تهجين خوارزمية *Salp Swarm* باستخدام الخوارزمية الجينية لتطوير منهجية جديدة لإيجاد شجرة القرار الأمثلية

الطالبة: هدى علي حبش

جامعة البعث – كلية العلوم – قسم الرياضيات

المشرف: د. زكريا زكريا – المشرف المشارك: د. عبد الله العمر

الملخص

تكمن أهمية نظم دعم اتخاذ القرار الناتجة عن تطوُّر تقنية المعلومات في تركيزها على توفير الدعم المناسب لتحسين جودة القرار، واعتمادها على تَوَفُّر المعلومات الكافية وملاءمة النماذج لتحليل المشكلة.

في بحثٍ سابق تمَّت دراسة أشجار القرار وطرق بنائها وتطويرها، وتمَّ اقتراح خوارزمية هجينة تَزِيد من موثوقية النتائج التي نحصل عليها عن طريق الوصول إلى شجرة القرار المثالية التي تتفوَّق على المنهجيات السابقة. وفي هذا البحث سنتابع العمل على هذا الموضوع واستنتاج خوارزمية هجينة عن طريق دمج الخوارزمية الجينية مع خوارزمية Salp Swarm (سرب الهلاميات البرميلية).

قمنا في هذا البحث بدمج خوارزمية Salp Swarm والخوارزمية الجينية لبناء أشجار قرار أمثلية، وقد أظهرت النتائج تَحَسُّن شجرة القرار التي حصلنا عليها من الخوارزمية الهجينة عن الخوارزميتين كلٍّ على حدة وفق معايير الأداء الخاصة بأشجار القرار والتي سنتحدث عنها بالتفصيل في هذا البحث.

الكلمات المفتاحية:

نظم دعم اتخاذ القرار، تحسين جودة القرار، الخوارزمية الجينية، شجرة القرار، خوارزمية Salp swarm.

Hybridization of the Salp Swarm algorithm using Genetic algorithm to develop a new methodology for finding the optimal decision tree

Huda Ali Habash

ALBaath University – College of Science – Mathematics department

Abstract:

Decision support systems (DSS) are considered the resultant of the significant advancement in information technology. As these systems provides the suitable support to improve the quality of human decision and depends on the availability of sufficient information as well as fitting the models to analyze the problem.

In earlier research we have studied decision trees along with different methods to build and develop, besides, we have proposed a hybrid approach that increased the quality of the results we got by reaching the optimal decision tree that outperformed the earlier methodologies.

In this research we go on with our efforts to develop a new hybrid algorithm that merges genetic algorithms with Salp swarm optimization approach.

In this research, we combined the Salp Swarm algorithm and the genetic algorithm to build optimal decision trees, and the results showed that the decision tree we obtained from the hybrid algorithm improved over the two algorithms separately according to the performance criteria of decision trees, which we will talk about in detail in this research.

Keywords:

Decision support systems, decision making improvement, genetic algorithms, decision tree, Salp swarm optimization

1. مقدمة:

تعتبر قواعد البيانات اليوم من المجالات الهامة والمستخدمة بشكل واسع في هذا العصر.. عصر المعلوماتية. ومع ازدياد قواعد البيانات وضخامة محتوياتها كان لابد من ظهور أدوات وخوارزميات تساعد في التنقيب في البيانات وإعطاء المستخدم المعلومات المفيدة منها.

تُعتبر نظم دعم اتخاذ القرار الذكية من تطبيقات الذكاء الصنعي، فهي تساعد على اتخاذ القرار في كافة المجالات، الاقتصادية والزراعية وفي مجال الطقس وغيرها... [1].

في الواقع تُعتبَر أشجار القرار من أهم خوارزميات بناء نظم دعم اتخاذ القرار، والتي تعتبر من أهم أدوات تحليل القرار التي تساعد على اتخاذ القرار بالتوازي مع النماذج الاحتمالية وخوارزميات الاختيار [2].

ذكاء السرب (Swarm Intelligence) هو مصطلح يصف السلوك الجماعي للنظم سواء كانت طبيعية أم صناعية. وقد قدمه جيراردو بيني (Gerardo Beni) وجينغ وانغ (Jing wang) في عام 1989 في سياق النظم الروبوتية [3].

بدأت الأبحاث في هذه الخوارزميات مع بداية التسعينات من القرن الماضي، فبدأت بمحاكاة الكائنات الحية الأقل ذكاءً والتي لها إمكانيّات محدودة كالنمل والطيور والأسماك والتي تُبدي سلوكاً جماعياً شديد الذكاء [4].

