

## التنقيب في البيانات ومعالجة الصور باستخدام تقنيات الذكاء الصناعي

أحمد الرزّو محمد أسعد

كلية العلوم - جامعة تشرين

### الملخص

في هذا البحث تم العمل على تحسين خوارزمية شبكة العنقدة كوهنين المضببة (FKCN) Fuzzy Kohonen Clustering Network التي تعتبر من أشهر الشبكات العصبية التي تستخدم للتنقيب في البيانات Data Mining. تهدف خوارزمية FKCN إلى تجميع البيانات المتماثلة مع بعضها البعض اعتماداً على خصائصها إلى K عنقود يسهل تحليلها ومن ثم التعامل معها. تتم عملية اختيار مراكز العناقيد (أوزان الشبكة) بشكل عشوائي ويفضل أن تكون هذه المراكز بعيدة عن بعضها البعض قدر الإمكان. تؤثر نقطة البدء العشوائية على فعالية عملية التجميع والنتائج وتعتمد عملية التقارب بشكل رئيسي على المراكز الابتدائية لذا تم اقتراح علاقة جديدة لاختيار أوزان الشبكة الابتدائية اعتماداً على البيانات المدروسة.

تم اختبار الخوارزمية المقترحة في مجال التنقيب في البيانات وتجزئة الصور الطبية (صور الرنين المغناطيسي للدماغ) باستخدام برنامج الماتلاب MATLAB. وتُظهر النتائج أن خوارزمية FKCN المحسنة أفضل من خوارزمية FKCN العادية مقارنة بالنسبة لعدد التكرارات وزمن التنفيذ.

**الكلمات المفتاحية:** الذكاء الصناعي، شبكة كوهنين، تنقيب البيانات، العنقدة، مراكز العناقيد.

### مقدمة :

تعتبر العنقدة Clustering من أهم خوارزميات التنقيب في البيانات التي تفيد في تجميع البيانات المتماثلة مع بعضها البعض في عنقايد يسهل التعامل معها وإجراء البحث ضمنها للوصول إلى الهدف المنشود، وتتم عملية العنقدة من خلال تقليل المسافات بين الكائنات ومركز العنقود. تعتبر خوارزمية العنقدة FKCN من أشهر الشبكات العصبية الصناعية التي تستخدم في تحليل البيانات نظراً لأدائها الحسابي المميز. حيث يتم تجميع البيانات في K عنقود (K هو عدد العناقيد والمحددة قبل البدء في الخوارزمية) وتتم عملية العنقدة عن طريق التقليل من مجموع المسافات المربعة (المسافات الإقليدية) بين العناصر ومركز العنقود (أوزان الشبكة) المقابل لها [1]. تختلف العنقدة Clustering عن التصنيف Classification كون أنه في التصنيف يتم تعيين الكائنات إلى فئات محددة مسبقاً، في حين أنه في العنقدة يتم تجميع الكائنات التي تتشابه في الخصائص. وتفيد أداة التحليل العنقودي في تنظيم البيانات المستخلصة، ويتم تشكيل العنقود بناءً على التشابه الأعظمي للكائنات في كل مجموعة وزيادة الاختلاف بين العناقيد والتي لم تكن معروفة قبل بدء عملية العنقدة. أما في التصنيف فيتم تصنيف الكائنات في مجموعات معروفة ومحددة مسبقاً [2]. يلعب مقياس التشابه دوراً مهماً في القيام بعملية العنقدة حيث تُجمع الكائنات في عدة عنقايد ويحوي كل عنقود كائنات متشابهة ومختلفة عن الكائنات في العناقيد الأخرى [3].

تحاول خوارزمية العنقدة العثور على مجموعات من البيانات بالاعتماد على التشابه بين هذه البيانات وإيجاد مركز كل مجموعة، ومن أجل تحديد اعضاء كل عنقود تقوم خوارزمية العنقدة بتقييم المسافة بين نقطة ما ومركز العنقود. وفي نهاية الخوارزمية نحصل على العناقيد المطلوبة وعدد الأعضاء المتشابهة في كل عنقود.

