

تحسين شبكة العنقدة كوهنين المضبية وتطبيقها على البيانات Iris وفي تجزئة الصور الطبية

الدكتور محمد أسعد *

تاريخ الإيداع 26 / 4 / 2017. قُبل للنشر في 27 / 8 / 2017

□ ملخص □

تم في هذا البحث اختبار شبكة كوهنين العصبية الصناعية المهجنة بالمنطق الضبابي بعد اقتراح علاقة جديدة لمعامل التضبيب m في إجراء عملية العنقدة. من خلال دراسة عمل هذه الخوارزمية تبين بأن العامل الأكثر تأثيراً في سلوك شبكة كوهنين المضبية هو معامل التضبيب m وذلك لأن تحديث معامل التضبيب في كل تكرار يتحكم بشكل ذاتي وبفعالية في توزيع معدل التعلم وتحديث أوزان الشبكة لجميع عقد البيانات. وتم اختبار شبكة العنقدة كوهنين المضبية المحسنة في هذا البحث على بيانات (Fisher's IRIS) [1] التي تعد بيانات قياسية و يتم اختبار مدى فعالية الخوارزميات الجديدة عليها وهي عبارة عن بيانات لثلاثة أصناف مختلفة من الأزهار. كما تم اختبار الخوارزمية المحسنة في مجال تجزئة الصور الطبية الممثلة بصور الرنين المغناطيسي للدماغ. لما لهذا التطبيق من أهمية كبيرة في المجال الطبي. لقد تم مقارنة نتائج الخوارزمية المحسنة في مجالي التطبيقين السابقين مع خوارزميتي العنقدة الشهيرتين K_means و (FuzzyC-Means, FCM) وشبكة العنقدة كوهنين المضبية. لقد أظهرت نتائج الخوارزمية المحسنة IFKCN أفضلية مقارنة مع الخوارزميات الثلاث السابقة من حيث عدد التكرارات ومراكز العناقيد ومقياس الخطأ (Mean Square Error, MSE).

الكلمات المفتاحية: شبكة العنقدة كوهنين المضبية، العنقدة، المنطق الضبابي، FCM، K_Means.

* مدرس - قسم الرياضيات/شعبة الرياضيات التطبيقية - كلية العلوم - جامعة تشرين (علوم حاسوب/ذكاء صناعي).

Improving Fuzzy Kohonen Clustering Network and Applying it on Iris Data and Medical Image Segmentation

Dr. Mohammad Assaad*

(Received 26 / 4 / 2017. Accepted 27 / 8 / 2017)

□ ABSTRACT □

In this research fuzzy kohonen clustering network has been tested after proposing a new relationship for the fuzzy parameter \mathbf{m} in a clustering process. By studying the work of this algorithm, it shows that the most influential factor in fuzzy kohonen network behavior was fuzzy parameter \mathbf{m} , This is because the update fuzzy parameter in each iteration automatically and effectively controls to distribution of learning rate and update the network weights to all data nodes at each. In this research the improved Fuzzy Kohonen Clustering Network (IFKCN) was tested on (Fisher's IRIS) data, which is the standard data are testing the effectiveness of new algorithms, which represent data for three different classes of flowers. The proposed algorithm has been tested in the field of medical image segmentation (magnetic resonance images of the brain) because of this application has great importance in the medical field. The proposed algorithm results were compared in the two areas of the former applications with two clustering famous K_means, (Fuzzy C-Means, FCM) and Fuzzy Kohonen Clustering Network (FKCN). The IFKCN algorithm results showed a good result comparing with the previous three algorithms in terms of the number of iterations and centers of clusters, mean square error (MSE).

Keywords: Fuzzy kohonen clustering network, Clustering, Fuzzy logic, FCM, K_Means.

*Assistant Professor, Mathematics Department/ Applied Mathematics Branch , Faculty of Science, Tishreen University (Computer Science / Artificial Intelligence).

مقدمة:

تعد شبكة كوهنين واحدة من الشبكات ذات التعلم الذاتي، أي التعلم دون معلم [2]، بمعنى أنه لا يوجد خرج محدد يعطى للشبكة كما في حالة التعلم مع معلم. ويمكن أن تعمم هذه الشبكة لتصميم شبكة كوهنين المضببة ويكون أداؤها أفضل من الشبكات العصبونية الصناعية الاعتيادية. حيث تتم عنقدة مدخلات الشبكة العصبونية باستخدام المسافة الاقليدية والتي تعتمد على أخذ أقل مسافة ناتجة بين كل دخل ومراكز العناقيد، و سينتج عن هذا عدة خلايا، حيث تعد الخلية الفائزة هي الخلية التي تحصل على أقل مسافة. ويعتمد تعلم شبكة كوهنين المضببة على دالة العضوية (membership function) للخلية الفائزة [3].

تمثل عملية العنقدة تقسيم للبيانات إلى مجموعة من الأصناف اعتماداً على اشتراكها بالخواص المتشابهة وهي تقسيم غير موجه للبيانات على عكس التصنيف الذي يستخدم بشكل واسع في حل الكثير من المشكلات من خلال تحليل مجموعة من البيانات ووضعها على شكل أصناف أو أقسام يمكن استخدامها فيما بعد لتصنيف البيانات المستقبلية باستخدام الطرائق التي تستخدم في تصنيف البيانات. أي أن العنقدة هي تنظيم البيانات في مجموعات أعضائها متشابهة إلى حد ما والعنقود هو مجموعة من البيانات المتشابهة فيما بينها وغير مشابهة للأعضاء المنتمية إلى عناقيد أخرى. فإذا كان لدينا مجموعة من العناقيد عندها يمكن تقسيم البيانات بينها حسب صفة التشابه الممثلة بالمسافة عندها نقول عن عنصرين ما إنهما ينتميان إلى نفس العنقود إذا كان أحدهما قريب من الآخر وفقاً لمسافة معينة. وتعد تقنية العنقدة مفيدة في عدة مجالات كالتجميع وصنع القرار ومعالجة الصور، والتتقيفي للبيانات وغيرها.

مشكلة البحث:

تعاني بعض خوارزميات العنقدة من مساوئ مختلفة فخوارزمية K_means تعاني من الاختيار العشوائي للمراكز الابتدائية و هذا ما يؤثر على استقرار الحل أثناء تنفيذ الخوارزمية [4]. كما تعاني خوارزمتي التضييب FCM و FKCN من إيجاد عملية فعالة في تحديد طريقة انتماء العنصر لأي عنقود [2]. لذلك تم العمل في هذا البحث على اقتراح علاقة لاختيار معامل التضييب والذي يعمل على تقليل تأثير القيم العضوية الصغيرة في كل تكرار والذي بدوره يعطي دقة عالية تقارب أسرع للحل مقارنة مع خوارزميات العنقدة الأساسية. وبالتالي يمكن صياغة المشكلة بالتساؤلات الآتية:

▪ هل من الممكن تحسين خوارزمية (FKCN) ومقارنتها مع خوارزميات العنقدة الأخرى

؟ (K_means FCM,)