ومن الجدير بالذكر أنَّ عملية إيجاد الشجرة الأفضل التي تُناسِب البيانات هي عملية NP (Nondeterministic Polynomial) والتي لا تُحَل بزمن نموذجي [5]، وتنقسم بدورها إلى NP Hard و NP Complete، وإنَّ موضوع بحثنا هذا يندرج تحت قسم NP Complete، لذلك فإنَّ النتائج التي سنحصل عليها ومقارنتها مع الطرق السابقة ستبين فعالية هذه الطريقة ودقة نتائجها تؤكد أفضليتها على الخوارزميات الأُخرى.

سنعرض في هذا البحث تفاصيل المنهجية المُتَّبَعَة ونوضح نتائجها على مجموعة بيانات حقيقية حيث تُظهِر المقارنة تَحَسُّن نتائج هذه الطريقة عن نتائج الطرق السابقة في مجال

بناء أشجار القرار. كما سنقارن في نهاية البحث نتائج الخوارزمية الحالية مع الخوارزميات المستخدمة سابقاً ومع الخوارزمية الهجينة في بحثي السابق [19].

وسنبين الهدف من البحث، كما سنعرض باختصار الدراسات السابقة والطرق المقترحة في مجال بناء أشجار القرار، ثمَّ سنطرح مشكلة البحث، وما هي المنهجية التي تمَّ اقتراحها لدمج الخوارزميتين Salp Swarm و Genetic، ثمَّ نعرض النتائج التي حصلنا عليها من الخوارزمية الهجينة والتي تبين تَفَوُّق الخوارزمية الجديدة المقترحة على الخوارزميات الأُخرى.

2. هدف البحث:

يهدف هذا البحث إلى تعديل خوارزمية الهلاميات البرميلية Salp Swarm ودمجها مع الخوارزمية الجينية Genetic Algorithm لتطوير طريقة جديدة يمكن من خلالها إيجاد شجرة القرار الأفضل المرتبطة بمجموعة بيانات وبحيث تكون شجرة القرار المُستَنتَجَة ذات قدرة أعلى من أشجار القرار المبنية بالطرق التقليدية الأُخرى، وبالتالي يكون الهدف هو الوصول إلى شجرة القرار المثالية التي تُضاهي بقدراتها المنهجيات المتبعة سابقاً.

3. الدراسات السابقة:

- تم اقتراح المبدأ الاساسي للخوارزمية الجينية (Genetic Algorithm) عام 1975 من قبل العالم هولاند، وذلك وفقاً لنظرية دارون والتي تعتمد على مبدأ البقاء للأصلح. بحيث تعمل الخوارزمية اعتماداً على مبدأ أن الأفراد الذين هم أكثر عرضة للبقاء على قيد الحياة لفترة أطول هم الأشخاص الأفضل من حيث سماتهم الوراثية [6].

على مدى أجيال عديدة، تنتشر هذه المادة الوراثية الملائِمة لعدد متزايد من الأفراد، حيث يمثِّل كل فرد –في الخوارزمية الجينية، بمعنى آخر كل "كروموسوم" Chromosome ينتمي إلى المجتمع– حلاً محتملاً للمشكلة، وتمر الخوارزمية

الجينية بعدة مراحل من أجل أداء وظيفتها، هذه المراحل هي: التهيئة والتقييم والاختيار والعبور (التزاوج) والطفرة [7].

