – Répartition des fonctions d'appartenance avec Soustractive clustering avec $r=0.8$

Pour un rayon $r=0.8$ le regroupement de données est réalisé sur deux 2 centres. Nous remarquons que la disposition des fonctions d'appartenance est presque superposée (figure 3.19), ce qui relève d'une mauvaise classification des patients. En essayant de varier le rayon (figures 3.20-3.21-3.22-3.23) nous avons constaté une augmentation du nombre de fonctions d'appartenance, mais les données demeurent non réparties de manière distincte.

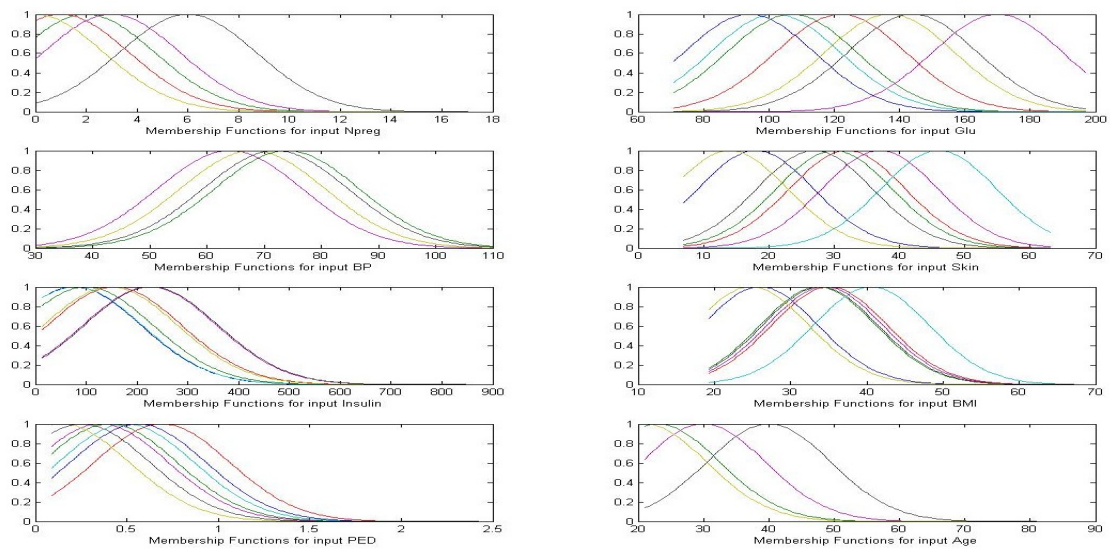


FIGURE 3.20 – Répartition des fonctions d'appartenance avec Soustractive clustering avec $r=0.45$

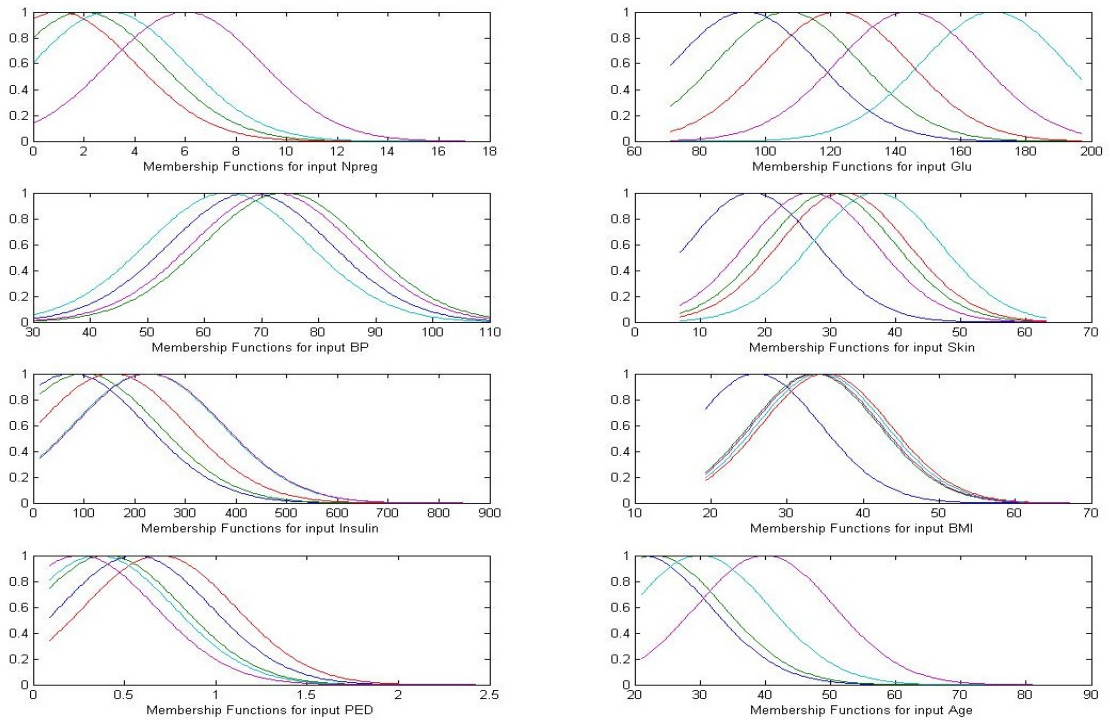


FIGURE 3.21 – Répartition des fonctions d'appartenance avec Soustractive clustering avec $r=0.5$

