

A Proposed Method for Feature Selection using a Binary Particle Swarm Optimization Algorithm and Mutual Information Technique

Mustafa Ayham Abed Alhafedh

mouhamed_milan@yahoo.com

Omar Saber Qasim

omar.saber@uomosul.edu.iq

College of Computer Science and Mathematics
University of Mosul, Mosul, Iraq

Received: 07/04/2019

Accepted: 30/06/2019

ABSTRACT

Feature selection is one of the most important issues in improving the data classification process. It greatly influences the accuracy of the classification. There are many evolutionary algorithms used for this purpose, such as the Particle Swarm Optimization (PSO) in discrete space through the Binary PSO concept. The BPSO optimization algorithm derives its mechanism from the default PSO algorithm but in discrete space. In this research, a hybrid approach was proposed between the BPSO algorithm and Mutual Information (MI) to obtain subsets of features through two basic phases: the first is to use the BPSO algorithm to determine the features affecting the data classification process by relying on an objective function. In the second phase, the MI method is used to reduce the number of features identified by the BPSO method. The results of the proposed algorithm have demonstrated efficiency and effectiveness by obtaining higher classification accuracy and using fewer features than default methods.

Keywords: Feature selection; particle swarm optimization; Mutual information technique; classification.

طريقة مقترحة لاختيار الميزات باستخدام خوارزمية أمثلة أسراب الجسيمات الثنائية وتقنية المعلومات المتبادلة

عمر صابر قاسم

مصطفى أيهم عبد الحافظ

قسم الرياضيات

كلية علوم الحاسوب والرياضيات

جامعة الموصل، الموصل، العراق

تاريخ قبول البحث: 2019\06\30

تاريخ استلام البحث: 2019\04\07

المخلص

تعد مسألة اختيار الميزات Feature selection من أهم المسائل المستخدمة في تحسين عملية تصنيف البيانات، إذ تؤثر بشكل كبير على دقة التصنيف، إذ يوجد العديد من الخوارزميات التطورية Evolutionary algorithms التي تستخدم لهذا الغرض، مثل خوارزمية تحسين أسراب الجسيمات Particle swarm optimization (PSO)، بعد تحويلها من الفضاء المستمر Continuous إلى الفضاء المنقطع Discrete من خلال مفهوم Binary PSO، إذ تستمد خوارزمية تحسين سرب الجسيمات الثنائية BPSO آلية عملها من خوارزمية PSO الاعتيادية ولكن في مساحة بحث متقطعة ثنائية. اقترحت في هذا البحث طريقة مهجنة بين خوارزمية BPSO مع تقنية المعلومات المتبادلة Mutual information (MI) للحصول على مجموعات فرعية من الميزات من خلال

مرحلتين أساسيتين: الأولى هي استخدام خوارزمية BPSO لتحديد الميزات المؤثرة في عملية التصنيف للبيانات بالاعتماد على دالة هدف Objective function مقترحة ، في المرحلة الثانية، تستخدم طريقة المعلومات المتبادلة (MI) لتقليص عدد الميزات التي تم تحديدها من خلال طريقة BPSO ، وقد اثبتت نتائج الخوارزمية المقترحة الكفاءة والفعالية من خلال الحصول على دقة تصنيف أعلى و استخدام ميزات قليلة مقارنة بالطرائق الافتراضية الاعتيادية. **الكلمات المفتاحية:** اختيار الميزة، تحسين سرب الجسيمات، تقنية المعلومات المتبادلة، التصنيف.

1. المقدمة Introduction

أجريت في السنوات الأخيرة العديد من الدراسات التي تعتمد على محاكاة الأنظمة الذكية الطبيعية Normal intelligent systems ومنها ما يسمى بذكاء السرب Swarm intelligence الذي يمثل الأنظمة التي تعتمد على تفاعل السلوك الجماعي للنظام مع بيئتها، وبهذا يمكن القول إن العالم بدأ تَوَّأ يرى الطاقة الهائلة والخيارات المتعددة التي يمكن الحصول عليها من السلوكيات البسيطة لبعض الحشرات والحيوانات الصغيرة التي تعيش في مجموعات أو أسراب. هنالك خوارزميات عديدة تحاكي هذه الأسراب ومن ضمنها خوارزمية Particle swarm optimization (PSO) وهي واحدة من خوارزميات الاستدلال التي اقترحها الباحثان Eberhardt و Kennedy في عام 1995 وجذبت انتباه العديد من الباحثين حول العالم [1]، إذ تعتمد هذه الخوارزمية على المجتمعات أو الأسراب Swarm التي شكلت بواسطة حلول عشوائية تسمى الجسيمات Particles ، إذ يرتبط كل جسيم في هذه الخوارزمية بالسرعة Velocity وتطير هذه الجسيمات عبر منطقة البحث بسرعات متفاوتة من منطقة إلى أخرى ويكون لكل جسيم سرعته وموقعه Position الخاص الذي يمكن ضبطه وفقاً لسلوك هذه الجسيمات [2]، إذ أن الجسيم لديه القدرة على تغيير موقعه بحثاً عن منطقة أفضل داخل عملية البحث [3, 4].