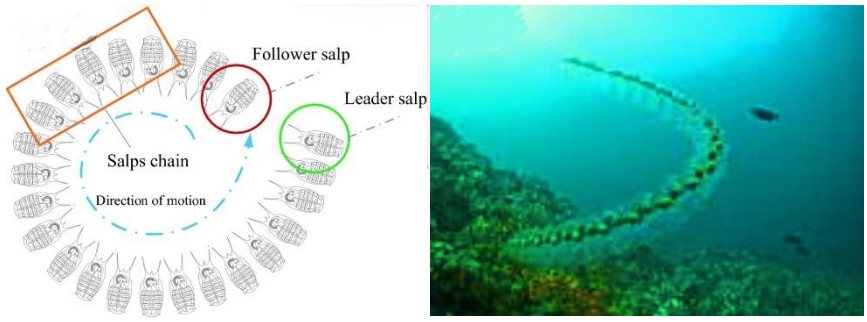
- في عام 1995 اقترح العالم الإيطالي ماركو دوريغو Marco Dorigo خوارزمية أمثلية مستعمرة النمل Ant Colony Optimization (ACO)، التي تُحاكي مستعمرات النمل [8] (والتي تمَّ تطوير منهجية معتمِدة عليها وتمَّ استخدامها والحديث عنها مفصلاً في بحثي السابق) [19].
- كما اقترح كلاً من العالمين Russell Eberhart و James Kenndy الخوارزمية Particle Swarm Optimization (PSO) والتي تعتمد بشكل رئيسي على محاكاة أسراب الطيور [9].
- الخوارزميتان ACO و PSO كانتا بداية لفرع جديد من فروع الذكاء الاصطناعي، هو ذكاء السرب SI (Swarm Intelligence) [10].
- تُعَدّ خوارزمية Salp Swarm واحدة من الخوارزميات الهامة التي أُقتُرِحَت والتي تحاكي تقنيات ذكاء الأسراب أو القطعان [11].
- يظهر ذكاء السرب عادة في مجتمعات من العملاء المحدودين والذين يتفاعلون مع بعضهم البعض ومع بيئتهم. وعلى الرغم من وجود متحكِّم مركزي يملي على الأفراد كيف يتصرفون. تؤدي التفاعلات المحلية، والعشوائية عموماً، بينهم إلى ظهور سلوك كلي ذكي غير ملحوظ على مستوى الأفراد [12].
- ومن الأمثلة الطبيعية على هذه الظاهرة مستعمرات النمل وأسراب الطيور وقطعان الماشية ونمو البكتريا وأفواج الأسماك [13].
- تعود خلفية هذه الخوارزميات إلى مصطلح الحياة الصنعية A–life (Artificial Life)، الذي يدرس نظماً من صنع الانسان وتملك خصائص الحياة الأساسية، مثل نمذجة حركة الطيور أو النمل في بناء المستعمرات، ويتفرَّع عن هذا المصطلح فرعان أساسيان:

– الدراسات المهتمة بتقنيات الحوسبة والمساعدة في نمذجة الظواهر البيولوجية.

- الدراسات المهتمة بالظواهر البيولوجية وكيفية توظيفها الصحيح في حل المشكلات الحاسوبية، مثل الخوارزميات الجينية والشبكات العصبونية [14].

4. مشكلة البحث:

تتحرك أسماك Salp (الهلاميات البرميلية) في البحر بشكل خط مستقيم، ويوجد قائد للمجموعة، ويتبعه بقيّة أفراد السرب [13]، كما هو موضح في الشكل(1).



الشكل(1)

تعتمد خوارزمية Salp Swarm على هذه الفكرة، حيث يتجه القائد باتجاه الحل الأمثل –الأفضل حتى الآن– وتتبعه العناصر الأخرى، ولكن تقوم بحركات عشوائية أثناء الحركة، ويتم استخدام هذه المُحاكاة بحيث يتم إجراء مسح لفضاء الحالات بشكل جيد، وبهذا نكون قد تحركنا باتجاه الحل الأفضل، ويكون احتمال إيجاده أكبر، وبهذه العملية يكون قد تحسن الحل [15].

وبما أن الخوارزمية الجينية تساهم في تحسين الحل أيضاً، لذلك يمكن إجراء تزاوج بين الحلول التي حصلنا عليها من خوارزمية Salp Swarm ، وبهذا نحصل على الحل الأفضل بشكل أسرع.

بعد ذلك نختار أحد الحلول المثلى التي حصلنا عليها ونجعله القائد الجديد، ونُعيد جعل بقية الحلول (أفراد السرب) يتبعونه، وهكذا. وبهذا نحصل على الحل الأفضل بشكل أسرع وأكثر دقة.

لدينا سرب من الهلاميات البرميلية (والتي هي هنا أشجار القرار أو العينات (Sample) والصفات (Attributes) في البيانات الجدولية التي سيتم تدريب النظام الخبير عليها) ينتشر في منطقة محددة بغرض البحث عن الطعام (البحث عن شجرة القرار المُثلَى التي نبحث عنها). وينتشر الطعام في هذه المنطقة عشوائياً، بالإضافة إلى أن الهلاميات لا تعلم مواضع الطعام على نحو واضح. فما هي الطريقة المُثلَى للبحث عن الطعام؟

تبدأ العناصر رحلة البحث من مناطق عشوائية، ومع كل إعادة تقترب العناصر من الحل الأمثل والأكثر دقَّة (المنطقة الممتلئة بالطعام).

إنَّ خوارزمية Salp Swarm (SS) والخوارزمية الجينية GA (Genetic Algorithm) تنحدر من سلالة الخوارزميات التطورية، إذ تمتلك الخوارزميتان بعض نقاط التشابه مثل توليد تجمعات عشوائية وحساب قيمة الملاءمة وتُحَدّث الخوارزميتان وفقاً لتقنيات عشوائية [16].

بينما توجد نقاط اختلاف بينهما، إذ أنَّ خوارزمية SS لا تمتلك معاملات التزاوج والطفرة التي تمتلكها GA. وبهذا نكون قد استفدنا من ميزات كلا الخوارزميتين معاً.

