Renforcement de l'Apprentissage Structurel pour la reconnaissance du Diabète

Nesma Settouti

Renforcement de l'Apprentissage Structurel pour la Reconnaissance du Diabète

Nesma Settouti[1]

[1]Département Génie Électrique et Électronique,
Laboratoire Génie Biomédical
Université Abou Bekr Belkaid – Tlemcen,
B.P.230- Tlemcen 13000, Algérie

Table des matières

Table des figures

Liste des tableaux

1 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.

2 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 l'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 floue (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'interprétablité et la lisibilité.

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.

3 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 interpretabilité en renforçant la connaissance extraite.

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.

4 Propositions

Dans ce rapport nous élaborons nos différentes contributions proposées en parcourant l'état de l'art, Pour cela nous divisons cette section 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 :

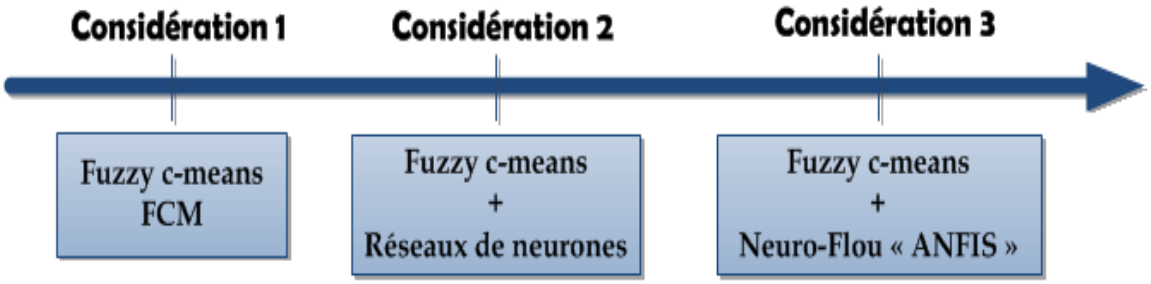

FIGURE 1 – Schéma représentatif de la procédure de classification du diabète.

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 mutlicouches 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 recourt aux logiciels : R 2.12.0 (open source) et Matlab 2010a.

5 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 [FA10] 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.

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'héridité).
8. Age : âge

Les figures 2 et 3 montrent respectivement les histogrammes et la répartition des diabétiques et non diabétiques des paramètres de la base de données :