### مشكلة البحث :

تعتبر عملية اختيار المراكز الأولية الصحيحة في خوارزمية FKCN هي العملية الرئيسية والأساسية، ومن السهل اختيار المراكز الأولية عشوائياً، ولكن قد لا نحصل على النتائج المرجوة من أول مرة. وبالتالي يمكن صياغة المشكلة بالتساؤلات التالية:

- 1- هل من الممكن إيجاد طريقة ناجحة لتحديد المراكز الابتدائية لخوارزمية FKCN ؟
- 2- هل من الممكن العمل على تقليل زمن تنفيذ خوارزمية FKCN ؟
- 3- هل نستطيع تقسيم البيانات المدروسة إلى مجموعات يسهل تحليلها ودراستها؟

### أهمية البحث أهدافه :

#### أهمية البحث:

تعتبر خوارزمية العنقدة FKCN من أهم خوارزميات العنقدة التي نتعامل من خلالها مع البيانات حيث استُخدمت هذه الخوارزمية في تحليل المكونات الرئيسية للبيانات الاحصائية [3]، وفي مجال تجزئة الصور image segmentation حيث تعتبر تجزئة الصور من الخطوات الأساسية في تحليل وتطبيقات معالجة الصور (المعالجة اللونية، تقليل الأبعاد...). إن هدف تجزئة الصورة الحصول على مناطق ذات معنى أي إيجاد المنطقة التي تمثل كياناً أو جزءاً من كيان داخل الصورة وهذا مهم قبل أي معالجة تحدث في المستوى الأعلى للصورة [4]. وفي مجال الطب فقد استخدمت خوارزمية FKCN بعد إجراء التعديلات عليها في معالجة صور الرنين المغناطيسي للدماغ حيث يعتبر تقسيم الصور الطبية خطوة حاسمة في إجراء العمليات الجراحية والعلاج حيث يمكن بعد إجراء عملية العنقدة اكتشاف واقتطاع المنطقة المصابة في الصورة ومن ثم حساب مساحتها [5,6]. كما تفيد عملية التنقيب في البيانات باستخدام العنقدة في تحليل البيانات دون فرضيات مسبقة من خلال ربطها مع تقنيات الذكاء الصناعي والعمليات الاحصائية بغرض إيجاد قواعد وأمثلة ونماذج تفيد اصحاب القرار في اتخاذ القرار الصحيح.

**أهداف البحث :**

يهدف البحث على تحسين أداء خوارزمية FKCN وذلك من خلال الخطوات التالية:

- 1- إيجاد علاقة رياضية مناسبة لحساب المراكز الابتدائية لخوارزمية FKCN للتخلص من تحديد المراكز الابتدائية بشكل عشوائي.
- 2- توفير تجمعات ذات كائنات متشابهة من خلال عملية العنقدة.
- 3- إيجاد عملية ناجحة لتحديد طريقة انتماء الكائن لأي عنقود.

**منهجية البحث:**

تم اتباع المنهج الوصفي التحليلي، ووفقاً للمنهج الوصفي تم وصف ظواهر الدراسة بالاعتماد على المراجع والدوريات العلمية والمقالات. كما تم تحليل البيانات المأخوذة والوصول الى النتائج من خلال استخدام نموذج مقترح معتمد على طرائق الذكاء الصناعي وبرنامج MATLAB.

**شبكة كوهنين التقليدية:**

جعلت عملية التدريب الشبكات العصبية الصناعية من أفضل السبل لتحقيق الكثير من الأهداف في العديد من التطبيقات. وتعتبر شبكة كوهنين واحدة من الشبكات ذات التعلم بدون مشرف أي من نوع التعلم الذاتي (لا يوجد خرج معين يعطى للشبكة).

تتكون شبكة كوهنين من طبقتين، طبقة الإدخال والتي تتألف من مجموعة من الخلايا العصبية (العقد) وكل خلية من خلايا المدخلات ترتبط بجميع الخلايا الموجودة في طبقة المخرجات أو ما يسمى بطبقة كوهنين (نسبة للعالم كوهنين) عن طريق أوزان الارتباطات بين الخلايا [7].

تعتمد خوارزمية شبكة كوهنين في تحديد الخلية الفائزة على حساب المسافة الاقليدية بين خلايا طبقة المدخلات وخلايا طبقة المخرجات. واعتبار الخلية في طبقة المخرجات التي تحمل أقل قيمة للمسافة هي الخلية الفائزة.