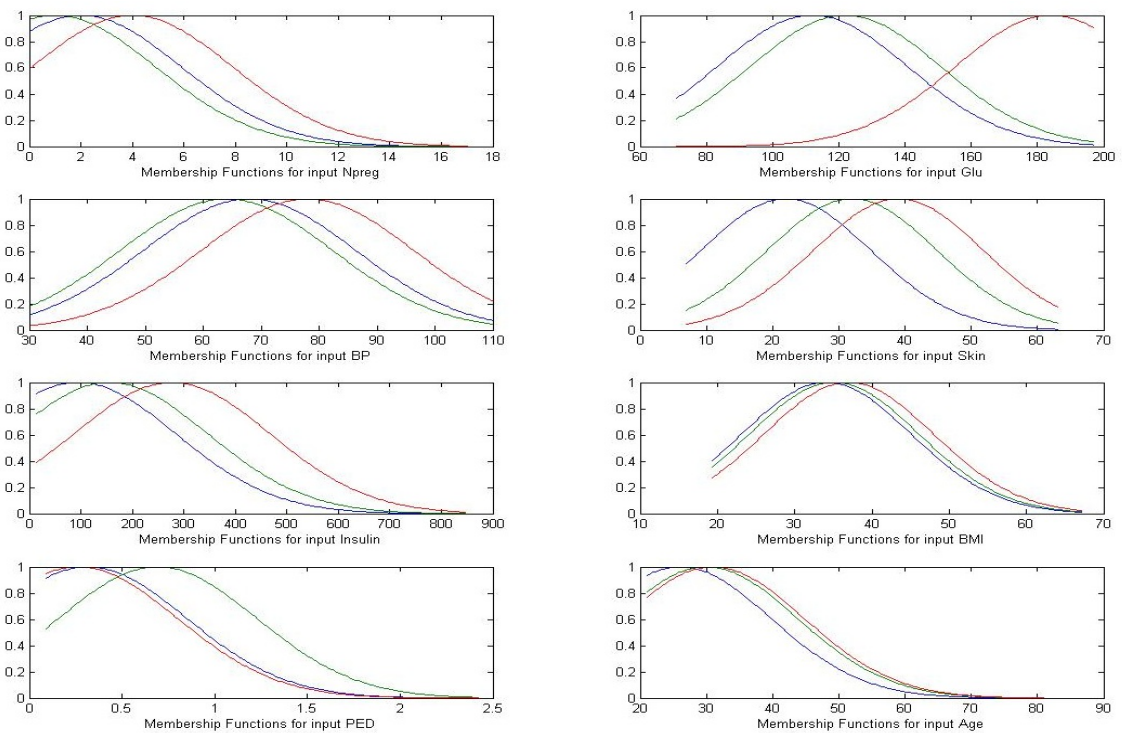


FIGURE 3.22 – Répartition des fonctions d'appartenance avec Soustractive clustering avec $r=0.65$

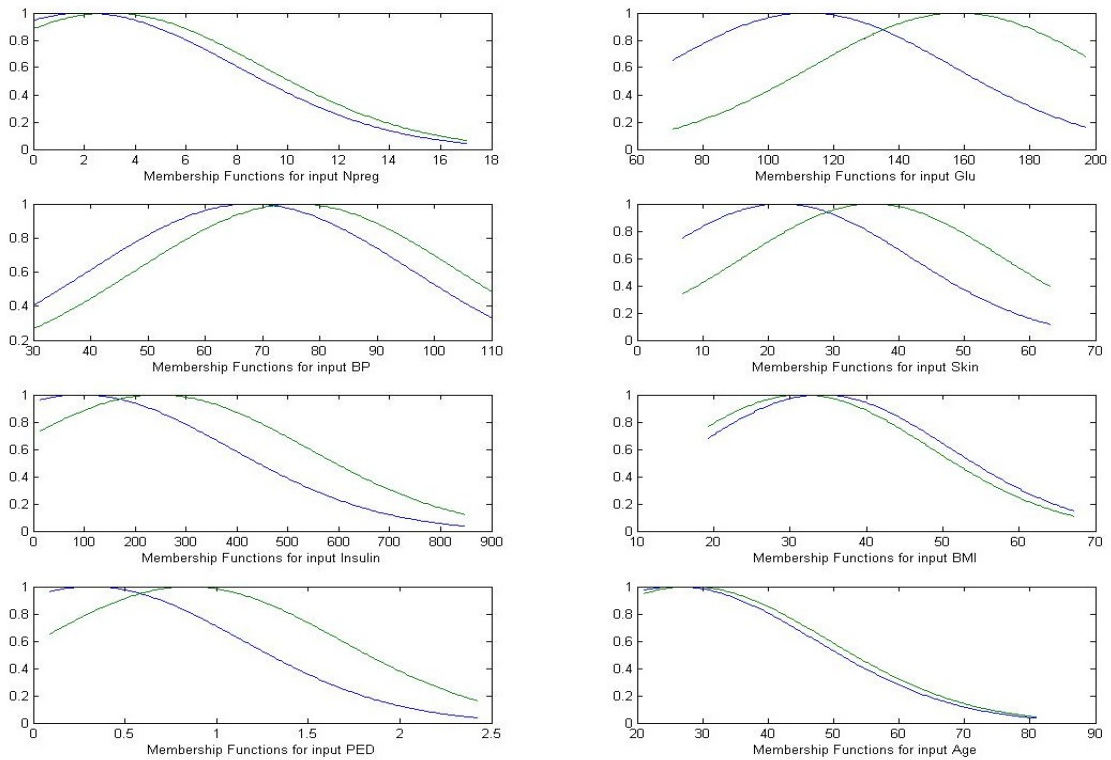


FIGURE 3.23 – Répartition des fonctions d’appartenance avec Soustractive clustering avec $r=1$

D’après les figures 3.20-3.21-3.22-3.23, l’automatisation de l’apprentissage structurel des fonctions d’appartenance avec l’algorithme soustractive clustering n’a pas d’intérêt. Cette hypothèse est confirmée en visualisant les résultats de son implémentation avec ANFIS voir table 3.7.

Rayons	TC%	Se%	Sp%	Nombre de règles
0.45	74.42	81.32	56.41	7
0.5	76.92	82.42	64.10	5
0.65	77.69	84.62	61.54	3
0.8	82.38	89.01	66.67	2
1	85.38	89.01	76.92	2

TABLE 3.7 – Performances du classifieur SCM-ANFIS avec différents rayons

La connaissance extraite du modèle SC-ANFIS

La connaissance extraite ne s’adapte pas avec les critères de diagnostic du diabète

Règle 1 : *If (Npreg is High) and (Glu is High) and (BP is High) and (Skin is Medium) and (Insulin is High) and (BMI is High) and (PED is High) and (Age is High) then (Class is Non Diabetic)*

Règle 2 : *If (Npreg is **Medium**) and (Glu is **Normal**) and (BP is **Low**) and (Skin is **High**) and (Insulin is **High**) and (BMI is **Medium**) and (PED is **Low**) and (Age is **Medium**) then (Class is **Diabetic**)*

Ces règles ne sont pas conformes au raisonnement des experts, l'interprétabilité présente est erronée. L'objectif visé qui est une bonne classification avec une amélioration de l'interprétabilité (une base de connaissance optimale) n'est pas vérifié.