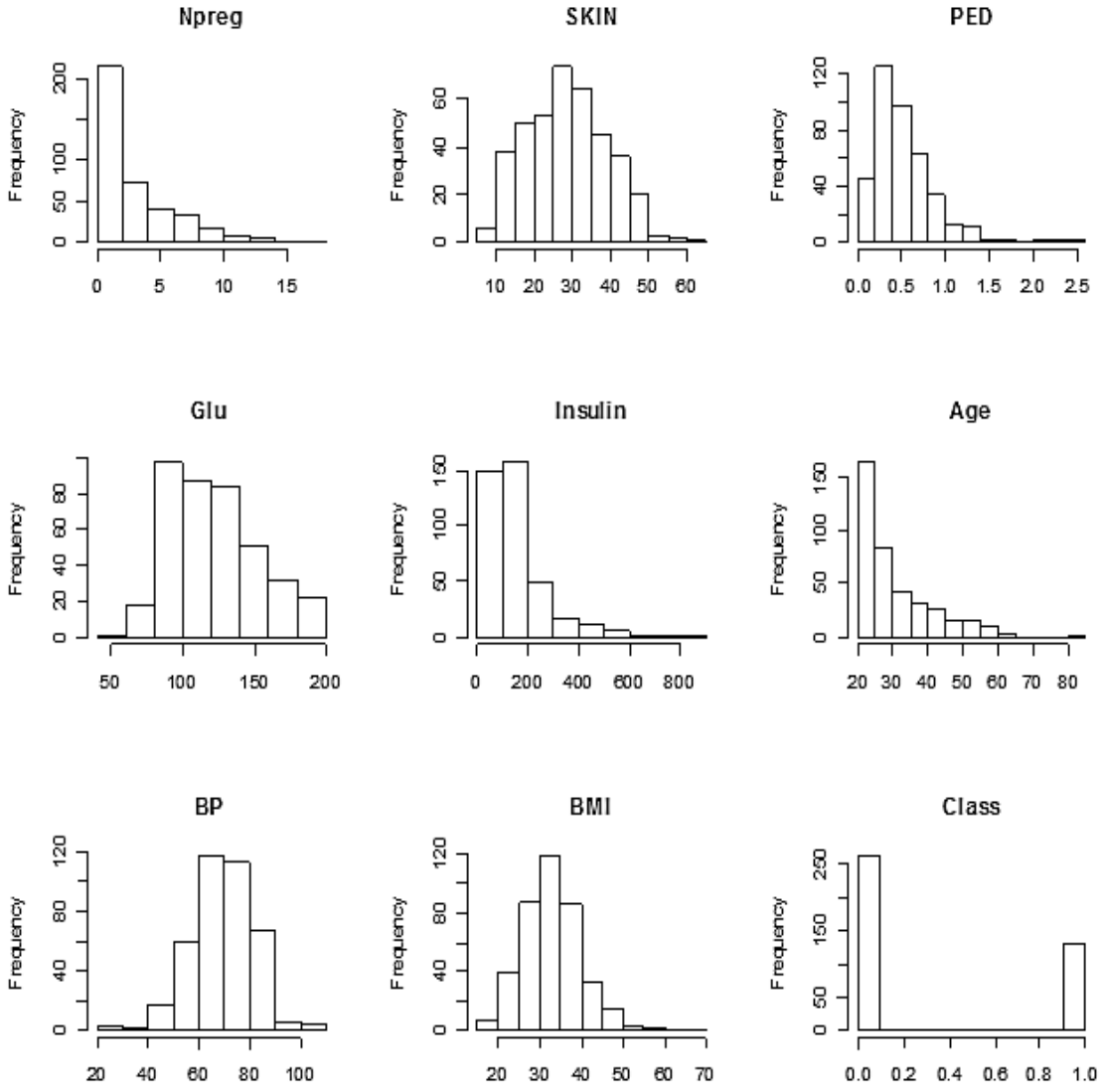

FIGURE 2 – Histogrammes des 9 paramètres de la base de données

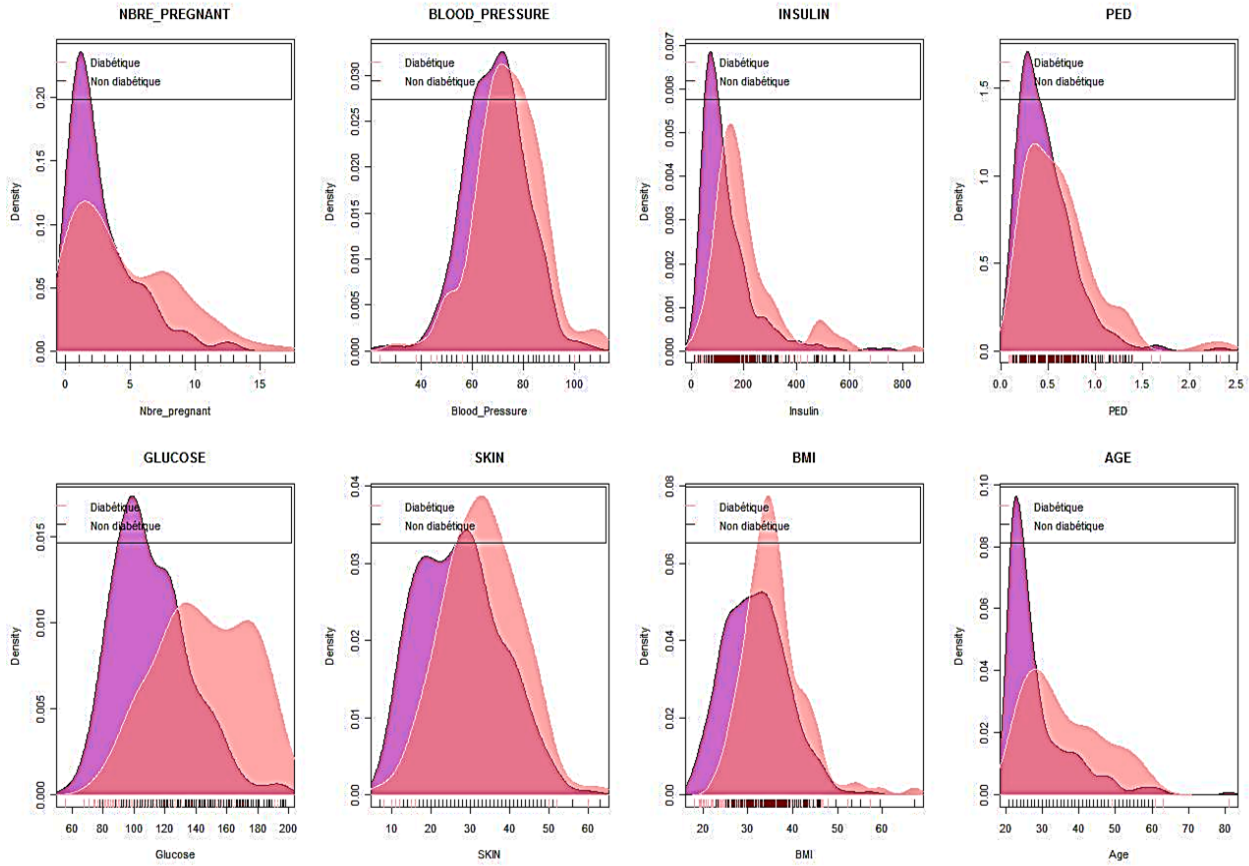

FIGURE 3 – Représentation de la base de données : distribution des diabétiques et non diabétiques dans les paramètres de la base de données.

6 Considération 1 : Classification par Fuzzy C-Means

6.1 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** $t = 1, 2, \ldots$ **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 :
$D_{ijA}^2 = (x_j – c_i)^T A (x_j – c_i), \qquad 1 \leq i \leq n_c, 1 \leq j \leq N.$
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).

6.2 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 $c = 2, \ldots, 10$ 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 $V_{S_{Dbw}}$ 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 a 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.

6.2.1 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 **Si** et **Où** de nouveaux centres de gravité devront apparaître.

La figure 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 4 correspond à c = 2, où la somme des carrés au sein

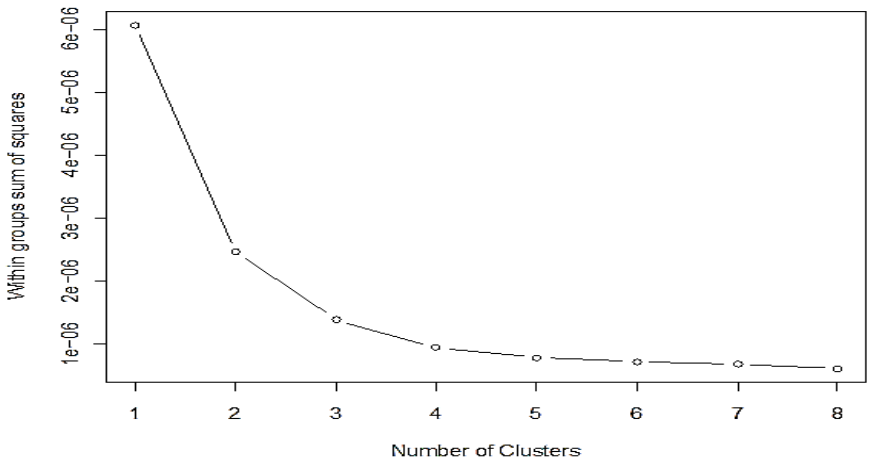

FIGURE 4 – Nombre de centres optimaux avec la technique x-means

des groupes est de 2e^{-8} (valeur relativement faible).

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

6.2.2 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 5) :