**خوارزمية شبكة كوهنين التقليدية: [8]**

**الخطوة الاولى:** تهيئة الأوزان  $w_{ij}$ ، تحديد عدد العناقيد  $C$ ، وضع قيمة ابتدائية لمعدل التعلم  $\alpha$ .

**الخطوة الثانية:** طالما أن شرط التوقف لم يتحقق نفذ الخطوات من الخطوة 3 حتى الخطوة 9.

**الخطوة الثالثة:** من أجل كل شعاع دخل، نفذ الخطوات من الخطوة 4 حتى الخطوة 6.

**الخطوة الرابعة:** من أجل كل خلية (عقدة)  $z$  احسب:

$$D(j) = \sum_i (w_{ij} - x_i)^2 \quad 2$$

**الخطوة الخامسة:** أوجد الدليل  $z$  الذي يكون من أجله  $D(j)$  أصغري.

**الخطوة الخامسة:** من أجل جميع الخلايا  $z$  التي تدخل الجوار الخاص لـ  $z$ ، ومن أجل كل الخلايا  $i$  أوجد:

$$w_{ij}(\text{new}) = w_{ij}(\text{old}) + \alpha[x_i - w_{ij}(\text{old})] \quad 3$$

**الخطوة السابعة:** حدّث معدل التعلم  $\alpha$ .

**الخطوة الثامنة:** اختبر شرط التوقف.

لقد تطور استخدام شبكة كوهنين بالنسبة لحل العديد من المشاكل ومن الأمثلة الكثيرة التي طبقت عليها هذه الشبكات مسألة توليد الموسيقى من الحاسب، ومسألة البائع الجوال التي استخدم العديد من الباحثون خرائط التنظيم الذاتي لـ Kohonen.

**شبكة كوهنين المصيبة:**

شبكة كوهنين يمكن أن تعمد لتصميم شبكة كوهنين المصيبة وذلك من خلال تهجينها مع المنطق الضبابي Fuzzy Logic (أو منطق الغموض) وهو أحد أشكال المنطق، يستخدم في بعض الأنظمة الخبيرة وتطبيقات الذكاء الصناعي، نشأ هذا المنطق عام 1965 على يد العالم " لطفى زادة " [9] من جامعة كاليفورنيا حيث طوره ليستخدمه كطريقة أفضل لمعالجة البيانات، لكن نظريته لم تلق اهتماماً حتى عام 1974 حيث استخدم المنطق الضبابي في تنظيم محرك بخاري، ثم تطورت تطبيقاته حتى وصلت لتصنيع "شريحة منطق ضبابي" والتي استعملت في العديد من المنتجات كآلات التصوير. هناك العديد من الدوافع التي وجهت العلماء إلى تطوير علم المنطق الضبابي ومع تطور الحاسوب والبرمجيات نشأت الرغبة في اختراع أو برمجة أنظمة يمكنها التعامل مع المعلومات الغير الدقيقة على غرار الإنسان لكن هذا ولد مشكلة حيث أن الحاسوب لا يمكنه التعامل إلا مع بيانات دقيقة ومحددة. وقد نتج عن هذا التوجه ما يعرف بالأنظمة الخبيرة أو الذكاء الصناعي ويعتبر علم المنطق الضبابي أحد النظريات التي يمكن من خلالها بناء مثل هذه الأنظمة.

المبادئ الأساسية للعنقدة المصيبة هو أن يكون لكل عنصر درجة انتماء جزئية في العناقيد المصيبة. ويتم اختيار قيمة درجة 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 وفق قيمة دالة العضوية (membership function).

خوارزمية شبكة العنقدة كوهنين المضببة FKCN: [10]

**الخطوة الأولى:** تكوين أوزان عشوائية بدائية لخلايا الشبكة  $(0)$  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  $w_1$  المرتبطة بالخلية وفق العلاقة التالية:

$$d_i = \sqrt{\sum_{j=1}^{n-1} (x_j - w_{1j})^2} \quad 4$$

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

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{\|x_j - w_1\|}{\|x_j - w_k\|} \right)^{\frac{1}{m_t-1}}} ; \quad \alpha_{ij} = (u_{ij})^{m_t} \quad 5$$

حيث تُعطى علاقة معامل التضبيب على الشكل:

$$m_t = m_0 - t * \left( \frac{m_0-1}{t_{max}} \right) , \quad t = 1, 2, \dots, t_{max} \quad 6$$

**الخطوة الرابعة:** تحديث أشعة الأوزان وفق العلاقة التالية:

$$w_i(t) = w_i(t-1) + \frac{\sum_{j=1}^n \alpha_{ij} (x_j - w_i)}{\sum_{s=1}^n \alpha_{is}} , \quad i = 1, 2, \dots, c \quad 7$$