De nombreuses études utilisant la structure ANFIS ont été réalisées dans le diagnostic du diabète. Polat & Günes [PG07] ont rapporté un taux de classification de 89,47% en appliquant l'analyse en composantes principales (ACP) et ANFIS. L'inconvénient majeur dans cette approche est que les composantes principales sont appliquées pour générer les règles floues, ce qui rend leur interprétation difficile, voir impossible. En outre, le résultat réalisé par [PG07] n'est pas reproductible, car Termurtas et al. [TNT09] qui recourent aux mêmes méthodes sur la même base de données ont obtenu 66,78% de reconnaissance des diabétiques. Vosoulipour et al. [VTM08] ont sélectionné quatre paramètres par l'implémentation de l'algorithme génétique (GA), à l'entrée d'un réseau de neurones et d'ANFIS. Ils ont rapporté respectivement 77,60% et 81,30%. Dogantekin et al. [DADL10], ont obtenu une précision de 84,61% à l'aide de l'analyse discriminante linéaire et ANFIS. La critique commune de ces méthodes est qu'elle ne fournit pas la connaissance explicite à l'utilisateur (l'interprétabilité des résultats reste absente).

Récemment Übeyli [Ü10] a appliqué ANFIS pour le diagnostic du diabète Indiens Pima. Elle est arrivée au meilleur résultat de classification avec 98,14%. Toutefois, cette amélioration pourra être discréditée à cause des valeurs atypiques dans la base de données Indiens Pima. Elle s'est basée sur la validation classique (une base d'apprentissage et une base de test) au lieu de K-fold technique de validation croisée.

Globalement ces travaux cités sont portés sur l'exactitude de classification au lieu de l'interprétabilité (participation des règles floues dans la décision finale). Des détails de ces études peuvent être vus dans Polat & Günes, [PG07].

Les résultats réalisés dans ce mémoire de Magister sont très intéressants par rapport aux études citées dans la littérature, puisque l'apport de Fuzzy C-Means a permis de réduire le nombre de règles floues qui peuvent être interprétées linguistiquement. Cela permettra au clinicien d'obtenir des informations sur le Diagnostic Automatique.

5.6 Conclusion

Cette étude présente un modèle de classification flou du diabète. Ici, deux critères sont utilisés pour évaluer la méthode proposée. Le premier est la bonne précision des performances du classifieur, le second touche la compréhensibilité des

résultats obtenus. Elle montre que l'algorithme Fuzzy C-Means (FCM) a réduit énormément la structure du modèle ANFIS, ainsi que le temps d'apprentissage, puisque le nombre de centres est égal au nombre de fonctions d'appartenance indépendamment de la taille des entrées. Avec l'hybridation de ces méthodes une extraction de connaissances a pu être faite avec 2 règles fiables, précises et suffisamment simples pour être comprises, le tout en améliorant les performances avec un taux de classification égale à 83,85%.

Conclusion

Un classifieur flou peut être un outil pratique dans le processus de diagnostic. Il se compose de règles linguistiques qui sont faciles à interpréter par l'expert humain. Ce classifieur n'est pas une boîte noire au sens où il peut contrôler la vraisemblance. Ceci est très important pour les systèmes de prise de décision, car les experts n'acceptent pas une évaluation sur ordinateur, à moins qu'ils comprennent pourquoi et comment une recommandation a été donnée.

Une règle floue (ou relation floue) peut être interprétée comme un prototype vague des données créées par le classifieur flou. Un résultat de classification est donné par les degrés de vérité (activations) de plusieurs règles. Les règles floues pour des cas similaires peuvent se chevaucher, delà un exemple peut appartenir à plusieurs classes à différents degrés d'appartenance. Cette information peut être exploitée pour évaluer la qualité du résultat de la classification.

Le résultat idéal serait qu'un modèle appartient clairement à une seule classe, i.e le degré d'appartenance à une classe est sensiblement plus grand que les degrés d'appartenance à d'autres classes. Toutefois, les cas limites n'ont généralement aucune relation avec les différentes classes. Nous pourrions rejeter une classification, ou dire qu'un exemple est inconnu, si les degrés d'appartenance à plusieurs classes sont très similaires, ou ne dépassent pas un certain seuil. Les informations fournies par les degrés d'appartenance nous permettent d'identifier les modèles qui sont mal classés, et les traiter séparément.

Les classifieurs flous ne résolvent pas les problèmes de classification mieux que d'autres approches, tel que : les statistiques, les arbres de décision ou les réseaux de neurones. Les avantages des classifieurs flous peuvent être principalement vus dans leur interprétabilité linguistique, la manipulation intuitive et la simplicité, qui sont des facteurs importants pour l'acceptation et l'utilisation d'une solution. Ce travail aborde le problème de la création d'un classifieur flou par apprentissage à partir des données, et notre objectif principal est d'obtenir une lisibilité du classifieur. L'application du Fuzzy C-Means a montré son intérêt dans la répartition des données et permet ainsi la régularisation des contraintes qui s'appliquent sur les paramètres des fonctions d'appartenance floues. Cet algorithme élimine la similarité par la fusion des partitions floues afin d'améliorer leurs distinctions, réduire le nombre de sous-ensembles flous et de minimiser le nombre de règles pour une connaissance ciblée. Delà donc automatiser et optimiser la structure et les paramètres des fonctions d'appartenance.

Fuzzy C-Means a fait ressortir aussi son efficacité dans sa capacité de classification des données non supervisées et ainsi exploiter cette caractéristique dans les réseaux de neurones pour un apprentissage plus rapide avec une structure réduite.