5. المنهجية المقترحة:

1.5. آلية تقييم شجرة القرار:

تتميَّز أشجار القرار عن بعضها بعدة معايير، هي:

(a) معايير قياس الأداء المرتبطة بالبيانات [6]، (حيث الأعلى هو الأفضل) (مثل: الدقة P (Precision)، الاستعادة R (Recall)، ومعيار f_1، حيث f_1 هو المتوسط التوافقي للدقة P والاستعادة R، ويُحسب بالعلاقة:

$$f_1 = \frac{1}{\frac{\frac{1}{P}+\frac{1}{R}}{2}} = \frac{2PR}{P+R} \quad \text{.....} \quad (1)$$

(والمتوسط التوافقي هو أحد المتوسطات (مقاييس النزعة المركزية) وهو المقابل للمتوسط الحسابي)

(b) ارتفاع الشجرة h (حيث الأقل هو الأفضل).

(c) عدد الأوراق L (حيث الأقل هو الأفضل).

لذلك، ومن أجل الحصول على جودة الشجرة، سنستخدم الجمع الموزون Weighted Sum [17] للقيم السابقة. فإذا كان لدينا شجرة ارتفاعها الأقصى h وارتفاعها الحقيقي d وعدد أوراقها L. وعند حساب معامل f_1 (العلاقة (1)) كانت قيمته f. عندئذٍ تكون جودة هذه الشجرة G، وهي:

$$G = a * f + b * \frac{h-d}{h} + c * \frac{2^{h-1}-L}{2^{h-1}} \quad \text{.....} \quad (2)$$

حيث a و b و c هي قيم اختيارية تعطي أفضلية لكل معيار (يمكن وضعها حسب خبير أو حسب حاجة المستخدم)، وتحقق الشرط $a+b+c=1$ ، ويفضل استخدام قيمة عالية لـ a لأن الأهمية هي للدقة.

سوف نستخدم التابع السابق G لقياس دقة شجرة القرار هنا أيضاً.

2.5. آلية العمل:

سيكون لدينا مجموعة من البيانات الجدولية، يُمَثِّل كلّ سطرٍ عنصراً من البيانات يسمى عينة (Sample). وتُمَثِّل الأعمدة –باستثناء العمود الأخير– الصِّفات (Attributes)،

بحيث يُعَبِّر كل عمود عن صفة للعينات، ويسمَّى العمود الأخير الهدف (Target)، ويُمَثِّل القرار المرتبط بكل عنصر عينة.

نريد بناء نموذج لشجرة القرار اعتماداً على هذه البيانات الموجودة بحيث تَعكِس هذه الشجرة ارتباط الهدف مع الخصائص بأحسن طريقة ممكنة، وتُحقِّق شروط الشجرة الأمثلية التي تحدثنا عنها سابقاً، وتكون قابلة للتعميم. أي أنَّ المسألة هي مسألة تعلُّم آلي (Machine Learning).

كل سمكة Salp (هلامية برميلية) تُمَثِّل حلاً عشوائياً، والذي يُمَثِّل جزءاً من شجرة قرار.

في البداية لدينا أشجار القرار المبدئية والتي هي هنا حلول، سنطبق الآن الخوارزمية الجينية بإجراء التزواج بين الحلول التي حصلنا عليها ونعيد هذه العملية أربعة أو خمسة أجيال حتى نحصل على حل جيد.

لاستخدام الخوارزمية الجينية يجب أن نقوم بتمثيل كل شجرة بسلسلة من الأرقام، ويجب أن تكون هذه السلاسل ذات نفس الطول لكي تعمل الخوارزمية الجينية.

لتمثيل الشجرة على شكل سلسلة نقوم بما يلي:

نُرقم الصفات (Attributes) بدءاً من 0 . الرمز الأول ضمن السلسلة هو رقم الصفة المرتبطة بالجذر.

من أجل كل شجرة فرعية نقوم بحذف الجذر وترقيم الصفات المتبقية من جديد وإعادة العملية.