**الخطوة الخامسة:** حساب تابع الخطأ:

$$E(t) = \| W(t) - W(t-1) \| \quad 8$$

**الخطوة السادسة:** إذا كان  $t+1 > t_{max}$  أو  $t < \epsilon$  يتوقف التكرار وإلا العودة للخطوة الثانية، حيث  $\epsilon$  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)**

إن الهدف الرئيسي لخوارزمية FKCN المحسنة هو تحسين أداء الخوارزمية السابقة من خلال التخلص من العشوائية في اختيار الأوزان الابتدائية في الخطوة الأولى من الخوارزمية العادية. وجعل كل النقاط لها فرص متساوية أن تكون نقاط مراكز أولية. ولأن خوارزمية FKCN حساسة لاختيار المراكز الأولية، سنجعل تقارب هذه النقاط أقرب للحل النهائي بالحدود الدنيا [11] مما يؤدي إلى تحسين أداء عملية العنقدة.

اعتمدت آلية العمل على تقسيم البيانات المدروسة على طول محور البيانات إلى  $K$  مجموعة ( $K$  عنقود) ومن ثم يتم حساب مراكز المجموعات المقسمة (متوسط المجموعات) فنحصل بذلك على مراكز أولية للبيانات المدروسة. حيث يتم تقسيم البيانات وفقاً للعلاقة المقترحة التالية:

$$C_i = \text{Mean} \left( \text{Data} \left( \text{Floor} \left( \frac{(i-1) * n}{k} \right) + 1 : \text{Ceil} \left( \frac{i * n}{k} \right) \right) \right) , \quad i = 1, 2, \dots, k \quad 9$$

حيث أن :

▪  $K$  عدد العناقيد،  $n$  عدد نقاط البيانات،  $C_i$  مراكز العناقيد.

- **Mean** دالة تقوم بإيجاد المتوسط الحسابي لكل مجموعة  $\bar{x}$  مقسمة.
- **Data** مجموعة البيانات المدروسة.
- **Floor** التقريب لأصغر عدد صحيح.
- **Ceil** التقريب لأكبر عدد صحيح.

وتم استدعاء الدوال Mean ، Floor ، Ceil في العلاقة المقترحة باستخدام برنامج الماتلاب [12].

يمثل المسقط الأول من العلاقة بداية البيانات المدروسة والمسقط الثاني نهاية بيانات العنقود الأول وذلك من أجل  $k=1$ ، ومن ثم إيجاد المتوسط الحسابي لهذه البيانات فنحصل على مركز العنقود الابتدائي الأول وهكذا حسب عدد العناقيد المختارة.

#### الجانب التطبيقي والعملي :

تم كتابة خوارزمية FKCN المحسنة باستخدام لغة الماتلاب (MATLAB R2013a) حيث تعتبر هذه اللغة من اللغات الحديثة، وتستخدم في المجالات التقنية لما تتمتع به من خصائص ومزايا توفر للمستخدم سهولة التعامل مع البيانات ومعالجتها وتضم عدداً كبيراً من الدوال والمكتبات الجاهزة حسب التطبيق المستخدم. تم اختبار الخوارزمية المقترحة في مجال التنقيب في البيانات وتجزئة الصور الطبية (صور الرنين المغناطيسي لدماع) في إجراء عملية العنقدة ومقارنتها مع الخوارزمية العادية من حيث عدد التكرارات وزمن التنفيذ.

**التنقيب في البيانات:** إن علم التنقيب في البيانات Data Mining كما يطلق عليه البعض علم استنباط البيانات وهو تطبيق هام جداً في جميع المجالات. في هذا الكم الهائل من البيانات كان لابد من ظهور أدوات تساعد على التعامل معها بكل يسر وسهولة واستنباط معلومات مخبئة ضمنها تفيد الحاضر والمستقبل.

تم إجراء عملية العنقدة على مجموعة من البيانات ذات الخصائص المختلفة [12]. وتم مقارنة النتائج بين الخوارزمية FKCN المحسنة وخوارزمية FKCN العادية من حيث عدد التكرارات وزمن التنفيذ، وذلك من أجل عدد العناقيد  $K=3$ . وفي التالي نبين شرحاً مبسطاً عن طبيعة هذه البيانات:

- **Cancer\_ dataset** : وهي تمثل مجموعة بيانات لمرض سرطان الثدي لأحد المرضى.
- **Iris\_ dataset** : تمثل بيانات لثلاثة أصناف مختلفة من الأزهار.
- **Chemical\_ dataset** : وهي مجموعة بيانات لمركبات كيميائية.
- **Bodyfat\_ dataset** : تمثل النسبة المئوية للدهن في الجسم لشخص ما.
- **Exchanger\_ dataset** : مجموعة بيانات لتقلبات حرارة الطقس.