Les résultats obtenus dans ce travail de recherche nous permettent d'apporter une valeur ajoutée dans les systèmes d'Aide au Diagnostic Médical avec comme objectif principal la transparence du modèle. Chaque cas classifié est justifié par une ou plusieurs règles floues distinctes. Cette qualité absente chez la majorité des classifieurs de données, peut convaincre les médecins à utiliser largement les outils d'aide au diagnostic Médical Intelligents.

Beaucoup d'approches restent en perspective pour l'amélioration des classifieurs flous et neuronaux, par le biais des méthodes d'optimisation ou de réductions des paramètres tel que les algorithmes génétiques, Optimisation par essais particuliers, Systèmes immunitaires artificiels, Algorithme des colonies de Fourmies, etc.

Bibliographie

- [Alw11] Ala Alwan. Aperçu régional. Technical report, Fédération internationale du diabète, <http://www.diabetesatlas.org/>, Access Mars 2011.
- [BAKF06] N. Benamrane, A. Aribi, L. Kraloua, and A. Freville. Interprétation des images médicales par une approche neuro-flou-genétique. In *Modélisation, Optimisation et Simulation des Systèmes : Défis et Opportunités. 6e Conférence Francophone de Modélisation et Simulation - MOSIM06 - Rabat - Maroc*, 3 au 5 avril 2006.
- [BB97a] H. Bersini and G. Bontempi. Fuzzy models viewed as multi-expert networks. In *IFSA '97 (7th International Fuzzy Systems Association World Congress, Prague)*, pages 354–359, Prague, 1997. Academia.
- [BB97b] H. Bersini and G. Bontempi. Now comes the time to defuzzify the neuro-fuzzy models. *Fuzzy Sets and Systems*, 90(2) :161–170, 1997.
- [Bel11] Christine Beller. Centre européen d'étude du diabète. Technical report, www.ceed-diabete.org, access april 2011.
- [Bez74] James C. Bezdek. Cluster validity with fuzzy sets. *Cybernetics and Systems*, pages 58–72, 1974.
- [Bez81] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 1981.
- [BH94] M. Brown and C. J. Harris. *Neurofuzzy adaptive modelling and control*. Prentice Hall, Hemel Hempstead, 1994.
- [Bon97] Piero P. Bonissone. Soft computing : the convergence of emerging reasoning technologies. *Soft Computing*, 1 :6–18, 1997.
- [BPKK99] James C. Bezdek, Mikhail R. Pal, James Keller, and Raghu Krishnapuram. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Kluwer Academic Publishers, Norwell, MA, USA, 1999.
- [CB07] I. Burhan Aktac Ramazan Douganay M. Mete Ceylan Celikyilmaz, Asli Türkçen and N. Bacak. A new classifier design with fuzzy functions. In *Proceedings of the 11th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, RSFDGrC '07*, pages 136–143, Berlin, Heidelberg, 2007. Springer-Verlag.

- [CCLC99] K. Chuang, M. Chiu, C. Lin, and J. Chen. Model-free functional mri analysis using kohonen clustering neural network and fuzzy c-means. *IEEE Trans Med Imaging*, 18 issue 8 :11 :17–28, 1999.
- [Che07] Toly Chen. Incorporating fuzzy c-means and a back-propagation network ensemble to job completion time prediction in a semiconductor fabrication factory. *Fuzzy Sets Syst.*, 158 :2153–2168, October 2007.
- [CLF06] X. Chang, W. Li, and J. Farrell. A c-means clustering based fuzzy modeling method. In *The Ninth IEEE International Conference on Fuzzy Systems*, Vol.2, pp. 937–940, BIME Journal, Vol. 06, Issue 1, Dec. 2006.
- [CLM96] K. S. Cheng, J. S. Lin, and C. W. Mao. The application of competitive hopfield neural network to medical image segmentation. *IEEE Trans. Med. Imaging*, 15, no. 4 :pp. 560–567, Aug. 1996.
- [CO07] Rahime Ceylan and Yüksel Özbay. Comparison of fcm, pca and wt techniques for classification ecg arrhythmias using artificial neural network. *Expert Syst. Appl.*, 33 :286–295, August 2007.
- [DADL10] Esin Dogantekin, Dogantekin Akif, Avcı Derya, and Avcı Levent. An intelligent diagnosis system for diabetes on linear discriminant analysis and adaptive network based fuzzy inference system : Lda-anfis. *Digit. Signal Process.*, 20 :1248–1255, July 2010.
- [ded194] Groupe d’ étude de l’OMS. La prévention du diabète sucré. Technical report, OMS, Serie de rapports techniques No.844, 1994.
- [Dia02] OMS Diabète. Le coût du diabète. Technical report, Aide-mémoire No.236, 2002.
- [Dia03] OMS Diabète. Diabète et maladies rénales : il est temps d’agir. Technical report, Fédération Internationale du Diabète, 2003.
- [Dia11] OMS Diabète. Diabète. Technical report, Aide-mémoire No.312, Janvier 2011.
- [DJZ06] Sun Dan, Meng Jun, and He Zongyuan. Diagnosis of inverter faults in pmsm dtc drive using time-series data mining technique. In *Advanced Data Mining and ADMA 2006 Xi’an China August 14-16 2006 Proceedings Applications*, Second International Conference, editors, *ADMA*, volume 4093 of *Lecture Notes in Computer Science*. Springer, 2006.
- [DK97] R. N. Dave and R. Krishnapuram. Robust clustering methods : a unified view. *IEEE Transactions on Fuzzy Systems*, 5(2) :270–293, May 1997.
- [DLE04] D. Devendra, E. Liu, and G.S Eisenbarth. Eurodiab. Technical report, *BMJ* 328 :750-754 <http://bmj.bmjournals.com/>, 2004.
- [DS06] K.-L. Du and M. N. S. Swamy. *Neural Networks in a Softcomputing Framework*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