يكون التمثيل النهائي هو دمج لتمثيل الجذر مع تمثيل الشجرة اليمنى ومن ثمَّ الشجرة اليسرى.

$$R(t)=\begin{cases} empty & t=empty \\ R(root)+R(left)+R(right) & otherwise \end{cases}$$

ولكن باعتبار أنَّ أشجار القرار تختلف بأعماقها وبُنيتها فإنَّ الطريقة السابقة لن تقوم بتوليد سلاسل ذات طول متساوي ولا يمكن أصلاً إعادة فك تشفيرها. لذلك سنقوم بما يلي:

نعتبر كل أشجار القرار هي أشجار كاملة (Perfect) [18]، أي أننا سنُضيف عقداً غير ضرورية للأوراق لتتوسع الأشجار ويصبح عدد العقد متساوي في كل الأشجار. مع ملاحظة أنَّ هذه الإضافة لن تتسبَّب في تغيير بنية الشجرة الحقيقية، لأنَّ الوصول إلى ورقة في شجرة القرار عندما لايتم استخدام كامل الصفات هو مُكافِئ لاستخدام أي صفة من هذه الصفات بعد هذه الورقة.

أصبح التمثيل لكل الأشجار متساوي الطول وأصبح التمثيل قابلاً للعكس، لأنَّ كل عقدة تُمَثَّل بدليل محدد، وهو ضِعف دليل الأب. (نظرية معروفة بتمثيل الأشجار الكاملة) [18].

بما أنَّ كل الأشجار أصبحت مُمثَّلة بشكل سلسلة من الأرقام وتُمَثِّل هذه السلاسل الكروموسومات في الخوارزمية الجينية، فيمكن تطبيق الخوارزمية الجينية التقليدية (من اختيار Selection وتزاوج Crossover وطفرة Mutation) ونُعِيدها حتى نصل للتقارب.

للاستفادة من الخوارزميتين معاً حيث تساعدنا خوارزمية Salp على التوليد العشوائي للأشجار بشكل يميل إلى الصحة، و تساعدنا الخوارزمية الجينية على تحسين الأشجار الموجودة، سنقوم بتكرار توليد الأشجار وتزاوجها بشكل متتالي للاستفادة من خصائص الخوارزميتين، حيث نقوم بتوليد مجموعة من الأشجار بشكل عشوائي بطريقة خوارزمية

Salp ثم نقوم بتحسين هذه الأشجار التي قمنا بتوليدها باستخدام الخوارزمية الجينية، وبعد عدد كافٍ من الدورات نقوم باختيار أفضل شجرة (حسب G). ونُعِيد العملية السابقة عدداً كافياً من المرات للوصول لأفضل شجرة.

من الجدير بالذكر أنَّنا استخدمنا الإجراءات التالية في خوارزميتنا:

(1) بالنسبة لتحريك الشجرة: نلاحظ أنه في خوارزمية SS فإن الهلاميات تتحرك سواء باتجاه العنصر الذي يسبقها أو باتجاه عشوائي، فهي تغيّر مكانها باستمرار.

وهنا فإن الهلامية تمثِّل شجرة، فكيف يمكن تحريك الشجرة؟

والجواب هو أننا نختار أحد العقد ونُغَيّرها بشكل عشوائي فنكون قد حرّكنا الشجرة في فضاء الأشجار لدينا، أي تمثِّل كل شجرة بمتجه، أبعاده هو عدد العقد، وكل محور هو عقدة، فعدد المحاور هو عدد العقد، والاحداثيات هي أرقام الصفات Attribute، لذا فالشجرة تُمَثَّل بنقطة، بحسب عدد العقد، في فضاء ذو n بُعداً.

(2) التطابق بين شجرتين: نقارن بين شجرتين من اليسار، وكلما كان التطابق أكثر كلما ازداد تطابق الشجرتين.

(3) تحريك (تقريب) شجرة باتجاه شجرة: إذا كان الجذرين مختلفين فإننا نجعلهما متماثلين، وقد استطعتُ في بحثي السابق[19] كتابة الشجرة على شكل متجه، أي أننا استطعنا تمثيل الشجرة رياضياً، ونستطيع من خلال هذا التمثيل الرياضي فعل كلا أمرين: تحريك شجرة بالنسبة لشجرة أخرى، وتحريك شجرة بشكل عشوائي، بالإضافة إلى تنفيذ الخوارزمية الجينية على ذلك التمثيل الرياضي.

ولتوضيح ذلك: لنفرض الشجرة $T(c_1\, c_2\, \ldots\, c_n)$

حيث: $n=2^{h+1}-1$

h ارتفاع الشجرة المطلوب

لتحريك الشجرة بشكل عشوائي نختار أحد c_i ونُغَيِّرُهُ، علماً أنَّ مجال كل c_i متغير، وحيث:

$$c_i < n - \log_2 i$$

حيث نختار رقماً عشوائياً من مجالها، وبهذا نكون قد حركنا الشجرة بشكل عشوائي. (وهذه الطريقة تمَّ تطبيقها في بحثي السابق في خاصية الطفرة (Mutation) الخاص بالخوارزمية الجينية).