في العام 1997، قام العالمان Eberhardt و Kennedy بتطوير خوارزمية PSO الأساسية، بتحويل خوارزمية البحث من مساحة البحث المستمر Continuous search space إلى مساحة البحث المنقطعة Discrete search space [5]. تعمل خوارزمية PSO المعدلة على مناطق البحث ثنائية الأبعاد two-dimensional، وتسمى الخوارزمية المعدلة الخوارزمية الثنائية Binary particle swarm optimization (BPSO)، إذ تمثيل الجسيمات في الفضاء الثنائي ويمكن أن تأخذ مواقع هذه الجسيمات قيمة ثنائية من (0,1) [6].

طريقة اختيار الميزة Feature selection هي طريقة لتقليل عدد الميزات غير الضرورية من مجموعة الميزات الأصلية، إذ تستخدم هذه الطريقة عندما يكون لدينا عدد كبير من الميزات في مجموعة البيانات، إذ يثبت عدد محدد من الميزات الضرورية لتحسين دقة التصنيف [7]. عادة ما تستخدم طريقتان أساسيتان في تحديد الميزة: الأولى هي طرائق التصفية Filtering approaches، والثانية هي طرائق الالتفاف Wrapper approaches. تعتمد طرائق التصفية على معايير محددة من خلال الحصول على معلومات لكل ميزة، كما أن اختيار الميزة يعتمد على أداء خوارزمية التصنيف [8]. المعلومات المتبادلة (MI) هي إحدى طرائق التصفية التي يتم حسابها بين متغيرين عشوائيين منفصلين باستخدام الإنتروبيا Entropy، إذ يقيس الإنتروبيا للمتغير العشوائي متوسط كمية المعلومات المطلوبة لوصف المتغير العشوائي التي بها ترتب الميزات على وفق أهميتها من الأعلى تأثيراً إلى الأقل تأثيراً في عملية التصنيف [9].

2. الاعمال السابقة

يتم في هذه الفقرة التركيز على أهم تفاصيل الأعمال البحثية المختلفة المتعلقة بهذه الدراسة ومن ضمنها خوارزمية تحسين سرعة الجسيمات Particle swarm optimization، واختيار الميزة Feature selection ونستعرضها على النحو الآتي:

فقد قام الباحثان David W. Aha، Richard LBankert. في العام 1993 باستخدام خوارزمية اختيار الميزة أدارة" للتعقب بالطقس مثل الضباب والاضطرابات الجوية، وحركة العواصف المدارية.

كما قام الباحثان Anil Jain، Douglas Zongker في عام 1997 بتوظيف اختيار الميزة المثلى لتصنيف الأراضي على أساس صور الأقمار الصناعية SAR باستخدام أربعة نماذج نسيج مختلفة التي توصف بكونها ميزات تجميع مستمدة من نماذج نسيج مختلفة، تليها ميزة اختيار النتائج في تحسين دقة التصنيف.

أما في العام 2003 فقد قام الباحثان Isabelle Guyon، Andre Elisseeff باستخدام تقنية اختيار الميزة لوصف متغير Variable معين يستخدم في معالجة النصوص من مجموعة متغيرات.

كما عمد الباحثان DW van der Merwe، AP Engelbrecht في العام 2003 الى استخدام خوارزمية تحسين سرب الجسيمات PSO لاختيار البيانات والحصول على دقة عالية في التصنيف.

أيضاً في 2009 قام الباحثون Pablo A. Estevez, Michel Tesmer, Claudio A. Perez, Jacek M. Zurada باستخدام طرائق التصفية لاختيار الميزة المعتمدة على المعلومات المتبادلة.

كذلك قام الباحثون Neveen Ghali، Hossam M. Zawbaa، Mona Nagy Elbedwehy، Aboul Ella Hassanien في عام 2012 باستخدام تشخيص بمساعدة الكمبيوتر للكشف عن مرض صمام القلب باستخدام خوارزمية تحسين سرب الجسيمات الثنائية (BPSO) Binary particle swarm optimization.

كما قام الباحثان Muhammad Aliyu Sulaiman، Jane Labadin في عام 2015 باستخدام خوارزمية اختيار الميزة على أساس المعلومات المتبادلة للتعلم الآلي.

وفي عام 2018 قام الباحثون Maryam Zaffar، Manzoor Ahmed Hashmani، K.S. Savita، Syed Sajjad Hussain Rizvi بدراسة خوارزميات اختيار الميزة للتعقب بالأداء الأكاديمي كما قدموا تحليلاً لأداء خوارزميات اختيار ميزة الفلتر وخوارزميات التصنيف في مجموعتي بيانات مختلفتين.