▪ هل من الممكن التحسين في علاقة دالة العضوية بدلالة معامل التضييب m ؟

أهمية البحث وأهدافه :**أهمية البحث:**

الأهمية العلمية: في ظل ما نعيشه من تطور هائل في جميع المجالات العلمية والاقتصادية وغيرها ظهرت الحاجة لإيجاد طرائق غير تقليدية نتعامل من خلالها مع البيانات بجميع أنماطها (النصية و المرئية و الصوتية وغيرها) و هذه البيانات ذات حجوم كبيرة. فكان لابد من إيجاد طرائق جديدة لاستكشاف المعرفة والمعلومات التي تقع ضمن هذا الكم الهائل من البيانات. فظهرت تقنية التقيب في البيانات و التي تمثل عملية تفتيش وبحث عن معلومات معينة و مفيدة في حجم كبير من البيانات و يتم ذلك من خلال عملية الربط بين تحميل هذه البيانات وطرائق الذكاء الصناعي لتصبح أكثر كفاءة في عملية التقيب والتي تمثل خطوه من خطوات استكشاف المعرفة من قواعد البيانات. تعد خوارزميات العنقدة K_means و FCM و FKCN من أشهر خوارزميات التقيب في البيانات والتي تهدف إلى تجميع العناصر

المتماثلة مع بعضها البعض في عناقيد اعتماداً على خصائصها. حيث تمر البيانات بعدة مراحل (معالجات مختلفة) من أجل التوصل في النهاية إلى الاستكشاف الأمثل والأفضل للمعلومات.

الأهمية العملية: من الممكن تطبيق الخوارزمية المحسنة في مجالات عدة كضغط الصور وفي استرجاع البيانات وفي معالجة الصور وغيرها. في هذا البحث تم تطبيق هذه الخوارزمية في مجال تجزئة الصور الطبية (صور الرنين المغناطيسي للدماغ) والذي يساعد في اكتشاف واقتطاع المناطق المصابة في الدماغ لما لهذا التطبيق من أهمية كبيرة في أيامنا الحالية في معظم مشافي القطر.

أهداف البحث:

- 1- يهدف البحث إلى تحسين أداء خوارزمية شبكة العنقدة كوهنين المضبية FKCN من خلال الخطوات الآتية:
- 2- اختبار شبكة كوهنين العصبية الصناعية المهجنة بالمنطق الضبابي.
- 3- دراسة عمل شبكة كوهنين وبيان أثر معامل التضبيب m.
- 4- اقتراح علاقة جديدة لتحديد قيمة معامل التضبيب وذلك لأن تحديث معامل التضبيب في كل تكرار يتحكم بشكل ذاتي في توزيع معدل التعلم وتحديث أوزان الشبكة لجميع عقد البيانات وهذا ما يساعد في إيجاد طريقة أفضل لتحديد طريقة انتماء العنصر للعنقود.
- 5- تقسيم البيانات المدروسة إلى مجموعة من العناقيد، كل عنقود يحوي عناصر متشابهة حيث يتم تعريف العنقود من خلال تحليل البيانات والتي يمكن أن نؤديها على العناقيد كمجموعات .

فرضيات البحث: تنطلق الدراسة من الفرضيات الآتية:

- يوجد فروق جوهرية بمقياس الخطأ (MSE) وعدد التكرارات بين خوارزميتي (K_means) وخوارزمية (FCM) على البيانات المطبقة.
- يوجد فروق جوهرية بمقياس الخطأ (MSE) وعدد التكرارات بين خوارزميتي (FCM) وخوارزمية (FKCN) على البيانات المطبقة.
- يوجد فروق جوهرية بمقياس الخطأ (MSE) وعدد التكرارات بين خوارزمية (FKCN) وخوارزمية (IFKCN) المحسنة.
- لا تحتاج خوارزمية (IFKCN) المحسنة إلى زمن كبير مقارنة مع بقية الخوارزميات المستخدمة.

منهجية البحث:

تم إتباع المنهج التحليلي و التجريبي في البحث حيث تم وضع دراسة تحليلية للمواضيع المتعلقة بالبحث بالاعتماد على المراجع والدوريات العلمية والمقالات. كما تم تحليل البيانات المأخوذة والوصول إلى النتائج من خلال استخدام نموذج مقترح معتمد على الشبكات العصبية الصناعية المضبية وبرنامج MATLAB، الحدود الزمانية للعام الدراسي -2016 2017 .

وفيما يلي شرح مفصل لآلية عمل كل خوارزمية من الخوارزميات المذكورة سابقاً:

1- خوارزمية K-means: [4]

هي الخوارزمية الأكثر استخداماً وذلك لبساطتها وتعد طريقة تصنيف غير مرشدة حيث تعطي مجموعة من القيم وتحاول أن تجزئها إلى مجموعة من العناقيد K، وتستعمل خوارزمية تكرارية تقلل مجموع المسافات المربعة من جميع النقاط في حيز العنقود، وتتلخص الخوارزمية بالخطوات التالية:

الخطوة الأولى: اختيار k مركز عنقود كقيم ابتدائية وبشكل عشوائي C_1, C_2, \dots, C_k .

الخطوة الثانية: توزيع كل نموذج $x \in S$ إلى العنقود $C_i (1 \leq i \leq k)$ الذي مركزه قريب من النموذج x باستخدام المسافة الإقليدية وحسب المعادلة التالية:

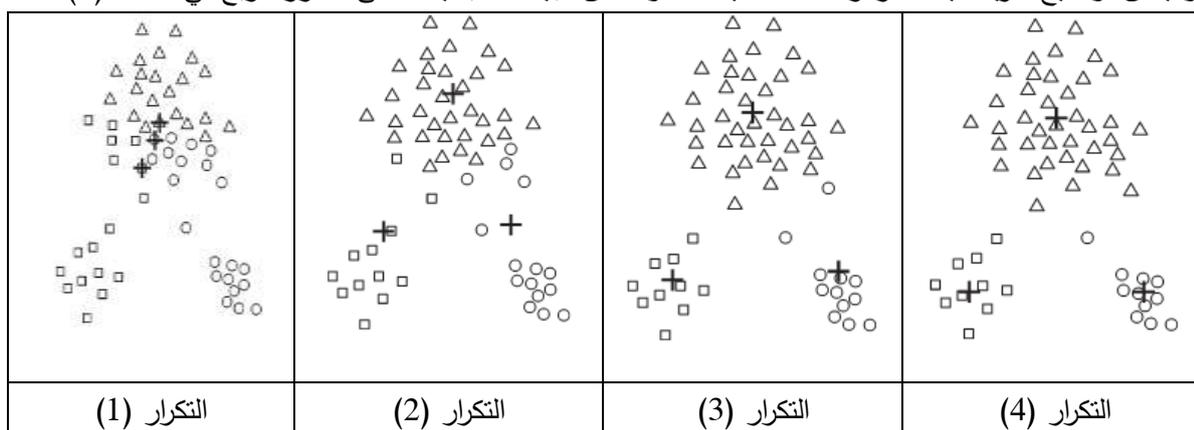
$$x \in S_j(t) \text{ if } \|x - C_i(t)\| < \|x - C_j(t)\| \text{ for all } i = 1, 2, \dots, k ; i \neq j \quad (1)$$

حيث تمثل S_j مجموعة النماذج (البيانات) التي تنتمي للعنقود j .