أما بالنسبة لتحريك شجرة باتجاه شجرة أُخرى فإننا نبحث عن أصغر i غير متطابقة في الشجرتين ونجعلهما متماثلتين.

فيما يلي نوجز خوارزميتنا المقترحة بالخوارزمية التالية:

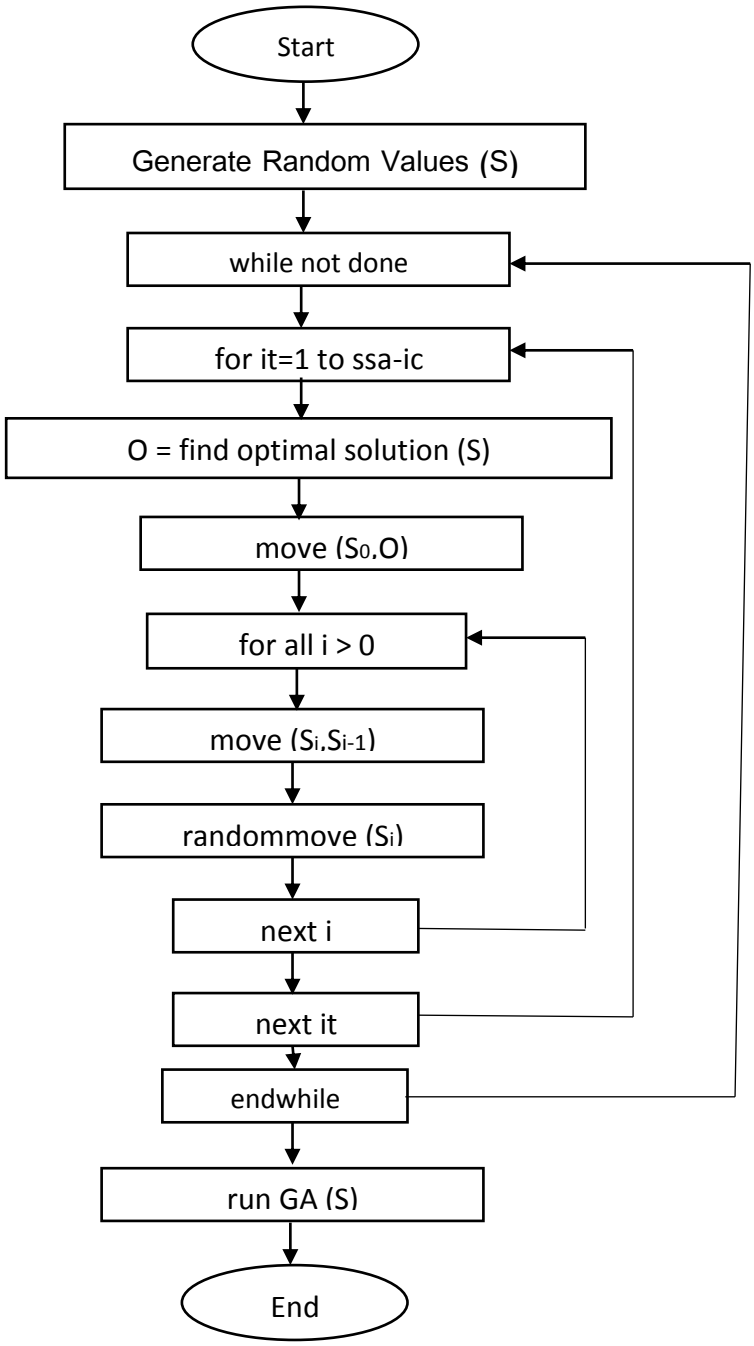
S set of size n= salp swarm count

Generate random values (S)

while not done

$$\begin{bmatrix} for\ it=1\ \ to\ \ ssa-ic \\ \begin{bmatrix} \\ \\ \\ O = find\ optimal\ solution\ (S) \\ move\ (S_0, O) \\ for\ all\ i>0 \\ \begin{bmatrix} move\ (S_i, S_{i-1}) \\ random\ move\ (S_i) \end{bmatrix} \\ \\ run\ GA\ (S) \end{bmatrix} \end{bmatrix}$$

يبين الشكل التالي المخطط التدفقي للخوارزمية الهجينة المقترحة:

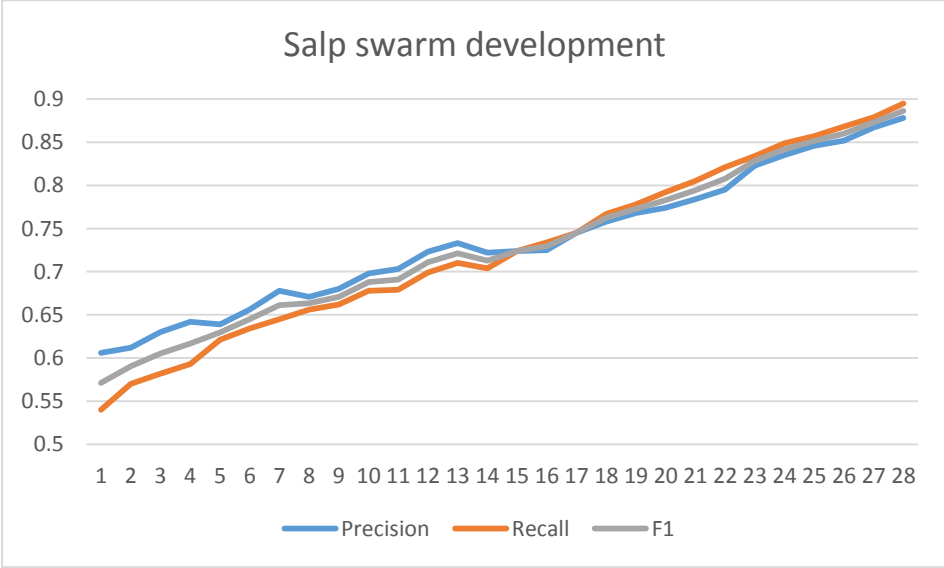


6. النتائج ومناقشتها:

قُمنا بتطبيق الخوارزمية السابقة باستخدام لغة بايثون وبرمجتها. لتجريب الخوارزمية السابقة حصلنا على مجموعة بيانات مرتبطة بأمراض القلب، الهدف منها تحديد ما إذا كان المريض المُحَدَّد بالصفات المطروحة مُرَشَّحاً للإصابة بقصور قلبي أم لا بحسب صفات طبية، مثل السن والضغط والتوتر الشرياني، ...

قمنا بتحويل كل البيانات إلى بيانات ثنائية (الضغط (مرتفع/ منخفض) – الجنس (ذكر/ أنثى)– العمر (كبير/ صغير) ...)، قمنا بتشغيل البرنامج وحصلنا على نتائج ممتازة.

يعرض الشكل (2) مخطَّط تَعَلُّم الخوارزمية ويُوَضِّح معايير القياس.



الشكل (2)

في الشكل (2)، كل خط يمثل إحدى الخوارزميات أثناء تنفيذ الخوارزمية خلال عدد الدورات. نلاحظ أنه على الرغم من تحسن الثقة في البداية في الخوارزمية الهجينة

المقترحة إلا أنها لم تكن الأعلى، لكن بعد فترة أصبحت الثقة في خوارزميتنا هي الأعلى وتفوقت على الخوارزميات الأخرى.

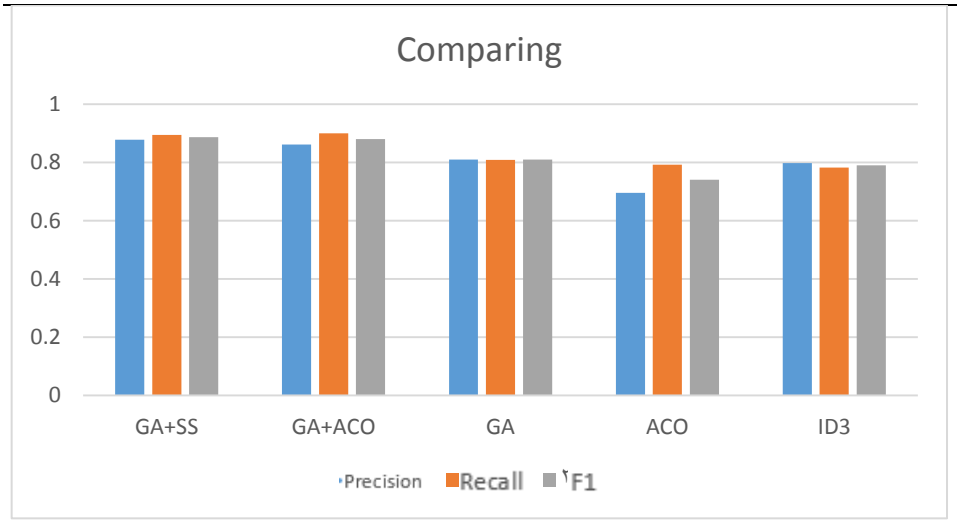
يعرض الجدول (1) النتائج المتعلِّقة بالشجرة المُستَقرَأة.