3. تحسين سرعة الجسيمات Particle swarm optimization

تعد خوارزمية PSO من الخوارزميات التي تحاكي سلوك المجموعات مثل مجموعات الأسماك Fish والحشرات Insects والطيور Birds التي تطير بحثاً عن الطعام بالتعاون بين أعضاء المجموعة، إذ طبق هذه الخوارزمية العالمان Kennedy و Eberhard في العام 1995، إذ تعتمد على مجموعة من القيم العشوائية التي تسمى الجسيمات Particles [10] و يمتلك كل جسيم فيها سرعة وموقعا" خاصا" به والذي من الممكن تعديله ، إذ تتحرك هذه الجسيمات داخل مساحة البحث العشوائي [1]. توصف المعادلات الأساسية للسرعة والموقع على النحو الآتي:

$$v_{id} = v_{id} + c1 \text{Rand}() (p_{id} - x_{id}) + c2 \text{Rand}() (p_{gd} - x_{id}) \dots (1)$$

$$x_{id} = x_{id} + v_{id} \dots (2)$$

إذ أن c_1 و c_2 هي ثوابت اختيارية موجبة ، وإن $\text{Rand}()$ و $\text{Rand}()$ هما دالتين عشوائيتين في مدى $[0,1]$

وتتمثل $X_i = (x_{i1}, x_{i2}, \dots, x_{id})$ والجسيمات وكذلك تمثل $P_i = (p_{i1}, p_{i2}, \dots, p_{id})$ أفضل وضع سابق (الموضع الذي يعطي أفضل قيمة لياقة (Fitness value) للجسيمات، كما يمثل الرمز g مؤشر أفضل جسيم (the index of the best particle) بين جميع الجسيمات في المجتمع كما يمثل $V_i = (v_{i1}, v_{i2}, \dots, v_{id})$ معدل تغير السرعة للجسيمات. توضح المعادلة (1) كيف يتم تحديث سرعة الجسيمات (الطيور، الأسماك، الحشرات، إلخ) بشكل حيوي وأن المعادلة (2) تصف كيفية تحديث الموقع للجسيمات [11]، إذ تتكون المعادلة (1) من ثلاثة أقسام:

1. يحدد القسم الأول أنه لا يمكن تغيير السرعة فجأة.
 2. القسم الثاني هو القسم المعرفي الذي يمثل تكبير الجسيم نفسه في تجربة الطيران.
 3. القسم الثالث هو القسم الاجتماعي الذي يمثل التعاون بين الجسيمات.
- في المعادلة (1)، إذا كان مجموع الأقسام الثلاثة على الجانب الأيمن يتجاوز القيمة الثابتة التي يحددها المستخدم، تحدد السرعة على هذا البعد إلى أن يكون $\pm V_{max}$ ، إذ أن سرعته قد وصلت إلى حد أقصى من V_{max} ، ويعدها المستخدم. لدى V_{max} ، جسيمات كبيرة لديها القدرة على الطيران بعيدًا للبحث عن مناطق حل أفضل بينما تمتلك جزيئات V_{max} ، الصغيرة إمكانية الوقوع في الحدود المحلية وبالتالي لا يمكنها الطيران والبحث عن منطقة حل أفضل للبقاء في منطقة الحل نفسها، خوارزمية PSO بسيطة في المفهوم، سهلة التنفيذ ولها كفاءة حوسبية عالية رغم عدم وجود أي تعقيد في معادلاتها [5]. إذ يمكن تطبيق خوارزمية PSO بالخطوات الآتية:
1. تحديد مجموعة من الجسيمات ذات المواقع العشوائية والسرعات الأولية (الابتدائية) في منطقة البحث.
 2. تحسين تقييم دالة اللياقة (Fitness function) وتحسين المتغيرات لكل جسيم.
 3. مقارنة تقييم دالة اللياقة (Fitness function) للجسيمات مع p_{best} . إذا كانت القيمة الحالية أفضل من p_{best} ، نقوم بتعيين p_{best} مساوية للقيمة الحالية، و P_i تساوي الموقع الحالي X_i في فضاء البحث.
 4. اختيار الجسيم الأفضل الذي يمتلك أفضل سرعة وموقع، وتعيين مؤشره إلى متغير g .
 5. تغيير سرعة ومكان الجسيم وفقًا للمعادلة (1) و (2).
 6. العودة إلى الخطوة (2) حتى يستوفي المعيار، وعادة ما تكون قيمة دالة اللياقة جيدة بما فيه الكفاية أو قد تصل إلى أقصى عدد من التكرارات [12, 13].