- [Dun74] J. Dunn. A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters. *Journal of Cybernetics*, 3 :32–57, 1974.
- [Elr06] Hassan M. Elragal. Neuro-fuzzy fabric defect detection and classification for knitting machine. In *The 23rd National Radio Science Conference (NRSC 2006) The 23 National March 14–16, 2006, Faculty of Electronic Engineering, Menoufiya University, Egypt., 2006*.
- [ETUA10] Narissara Eiamkanitchat, Nipon Theera-Umpon, and Sansanee Auephanwiriyaikul. A novel neuro-fuzzy method for linguistic feature selection and rule-based classification. *IEEE ASSP Magazine*, 2 :247–252, 2010.
- [FA10a] Mohammad Reza Faieghi and S. Mohammad Azimi. Design an optimized pid controller for brushless dc motor by using pso and based on narimax identified model with anfis. In *Proceedings of the 2010 12th International Conference on Computer Modelling and Simulation, UKSIM '10*, pages 16–21, Washington, DC, USA, 2010. IEEE Computer Society.
- [FA10b] A. Frank and A. Asuncion. UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. University of California, Irvine, CA : School of Information and Computer Science.(access 06-02-2011), 2010.
- [FCSFB01] A Fagot-Campagna, JB Saaddine, KM Flegal, and GL. Beckles. Diabetes, impaired fasting glucose, and elevated hba1c in u.s. adolescents : the third national health and nutrition examination surve. *adolescents : the Third National Health and Nutrition Examination Surve Diabetes Care*, 24 (5) :834–837, 2001.
- [FS89] Y. Fukuyama and M. Sugeno. A new method of choosing the number of clusters for fuzzy c-means method. *Fuzzy System Symposium*, pages 247–250, 1989.
- [ftDPPRG02] Knowler WC for the Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. Technical report, *N Engl J Med* ; 346 : 393–403., 2002.
- [GT00] R. Gilleron and M. Tommasi. Découverte de connaissances à partir de données. Technical report, Technical report, Grappa - Université de Lille 3, 2000.
- [Gui01] S. Guillaume. Designing fuzzy inference systems from data : An interpretability-oriented review. *IEEE Transactions on Fuzzy Systems*, 9 :426–442., 2001.
- [Gur97] Kevin Gurney. *An Introduction to Neural Networks*. Taylor & Francis, Inc., Bristol, PA, USA, 1997.
- [GVGB09] Jonas Guzaitis, Antanas Verikas, Adas Gelzinis, and Marija Ba-causkiene. A framework for designing a fuzzy rule-based classifier. In *Proceedings of the 1st International Conference on Algorithmic*

- Decision Theory*, ADT '09, pages 434–445, Berlin, Heidelberg, 2009. Springer-Verlag.
- [Hai98] S. Haikin. *Neural Networks : A Comprehensive Foundation*. Pearson Education, NY, 2nd edition, 1998.
- [HAJ09] Haseena Hassan, Mathew Abraham, and Paul Joseph. Fuzzy clustered probabilistic and multi layered feed forward neural networks for electrocardiogram arrhythmia classification. *Journal of Medical Systems*, July 2009.
- [HBC+96] CL Hanis, E Boerwinkle, R Chakraborty, DL Ellsworth, P Concannon, and B Stirling. A genomewide search for human non-insulin-dependent (type 2) diabetes genes reveals a major susceptibility locus on chromosome 2. *Nat Genet*, 13 (2) :161–166., 1996.
- [HBV02] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. Clustering validity checking methods : part ii. *SIGMOD Rec.*, 31 :19–27, September 2002.
- [HEHM07] Yasushi Hasegawa, Yasunori Endo, Yukihiro Hamasuna, and Sa-daaki Miyamoto. Fuzzy c-means for data with tolerance defined as hyper-rectangle. In *Proceedings of the 4th international conference on Modeling Decisions for Artificial Intelligence*, MDAI '07, pages 237–248, Berlin, Heidelberg, 2007. Springer-Verlag.
- [HJM10] H. H. Haseena, P. K. Joseph, and A. T. Mathew. Classification of arrhythmia using hybrid networks. *J Med Syst*, 2010.
- [HK00] Jiawei Han and Micheline Kamber. *Data Mining : Concepts and Techniques (The Morgan Kaufmann Series in Data Management Systems)*. Morgan Kaufmann, 1st edition, September 2000.
- [HK01] F. Höppner and H.F. Klawonn. A new approach to fuzzy partitioning. In *In Proc. of the Joint 9th IFSA World Congress and 20th NAFIPS Int. Conf*, pages 1419–1424, 2001.
- [HKK+07] Sang-Hyun Hwang, Dongwon Kim, Tae-Koo Kang, , and Gwi-Tae Park. Medical diagnosis system of breast cancer using fcm based parallel neural networks. In *Proceedings of the 3rd International Conference on Intelligent Computing : Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, ICIC '07, pages 712–719, Berlin, Heidelberg, 2007. Springer-Verlag.
- [HKKT99] F. Höppner, H. F. Klawonn, R. Kruse, and Runkler T. *Fuzzy Cluster Analysis : Methods for Classification, Data Analysis and Image Recognition*. John Wiley & Sons, 1., auflage edition, 1999.
- [HLL05] Hai Tao He, Xiao Nan Luo, and Bao Lu Liu. *Detecting Anomalous Network Traffic with Combined Fuzzy-Based Approaches*. ICIC 2005, Part II, LNCS 3645, pp. 433–442,, 2005.
- [HN99] Harald Hruschka and Martin Natter. Comparing performance of feedforward neural nets and k-means for cluster-based market segmentation. *European Journal of Operational Research*, 114(2) :346–353, April 1999.