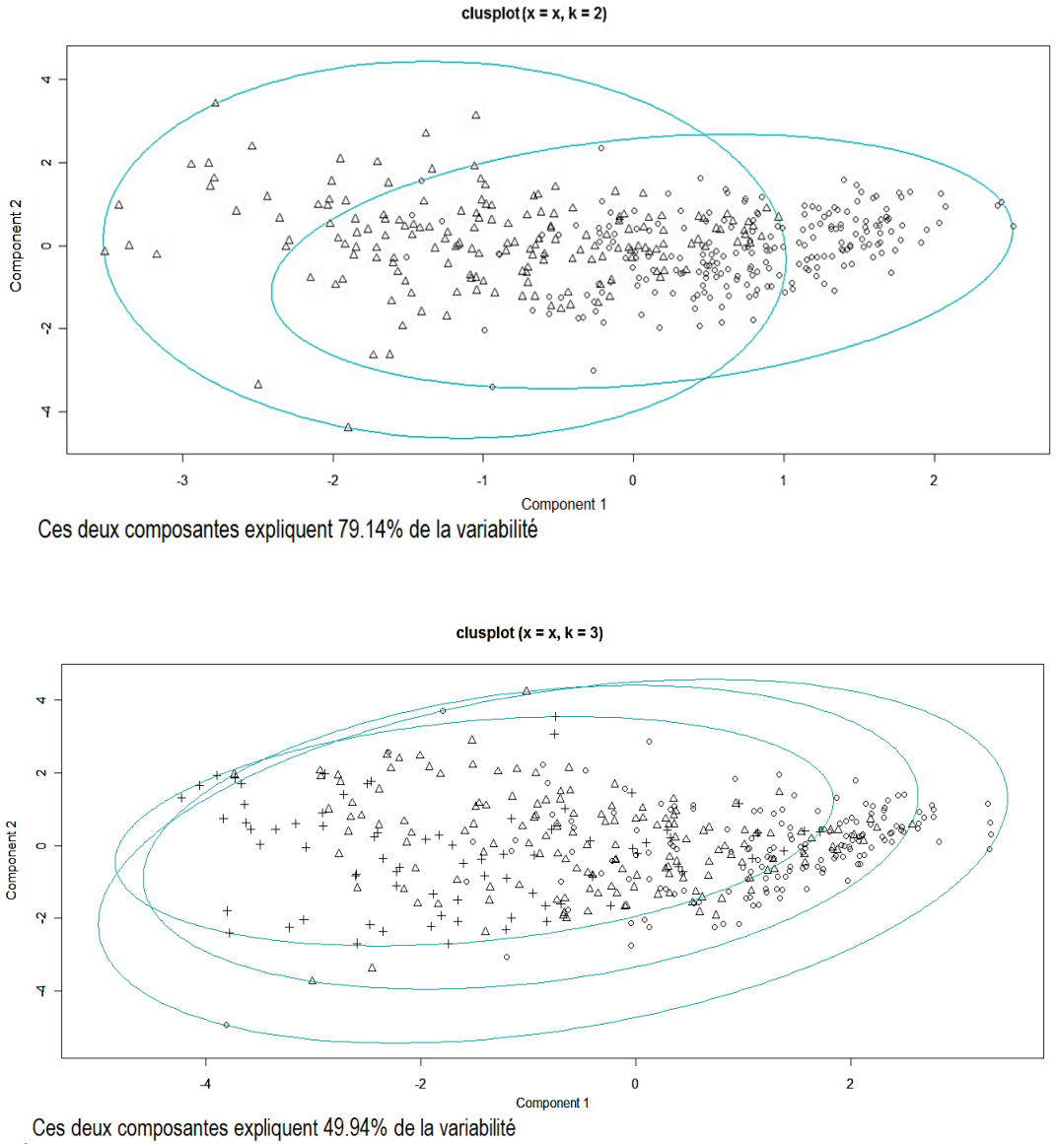

FIGURE 5 – Répartition des exemples sur 2 et 3 centres

Avec un nombre de centre égal à 2 les deux composantes obtenues comportent plus de 79% de la variabilité des points (voir figure 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 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.

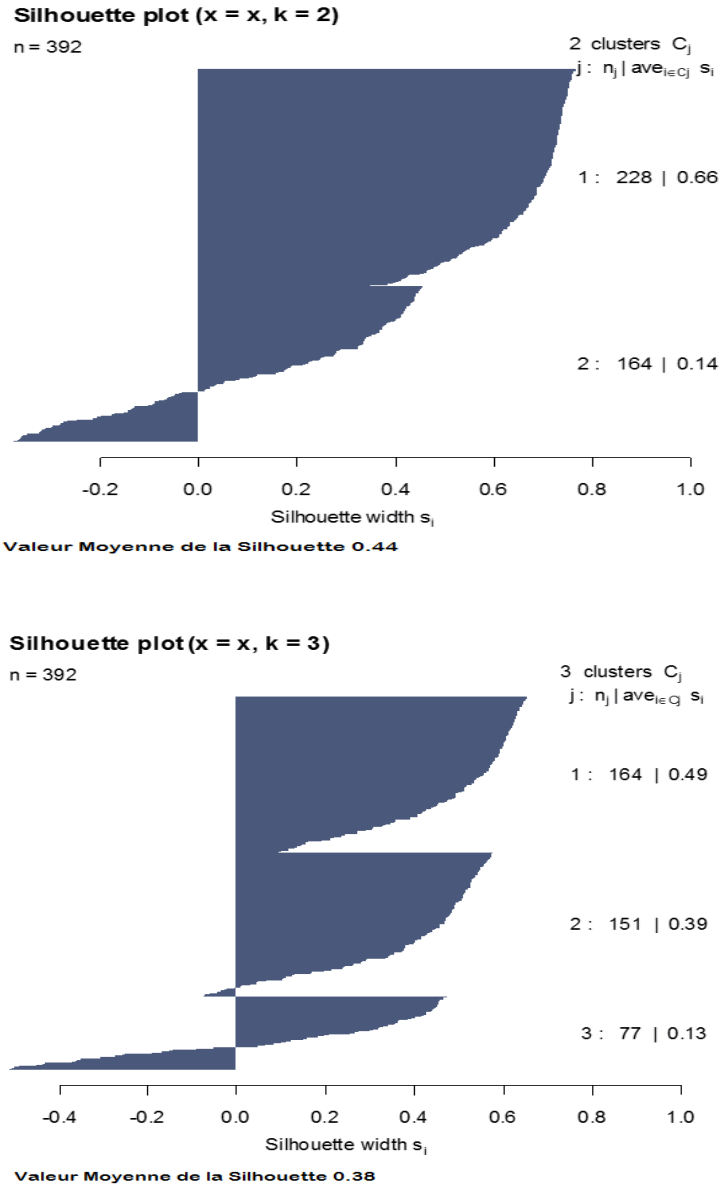

FIGURE 6 – Largeur maximale de la silhouette globale moyenne pour 2 et 3 centres

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 6), ce nombre de centres peut être considéré comme bien classé en observant les chevauchements dans les exemples de la base (figure 3).

6.3 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 à priori, ce qui n'est pas le cas du regroupement, alors nous pourrons dire que le clustring 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

Paramètres	*Centre1*	*Centre2*
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 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 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 8.

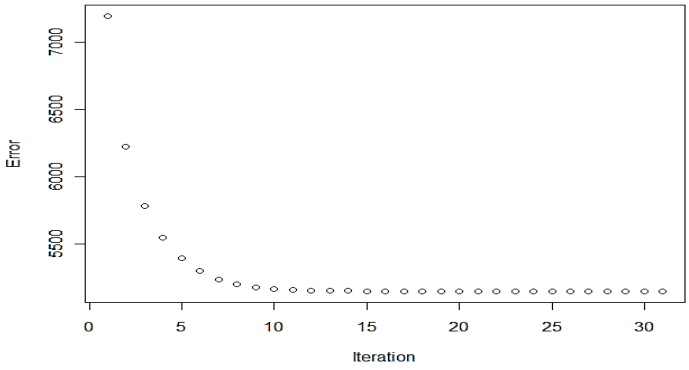

FIGURE 7 – Schéma de convergence de la fonction objective

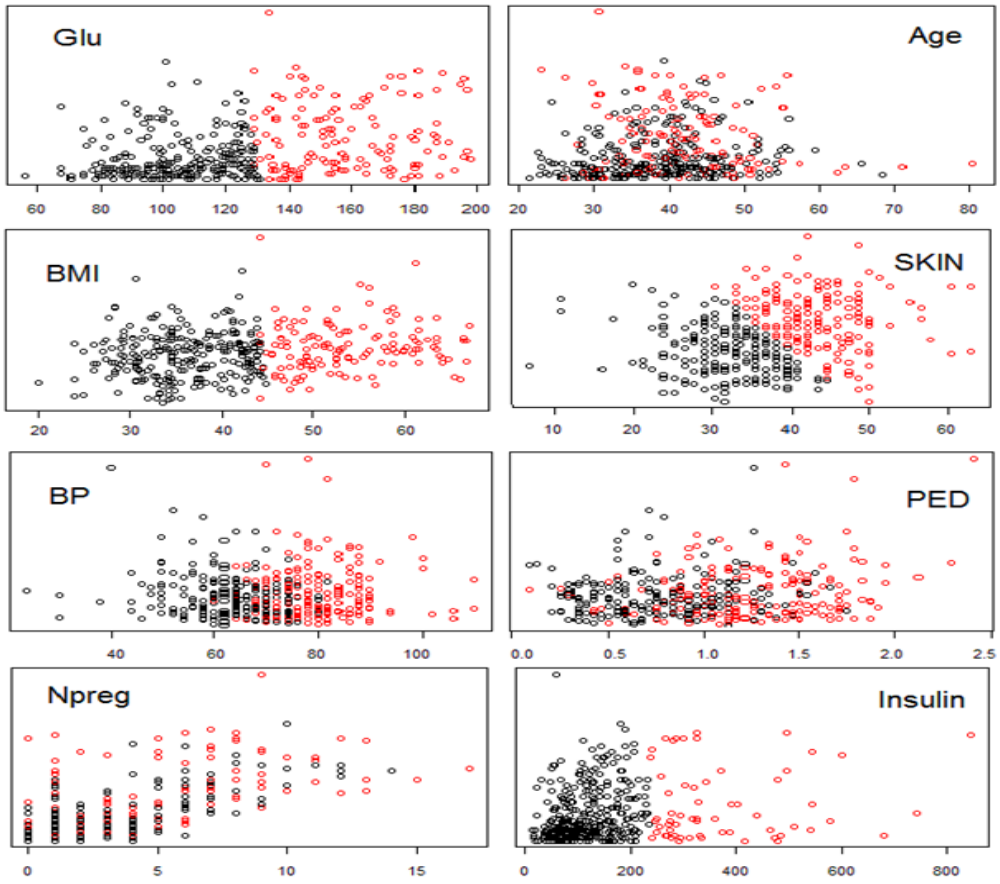

FIGURE 8 – Répartition des 8 paramètres pour c=2

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 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 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.