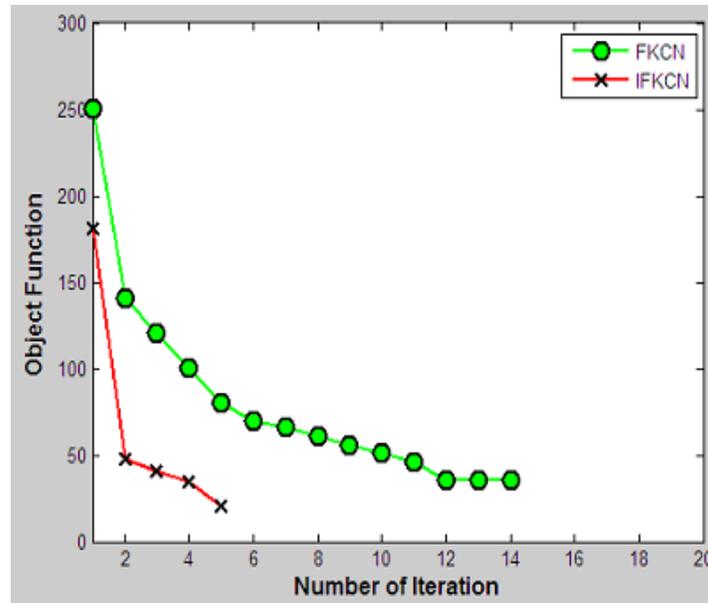
والجدول التالي يوضح عملية المقارنة :

**الجدول (1):** عملية المقارنة بين الخوارزمية المحسنة والخوارزمية العادية من حيث عدد التكرارات وزمن التنفيذ

Data	التكرارات		زمن التنفيذ (بالثانية)	
	FKCN	IFKCN	FKCN	IFKCN
<b>Cancer_ dataset</b>	<b>14</b>	<b>5</b>	<b>0.150</b>	<b>0.077</b>
<b>Iris_ dataset</b>	<b>14</b>	<b>8</b>	<b>0.036</b>	<b>0.026</b>
<b>Chemical_ dataset</b>	<b>22</b>	<b>14</b>	<b>0.172</b>	<b>0.105</b>
<b>Bodyfat_ dataset</b>	<b>15</b>	<b>8</b>	<b>0.66</b>	<b>0.047</b>
<b>Exchanger_ dataset</b>	<b>19</b>	<b>10</b>	<b>0.748</b>	<b>0.662</b>

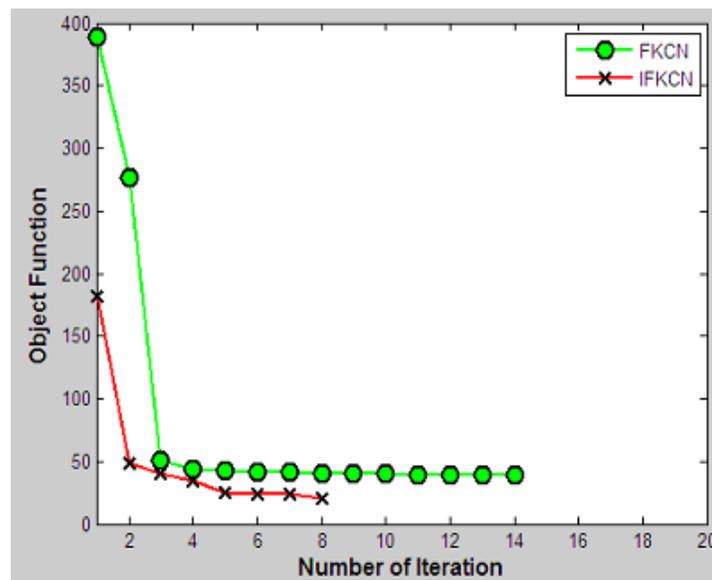
من خلال النتائج يمكن ملاحظة أن الخوارزمية المحسنة تحتاج إلى عدد أقل من التكرارات الذي يقلل بدوره من وقت التجميع.