- [HP10] Shrishail Hiremath and Sarat Kumar Patra. Transmission rate prediction for cognitive radio using adaptive neural fuzzy inference system. In *Fifth International Conference on Industrial and Information Systems*, 2010.
- [HSA09] Ude D. Hemanth, D. Selvathi, and J. Anitha. Effective fuzzy clustering algorithm for abnormal mr brain image segmentation. In *International/Advance Computing Conference (IACC 2009),IEEE*, 609–614, 2009.
- [IKM10] S. Ibrahim, N. E. Abdul Khalid, and M. Manaf. Seed-based region growing (sbrg) vs adaptive network-based inference system (anfis) vs fuzzy c-means (fcm) : Brain abnormalities segmentation. *International Journal of Electrical and Computer Engineering*, 5 :2–5, 2010.
- [Jan93] J. S. R. Jang. ANFIS : adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3) :665–685, August 1993.
- [JD88] Anil K. Jain and Richard C. Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.
- [Jin03] Yaochu Jin. *Advanced Fuzzy Systems Design and Applications*. Physica-Verlag, 2003.
- [JM96] J.-S. R. Jang and E. Mizutani. Levenberg-marquardt method for anfis learning. In *Biennial Conference of the North American Fuzzy Information Processing Society*, 87–91., June 1996.
- [JM08] Nahla Ibraheem Jabbar and Monica Mehrotra. Application of fuzzy neural network for imagetumor description. *World Academy of Science, Engineering and Technology*, 44 :575–577, 2008.
- [JNM+03] Wei Ji, Raouf N.G. Naguib, John Macall, Dobrila Petrovic, Elena Gaura, and Mohamed Ghoneim. Prognostic prediction of bilharziasis-related bladder cancer by neuro-fuzzy classifier. In *ITAB*, 2003.
- [JS95] J. S. R. Jang and C. T. Sun. Neuro-Fuzzy modeling and control. *Proceedings of The IEEE*, 83(3) :378–406, March 1995.
- [JSM97] J. S. R. Jang, C. T. Sun, and E. Mizutani. *Neuro-Fuzzy and Soft Computing*. Matlab Curriculum Series. Prentice Hall, 1997.
- [JYAD96] A. Jana, P. H. Yang, D. M. Auslander, and R. N. Dave. Real time neuro-fuzzy control of a nonlinear dynamic system. In *Biennial Conference of the North American Fuzzy Information Processing Societ*, 210–214, June 1996.
- [KA08] Humar Kahramanli and Novruz Allahverdi. Design of a hybrid system for the diabetes and heart diseases. *Expert Syst. Appl.*, 35 :82–89, July 2008.
- [Kan08] S. R. Kannan. A new segmentation system for brain mr images based on fuzzy techniques. *Appl. Soft Comput.*, 8 :1599–1606, September 2008.

- [KLG⁺01] S. S. Kim, D.J. Lee, K. C. Kwark, J. S. Kim, J.W. Ryu, and S.H. Lee. Speech recognition using the metric defines by intregra normalizer. *ISIE IEEE*, 2001.
- [KoTA03] B. Karlik, M. o. Tokhi, and M. Alci. A fuzzy clustering neural network architecture for multifunction upper-limb prosthesis. In *Bio-medical Engineering*, 2003.
- [KR90] L. Kaufman and P.J. Rousseeuw. *Finding Groups in Data An Introduction to Cluster Analysis*. Wiley Interscience, New York, 1990.
- [KSC06] S. Chen J. Wu-T. J. Chen K. S. Chuang, H. L. Tzeng. Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics*, vol 30 :9–15, 2006.
- [KY03] K. Kayaer and T. Yildirim. Medical diagnosis on pima indian diabetes using general regression neural networks. In *the international conference on artificial neural networks and neural information processing (ICANN/ICONIP)*, 2003.
- [Lau08] P. Laurent. *L'hypertension artérielle en 200 questions*. De Vecchi, Paris, 2008.
- [LCT09] Shing-Hong Liu, Kang-Ming Chang, and Chu-Chang Tyan. *Fuzzy C-Means Clustering for Myocardial Ischemia Identification with Pulse Waveform Analysis*. ICBME 2008, pp. 485–489, 2009.
- [LDZ00] A. Lorette, X. Descombes, and J. Zerubia. Texture analysis through a markovian modelling and fuzzy classification : Application to urban area extraction from satellite images. *International Journal of Computer Vision*, 36(3) :221–236, 2000.
- [LS99] Zhi-Qiang Liu and R. Satur. Contextual fuzzy cognitive map for decision support in geographic information systems. *Fuzzy Systems, IEEE Transactions on*, 7 Issue :5 :495–507, 1999.
- [LVY06] Seung-Heon Lee, Murlikrishna Viswanathan, and Young-Kyu Yang. Link speed estimation and incident detection using clustering and neuro-fuzzy methods. In *Computational Science - ICCS 2006, 6th International Conference, Reading, UK,, May 28-31, 2006, Proceedings, Part III 2006*.
- [LWA⁺09] Chang-Shing Lee, Mei-Hui Wang, Giovanni Acampora, Vincenzo Loia, and Chin-Yuan Hsu. Ontology-based intelligent fuzzy agent for diabetes application. pages 16–22, 2009.
- [MBG06] Boyd E. Metzger, Susan A. Biastre, and Beverly Gardner. National diabetes information clearinghouse. Technical report, NIH Publication No.06–5129, April 2006.
- [ML00] Y. Miao and Z.Q. Liu. On causal inference in fuzzy cognitive maps. *IEEE Trans. Fuzzy Syst.*, 8, No. 1 :107–119, Feb. 2000.
- [MP43] W. S. Mcculloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysic*, 5 :115–133, 1943.