6.4 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 attirer 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.

7 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 améliora considérablement le temps d'apprentissage.

7.1 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 9.

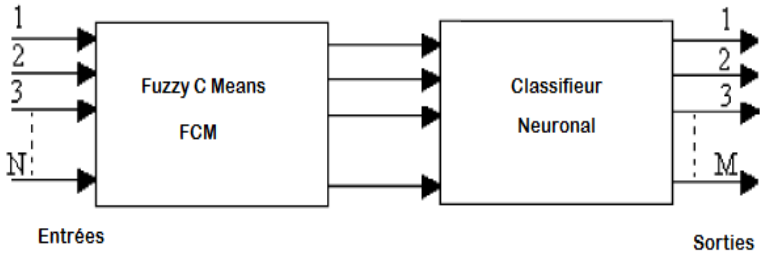

FIGURE 9 – Architecture de FCNN [OCK06]

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.

7.2 Réseau de neurones Mutlicouches RNMC

Dans cette partie, un réseau de neurones mutlicouchess 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 10 montre une structure générale de RNMC. L'apprentissage par rétro propagation fait appel a 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].

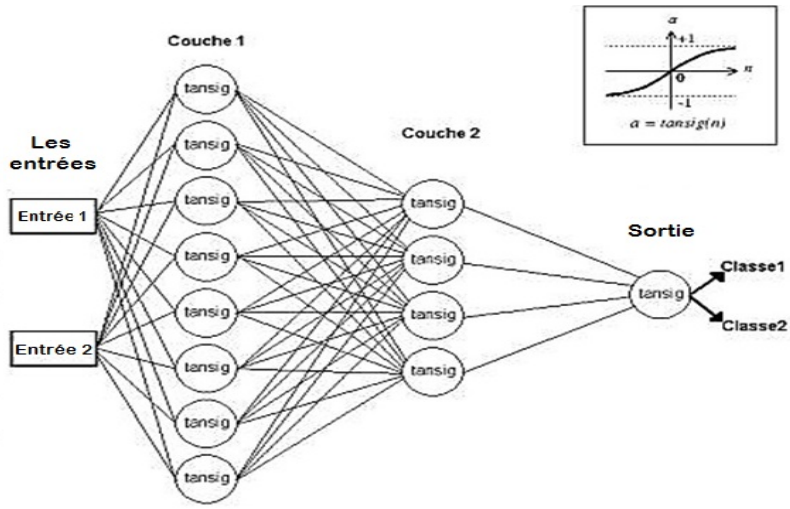

FIGURE 10 – Structure générale d'un réseau de neurones mutlicouches RNMC

7.2.1 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: **Presentation des entrees et la sortie desiree** :
Présenter le vecteur d'entrée $x(1), x(2), \ldots, x(N)$ et leurs correspondantes sorties désirées $d(1), d(2), \ldots, d(N)$, une paire à la fois, où N est le nombre d'exemple d'apprentissage.
3: **Calcul des sorties reelles** : calcule les sorties $y_1, y_2, \ldots, y_{NM}$

$$y_i = \phi\Big(\sum_{j=1}^{N_{M-1}} w_{ij}^{(M-1)} x_j^{(M-1)} + b_i^{(M-1)} \Big), \qquad i = 1, \ldots, N_{M-1}.$$

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

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

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 μ=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.

7.3 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.

7.3.1 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 392x2. 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 11 avec 2 :4 :1.

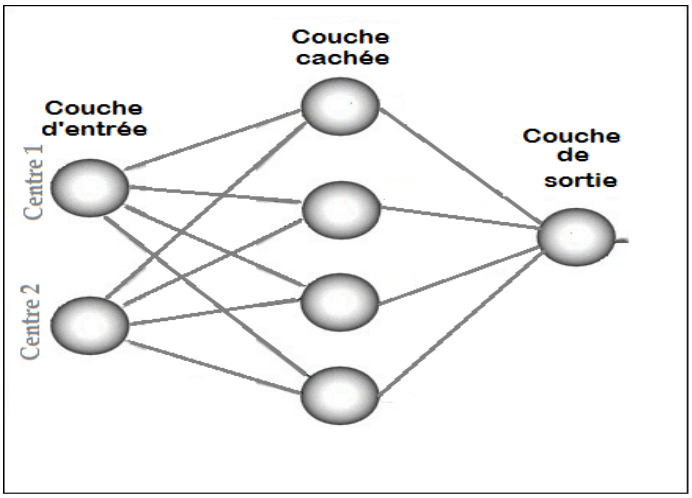

FIGURE 11 – L'architecture optimale du réseau de neurones obtenue

7.3.2 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.

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 – Performances des classifieurs RNMC et FCNN

Le perceptron mutlicouche 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.

7.4 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.

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

8.1 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.

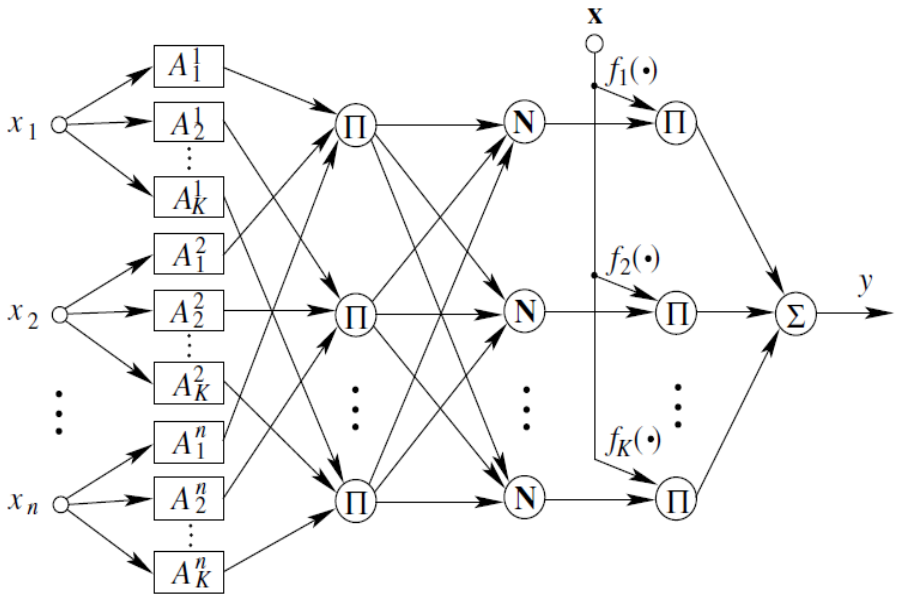

FIGURE 12 – Architecture d'ANFIS

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 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.

