

هندسة برمجية لخوارزمية الاختيار النسيلى المناعية والمهجنة بالخوارزمية الجينية لتميز خط اليد

د. جمال صلاح الدين سيد مجيد

قسم هندسة البرمجيات

كلية علوم الحاسبات والرياضيات

جامعة الموصل

شيرين معتز محمد صديق

طالبة دراسات عليا قسم هندسة البرمجيات

كلية علوم الحاسبات والرياضيات

جامعة الموصل

الاستلام

2011 / 06 / 19

القبول

2011 / 10 / 19

Abstraction:

The work explores the potentiality of a clonal selection algorithm and it's hybridizing with the genetic algorithm GA in cursive and discrete handwritten English character recognition. In particular, a retraining scheme for the clonal selection algorithm is formulated for better recognition rates. Empirical study with a dataset (which contains about 100 handwritten samples for 26 characters taken from 30 persons) shows that the proposed approach exhibits very good generalization ability, such that results reported recognition accuracy reached to 100% for the recognition of characters that have been used in building database, and an average recognition accuracy of about 94% for other characters.

الملخص:

يعرض العمل قابلية استخدام خوارزمية الاختيار النسيلى clonal selection المناعية وتهجينها مع الخوارزمية الجينية GA في تمييز أحرف اللغة الانكليزية المتصلة المكتوبة بخط اليد. فقد استخدم على وجه التحديد هيكل معاد الترتيب لخوارزمية الاختيار النسيلى clonal selection لتحقيق أفضل نسبة تمييز ممكنة. فمن خلال دراسة تجريبية أجريت على مجموعة مكونة من 100 نموذج sample من الأحرف الانكليزية المكتوبة بخط اليد لأكثر من 30 شخصاً أثبتت هذه الخوارزمية فعاليتها العالية في تحقيق نسبة تمييز تصل إلى 100% بالنسبة

لتمييز الخطوط المستخدمة في تكوين قاعدة البيانات, ومعدل نسبة تمييز وصلت لـ 94% للخطوط التي لم تدخل في تكوين قاعدة البيانات.

Introduction

1. المقدمة :

تم في الوقت الحالي استخدام العديد من الاستعارات المناعية في تصميم أنظمة مناعية اصطناعية Artificial Immune Systems (AIS) [1]. تصنف هذه الأساليب بشكل عام إلى ثلاث مجاميع: نماذج الشبكات المناعية immune network models [2], خوارزميات الاختيار النسيلى clonal selection algorithms CSA [3] وخوارزميات الاختيار السلبي negative selection algorithms [4]. يناقش هذا البحث منهج استخدام جديد لخوارزمية الاختيار النسيلى CSA وتطبيقها في تمييز أحرف اللغة الانكليزية المكتوبة بخط اليد.