- [MR06] S. Murugavalli and V. Rajamani. A high speed parallel fuzzy c-mean algorithm for brain tumor segmentation. *ICGST International Journal on Bioinformatics and Medical Engineering, BIME*, 06 :29–34, 2006.
- [NA09] Deok Hee Nam and Edward Asikele. Data mining of image segments data with reduced neurofuzzy system. In *Proceedings of the 2nd International Conference on Digital Human Modeling : Held as Part of HCI International 2009, ICDHM '09*, pages 710–716, Berlin, Heidelberg, 2009. Springer-Verlag.
- [NKK97] Detlef Nauck, Frank Klawonn, and Rudolf Kruse. *Foundations of Neuro-Fuzzy Systems*. John Wiley & Sons, Inc., New York, NY, USA, 1997.
- [Now09] Robert Nowicki. Rough neuro-fuzzy structures for classification with missing data. *Trans. Sys. Man Cyber. Part B*, 39 :1334–1347, December 2009.
- [NSK⁺87] B Newman, JV Selby, MC King, C Slemenda, R Fabsitz, and GD. Friedman. Concordance for type 2 (noninsulin-dependent) diabetes mellitus in males twins. Technical report, *Diabetologia*, 30 :pp. 763–768., 1987.
- [OCK06] Yüksel Özba, Rahime Ceylan, and Bekir Karlik. A fuzzy clustering neural network architecture for classification of ecg arrhythmias. *Comput. Biol. Med.*, 36 :376–388, April 2006.
- [OMS03] OMS/FAO. Diet, nutrition and the prevention of chronic diseases. Technical report, Geneva : WHO TRS 91, 2003.
- [OMS07] OMS. Service des maladies chronique. Technical report, *Diabetes Care* ; vol 28 : pp. 2130–2135, septembre 2007.
- [OY07] Mohd Fauzi bin Othman and Thomas Moh Shan Yau. Neuro fuzzy classification and detection technique for bioinformatics problems. In *Proceedings of the First Asia International Conference on Modelling & Simulation*, pages 375–380, Washington, DC, USA, 2007. IEEE Computer Society.
- [Par06] Pole Santé Parisien. Diabète. Technical report, Dossier d'Actualités Diabète No.12, Novembre 2006.
- [PDP10] Dipti Patra, Manab Kumar Das, and Smita Pradhan. Integration of fcm, pca and neural networks for classification of ecg arrhythmia. *IAENG International Journal of Computer Science*, 36 :3, 3 February 2010.
- [PG07] Kemal Polat and Salih Güneş. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digit. Signal Process.*, 17 :702–710, July 2007.
- [PKKS09] Nikolaos Papageorgiou, Elpiniki I. and Papandrianos, Georgia Karagianni, G. Kyriazopoulos, and D. Sfyra. Fuzzy cognitive map based approach for assessing pulmonary infections. In *Proceedings*

- of the 18th International Symposium on Foundations of Intelligent Systems, ISMIS '09, pages 109–118, Berlin, Heidelberg, 2009. Springer-Verlag.
- [PM00] Dan Pelleg and Andrew W. Moore. X-means : Extending k-means with efficient estimation of the number of clusters. In *Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00*, pages 727–734, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.
- [PSS⁺06] E.I. Papageorgiou, P.P. Spyridonos, C.D. Stylios, P. Ravazoula, P.P. Groumos, and G.N. Nikiforidis. Advanced soft computing diagnosis method for tumour grading. *Artif. Intell. Med.*, 36 :59–70, January 2006.
- [PY98] K. M. Passino and S. Yurkovich. *Fuzzy control*. Addison Wesley Longman, Inc., 1998.
- [RHW88] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. *Learning representations by back-propagating errors*, pages 696–699. MIT Press, Cambridge, MA, USA, 1988.
- [RLCB86] J V. Dave R L. Cannon and J C. Bezdek. Efficient implementation of the fuzzy c-means clustering algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8 :248–255, March 1986.
- [RP03] M. Ceylan B. Karlik R. Pektatli, Y. Ozbay. Classification of ecg signals using fuzzy clustering neural networks (fcnn). In *Proceedings of the International XII, TAINN'03, vol. 1(1), Canakkale, Turkey, pp. 105–108.*, 2003.
- [RTK95] P. J. Rousseeuw, E. Trauwaert, and L. Kaufman. Fuzzy clustering with high contrast. *J. Comput. Appl. Math.*, 64 :81–90, November 1995.
- [Rus69] E. H. Ruspini. A new approach to clustering. *Information and Control*, 15(1) :22–32, 1969.
- [RV10] V.S. Rao and S. Vidyavathi. Comparative investigations and performance analysis of fcm and mfpcm algorithms on iris data. *Indian Journal of Computer Science and Engineering*, 1 No. 2 :145–151, 2010.
- [SABAAG00] M.N. Syed-Ahmad, Ahmed Bensenouci, Saleh A. Alghamdi, and A.M. Abdel Ghany. Short-term load forecasting using adaptive neuro-fuzzy inference system (anfis) application to aleppo load demand. *Aleppo Research magazine (Eng. Sciences)*, No 28, 2000.
- [SBAS09] Ahmad Shihi, Rodziah Binti Atan, and Nasir Sulaiman. An effective fuzzy c-mean and type-2 fuzzy logic for weather forecasting. *Journal of Theoretical and Applied Information Technology*, 5. No5. :556–567, 2009.
- [SBKvN98] M. Setnes, R. Babuska, U. Kayamak, and van Nauta. Similarity measures in fuzzy rule base simplification. *IEEE Transactions on Systems, Man and Cybernetics - Part B : Cybernetics*, 28 :376–386, 1998.

- [SCF09] Nesma Settouti, Amine Chikh, and Tadj Eddine Fandi. Reconnaissance par Neuro-Floue du Diabète. Research report, hal.inria.fr, 07 2009. Ce rapport de recherche fait parti du projet fin d'étude d'ingénieur d'état en Génie BioMédical.
- [SED⁺88] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. pages 261–265, 1988.
- [Sin08] Raj Mohan Singh. Fuzzy and neural network models for pollution concentration predictions in streams. In *The 12th International Conference of International Association for Computer Methods and Advances in Geomechanics (IACMAG)*. Goa, India., 1-6 October, 2008.
- [SK88] M. Sugeno and G. T. Kang. Structure identification of fuzzy model. *Fuzzy Sets Syst.*, 28 :15–33, October 1988.
- [SK10] Deepika Sirohi and Pawan Kumar. Comparative analysis of fcm and hcm algorithm on iris data set. *International Journal of Computer Applications*, 5(2) :33–37, August 2010. Published By Foundation of Computer Science.
- [Sky93] J. Skyler. Immune intervention in type i diabetes mellitus. diabetes/metabolism review. Technical report, vol1 : pp. 15–42., 1993.
- [SM00] Y. Hayashi S. Mitra. Neuro-fuzzy rule generation : survey in soft computing framework. *IEEE Trans. on Neural Networks*, 11(3) :748–768, 2000.
- [SPGN06] P. Spyridonos, E.I. Papageorgiou, P.P. Groumpos, and G. Nikiforidis. Integration of expert systems with image analysis techniques for medical diagnosis. In : *Proc ICIAR 2006, Lecture Notes in Computer Science*, 4142, Springer-Verlag :110–121, 2006.
- [Sun94] C.-T Sun. Rulebase structure identification in an adaptive-network-based fuzzy inference system. *IEEE Trans. Fuzzy Systems*, 2 (1) :64–73, 1994.
- [TNT09] Hasan Temurtas, Yumusak Nejat, and Feyzullah Temurtas. A comparative study on diabetes disease diagnosis using neural networks. *Expert Syst. Appl.*, 36 :8610–8615, May 2009.
- [TS85] Tomohiro Takagi and Michio Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(1) :116–132, February 1985.
- [Ü08] Elif Derya Übeyli. Modified mixture of experts for diabetes diagnosis. *J. Med. Syst.*, 33 :299–305, August 2008.
- [Ü10] Elif Derya Übeyli. Automatic diagnosis of diabetes using adaptive neuro-fuzzy inference systems. *Expert Systems*, 27, Issue 4, :259–266, September 2010.
- [VTM08] A. Vosoulipour, M. Teshnehlab, and H. A. Moghadam. Classification on diabetes mellitus data-set based-on artificial neural networks and anfis. 21 :27–30, 2008.