الخطوة الثالثة: يتم حساب مراكز العناقيد الجديدة أي حساب المركز لكل عنقود $C_j (1 \leq j \leq k)$ على اعتبار أنه ممكن حصول إضافة أو حذف للنماذج بكل عنقود حسب العلاقة التالية:

$$C_j(k+1) = \frac{1}{S_j} \sum_{x \in C_j(k)} x \quad (2)$$

و يمكن توضيح طريقة إيجاد مراكز ثلاثة عنقايد لمجموعة من البيانات البسيطة حتى التكرار الرابع في الشكل (1):



الشكل (1): مثال يوضح إيجاد مراكز ثلاثة عنقايد باستخدام خوارزمية K-means

الخطوة الرابعة: العودة إلى التكرار بدءاً من الخطوة الثانية إلى أن يتحقق معيار الاقتراب من الحل و المتمثل بـ (عدم تغيير النموذج للعنقود الذي ينتمي له، أو تناقص المسافة الإقليدية، أو الوصول إلى القيمة العظمى للتكرارات). وتتغير هذه الخوارزمية تبعاً للآتي:

أ- عدد العناقيد ب- اختيار القيم الابتدائية لمراكز العناقيد ج- خصائص البيانات.

2- خوارزمية FCM: [5]

طُرحت فكرة المجموعات الضبابية والمنطق الضبابي من قبل العالم الأذربيجاني (لطي زاده) حيث طور أسلوباً لغوياً للتعامل مع المعلومات اللغوية الغامضة المبنية على أساس المجموعات الضبابية والمنطق الضبابي [6]. إن المبدأ الأساسي للعنقدة المضببة هو أن يكون لكل عنصر درجة عضوية جزئية في العناقيد المضببة ويتم اختيار قيمة

درجات العضوية ضمن المجال [0,1] لكل عينة في كل عنقود cluster ويكون الاختيار مبني على قياسات معينة، ويتم تحديد انتماء النموذج بدرجة عضوية إلى كل عنقود.

وتعد خوارزمية الـ FCM من خوارزميات العنقدة المضببة الأكثر شيوعاً وتُعرف على أنها تقنية عنقدة شائعة الاستخدام تمثل المبدأ الأساسي لخوارزمية means_K، حيث تسمح هذه الخوارزمية بتقسيم البيانات اعتماداً على نظرية المجموعات المضببة. وكما في خوارزميات K_means تكون خوارزمية FCM مبنية على تصغير الدالة القياسية .

يتم اختيار الدالة القياسية في خوارزمية FCM بإضافة قيم العضوية ومعامل التضييب m إلى دالة K_means القياسية، على الشكل التالي:

$$J(u, v) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m |x_k - v_i|^2 \quad (3)$$

حيث :

- x_1, x_2, \dots, x_n تمثل نقاط العينة .
- $V = \{v_1, v_2, \dots, v_c\}$ مراكز العناقيد .
- $U = \{u_{ik}\}$ هي مصفوفة من القياس $c \times n$ ، حيث u_{ik} يمثل القيمة العضوية للعنصر x_k ذي التسلسل k^{th} ضمن العنقود i^{th} ، ويمثل m معامل التضييب (العامل الأسي للوزن) ويكون $m > 1$. يقوم معامل التضييب على تقليل تأثير قيم العضوية الصغيرة. و لا يوجد قاعدة ثابتة لاختيار هذا المعامل ولكن كخيار شائع للاستخدام يأخذ القيمة $m=2$. وتُحدّث هذه الخوارزمية مراكز العناقيد بشكل تكراري.

تتلخص خوارزمية FCM بالخطوات التالية:

الخطوة الأولى: اختيار مراكز عنقايد ابتدائية ممثلة بـ $V^{(0)}$ عشوائياً بعد تحديد عدد مراكز العناقيد، ووضع قيم ابتدائية $U^{(0)}$ لدالة العضوية. اختيار m قيمة معامل التضييب، وكذلك إسناد قيمة ابتدائية مقدارها العدد خطوات التنفيذ $t = 1$.

الخطوة الثانية: حساب مراكز العناقيد $V^{(t)}$ بالاعتماد على قيم دالة العضوية $U^{(t)}$ حسب المعادلة (4):

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m} \quad i = 1, 2, 3, \dots, c \quad (4)$$

الخطوة الثالثة: تحديث قيم دالة العضوية للحصول على $U^{(t+1)}$ بالاعتماد على قيمة $V^{(t)}$ حسب المعادلة التالية:

$$u_{ik} = \frac{\left[\frac{1}{(x_k - v_i)^2} \right]^{\frac{1}{m-1}}}{\sum_{j=1}^c \left[\frac{1}{(x_k - v_j)^2} \right]^{\frac{1}{m-1}}}, \quad i = 1, 2, \dots, c; k = 1, 2, \dots, n \quad (5)$$

الخطوة الرابعة: تتوقف الخوارزمية عندما يتحقق الشرط $|u_{ik}^{(t+1)} - u_{ik}^{(t)}| \leq \varepsilon$ حيث ε عدد صغير موجب و إلا زيادة عدد الخطوات $t = t + 1$ والعودة إلى التكرار بدءاً من الخطوة الثانية.

3- شبكة العنقدة كوهنين المضببة: [2]

Fuzzy Kohonen Clustering Network (FKCN)

شبكة كوهنين هي واحدة من الشبكات ذات التعلم دون معلم أي من نوع التعلم الذاتي بمعنى أنه ليس معلوم مسبقاً مخرجات الشبكة كما في حالة التعلم عن طريق معلم. ويمكن أن تعمم هذه الشبكة لتصميم شبكة كوهنين المضببة ويكون أدائها أفضل من الشبكات الاعتيادية.

آلية عمل هذه الشبكة :

تتم عنقدة النقاط المدخلة باستخدام المسافة الإقليدية حيث يتم أخذ أقل مسافة ناتجة بين كل دخل ومركز العنقود وبذلك سينتج لدينا عدة خلايا، وتعد الخلية التي تحوي على أقل مسافة هي الخلية الفائزة. إن تعلم شبكة كوهنين المضببة يعتمد على دالة العضوية (membership function) للخلية الفائزة.

من خلال دراسة العوامل المؤثرة في سلوك هذه الشبكة نلاحظ إن معامل التضبيب m هو العامل الأكثر تأثيراً في سلوك شبكة كوهنين المضببة وذلك لأن تحديث معامل التضبيب في كل تكرار يتحكم بشكل ذاتي في توزيع معدل التعلم وتحديث أوزان الشبكة لجميع عقد البيانات بفعالية دون الاعتماد على أقل مسافة ناتجة كما في شبكة كوهنين التقليدية، وذلك لأن معدل التعلم يرتبط بمعامل التضبيب وفق علاقة تكون قيمها بين $[0,1]$ بترتيب تصاعدي في كل تكرار.

خوارزمية شبكة العنقدة كوهنين المضببة FKCN: [2]

الخطوة الأولى: تكون أوزان عشوائية بدائية لخلايا الشبكة $W^{(0)}$ بعد تحديد عدد مراكز العناقيد c .
ثم اختيار قيمة ابتدائية لمعامل التضبيب $m_0 \in [0,1]$ وكذلك وضع قيمة ابتدائية مقدارها 1 لعدد خطوات التنفيذ $t=1$.