$R^{(r)}$: SI **x** est A^r

ALORS $y_r = f^{(r)}(x_1, x_2, \ldots, x_n)$.

Le mécanisme de raisonnement flou de TSK est résumé dans la figure 13 [JM96]. Les moyennes pondérées sont utilisées afin d'éviter les calculs complexes dans le processus de défuzzification.

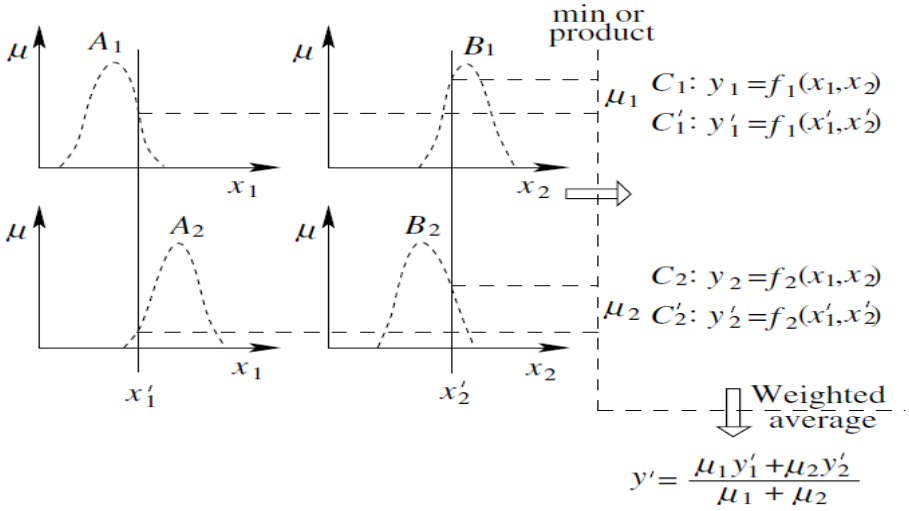

FIGURE 13 – Modèle flou de Sugeno du Premier ordre

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

1: **1ère couche** : Fuzzyfication
   $O_i^1 = \mu_{Ai}(X)$ *avec* : $i = 1, 2$
   $O_i^1 = \mu_{Bi-2}(Y)$ *avec* : $i = 3, 4$
2: **2ème couche** : Pondération des règles floues
   $O_i^2 = w_i = \mu_{Ai}(X)\mu_{Bi-2}(Y)$ $i = 1, 2$
3: **3ème couche** : Normalisation
   $O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}$ $i = 1, 2$
4: **4ème couche** : Défuzzyfication
   $O_i^4 = w_i f_i$ *avec* $f = a_i.X + b_i.Y + c_i$ *où* $i = 1, 2$
5: **5ème couche** : Calcul de la sortie
   $O_i^5 = \sum_{i=1}^{2} w_i f_i = \frac{\sum_{i=1} w_i f_i}{w_1 + w_2}$

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

8.2 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.

8.3 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 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].

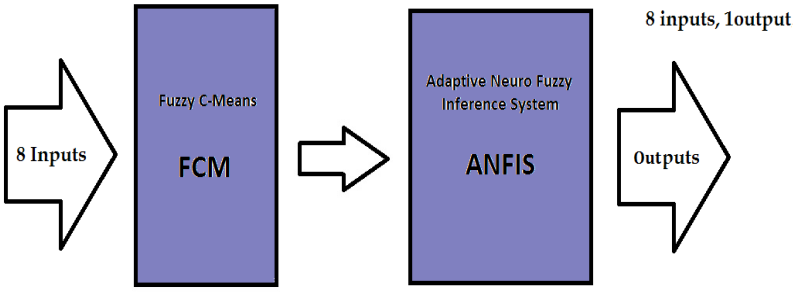

FIGURE 14 – L'approche réalisée FCM-ANFIS

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.

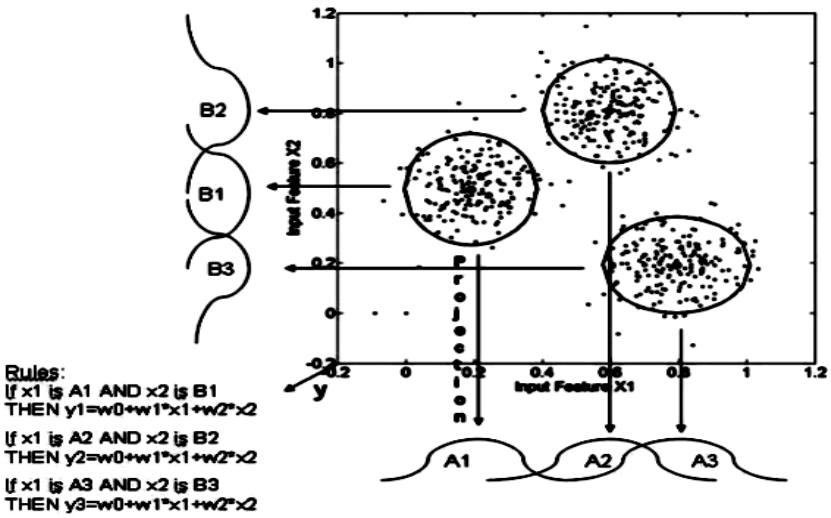

FIGURE 15 – Schéma représentatif de la répartition des clusters en fonctions d'appartenance avec FCM

On distingue dans la figure 16, la répartition obtenue sur deux centres (deux fonctions d'appartenance) pour les variables de la base de données Indiens Pima.

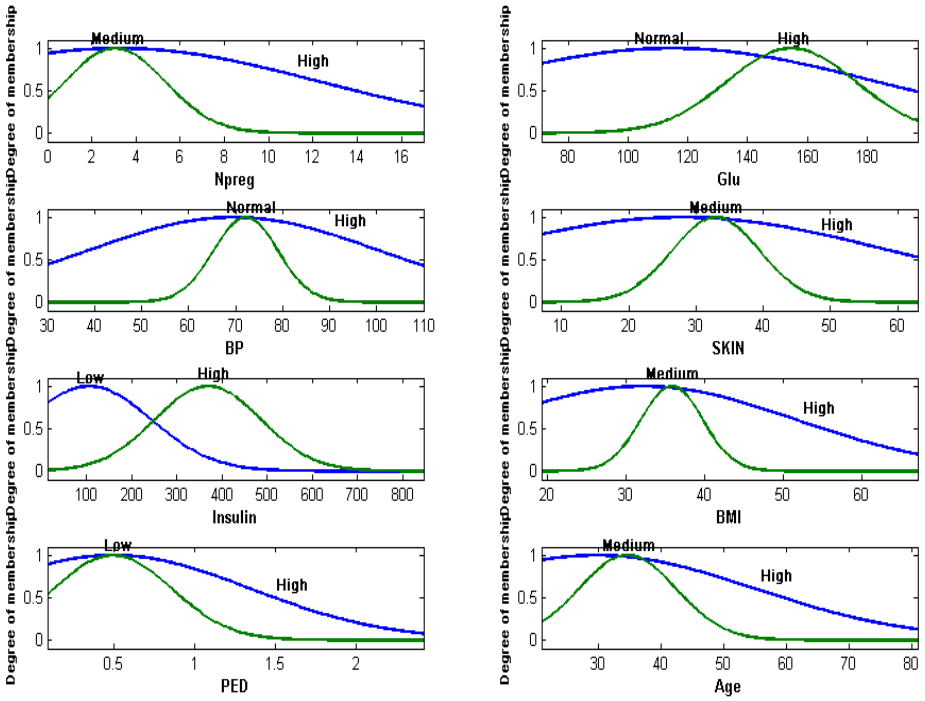

FIGURE 16 – Les fonctions d'appartenance de la base de données Pima avec FCM

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^{ere} 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 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).