سنوضح في الشكل (1) العلاقة بين عدد التكرارات والدالة القياسية لبيانات سرطان الثدي في الجدول (1) بالنسبة للخوارزمية المحسنة والخوارزمية العادية.



الشكل (1) : العلاقة بين عدد التكرارات والدالة القياسية للخوارزمية المحسنة والخوارزمية العادية

كما يوضح الشكل(2) المقارنة بين الخوارزمية العادية والخوارزمية المحسنة بالنسبة لعدد التكرارات وقيمة الدالة القياسية لبيانات Iris data :



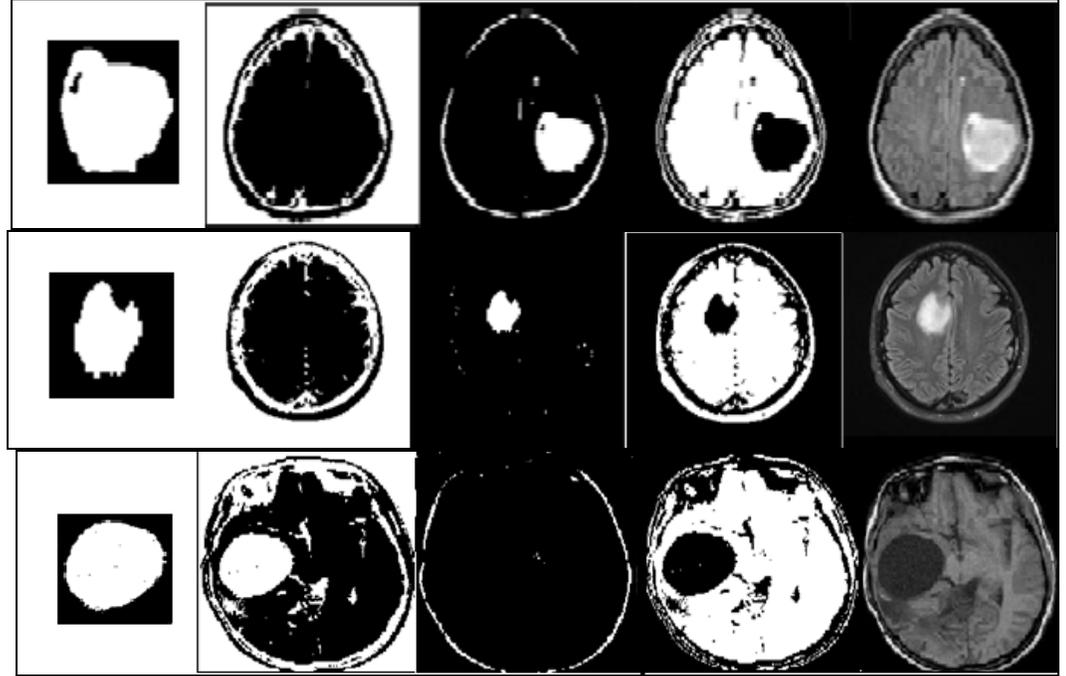
الشكل(2): العلاقة بين عدد التكرارات والدالة القياسية للخوارزمية المحسنة والخوارزمية العادية

### تجزئة الصور: (image segmentation) [13]

تعتبر تجزئة الصور من الخطوات الأساسية في تحليل وتطبيقات معالجة الصور. إن هدف تجزئة الصورة 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. تعتبر عملية العنقدة طريقة لتجميع البيانات المتشابهة في مجاميع أو عنقود 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. عملية تجزئة الصور ذات التدرج الرمادي تكون نتيجتها تجزئة المستوي الرمادي إلى عدد من المجاميع أو العناقيد وإن المستوي الرمادي داخل العنقود الواحد يكون متشابهاً بالنسبة إلى المستوي الرمادي في عنقود آخر.

تم اختبار الخوارزمية المقترحة في مجال تجزئة الصور الطبية الممتلئة في صور الدماغ بالرنين المغناطيسي [14] حيث يعتبر تقسيم الصور خطوة حاسمة في إجراء العمليات الجراحية والعلاج حيث تُجزأ صور الانسجة 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 النتائج مع خوارزمية FKCN العادية من حيث عدد التكرارات وزمن التنفيذ. يوضح الشكل (3) نتائج العنقدة للخوارزمية المقترحة على ثلاث صور مصابة بحالات مرضية مع اختيار عدد العناقيد  $K=3$ . واقتطاع المنطقة المصابة (الورم) من هذه الصور بعد إجراء عملية العنقدة.