4. خوارزمية تحسين سرب الجسيمات الثنائية (BPSO) Binary particle swarm optimization

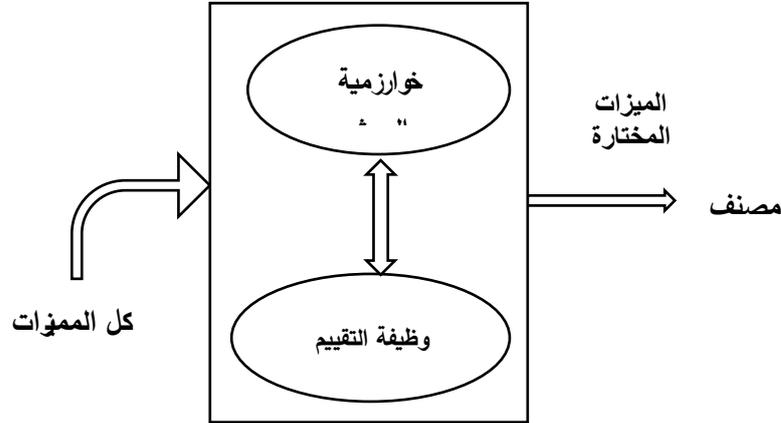
اقترح العالمان Kennedy و Eberhard خوارزمية الأسراب الثنائية (BPSO) في العام (1997)، التي يمكن استخدامها في منطقة بحث في الفضاء المنفصل أو المنقطع. لقد صممت خوارزمية PSO لحل المشكلات في الفضاء المتصل Continuous search space، وطورت خوارزمية PSO للتعامل مع المشاكل الثنائية في الفضاء المنفصل Discrete search space [14]. لتوسيع قيمة PSO الحقيقية إلى الفضاء الثنائي، فإن الجزء المهم هنا هو تطور المفاهيم الأساسية المتمثلة بالموقع Position، والسرعة Velocity في الفضاء الثنائي المنفصل Discrete search space. استخدام السرعة كاحتمال لتحديد ما إذا كان x_{id} في حالة واحدة أو في حالة أخرى (صفر أو واحد). يضغظ باستخدام الدالة اللوجستية $(v) = 1 / (1 + \exp(-v))$ بينما تحسب السرعة باستخدام المعادلة نفسها مثل المعادلة (1). إذا كان الرقم الذي أنشئ عشوائيًا في الداخل [0, 1] أقل من $s(V_{id})$ ، X_{id} سيعين إلى 1، وإلا سيعين إلى 0 [14-17].

إذ يمكن أن نلخص خطوات خوارزمية BPSO على النحو الآتي:

1. نحسب الموقع الابتدائي Initial position والذي يرمز له $X(0)$ والسرعة الابتدائية Initial velocity والذي يرمز لها $V(0)$ لكل جسيم Particle في السرب Swarm بكامله.
2. نقوم بوضع عداد للتكرار Iteration ونحدد عدد التكرارات المطلوبة وليكن العداد $T=0$ ونكرر العملية.
3. نقارن بين لياقة السرب Fitness swarm بكامله ونأخذ أعلى لياقة Fitness.
4. نقوم بتنفيذ الخوارزمية الثنائية (BPSO) Binary particle swarm optimization.
5. نحسب سرعة الجسيمات Particles velocity في كل تكرار والتعديل عليه من المعادلة (1).
6. نحدد مواقع الجسيمات Particles position في كل تكرار والتعديل عليه من المعادلة (2).

5. اختيار ميزة Feature selection

- يحدد بطريقة اختيار الميزة أهم الخصائص أو الميزات Features المستخدمة في وصف مجموعة البيانات Dataset ، إذ يكون الهدف من اختيار الميزة هو زيادة دقة التصنيف وتقليل عدد السمات المستخدمة في خوارزمية التصنيف [18]. تتكون معظم خوارزميات تحديد الميزة من ثلاثة أساليب (أجزاء) رئيسية:
1. خوارزمية البحث Search algorithm: يبحث في مجموعة فرعية من الخصائص التي تحتوي على عدد من الميزات الأصلية.
 2. وظيفة التقييم Evaluation function: تقوم هذه الوظيفة على أساس تقييم الميزات المهمة والأساسية وتميزها عن الميزات الأقل أهمية.
 3. المصنف Classifier: إن الوظيفة الأساسية للمصنف بعد تقييم الميزات سوف يفصل بين الميزات الأساسية المهمة وغير المهمة [19].



الشكل (1): يمثل مخطط عام لعملية الالتفاف Wrapper.

إذ تعتمد أساليب الالتفاف Wrapper وهي إحدى أقسام اختيار الميزة على تقييم جودة وكفاءة خصائص معينة من البيانات [20] بالاعتماد على دالة اللياقة Fitness function كامابين في الشكل (1) واستنادًا إلى خوارزمية تعلم محددة (مثل خوارزمية أسراب الطيور PSO)، إذ تعتمد هذه الطريقة على خطوتين: (1) البحث عن مجموعة فرعية من الخصائص للبيانات الأصلية (2) تقييم هذه الميزات والسمات المحددة بدالة اللياقة. تكرر الخطوات (1) و (2) حتى يتم الوصول إلى بعض معايير التوقف التي تحدد مسبقًا.