- [WBG97] Patricia M. West, Patrick L. Brockett, and Linda L. Golden. A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, 16(4) :370–391, 1997.
- [Wer74] P. Werbos. *Beyond Regression : New Tools for Prediction and Analysis in the Behavioral Sciences*. PhD thesis, Harvard University, Cambridge, MA, 1974.
- [WMBOLS04] Axel Wismüller, Anke Meyer-Bäse, Maximilian F. Reiser Oliver Lange, and Dorothee Auer, and DeWitt Sumners. Model-free functional mri analysis based on unsupervised clustering. *J. of Bio-medical Informatics*, 37 :10–18, February 2004.
- [WVB+00] P.A Weill, N Vallier, R Bourrel, P Fender, and H. Allemand. Lépidémiologie du diabète en france métropolitain. Technical report, *Diabetes Metab ;26(6) :pp. 11–24.*, 2000.
- [WY02] Kuo-Lung Wu and Miin-Shen Yang. Alternating c-means clustering algorithm. *Pattern recognition*, 35 :267–278, 2002.
- [WY05] Kuo-Lung Wu and Miin-Shen Yang. A cluster validity index for fuzzy clustering. *Pattern Recogn. Lett.*, 26 :1275–1291, July 2005.
- [WZYL05] Yuan-zhen Wang, Zhi-bing Zhang, Bao-li Yi, and Hua-yang Li. Data mining based on computational intelligence. *Wuhan University Journal of Natural Sciences*, 10, No.2 :371–374, 2005.
- [XB91] Xuanli Lisa Xie and Gerardo Beni. A validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13 :841–847, August 1991.
- [XH08] Hong-Jie Xing and Bao-Gang Hu. An adaptive fuzzy c-means clustering-based mixtures of experts model for unlabeled data classification. *Neurocomput.*, 71 :1008–1021, January 2008.
- [Yag08] Ronald R. Yager. S-mountain method for obtaining focus points from data. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 16(6) :815–828, 2008.
- [Zad65] L.A. Zadeh. Fuzzy sets. *Information Control*, 8 :338–353, 1965.
- [ZJW08] Hui-zhe Zhang and and Zi-hui Ren Jian Wang. Rough sets and fcm-based neuro-fuzzy inference system for traffic incident detection. In *Proceedings of the 2008 Fourth International Conference on Natural Computation - Volume 07, ICNC '08*, pages 260–264, Washington, DC, USA, 2008. IEEE Computer Society.

Résumé

L'intelligibilité représente la force motrice la plus importante derrière la mise en œuvre des classifieurs à base flous pour les problèmes d'application médicale. L'expert devrait être capable de comprendre le classifieur et d'évaluer ses résultats. Les modèles à base de règles floues sont particulièrement adaptés, car ils sont constitués de simples règles linguistiques interprétables. Dans la littérature, la majorité des algorithmes basés sur un système d'inférence neuro-flous adaptatifs (ANFIS) ne fournissent pas suffisamment d'explications sur la façon d'obtenir les résultats d'inférence. Ce mémoire de Magister traite la possibilité d'augmenter l'interprétabilité du classifieur ANFIS avec l'apport de la méthode de clustering Fuzzy C-Means . Il montre comment un classifieur neuro-flou interprétable peut être obtenu par un processus d'apprentissage et comment les règles floues extraites peuvent améliorer son interprétation. Les résultats expérimentaux appliqués sur la base de données du diabète (UCI Machine Learning) montrent de fortes similitudes avec les règles appliquées par les experts. Les résultats sont comparés à d'autres travaux dans la littérature. L'approche proposée est simple et efficace pour clarifier la décision finale du classifieur, tout en préservant sa précision à un niveau satisfaisant.

Mots clés

Classification interprétable ; Règles floues ; FCM ; Neuro-Flou ; Base de données du dépôt UCI Machine Learning.

Abstract

Interpretability represents the most important driving force behind the implementation of fuzzy-based classifiers for medical application problems. The expert should be able to understand the classifier and to evaluate its results. Fuzzy rule based models are especially suitable, because they consist of simple linguistically interpretable rules. The majority of classifiers based on an adaptive neuro-fuzzy inference system (ANFIS) used in literature do not provide enough explanation of how their inference results have been obtained. This Magister thesis discusses the possibility to increase the interpretability of ANFIS classifier by the hybridation with the clustering method Fuzzy C means. It is shown how a readable neuro-fuzzy classifier can be obtained by a learning process and how fuzzy rules extracted can enhance its interpretability. Experimental results show that the proposed fuzzy classifier can achieve a good tradeoff between the accuracy and interpretability. Also the basic structure of the fuzzy rules which were automatically extracted from the UCI Machine learning database shows strong similarities to the rules applied by human experts. Results are compared to other approaches in the literature. The proposed approach gives more compact, interpretable and accurate classifier.

Keywords

Interpretable classification ; fuzzy rules ; FCM ; neuro-fuzzy ; UCI Machine Learning Database.