الخطوة الثانية: حساب المسافة الإقليدية ما بين المدخلات x_k والأوزان w_i المرتبطة بالخلية وفق العلاقة التالية:

$$d_i = \sqrt{\sum_{i=0}^{n-1} (x_i - w_i)^2} \quad (6)$$

الخطوة الثالثة: حساب دالة العضوية u_{ij} للخلية الفائزة بالاعتماد على مقياس المسافة d_i ، ثم حساب معدل تعلم الشبكة α_{ij} كالتالي:

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|x_i - w_i\|}{\|x_j - w_k\|} \right)^{\frac{1}{m_t - 1}}}; \quad \alpha_{ij} = (\alpha_{ij})^{m_t} \quad (7)$$

حيث تُعطى علاقة معامل التضبيب على الشكل التالي وفق العلاقة (8):

$$m_t = m_0 - t * \left(\frac{m_t - 1}{t_{max}} \right), \quad t = 1, 2, \dots, t_{max} \quad (8)$$

الخطوة الرابعة: تحديث أشعة الأوزان وفق العلاقة (9):

$$w_i(t) = w_i(t-1) + \frac{\sum_{j=1}^n \alpha_{ij}(x_j - w_i)}{\sum_{s=1}^n \alpha_{js}}, \quad i = 1, 2, \dots, c \quad (9)$$

الخطوة الخامسة: حساب تابع الخطأ:

$$E(t) = \|V(t) - V(t-1)\| \quad (10)$$

الخطوة السادسة: إذا كان $t > t_{max} + 1$ أو $E(t) < \epsilon$ يتوقف التكرار وإلا العودة للخطوة الثانية، حيث ϵ عدد صغير موجب.

وبعد توضيح أداء عمل شبكة العنقدة كوهنين المضببة سيتم فيما يلي شرح الخطوات التي تم اقتراحها كتحسين على هذه الخوارزمية.

خوارزمية شبكة العنقدة كوهنين المضببة المحسنة :**Improved Fuzzy Kohonen Clustering Network Algorithm (IFKCN)**

لقد تم تحسين خوارزمية شبكة العنقدة كوهنين المضببة باقتراح علاقة جديدة لمعامل التضبيب m من خلال دراسة علاقة معامل التضبيب في الخطوة الثالثة من الخوارزمية والتي تمت بناءً على دراسة رياضية تُثبت مدى فعالية هذا

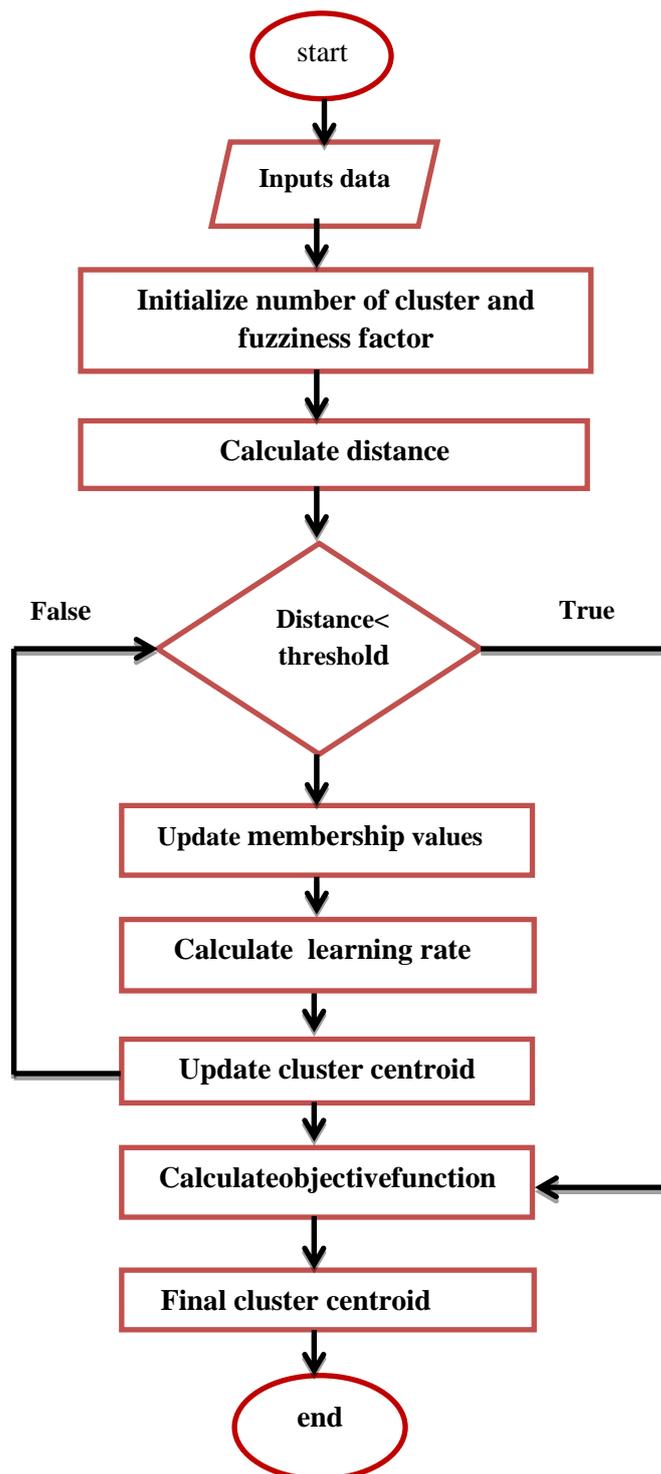
الاقتراح المتمثل بدالة تجيب الزاوية (\cos) التي تكون بدلالة m_0 وعدد التكرارات t . إن قيم هذه الدالة تقع ضمن المجال $[-1,1]$. و عند أخذ قيمة ابتدائية لمعامل التضبيب $m_0 = 0.5$ عندها يكون مجال قيم العلاقة المحسنة لمعامل التضبيب ضمن المجال $[0.5,1]$ [6]، فكانت العلاقة على الشكل التالي:

$$m_t = \cos\left(\frac{m_0}{t^2}\right), t = 1, 2, \dots, t_{max} \quad (11)$$

لقد أعطت هذه العلاقة الجديدة نتائج أفضل من حيث عدد التكرارات وقيم لدالة العضوية الممثلة بالقيمة (1) للانتماء التام للبيانات إلى العناقيد والقيمة (0) لعدم الانتماء. وتكون خوارزمية FKCN أيضاً مبنية على تصغير الدالة القياسية الموضحة في خوارزمية FCM:

$$J(u, v) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m |x_k - v_i|^2 \quad (12)$$

ويوضح الشكل (2) مخطط خطوات عمل الخوارزمية المحسنة:



الشكل (2): مخطط يوضح خطوات خوارزمية IFKCN

لقد تم اختبار أداء العلاقة المقترحة الممثلة بمعامل التضييب m على خوارزمية الـ FKCN و IFKCN، حيث نوضح هذه النتائج في الجدولين (1) و (2) هذه النتائج :

الجدول (1): يوضح قيم معامل التضبيب m لخوارزمية FKCN حتى التكرار $t_{max} = 20$ بقيمة ابتدائية $m_0 = 0.5$

t	1	2	3	4	5	6	7	8	9	10
m	0.5	0.505	0.515	0.530	0.548	0.571	0.597	0.625	0.655	0.686

t	11	12	13	14	15	16	17	18	19	20
m	0.717	0.748	0.779	0.807	0.834	0.859	0.882	0.902	0.920	0.935

الجدول (2): يوضح قيم معامل التضبيب m لخوارزمية IFKCN المحسنة حتى التكرار $t_{max} = 20$ بقيمة ابتدائية $m_0 = 0.5$

t	1	2	3	4	5	6	7	8	9	10
m	0.5	0.878	0.901	0.955	0.972	0.981	0.987	0.990	0.992	0.994

t	11	12	13	14	15	16	17	18	19	20
m	0.995	0.996	0.997	0.997	0.998	0.998	0.998	0.998	0.999	0.999

بعد مقارنة النتائج بين علاقة معامل التضبيب المحسنة وعلاقة معامل التضبيب الأساسية في الجدولين (1) و (2) تبين أن قيم معامل التضبيب المقترح تتقارب من الواحد بعدد تكرارات أقل من عدد التكرارات في العلاقة الأساسية.