6. المعلومات المتبادلة Mutual Information

تعد تقنية المعلومات المتبادلة (MI) Mutual Information من ضمن طرائق اختيار الميزة من نوع التصفية (Filtering) التي يقاس بها كم المعلومات التي يمكن الحصول عليها من مصدر معلومات معين ومقارنتها مع معلومات تم الحصول عليها من مصدر معلومات آخر وتصفيتها حسب المعلومات المشتركة بين المصدرين [9]. مثل شبكة الاتصالات إذ يساعد على فهم مقدار المعلومات المتبادلة في زيادة كم المعلومات المتبادلة بين الإشارات المرسله والإشارات المستقبلة [21]. وتقاس كمية المعلومات المتبادلة بين متغير عشوائي X ومتغير عشوائي آخر Y طبقاً للمعادلة:

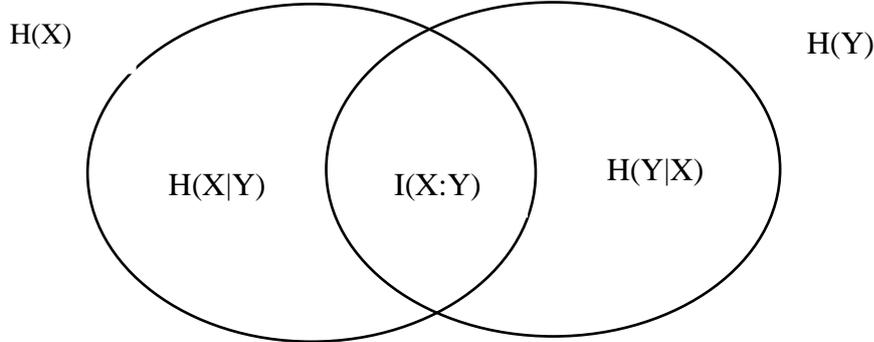
$$I(X; Y) = \sum P(X, Y) \log \frac{P(X, Y)}{P(X)P(y)} \quad \dots (3)$$

$$\begin{aligned} &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X, Y) \quad \dots (4) \end{aligned}$$

إن طريقة المعلومات المتبادلة يمكن تطبيقها لتصفية أعداد الميزات Features في البيانات الأصلية التي يمكن ترتيبها حسب الأهمية من خلال قياس العلاقة بين ميزات معينة مع تصنيفات الفئة. هناك استخدام كلاسيكي لنظرية المعلومات في العديد من مقاييس ترتيب السمات وان احدى أهم مساهمات نظرية المعلومات MI تكون في مسألة اختيار الميزة Feature selection، إذ تستخدم المعلومات المتبادلة MI في تقييم الميزات بترتيبها تنازلياً حسب أهميتها وصياغة مجموعة الميزات طبقاً للمعادلة:

$$I(X, Y) = \iint f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dx dy \quad \dots (5)$$

ومن الممكن تمثيل المعلمات المتبادلة كما في الشكل (2)



الشكل (2): تمثيل بسيط يوضح الية طريقة المعلومات المتبادلة

أن الطريقة المستخدمة لحساب المعلومات المتبادلة هي قياس الإنتروبيا Entropy واستبدالها في شكل تبادل المعلومات. كذلك عندما تكون المعلومات المتبادلة معياراً مناسباً لاختيار العناصر (الميزات) إذ تعمل دالة اللياقة على زيادة قيمة المعلومات المتبادلة [22].

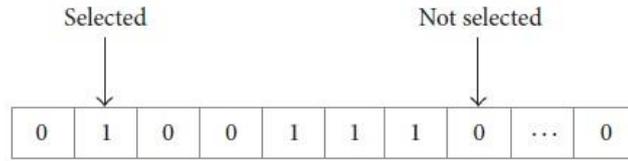
7. الخوارزمية المقترحة BPSO_MI

تتكون الخوارزمية المقترحة في هذا البحث من مرحلتين أساسيتين، تتمثل المرحلة الأولى باستخدام طريقة أمثلة أسراب الجسيمات الثنائية BPSO لتصفية البيانات المتكررة غير الهامة في عملية التصنيف بالتركيز على الميزات الأكثر أهمية التي يختار بالاعتماد على دالة اللياقة (Fitness function (FF)، إذ يختار طبقاً للمعادلة:

$$FF = 0.2 * C + 0.8 * \left(\frac{q}{p - q} \right) \quad \dots (6)$$

إذ تمثل C دقة النموذج المستخدم في التصنيف، q يمثل المجموعة الفرعية التي اختيرت من الميزات، كما أن p يمثل العدد الإجمالي أو الكلي للميزات الأصلية.