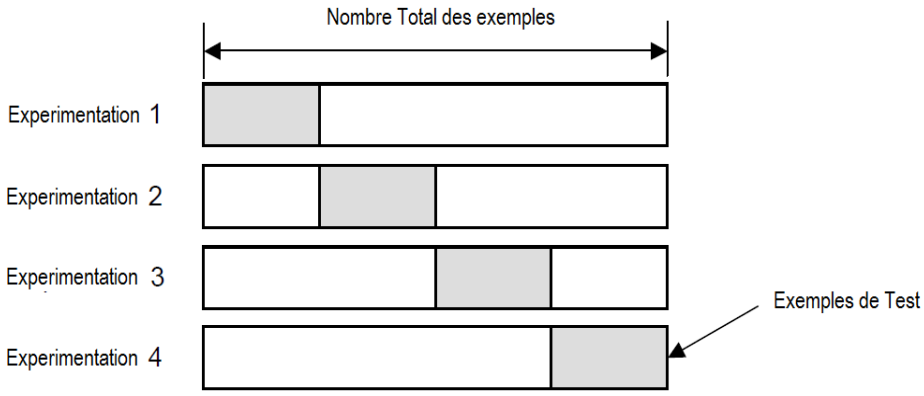

FIGURE 17 – Schéma de principe du K-fold cross validation

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 4 – Les points modaux finaux avec FCM-ANFIS

8.3.1 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* = $\frac{\text{Nombre de personnes qui activent cette regle a plus de 50\%}}{\text{Le nombre total des personnes de chaque classe (diabetiques et non diabetiques)}}$
- *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 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 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 5).

8.4 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)

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

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

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

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

Specificité

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

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 6)

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 6), tout en donnant un nombre réduit de cas mal classés (FP=42 et FN=21).

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 6 – Performances du classifieur FCM-ANFIS avec différentes entrées

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.

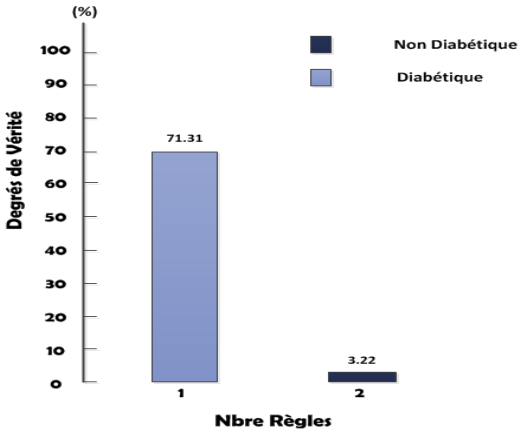

FIGURE.1 Les règles floues activées pour le patient N°107 correctement classé diabétique cas (TP)

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.

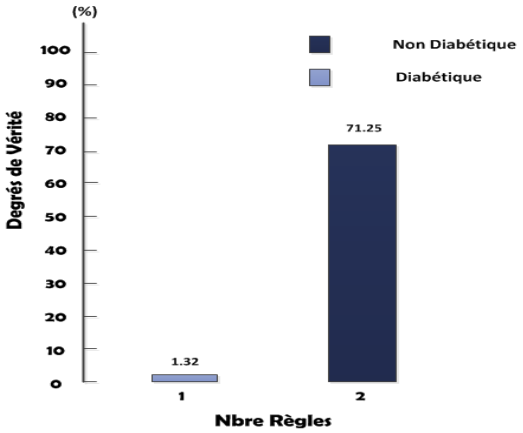

FIGURE.2 Les règles floues activées pour le patient N°28 correctement classé non diabétique cas (TN)

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.

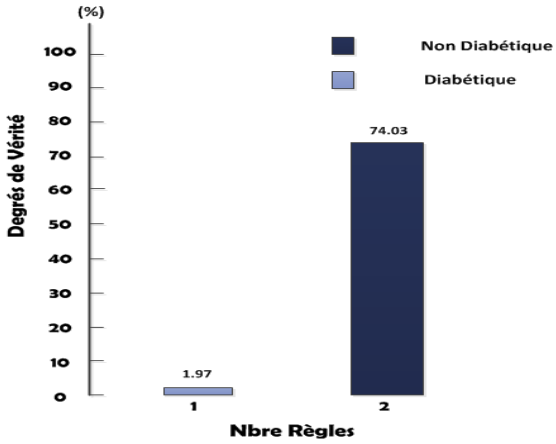

FIGURE.3 Les règles floues activées pour le patient N°102 (diabétique) classé non diabétique (FN)

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 hyperglycé-

mie. 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.

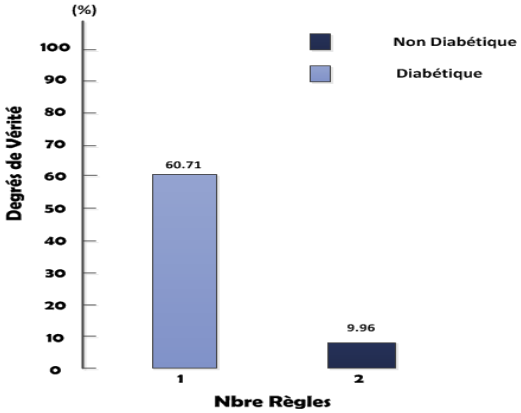

FIGURE.4 Les règles floues activées pour le patient N° 68 (non diabétique) classé diabétique (FP)

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.

8.5 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 performnces d'ANFIS.

8.5.1 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 $1^e r$ 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 pourrons 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 18), sachant que chaque centre est un point des données avec le plus grand potentiel.