الجانب التطبيقي والعملي :

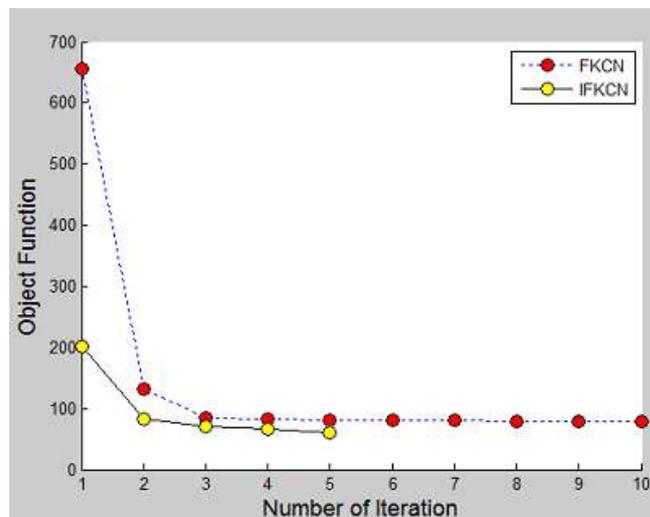
تم تنفيذ خوارزمية شبكة كوهنين العصبية الصناعية المضطربة باستخدام لغة الماتلاب (MATLAB R2013a) بالإصدار (8.1.0.0.604) وتعد هذه اللغة من اللغات الحديثة حيث أنها تستخدم في المجالات التقنية لما تتمتع به من خصائص ومزايا وتوفر للمستخدم سهولة التعامل مع البيانات ومعالجتها وتضم عدداً كبيراً من الدوال والمكتبات الجاهزة حسب التطبيق المستخدم وقد تم اختبار الخوارزمية المحسنة على المسألتين ([1][9]):

المسألة الأولى: على بيانات (Fisher's IRIS) [1] والتي تمثل بيانات قياسية يتم استخدامها لاختبار مدى فعالية الخوارزميات الجديدة المحسنة. حيث تتألف هذه البيانات من ثلاثة أصناف مختلفة من الأزهار وكلزهرة تقاس بواسطة طول الورقة وعرضها وطول الساق وعرضه، وتتألف هذه البيانات من 600 عينة اختبار ويمكن أن توسع لتصبح 6000 عينة اختبار. حيث قام كلاً من (Hathway and Bezdek 1995) بإيجاد المراكز الفعلية لهذه البيانات [1] والتي تمثل مراكز العناقيد الثلاثة، حيث أن كل عنقود يمثل بيانات زهرة. في الجدول (3) تم وضع مقارنة المراكز الفعلية لقاعدة البيانات (Fisher's IRIS) مع المراكز التي تم إيجادها بواسطة خوارزمتنا المحسنة. و تم مقارنة النتائج بين خوارزميات العنقدة الشهيرة مع الخوارزمية IFKCN المحسنة وبنفس الوسطاء اعتماداً على المراجع ([1] [4] [5] [8]) والنتائج التي تم الحصول عليها كانت باستخدام برنامج الماتلاب.

الجدول (3) : مقارنة النتائج بين خوارزميات العنقدة الشهيرة مع الخوارزمية IFKCN المحسنة وبنفس الوسيط

الخوارزمية المقترحة IFKCN	FKCN [8]	FCM[5]	K_means[4]	
$C=3 ; t_{max}=50$ $m=0.5 ; \varepsilon=0.001$	$C=3 ; t_{max}=50$ $m=3 ; \varepsilon=0.001$	$C=3 ; t_{max}=50$ $m=3 ; \varepsilon=0.001$	$c=3$	الوسطاء
5	10	23	8	عدد التكرارات
5.006 5.9366.588 3.4282.7702.974 1.4624.260 5.552 0.2461.3262.026				مراكز العناقيد الفعلية [1]
5.0075.8186.623 3.3992.7462.975 1.4944.1955.485 0.2601.3091.967	5.04 5.976.61 3.37 2.793.00 1.604.45 5.41 0.301.44 1.94	5.003 5.889 6.775 3.403 2.761 3.052 1.485 4.364 5.646 0.2511.397 2.053	5.00 5.906.85 3.41 2.743.07 1.46 4.39 5.74 0.24 1.43 2.07	مراكز العناقيد
0.0027	0.0088	0.0059	0.0123	مقياس الخطأ (MSE)

يبين الجدول (3) مقارنة بين الخوارزمية المقترحة وخوارزميات FCM و K_means الشهيرتان وخوارزمية FKCN. وقد تم استخدام نفس الوسيطاء في عملية المقارنة. حيث تم إيجاد مراكز العناقيد باستخدام هذه الخوارزميات و مقارنتها مع مراكز العناقيد الفعلية [1]. وتبين النتائج أن الخوارزمية المقترحة تعطي عدد أقل من التكرارات مقارنة مع الخوارزميات الأخرى وهذا يقلل من وقت التجميع. وتم أيضاً حساب مقياس الخطأ (MSE) لهذه الخوارزميات عن طريق إيجاد مجموع مربعات الفروق لمراكز العناقيد لكل خوارزمية مع المراكز الفعلية [1] وقد تبين أن مقياس الخطأ في الخوارزمية المقترحة أصغر من مقياس الخطأ للخوارزميات الأخرى، وهذا ما يعكس أفضلية للخوارزمية المقترحة.