الشكل (3) يبين عملية تحديد الميزات من خلال خوارزمية BPSO التي تتم عن طريق متجه يحتوي على سلسلة من القيم الثنائية من الأصفر والوحدات، إذ يمثل العنصر الواحد عملية اختيار الميزة المقابلة للمتجه والقيمة صفر تعني ان الميزة المقابلة لها لم يتم اختيارها



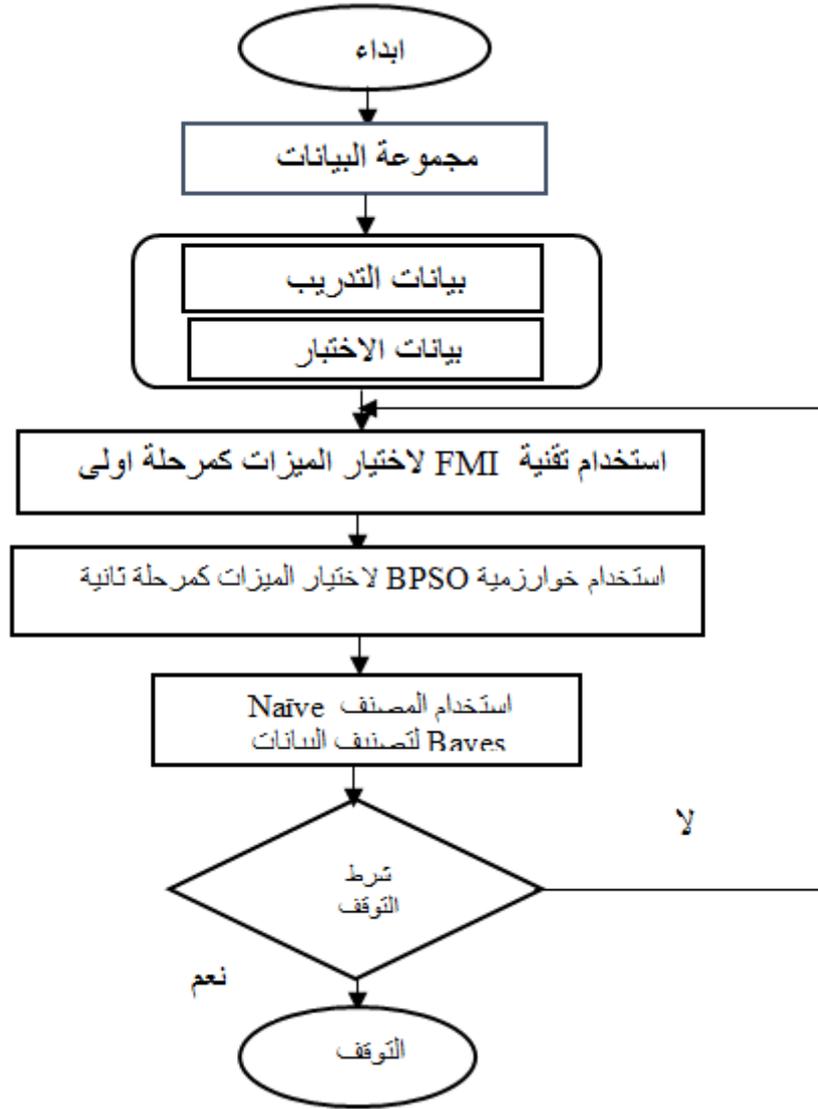
الشكل (3): عينة من حل مجموعة فرعية الميزة.

أما المرحلة الثانية من الخوارزمية المقترحة BPSO_MI فتركز على تقنية المعلومات المتبادلة MI التي يمكن اعتبارها إحدى طرائق المرشح Filtering المستخدمة في اختيار الميزات، إذ يتم إدخال مجموعة الميزات المختارة من المرحلة الأولى (من خلال طريقة BPSO) إلى طريقة MI ليتم ترشيحها وتخفيضها كمرحلة ثانية وفقاً لأهميتها من خلال مفهوم طريقة MI. تم اعتماد عدد قليل من الميزات في طريقة MI مقارنة بالميزات الأصلية للبيانات، اعتمدت إحدى طرائق المصنفات وتحديدًا طريقة Naive Bayes في عملية التصنيف النهائية.

لقد قورنت الطريقة المقترحة BPSO_MI التي تعتمد عدداً أقل من الميزات Features من خلال تقنية المعلومات المتبادلة MI وذلك من خلال التطبيق على خمس مجموعات من البيانات Datasets التي تمتلك عدداً كبيراً من الخصائص، حيث تم الحصول على هذه البيانات من خلال الموقع العالمي (UC Irvine Machine Learning Repository)، كما تم استخدام طريقة الـ Cross-Valid في توزيع مجاميع الاختبار Testing والتدريب Training للبيانات المختارة وبمعدل (20) تكرار، إذ توزع في كل تكرار البيانات بشكل مختلف بين مجموعات التدريب والاختبار ليتم التأكد من مدى فعالية الخوارزمية المقترحة BPSO_MI في الحصول على نتائج أفضل من خلال معيار متوسط مربعات الخطأ MSE مقارنة بالخوارزمية الاعتيادية. أن أهم الخطوات المستخدمة لتصنيف مجموعات البيانات باستخدام الطريقة المقترحة BPSO_MI، تكون على النحو الآتي:

1. تهيئة بيانات مصفوفة الإدخال بشكل مصفوفة X صفوفها تمثل عدد الحالات أو الأنماط (Patterns) الكلية وأعمدها تمثل عدد الميزات Features.
2. استخدام طريقة الـ Cross Valind لتقسيم البيانات المرضية إلى مجموعات للتدريب Training وأخرى للاختبار Testing.