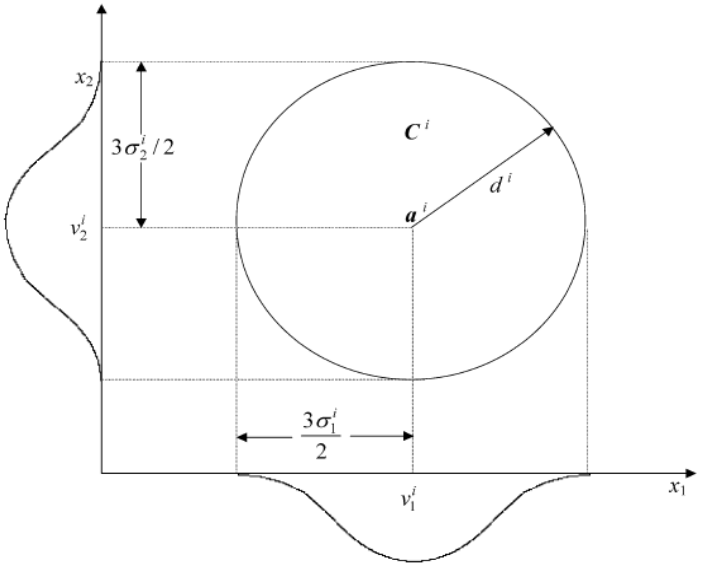

FIGURE 18 – Schéma représentatif de la répartition des clusters en des fonctions d'appartenance avec SC

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

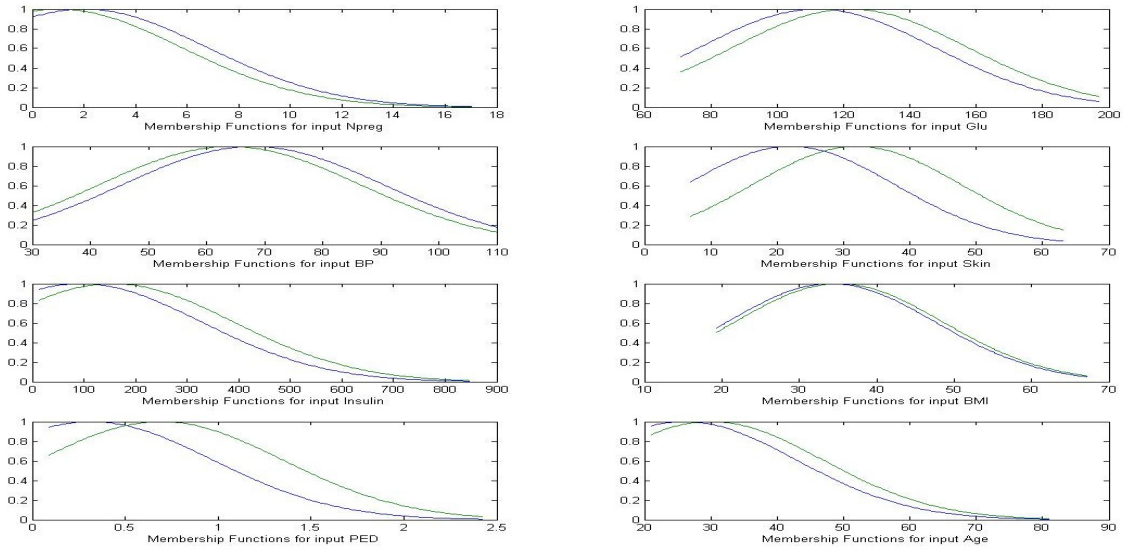

FIGURE 19 – Répartiton des fonctions d'appartenance avec Soustractive clustering avec r=0.8

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 19), ce qui relève d'une mauvaise classification des patients. En essayant de varier le rayon (figures 20-21-22-23) nous avons constaté une augmentation du nombre de fonctions d'appartenance, mais les données demeurent non réparties de manière distincte.

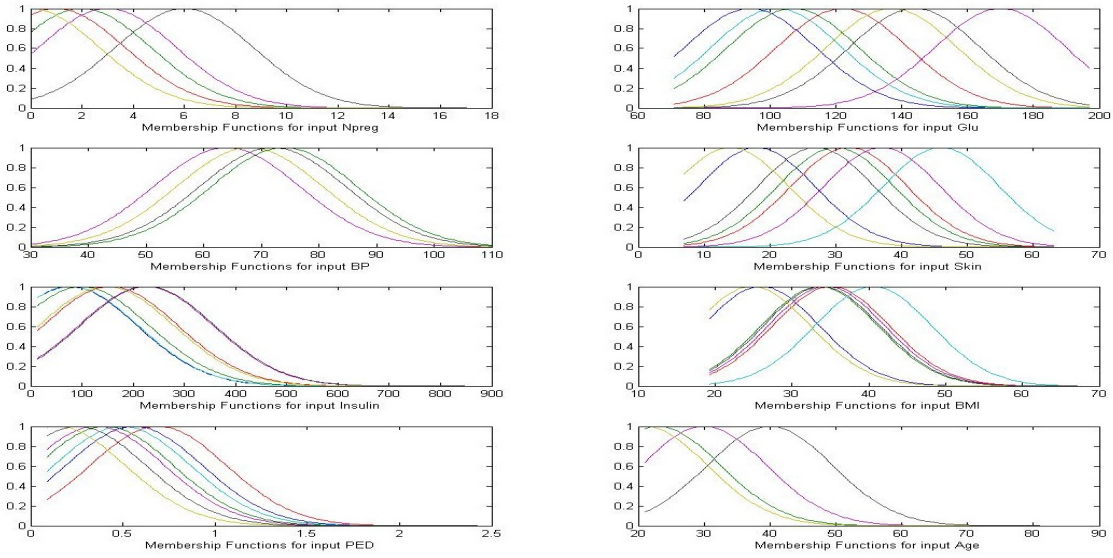

FIGURE 20 – Répartiton des fonctions d'appartenance avec Soustractive clustering avec r=0.45

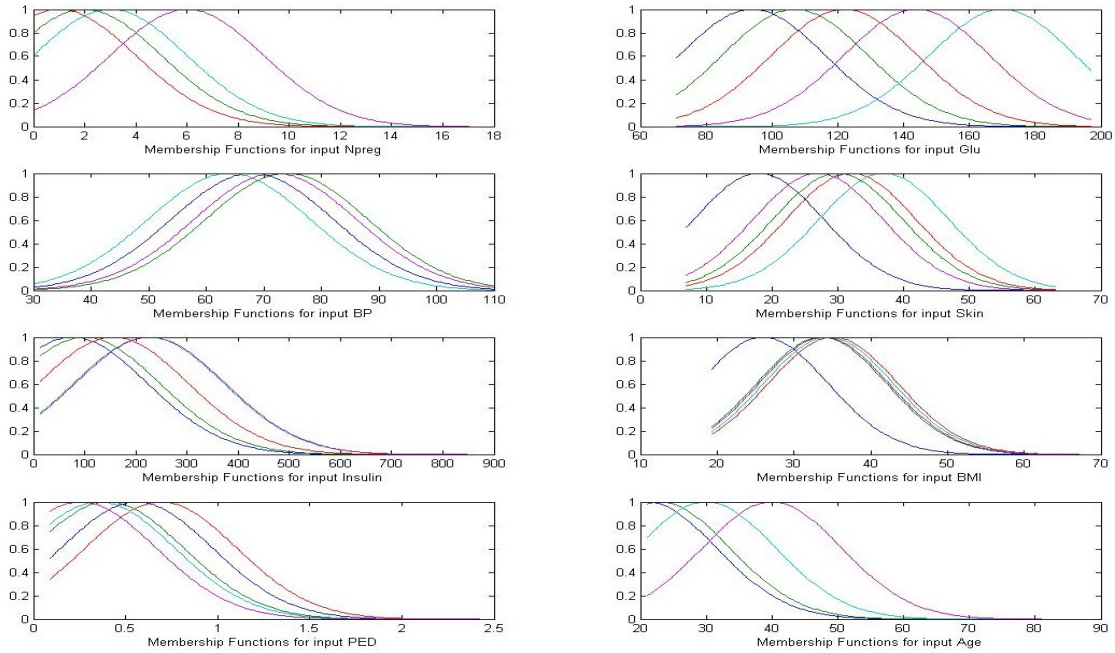

FIGURE 21 – Répartiton des fonctions d'appartenance avec Soustractive clustering avec r=0.5