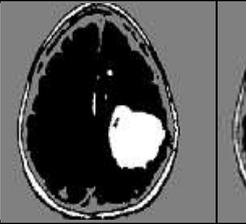
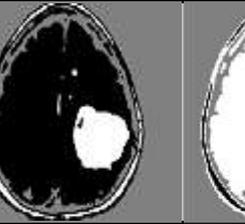
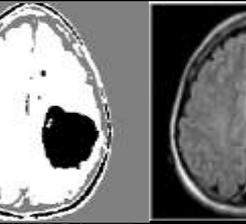
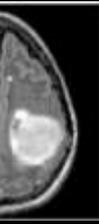
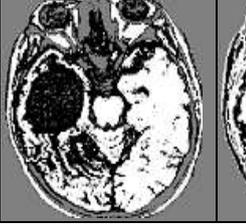
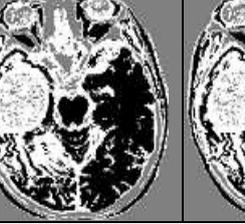
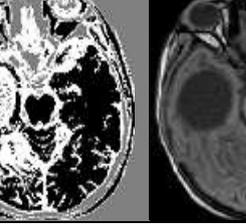
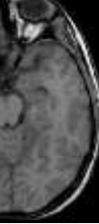


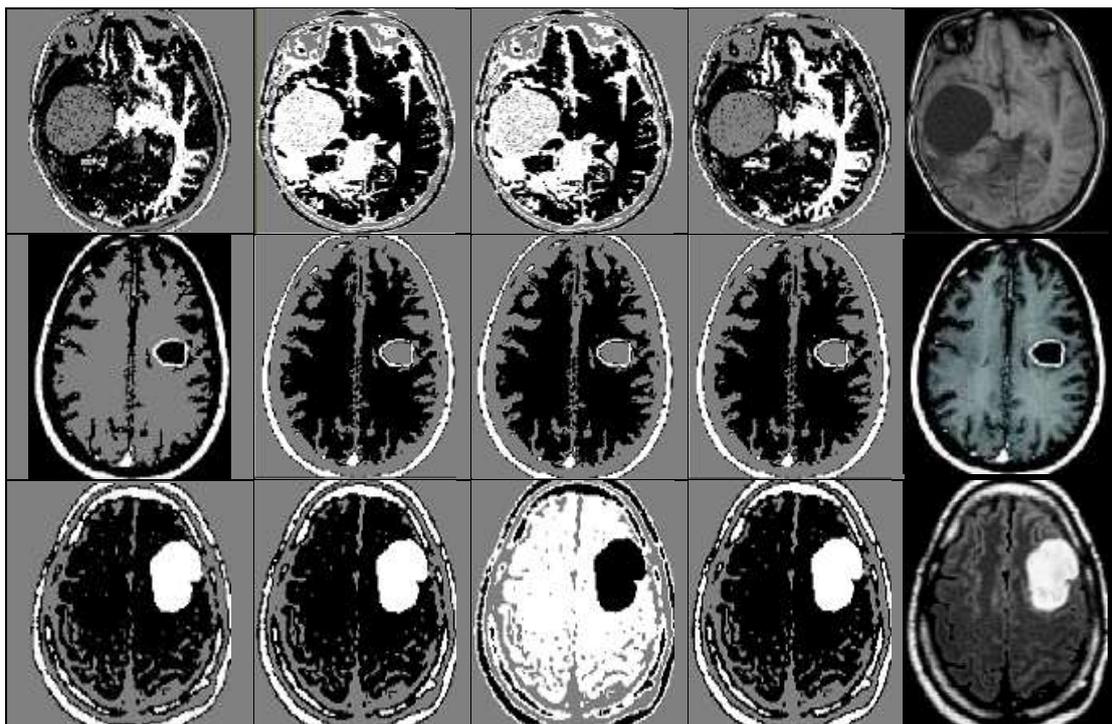
الشكل (3): يوضح العلاقة بين عدد التكرارات والدالة القياسية

يوضح الشكل (3) مقارنة بين الدالة القياسية في خوارزمية FKCN العادية وخوارزمية IFKCN المحسنة حيث نجد بأن الدالة القياسية في الخوارزمية المحسنة تتناقص بشكل واضح من القيمة 200 الى القيمة 65 تقريباً خلال التكرارات الخمسة الأولى وتسنقر عند التكرار الخامس بينما الخوارزمية العادية تبدأ الدالة القياسية من القيمة 650 تقريباً وتتناقص إلى القيمة 100 بعد 10 تكرارات وهذا ما يوضح أفضلية لخوارزمتنا المحسنة من حيث عدد التكرارات وزمن التنفيذ. **المسألة الثانية:** في مجال تجزئة الصور (imagesegmentation) [9]. سيتم إيضاح مفهوم تجزئة الصور قبل عرض نتائج الخوارزمية المحسنة على هذه المسألة.

تعد تجزئة الصور من الخطوات الأساسية في تحليل وتطبيقات معالجة الصور. إن هدف تجزئة الصورة هو الحصول على مناطق ذات معنى ونعني بذلك إيجاد المنطقة التي تمثل كياناً أو جزءاً من كيان داخل الصورة وهذا مهم قبل أي معالجة تحدث في المستوي الأعلى للصورة [9,10,13]. تعد عملية العنقدة طريقة لتجميع البيانات المتشابهة في مجاميع أو عناقيد مختلفة وكل مجموعة لابد أن تكون متشابهة وتصنف هذه التقنية على أنها طريقة غير مرشدة في تصنيف البيانات كما أن اختيار معيار التجانس أو التشابه يلعب دوراً مهماً في دقة نتيجة التقسيم [9]. إن عملية تجزئة الصور ذات التدرج الرمادي باستخدام خوارزمية العنقدة على هذا النوع من الصور تكون نتيجتها تجزئة المستوى الرمادي إلى عدد من المجاميع أو العناقيد، وإن المستوى الرمادي داخل العنقود الواحد يكون متشابهاً بالنسبة إلى المستوى الرمادي في عنقود آخر [11].

لقد تم اختبار الخوارزمية المحسنة في مجال تجزئة الصور الطبية الممثلة في صور الدماغ بالرنين المغناطيسي [12] حيث يعتبر تقسيم الصور خطوة حاسمة في إجراء العمليات الجراحية والعلاج حيث تُجزأ صور الأنسجة الطبيعية إلى صور المادة البيضاء وصور المادة الرمادية وصور سائل النخاع الشوكي باعتبارها طبيعة التشخيص وإلى صور الأورام باعتبارها صور مصابة بحالات مرضية، وتمت مقارنة نتائج الخوارزمية المحسنة IFKCN مع نتائج الخوارزميات القياسية K_means و FCM وخوارزمية FKCN قبل التعديل من حيث عدد التكرارات ([11] [9]). [12].

الخوارزمية المقترحة IFKCN	FKCN [8]	FCM [5]	K_means [4]	الصورة الأصلية
				
				



الشكل (4) : نتائج تجزئة لخمس صور رنين مغناطيسي لدماع باستخدام الخوارزميات $FKCN$, FCM , K_means وخوارزمية $IFKCN$ المحسنة .

يظهر الجدول (4) مقارنة بين الخوارزميات الأربع السابقة الذكر وذلك على الصور الخمس المعروضة في الشكل (4) .

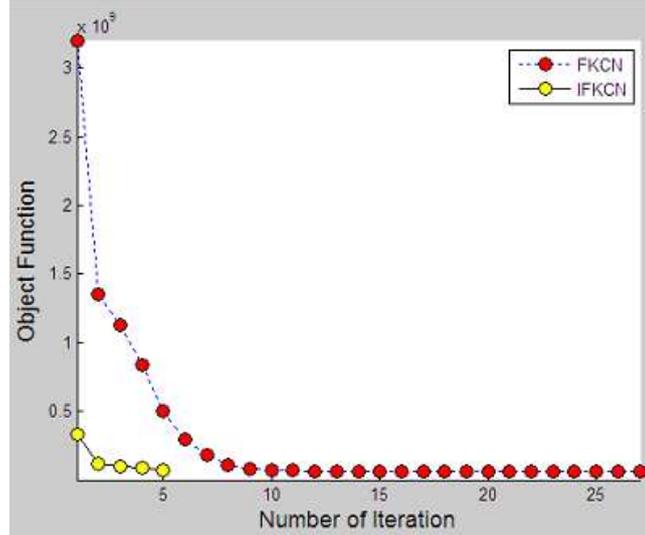
الجدول (4) : مقارنة بين الخوارزميات ($FKCN$, FCM , K_means) والخوارزمية المحسنة المطبقة على الصور في الشكل (4) من حيث عدد التكرارات.