3. استخدام خوارزمية أمثلة اسراب الجسيمات الثنائية BPSO ومن خلال مفهوم دالة اللياقة Fitness function (FF) في الحصول على أفضل مجموعات جزئية من الميزات.
4. استخدام تقنية المعلومات المتبادلة MI، لتقليل عدد الميزات التي تم اختيارت بخوارزمية BPSO .
5. اختيار المصنف Naive Bayes لأجل تصنيف بيانات التدريب والاختبار.
6. استخدام معيار MSE لمقارنة نتائج الطريقة المقترحة BPSO_MI مع الطريقة الاعتيادية BPSO.



الشكل (4): مخطط انسيابي يوضح الخوارزمية المقترحة

8. النتائج التجريبية Experimental results

طبقت الخوارزمية المقترحة BPSO_MI على خمس مجموعات من البيانات Dataset ذات الابعاد الكبيرة (او الميزات المتعددة)، والتي تم الحصول عليها من مستودع UCI العالمي، إذ تشترك جميع هذه البيانات المستخدمة في كونها ثنائية التصنيف، إذ يمكن من خلال الجدول (1) عرض وصفٍ لهذه البيانات المستخدمة.

الجدول (1): يبين وصف عام لمجموعات البيانات المستخدمة.

عدد الميزات المختارة بالطريقة المقترحة Features selection	عدد الميزات الكلية للبيانات الاصلية All features	مجموعات البيانات Dataset
10	2000	Colon
3	15	Japanese
3	8	Pima
3	195	park
3	14	Australian

لقد طبقت الخوارزمية المقترحة BPSO-MI على جميع البيانات الموصوفة في الجدول (1) ومقارنتها مع الخوارزمية الاعتيادية BPSO دون اختيار ميزة، وذلك للتحقق من النتائج بشكل دقيق، وبما أن هاتين الخوارزمتين تعملان في البيئة نفسها، والمعلومات، ومجموعات البيانات، لذا فإن النتائج قابلة للمقارنة تمامًا. ويبين الجدول (2) آلية هذه المقارنة على النحو الآتي:

الجدول (2): يبين مقارنة بين الخوارزمية المقترحة BPSO-MI والخوارزمية الاعتيادية BPSO من خلال معيار MSE بالنسبة لبيانات التدريب.

الخوارزمية الاعتيادية BPSO	الخوارزمية المقترحة BPSO- MI	مجموعات البيانات Dataset
73.409%	81.136%	Colon
80.559%	82.174%	Japanese
75.985%	76.487%	Pima
70.766%	82.591%	park
80.807%	81.863%	Australian

الجدول (3): يبين مقارنة بين الخوارزمية المقترحة BPSO-MI والخوارزمية الاعتيادية BPSO من خلال معيار MSE بالنسبة لبيانات الاختبار.

الخوارزمية الاعتيادية BPSO	الخوارزمية المقترحة BPSO- MI	مجموعات البيانات Dataset
84.722%	%87.222	Colon
82.222%	%83.527	Japanese
77.522%	%77.696	Pima

72.414%	%84.655	park
82.295%	%83.430	Australian

نلاحظ من خلال النتائج في الجداول (1,2,3) أعلاه أن الخوارزمية المقترحة BPSO-MI أعطت ميزات أقل ونتائج أفضل من الخوارزمية الاعتيادية BPSO على جميع البيانات الخمسة المستخدمة والتي اعتمدت كبيانات تدريب أو اختبار، مما يؤكد كفاءة الخوارزمية المقترحة وذلك بمقياس متوسط مربع الخطأ MSE.

9. الاستنتاجات والتوصيات Conclusions and Recommendation

تضمنت هذه الدراسة اقتراح خوارزمية أمثلة أسراب الجسيمات الثنائية BPSO مع تقنية المعلومات المتبادلة MI كطريقة لتحسين أداء التصنيف بالاعتماد على مجموعات فرعية من الميزات Features المهمة لمجموعة البيانات. استخدم مصنف Naive Bayes لتصنيف البيانات التي تم الحصول عليها من الخوارزمية المقترحة BPSO-MI فضلاً عن الخوارزمية الافتراضية. تظهر مقارنة النتائج التي تم الحصول في الجداول (2) و (3) أن الخوارزمية المقترحة BPSO-MI لتقليل عدد الميزات لديها دقة تصنيف أفضل من الخوارزمية الافتراضية (بالنسبة لبيانات الاختبار والتدريب)، أزيلت الميزات غير الضرورية للطريقة المقترحة BPSO_MI من خلال مرحلتين وبالتالي تحسين نتائج التصنيف الشاملة.