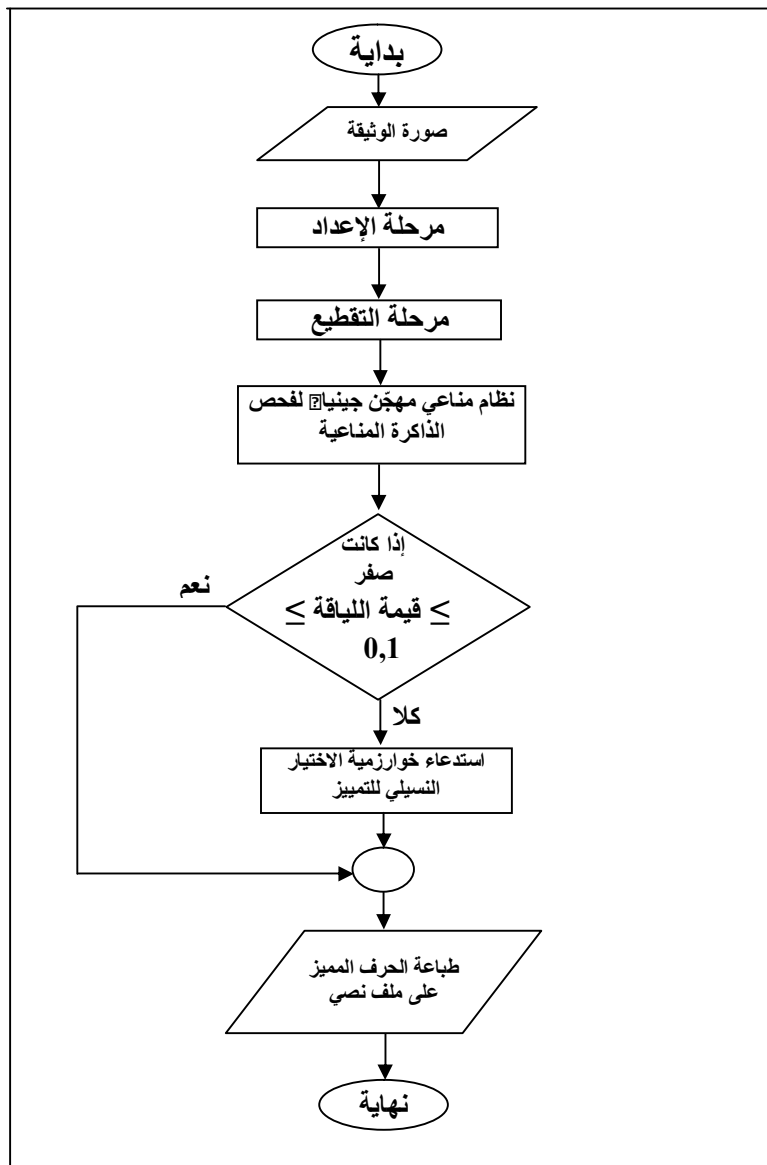
يعرض هذا البحث أسلوب جديد من التمييز معتمد على حساب مقدار الصلة affinity ما بين المستضدات antigens والتي تمثل الصور المراد تمييزها والأجسام المضادة antibodies والتي تمثل الخلايا المناعية التي سوف تقوم بعملية التمييز بثلاث مراحل. تشمل المرحلة الأولى والثالثة حساب مقدار معامل الارتباط الذاتي autocorrelation coefficient, أما المرحلة الثانية فتشمل حساب مقدار التطابق في عدد معابر الصفوف row crossings ما بين صورة الحرف الداخلة والصور المخزنة في قاعدة البيانات. وبسبب التطبيقات المتعددة لأتمتة البريد postal automation, قراءة شيكات البنوك... الخ من التطبيقات قام باحثي تحليل صور الوثائق بإجراء العديد من الدراسات في هذا المجال في السنوات الأخيرة وقد تم اقتراح العديد من أساليب التمييز. ففي الوقت الذي تكون فيه بعض هذه الأساليب مستوحاة حيويًا, مثل الشبكات العصبية [5], الخوارزميات الجينية GA [6], ما يزال النظام المناعي الاصطناعي AIS جديد نوعًا ما على هذا النوع من التطبيقات, ومع هذا استخدمت بعض من تقنيات AIS في عدد من تطبيقات تمييز الأنماط pattern recognition [7].

يتضمن هذا البحث ما يلي: الفقرة الثانية نعرض فيها النموذج العام للتمييز, وتضم هذه الفقرة الفقرات الآتية: مرحلة الإعداد, ومرحلة التقطيع, والنظام المناعي الحيوي, والنظام المناعي الاصطناعي, ومرحلة التمييز, وتحليل مرحلة التمييز, وتصميم مرحلة التمييز, برمجة مرحلة التمييز, واختبار مرحلة التمييز.

2. النموذج العام لنظام التمييز : General model of Recognition system

يتضمن نظام التمييز المقترح ثلاث مراحل رئيسية تمر بها صورة الوثيقة وصولاً إلى مرحلة التمييز. ومن الجدير بالذكر هنا بأن هذا النظام قد تم تطبيقه بأسلوب هندسة البرمجيات

software engineering, حيث تم اعتماد نموذج عملية برمجية software process model من النوع التطوري evolutionary, وبالتحديد تم استخدام النموذج التزايدي Incremental process model لتطوير هذا المشروع, حيث تم تطبيق أطوار التحليل analysis, والتصميم design, والبرمجة coding والاختبار testing على كل مرحلة من مراحل المشروع, وبسبب محدودية عدد صفحات البحث, فسوف يتم عرض مرحلة التمييز فقط بأسلوب هندسة البرمجيات وذلك لكون تطبيق هذا الأسلوب على جميع مراحل البحث يحتاج للكثير من الشرح والتفصيل, لذلك سيتم عرض بقية المراحل بالأسلوب الاعتيادي, إنَّ مراحل النظام موضحة في المخطط ذي الرقم (1):