بيانات الصور	$K_means[4]$	$FCM[5]$	$FKCN[8]$	الخوارزمية المقترحة $IFKCN$
الأولى	9	43	27	5
الثانية	7	100	36	7
الثالثة	14	81	38	11
الرابعة	15	33	39	15
الخامسة	18	29	38	15

من خلال ملاحظة نتائج العنقدة التي أُجريت على مجموعة من الصور المصابة بحالات مرضية تبين الآتي: بالنسبة لخوارزمية K_means كانت النتائج جيدة من حيث عدد التكرارات ولكن لا يوجد استقرار في الخوارزمية في إيجاد مراكز العناقيد وهذا واضح بالنسبة لمعالم الصورة حيث تحتاج هذه الخوارزمية إلى عدد إضافي من مرات التنفيذ. و في خوارزمية العنقدة المضطربة FCM كانت النتائج جيدة من حيث استقرار الخوارزمية في إيجاد المراكز، بينما كان عدد التكرارات كبير مقارنة مع بقية الخوارزميات.

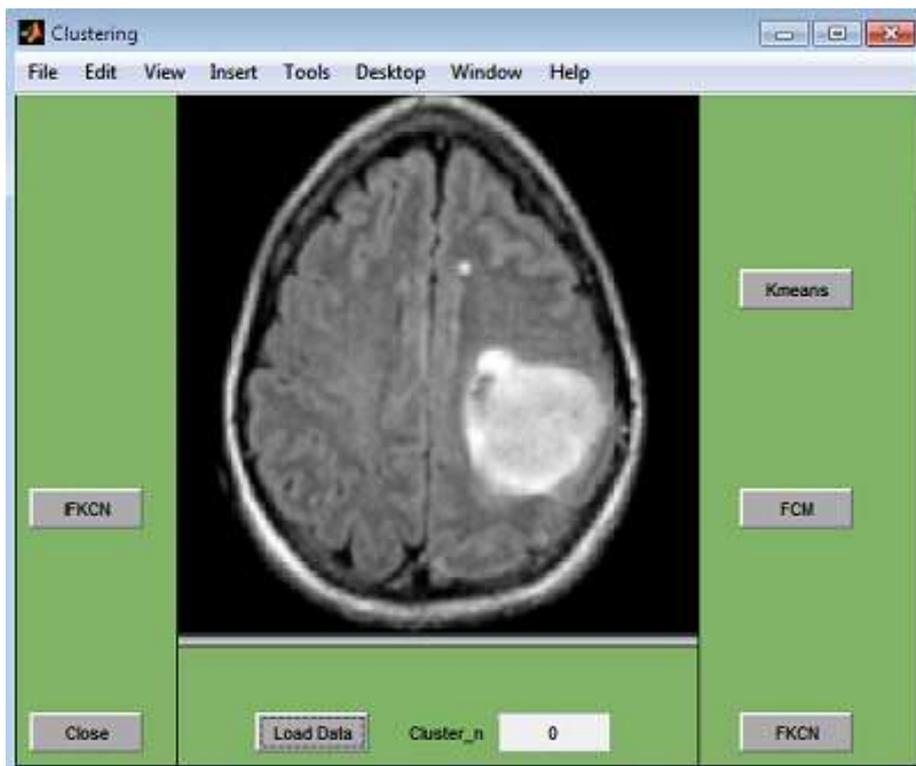
بينما خوارزمتي العنقدة المعتمدة على الشبكات العصبية الصناعية المضطربة $FKCN$ العادية و $IFKCN$ المحسنة قدمتا نتائج أفضل من حيث قيم مراكز العناقيد. وأبدت الخوارزمية المحسنة أفضلية على خوارزمية $FKCN$ العادية من حيث عدد التكرارات.

وكننتيجة نهائية نجد بأن خوارزمية IFKCN المحسنة أفضل من خوارزمية FKCN من حيث إيجاد مراكز العناقيد وعدد التكرارات وأظهرت دقة في تجزئة الصورة إلى مناطق فعّالة بحسب المستويات الرمادية الموجودة في الصورة. يمكن توضيح سرعة تقارب الخوارزمية المحسنة IFKCN مقارنة مع الخوارزمية العادية FKCN في الشكل (5).



الشكل (5) : العلاقة بين عدد التكرارات وقيم الدالة القياسية عند إجراء عملية العنقدة لبيانات الصورة الأولى .

يبين الشكل (5) أن الدالة القياسية للخوارزمية المحسنة IFKCN تتناقص بشكل واضح حتى التكرار 5 بينما الدالة القياسية للخوارزمية العادية تتناقص حتى التكرار 27 وهذا يعكس أفضلية للخوارزمية المحسنة من حيث زمن التنفيذ. ولتسهيل التعامل واستخدام هذا التطبيق فقد تم بناء واجهة باستخدام برنامج الماتلاب وفيما يلي نوضح الواجهة الرئيسية (GUI) لإجراء عملية العنقدة في مجال تجزئة الصور للخوارزميات الأربع السابقة وهي (خوارزمية العنقدة التقليدية K_means وخوارزمية العنقدة المضببة FCM وخوارزمية FKCN العادية وخوارزمتنا المحسنة IFKCN) وذلك لتسهيل عملية العنقدة ومقارنة النتائج :



الشكل (6) : الواجهة المستخدمة لإجراء عملية العنقدة على الصورة

حيث تتضمن الواجهة المستخدمة الأزرار التالية:

الزر الأول **LoadData**: بعد الضغط على هذا الزر يتم تحميل الصورة.

ومن ثم يتم تحديد عدد العناقيد في المربع المسمى **Cluster_n**.

الزر الثاني **K_means**: خاص بإجراء عملية العنقدة على الصورة باستخدام خوارزمية **K_means**.