كما نوصي بدراسة تطوير خوارزمية أمثلة أسراب الجسيمات الثنائية BPSO مع تقنيات أخرى مثل Fuzzy logic أو خوارزميات ذكاء السرب Intelligence Swarm من خلال إيجاد أفضل الميزات Features المستخدمة لتحسين أداء التصنيف Classification، كما يمكن اعتماد الخوارزمية المقترحة BPSO-MI في تصنيف العديد من البيانات ذات الأبعاد الكبيرة مثل الجينية وغيرها.

المصادر

- [1] T. S. Babu, J. P. Ram, T. Dragičević, M. Miyatake, F. Blaabjerg, and N. Rajasekar, "Particle swarm optimization based solar PV array reconfiguration of the maximum power extraction under partial shading conditions," *IEEE Transactions on Sustainable Energy*, vol. 9, pp. 74-85, 2018.
- [2] S. Cheng, H. Lu, X. Lei, and Y. Shi, "A quarter century of particle swarm optimization," *Complex & Intelligent Systems*, 2018.
- [3] S. Dutta and S. Singh, "Optimal rescheduling of generators for congestion management based on particle swarm optimization," *IEEE Transactions on Power Systems*, vol. 23, pp. 1560-1569, 2008.
- [4] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on, 1995*, pp. 39-43.
- [5] S. Rafi, A. Kumar, and G. K. Singh, "An improved particle swarm optimization method for multirate filter bank design," *Journal of the Franklin Institute*, vol. 350, pp. 757-769, 2013.
- [6] M. Abdul-Razek, "Detection of Heart Disease using Binary Particle Swarm Optimization," *Damietta University Publication*, vol. 1, 2014.
- [7] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, pp. 309-313, 2015.
- [8] M. Mafarja and S. Mirjalili, "Whale optimization approaches for wrapper feature selection," *Applied Soft Computing*, vol. 62, pp. 441-453, 2018.
- [9] E. Hancer, B. Xue, and M. Zhang, "Differential evolution for filter feature selection based on information theory and feature ranking," *Knowledge-Based Systems*, vol. 140, pp. 103-119, 2018.
- [10] A. Rezaee Jordehi, "Enhanced leader particle swarm optimisation (ELPSO): An efficient algorithm for parameter estimation of photovoltaic (PV) cells and modules," *Solar Energy*, vol. 159, pp. 78-87, 2018.
- [11] Z. Wu, W. Gu, R. Wang, X. Yuan, and W. Liu, "Economic optimal schedule of CHP microgrid system using chance constrained programming and particle swarm optimization," in *Power and Energy Society General Meeting, 2011 IEEE, 2011*, pp. 1-11.

-
- [12] Y. Shi and R. C. Eberhart, "Empirical study of particle swarm optimization," in *Evolutionary computation*, 1999. CEC 99. Proceedings of the 1999 congress on, 1999, pp. 1945-1950.
- [13] R. Thangaraj, M. Pant, A. Abraham, and P. Bouvry, "Particle swarm optimization: hybridization perspectives and experimental illustrations," *Applied Mathematics and Computation*, vol. 217, pp. 5208-5226, 2011.
- [14] M. S. Mohamad ,S. Omatu, S. Deris, M. Yoshioka, A. Abdullah, and Z. Ibrahim, "An enhancement of binary particle swarm optimization for gene selection in classifying cancer classes," *Algorithms for Molecular Biology*, vol. 8, p. 15, 2013.
- [15] M. N. Elbedwehy, H. M. Zawbaa, N. Ghali, and A. E. Hassanien, "Detection of heart disease using binary particle swarm optimization," in *Computer Science and Information Systems (FedCSIS)*, 2012 Federated Conference on, 2012, pp. 177-182.
- [16] A. H. El-Maleh, A. T. Sheikh, and S. M .Sait, "Binary particle swarm optimization (BPSO) based state assignment for area minimization of sequential circuits," *Applied Soft Computing*, vol. 13, pp. 4832-4840, 2013.
- [17] I. Jain, V. K. Jain, and R. Jain, "Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification," *Applied Soft Computing*, vol. 62, pp. 203-215, 2018.
- [18] Y. Zhang, D. Gong, Y. Hu, and W. Zhang, "Feature selection algorithm based on bare bones particle swarm optimization," *Neurocomputing*, vol. 148, pp. 150-157, 2015.
- [19] M. Zaffar, K. Savita, M. A. Hashmani, and S. S. H. Rizvi, "A Study of Feature Selection Algorithms for Predicting Students Academic Performance," *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, vol. 9, pp. 541-549, 2018.
- [20] B. Liu, E. Blasch, Y. Chen, D. Shen, and G. Chen, "Scalable sentiment classification for big data analysis using naive bayes classifier," in *Big Data*, 2013 IEEE International Conference on, 2013, pp. 99-104.
- [21] Y. Prasad, K. K. Biswas, and M. Hanmandlu, "A recursive PSO scheme for gene selection in microarray data," *Applied Soft Computing*, vol. 71, pp. 213-225, 2018.
- [22] M. A. Sulaiman and J. Labadin, "Feature selection based on mutual information," in *IT in Asia (CITA)*, 2015 9th International Conference on, 2015, pp. 1-6.