الصورة الأصلية العنقود الأول العنقود الثاني العنقود الثالث منطقة الورم



الشكل (3): إجراء عملية العنقدة على ثلاث صور رنين مغناطيسي لدماغ ، حيث عدد العناقيد  $k=3$

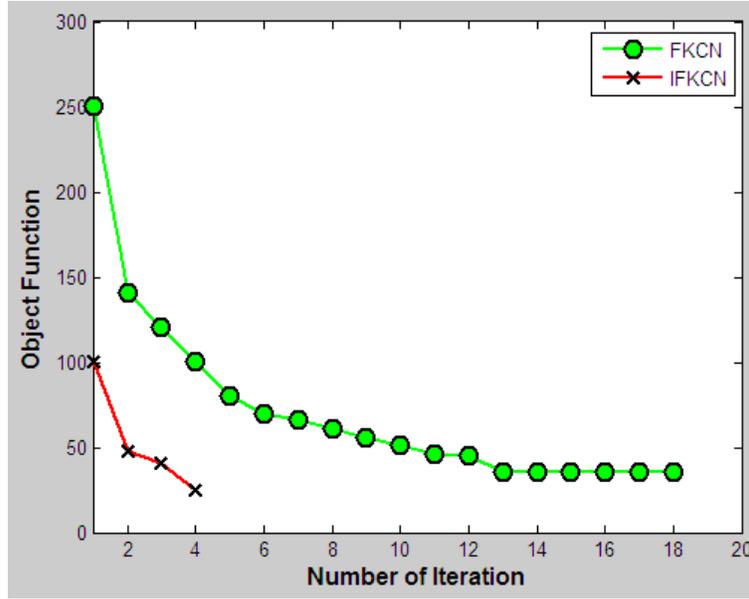
يمكن توضيح نتائج عملية المقارنة في الجدول (2) على الترتيب بين الخوارزمية العادية والخوارزمية المحسنة بعد إجراء عملية العنقدة على مجموعة الصور في الشكل (3) من حيث عدد التكرارات وزمن التنفيذ.

الجدول (2) : مقارنة بين الخوارزمية العادية والخوارزمية المحسنة من حيث عدد التكرارات وزمن التنفيذ

زمن التنفيذ (بالثانية)		التكرارات		البيانات
IFKCN	FKCN	IFKCN	FKCN	
0.359	0.611	4	18	الصورة الأولى
0.836	1.661	20	28	الصورة الثانية
0.303	0.747	7	10	الصورة الثالثة

من خلال النتائج يمكن ملاحظة أن الخوارزمية المحسنة تحتاج إلى عدد أقل من التكرارات وزمن تنفيذ أقل أيضاً مما يعطي أفضلية للخوارزمية المحسنة.

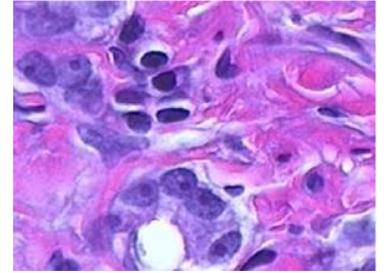
يمكن مقارنة العلاقة بين عدد التكرارات والدالة القياسية في الشكل (4) للخوارزمية المحسنة والخوارزمية العادية للصورة الأولى:



الشكل (4) : مقارنة بين عدد التكرارات والدالة القياسية للصورة الأولى.

من خلال ملاحظة الشكل نستنتج أن الخوارزمية المقترحة تتقارب بسرعة من الحل وبعدها تكرارات أقل مقارنة مع الخوارزمية العادية.

كما تم تطبيق الخوارزمية المقترحة على مجموعة من الصور الطبية المصابة بحالة مرضية بعد تحويل الصورة الأصلية لمستوى رمادي إذ إن الصورة في الشكل (5) تمثل نسيج لخلية مصابة بحالة مرضية لجسم إنسان، والشكل (6) يمثل صورة رنين مغناطيسي لدماع.