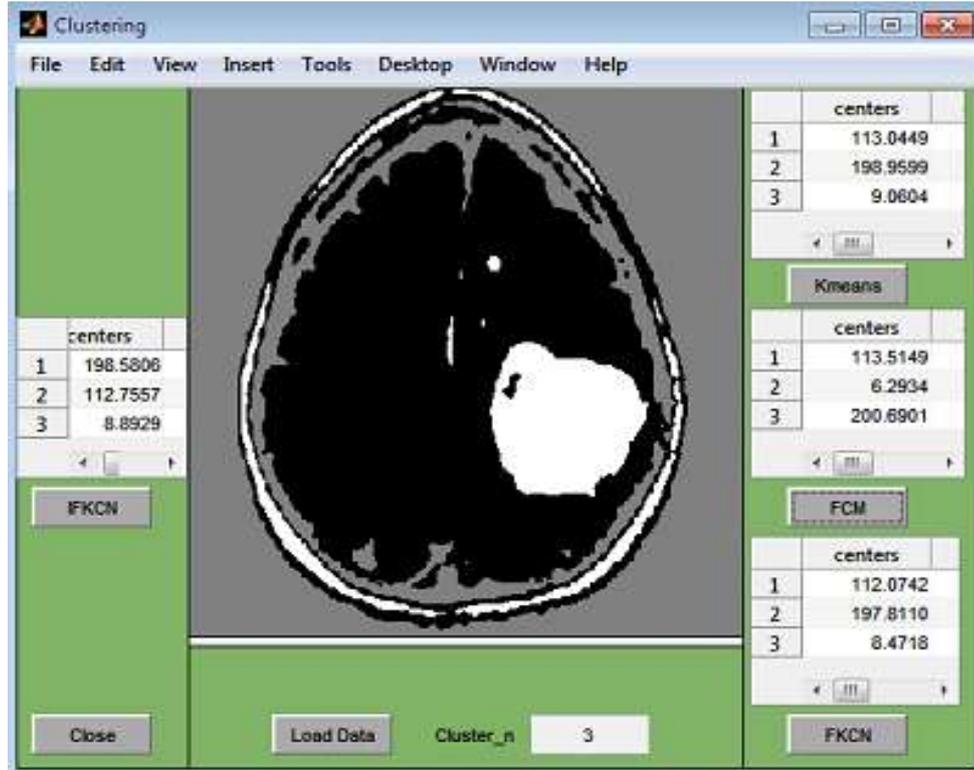
الزر الثالث **FCM**: إجراء العنقدة باستخدام الخوارزمية **FCM** على نفس الصورة أو بعد تحميل صورة جديدة .

الزر الرابع **FKCN**: إجراء عملية العنقدة باستخدام الخوارزمية **FKCN** العادية.

الزر الخامس **IFKCN**: إجراء العنقدة باستخدام خوارزمتنا المحسنة **IFKCN** .

الزر السادس **Close**: زر إغلاق الواجهة.

وبعد تفعيل جميع هذه الأزرار تكون واجهة البرنامج مبيّنه بالشكل (7):



الشكل (7) : الواجهة بعد إجراء عملية العنقدة على الصورة

نكون قد قمنا بإجراء عملية العنقدة على الصورة المصابة باستخدام جميع الخوارزميات السابقة الموضحة في الواجهة، حيث نتوضع نتيجة كل خوارزمية (مراكز العناقيد) فوق كل زر من أزرار الخوارزميات المسماة على الأزرار. و تمثل الصورة النهائية الموضحة في الشكل (7) نتيجة إجراء عملية العنقدة باستخدام خوارزمتنا المحسنة IFKCN.

الاستنتاجات والتوصيات :

الاستنتاجات :

على البيانات Iris:

- 1- قيمة مقياس الخطأ لخوارزمية FCM أصغر من قيمة مقياس الخطأ لخوارزمية K_means.
- 2- نتائج خوارزمية FKCN أفضل من نتائج خوارزمية FCM، حيث تبين أن عدد التكرارات بالنسبة لـ FKCN أقل من عدد تكرارات خوارزمية FCM. كما أن الفرق في مقياس الخطأ قد بلغ 33% وهي نسبة متقاربة.
- 3- نتائج خوارزمية IFKCN أفضل من نتائج خوارزمية FKCN من حيث عدد التكرارات حيث تبين أن هناك تطور ملحوظ بعدد التكرارات بمقدار 40% بالنسبة للخوارزمية المحسنة، كما تبين أنه يوجد تحسن واضح بمقياس الخطأ لصالح الخوارزمية المحسنة بمقدار 70% وهذا ما يثبت صحة الاختبار.

على تجزئة الصور:

زمن تنفيذ الخوارزمية المحسنة IFKCN أفضل من زمن تنفيذ باقي الخوارزميات المذكورة سابقاً وذلك بالاعتماد على عدد التكرارات.

وبغض النظر عن البيانات المستخدمة نستنتج أن:

1. خوارزميات العنقدة التقليدية لا يمكنها إجراء عملية العنقدة بشكل دقيق وبالتالي الاعتماد على طرائق أكثر فعالية -تقنيات الذكاء الصناعي- يعطي نتائج موثوقة للبحث.
2. تحديث معامل التضييب في كل تكرار يتحكم بشكل ذاتي في توزيع معدل التعلم وتحديث أوزان الشبكة لجميع عقد البيانات بفعالية.
3. الشبكات العصبية الصناعية المهجنة بالمنطق الضبابي تعطي نتائج أفضل من الشبكات العصبية الصناعية الاعتيادية.

التوصيات :

- 1- العمل على تطوير الخوارزمية المحسنة بتهجينها مع تقنيات أخرى كالخوارزميات الجينية.
- 2- تطوير النموذج السابق على أكثر من منصة ليكون موجوداً ومطبّقاً في جميع مشافي القطر .

المراجع :

1. HATHWAY, J. R., BEZDEK, J. C. "Optimization of Clustering Criteria by Reformulation". IEEE, 1995, PP 241 – 245.
2. BEZDEK, J.C., TSAO, E.C-K., PAL, N.R. "Fuzzy Kohonen Clustering Networks". IEEE, 1992, PP 1035-1043.
3. QIANMU, L., MANWU, X. "A Novel Classification Method and its Application in Flow Forecast". IJCSNS International Journal of Computer Science and Network Security, Vol.6, NO.9, 2006, PP.54-59.
4. MACQUEEN, J.B. "Some methods for classification and analysis of multivariate observations". Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Berkeley, University of California Press, 1, 1967, PP 281-297.
5. BEZDEK, J.C. "A Convergence Theorem for the Fuzzy C-Means Clustering Algorithm". IEEE, 1980.
6. NAES, T., MEVIK, B.H. "The flexibility of fuzzy clustering illustrated by examples". Journal of Chemometrics, 13, 1999, PP 435-444.
7. FISHER, R.A. "The Use of Multiple Measurements in Taxonomic Problems. Annual Eugenics". 7, Part II, 1936, PP 179-188.
8. IRWANSYAH, E., FAISAL, M., PRIMADINI, A. "Does Efficient Fuzzy Kohonen Clustering Network Algorithm Really Improve Clustering Data Result?". Journal of Theoretical and Applied Information Technology, Vol.71, No.1, 2015, PP 144-148.
9. SIVAKUMAR, K. "Image Segmentation". School of Electrical Engineering & Computer Science (School of EECS), 2002.
10. UMBAUGH, S.E. "Computer Vision and Image Processing". Prentice HALL PTR, USA, 2006.
11. COLEMAN, G., ANDREWS, H. "Image segmentation by Clustering". Addison Wesley, Proc. IEEE. 1979.
12. HALL, L.O. et al. "A Comparison of Neural Network and Fuzzy Clustering Techniques in Segmenting Magnetic Resonance Image of the Brain". IEEE Trans. Neural Networks, Vol.3, Iss.5, 1992, PP 672-682.
13. BODYANSKIY, YE. "Medical Online Neuro-Fuzzy Diagnostics System with Active Learning". IJ. of Advances in Computer and Electronics Engineering, Vol. 2,7, 2017, PP 1-10.