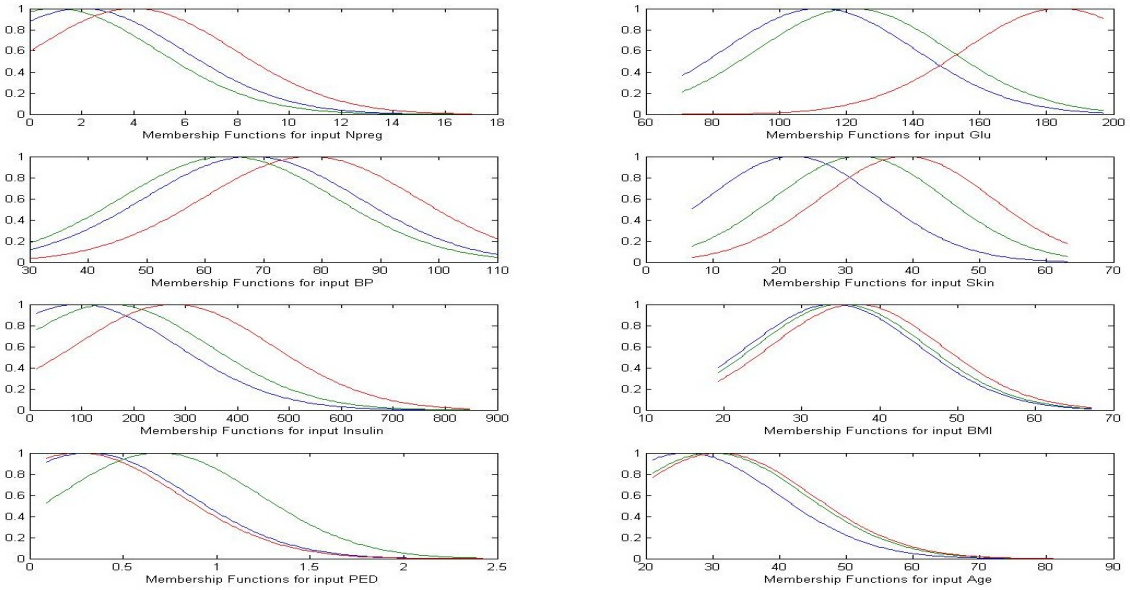

FIGURE 22 – Répartiton des fonctions d'appartenance avec Soustractive clustering avec r=0.65

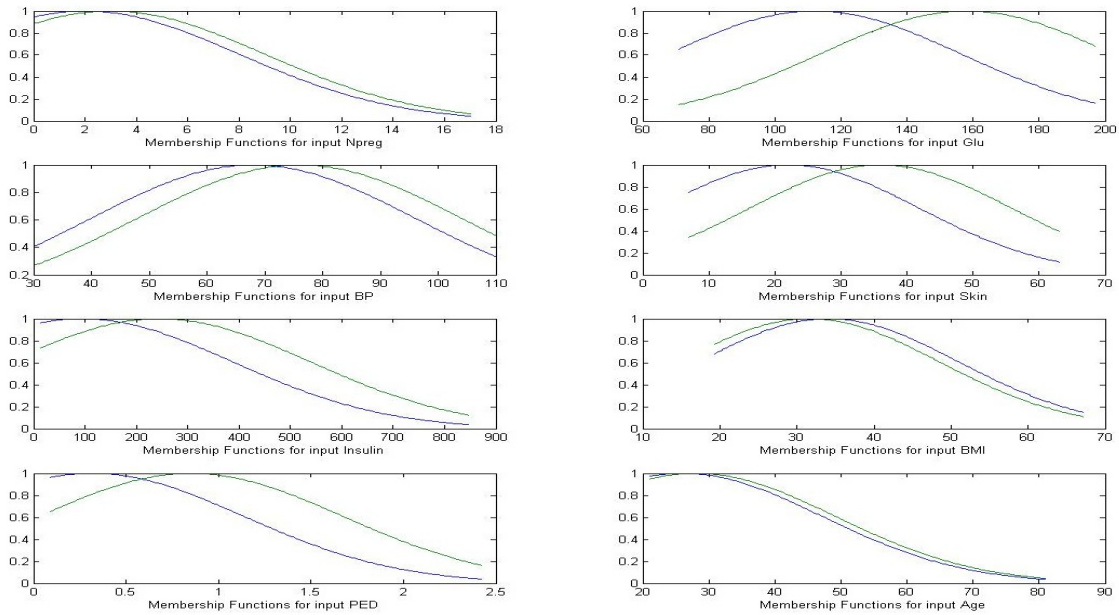

FIGURE 23 – Répartiton des fonctions d'appartenance avec Soustractive clustering avec r=1

D'après les figures 20-21-22-23, 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 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 7 – Performances du classifieur SCM-ANFIS avec différents rayons

8.5.2 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 permettera au clinicien d'obtenir des informations sur le Diagnostic Automatique.

8.6 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 crées 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 pourrons 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 mals 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.

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%.

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 essaims particulaires, Systèmes immunitaires artificiels, Algorithme des colonies de Fourmies, etc.

Références

[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 Sytems*, 90(2) :161–170, 1997.

[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, Mikhil R. Pal, James Keller, and Raghu Krisnapuram. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Kluwer Academic Publishers, Norwell, MA, USA, 1999.

[DADL10] Esin Dogantekin, Dogantekin Akif, Avci Derya, and Avci 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.

[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.

[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.

[FA10] 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.

[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.

[Gui01] S. Guillaume. Designing fuzzy inference systems from data : An interpretability-oriented review. *IEEE Transactions on Fuzzy Systems*, 9 :426–442., 2001.

[Hai98] S. Haikin. *Neural Networks : A Comprehensive Foundation*. Pearson Education, NY, 2nd edition, 1998.

[HBV02] Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. Clustering validity checking methods : part ii. *SIGMOD Rec.*, 31 :19–27, September 2002.

[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.

[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.

[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.

[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.

[JS95] J. S. R. Jang and C. T. Sun. Neuro-Fuzzy modeling and control. *Proceedings of The IEEE*, 83(3) :378–406, March 1995.

[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.

[KoTA03] B. Karlik, M. o. Tokhi, and M. Alci. A fuzzy clustering neural network architecture for multifunction upper-limb prosthesis. In *Biomedical Engineering*, 2003.

[KR90] L. Kaufman and P.J. Rousseeuw. *Finding Groups in Data An Introduction to Cluster Analysis*. Wiley Interscience, New York, 1990.

[NKK97] Detlef Nauck, Frank Klawonn, and Rudolf Kruse. *Foundations of Neuro-Fuzzy Systems*. John Wiley & Sons, Inc., New York, NY, USA, 1997.

[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.

[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.

[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.

[PY98] K. M. Passino and S. Yurkovich. *Fuzzy control*. Addison Wesley Longman, Inc., 1998.

[RTK95] P. J. Rousseeuw, E. Trauwaert, and L. Kaufman. Fuzzy clustering with high contrast. *J. Comput. Appl. Math.*, 64 :81–90, November 1995.

[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.

[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.

[SK88] M. Sugeno and G. T. Kang. Structure identification of fuzzy model. *Fuzzy Sets Syst.*, 28 :15–33, October 1988.

[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.

[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.

[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.

[XB91] Xuanli Lisa Xie and Gerardo Beni. A validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13 :841–847, August 1991.

[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.