الصورة الأصلية



الشكل (5): الصورة الأصلية مع نتائج العنقدة باختيار عدد العناقيد  $k=3$ .  
 العنقود الأول      العنقود الثاني      العنقود الثالث



الصورة الاصلية



العنقود الثالث

العنقود الثاني

العنقود الأول

الشكل (6): الصورة الأصلية لدماع مصاب مع نتائج العنقدة باختيار عدد العناقيد  $k=3$ . الاستنتاجات والتوصيات :

#### الاستنتاجات :

- يتأثر سلوك خوارزمية FKCN العادية في عملية اختيار نقاط البدء العشوائية (المراكز الابتدائية) بالنسبة لعدد التكرارات وإيجاد المراكز النهائية لمجموعة العناقيد المختارة.
- إن معالجة قاعدة البيانات متعددة الأبعاد وباختلاف أنماط البيانات وتحليلها عاملياً قبل تنفيذ خوارزمية عليها يعطي نتيجة أفضل في العنقدة.
- الخوارزمية المحسنة تعطي الحل من الخطوة الأولى من تنفيذ الخوارزمية.
- الخوارزمية المحسنة لا تحتاج إلى عدد كبير من التكرارات في عملية التنقيب في البيانات مقارنة مع الخوارزمية العادية كما هو واضح في الشكل (1) و (2).
- من خلال ملاحظة نتائج العنقدة التي أجريت في مجال تجزئة الصور استطاعت الخوارزمية المحسنة من اكتشاف المناطق السليمة والمناطق المصابة بحالات مرضية بدقة عالية وبعدد تكرارات أقل من الخوارزمية العادية كما هو واضح في الشكل (4).

#### التوصيات :

- العمل على تطوير الخوارزمية المحسنة وذلك من خلال تحسين معامل التصبيب  $m$  الذي يتحكم في سلوك الخوارزمية.
- العمل على تهجين الخوارزمية المقترحة مع خوارزميات أخرى.

#### Abstract

as far away from each other as possible. The random start point affects the effectiveness of the aggregation process and results. The convergence process is mainly dependent on the primary centers, so a new relationship was suggested for selecting the primary network weights based on the data studied.

The proposed algorithm was studied in the field of data mining and medical image segmentation (magnetic resonance images of the brain) "MRI images" using MATLAB. The results showed that the improved FKCN algorithm is better than the normal FKCN algorithm compared to the number of iterations and execution time.

**References**

1. NG,R., HAN,J. "Efficient and Effective Clustering Methods for Spatial Data Mining". Conf, ,2008, P144.
2. HAND,D. MANNILA,H. SMYTH,R. " Principles of Data Mining". MIT Press, London, P285,2001.
3. KAUFMAN,L. ROUSSEEUW,P." Finding Groups in Data: an Introduction to Cluster Analysis". John, ,2010,P170.
4. SIVAKUMAR K., "Image Segmentation ".School of Electrical Engineering & Computer Science (School of EECS),2002.
5. UMBAUGH S.E. "Computer Vision and Image Processing" . prentics HALL PTR . USA,2006.
6. AHIRWAR A., "Study of Techniques used for Medical Image Segmentation and Computation of Statistical Test for Region Classification of Brain MRI",Information Technology and Computer Science, Vol. 5, PP.44-53, 2013.
7. ACHARYA T., RAY A.K.," Image Processing Principles and Application", Wiley & Sons Inc.,Hoboken, New Jersey, 2005.
8. BEZDEK, J.C., TSAO, E.C-K., PAL, N.R." Fuzzy Kohonen Clustering Networks". IEEE ,1992, PP 1035-1043.
9. NAES, T., MEVIK, B. H. "The flexibility of fuzzy clustering illustrated by examples". Journal of Chemometrics,13, 1999, PP 435-444.
10. IRWANSYAH, E.,FAISAL, M., PRIMADINI, A." Does Efficient Fuzzy Kohonen Clustering Network Algorithm Really Improves Clustering Data Result?". Journal of Theoretical and Applied Information Technology, Vol.71, No.1, 2015, PP 144-148.
11. DE SOUZA, R.M.C.; DE CARVALHO, F.A.T."A Clustering Method for Mixed Feature-Type Symbolic Data using Adaptive Squared Euclidean Distances". 7<sup>th</sup> International Conference on Hybrid Intelligent Systems (HIS),2007,P 173.
12. THE MATH WORKS ,INC. "Matlab Toolbox", The Math Works Inc.1999.
13. DHARMENDRA K. et al, "Review on different Techniques of Image Segmentation using MATLAB", International Journal of Science Engineering and Technology. Vol. 5, Issue 2, 2017, PP 41-44.
14. HALL, L. O., BENSALID, A. M., CLARKE, L. P., VELTHUIZEN, R. P., SILBIGER M.S., BEZDEK J.C."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.