الجدول (1)

Method	Precision	Recall	F1 score	Leaf count	Height
GA+SS	0.878	0.895	0.8864185	8	4
GA+ACO	0.86122	0.8999	0.88013523	8	4
GA	0.81023	0.8092	0.809714672	9	4
ACO	0.69543	0.7923	0.74071127	10	4
ID3	0.79833	0.7822	0.790182693	8	4

تدل القيم المذكورة في الجدول (1) على فعَّالية الخوارزمية المقترحة.

لتوضيح كفاءة الخوارزمية المقترحة يجب مقارنتها مع المنهجيات السابقة، لذلك قمنا بتطبيق ID3 وخوارزمية Salp Swarm والخوارزمية الجينية كلاً على حدة، وتمت المقارنة معهم وفق معايير الأداء المقترحة. بالإضافة إلى مقارنتها بالخوارزمية المُقتَرَحة في بحثي السابق، فوجدنا النتائج الموضحة بالشكل (3):



الشكل (3)

كما ذكرنا سابقاً، يعرض الشكل (3) نتائج تطبيق عدة خوارزميات على نفس مجموعة البيانات التي قمنا بالتجريب عليها، من حيث معايير قياس الأداء المُرتَبِطة بالبيانات، ويُظهِر تفوُّق الخوارزمية المُطَوَّرة الجديدة بشكل ملحوظ وفق كل المعايير. نستنتج مما سبق أن الخوارزمية المُقتَرَحَة فعالة وأعطتنا نتائج أفضل.

7. الاستنتاجات والتوصيات:

عرضنا في هذه الورقة البحثية منهجية هجينة تمَّ فيها دمج خصائص خوارزمية Salp Swarm في توليد أشجار قرار واختيار الجيد منها عن طريق مسح لفضاء الحالات واختيار الأفضل، وخصائص الخوارزمية الجينية في وراثة الأجزاء الجيدة. وأظهرنا النتائج التي حصلنا عليها وتمَّت مقارنتها مع الخوارزميات السابقة كلاً على حدة، وكما يبدو من النتائج أن الخوارزمية الهجينة المُقتَرَحَة جيدة وفعَّالة. ويجب اختبارها على نتائج أُخرى وتطويرها أكثر ربَّما بدمجها مع تقنيات أُخرى لخوارزميات أُخرى أيضاً.

8. المراجع:

1. Z. Bandar, H. Al-Attar, K. Crockett, 2010- Genetic Algorithms For Decision Tree Induction.

2. S. Rzheutskaya, A. Rzheutskiy, R. R. Salikhova, 2020- Applying a genetic algorithm to build a classification tree.

3. Sousa T., Silva A., and Neves A.,2011- Particle Swarm based Data Mining Algorithms for classification tasks, Parallel Computing.

4. R. C. Eberhart and Y. Shi, 2018- Tracking and optimizing dynamic systems with particle swarms.

5. Eberhart, R. C. and Shi, 2001- Particle swarm optimization: developments, applications and resources. Proc. congress on evolutionary computation 2001 IEEE service center.

6.V. Tandon, 2000- Closing the gap between CAD/CAM and optimized CNC end milling.

7. A. Carlisle and G. Dozier, 2001- An off-the-shelf PSO, Proceedings of the Workshop on Particle Swarm Optimization.

8. Dorigo M, Birattari M, Stutzle T., 2006- Ant colony optimization – IEEE Journals & Magazine [Internet].

9. Holden N. P. and Freitas A. A., 2007- A Hybrid Pso/Aco Algorithm for Classification, GECCO Workshop on Particle Swarms.

10. Gambardella L, Dorigo M., 2000- An Ant Colony System Hybridized with a New Local Search for the Sequntial Ordering Problem| INFORMS Journal on Computing [Internet].

11. D. Oliva, R. Ibrahim, 2019- Improved salp swarm algorithm based on particle swarm optimization for feature selection.

12. H. Faris, S. Mirialili, 2018- An efficient binary swarm algorithm with crossover scheme for feature selection problems.

13. G. Sayed, G. Khoriba, 2018- A novel chaotic salp swarm algorithm for global optimization and feature selection.

14. J. Kennedy, K. C. Eberhart and Y. Shi, 2019- Swarm intelligence.

15. Springer, 2020- Article Slap Swarm Algorithm: a comprehensive survey.

16. H. Liu, G. Zhou, 2009- Minimum Spanning Tree Problem Research Based on Genetic Algorithm.

17. Saimanvitha, 2021- Weighted Sum Method, Multi Criteria Decision Making.

18. Sahiti Kappagantula, 2015- How To Create A Perfect Dicision Tree? .

19. H. Habash,2021- تطوير منهجية جديدة لبناء شجرة القرار باستخدام طريقة هجينة تعتمد على الخوارزمية الجينية وخوارزمية النمل.