المخطط ذو الرقم (1): المخطط الانسيابي لنظام التمييز التمييز

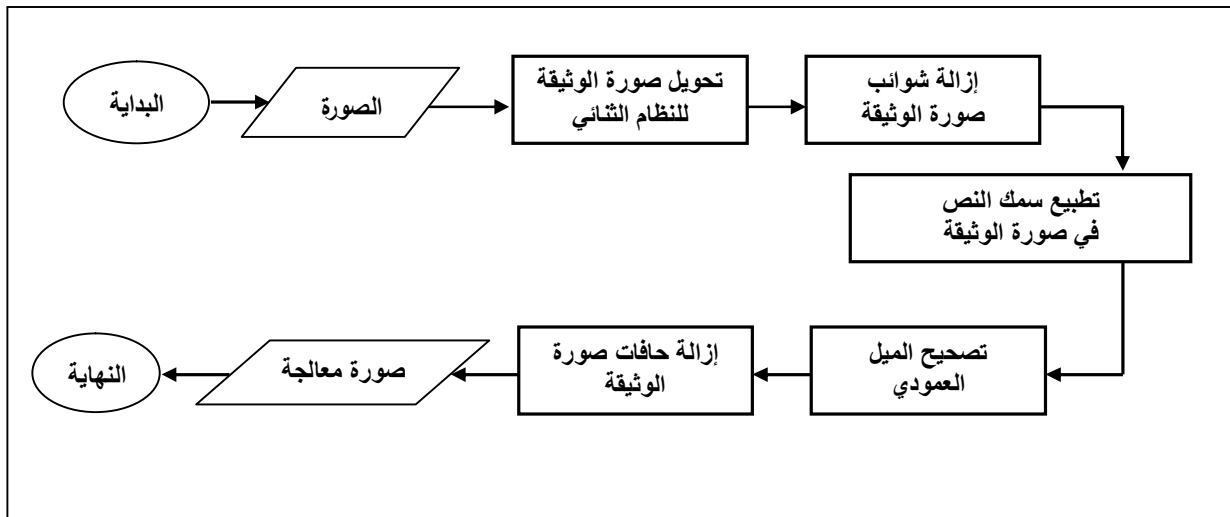
Preprocess Stage

1.2: مرحلة الإعداد:

تتضمن هذه المرحلة تنفيذ مجموعة من العمليات التحضيرية التي تطبق على صورة الوثيقة الداخلة لتهيئتها للمراحل اللاحقة، وهذه العمليات تعمل على تحسين نسبة التمييز بقدر الإمكان، لهذا فإن هذه المرحلة تختلف من مشروع وآخر. فبالنسبة لمشاريع تمييز الأحرف المشابهة لبحثنا الحالي فعدد ونوع هذه العمليات تعتمد على نوع الخوارزمية المستخدمة في التمييز ومدى حاجتها لهذه العمليات التحضيرية [8].

فبالإضافة إلى عملية التمييز فإن برنامج تمييز الأحرف المكتوبة يدويا يجب أن ينفذ مجموعة من مهام الإعداد preprocessing للتأكد من كون الخط بحالة ملائمة للتمييز. إن عملية التطبيع (إعادة الموازنة) normalization عملية ضرورية للأنظمة التي تعمل على تمييز الخطوط المكتوبة يدويا وذلك لكون مجال العشوائية فيها واسعا، ويسبب اختلاف أسلوب الكتابة من شخص لآخر مما ينتج عنها عدة اختلافات في الشكل، الحجم، الميل، الانحراف، الارتفاع والعرض [9] [10] لهذا فإننا نحتاج لعمليات الإعداد لمحاولة التقليل من هذه الاختلافات ومحاولة الحصول على هيكل عام تقريبا للنصوص الداخلة لتسهيل عملية التمييز بقدر الإمكان.

للأنظمة التي تعمل بصورة غير متصلة off-line وهي الأنظمة المشابهة لنظامنا، أما بالنسبة لمشاريع التي تعمل بشكل متصل online فهي تقوم بعملية التمييز أثناء عملية كتابة النص في الأنظمة المستخدمة لأجهزة الإدخال المعتمدة على الأقلام مثل الأقلام الرقمية digital pens, الألواح الإلكترونية electronic tablets, الألواح التفاعلية interactive whiteboard وأجهزة الحاسوب الشخصية ذو شاشات اللمس touch screen PC، فإن عمليات الإعداد تشمل [11]:



المخطط نو الرقم (2): المخطط الانسيابي لمرحلة الإعداد

1. تحويل الصورة إلى النظام الثنائي (Image Binarization):

تتضمن تمثيل قيم النقاط الضوئية pixels للصورة المراد تمييزها بقيمتين فقط إما صفر أو واحد. حيث تعطي اغلب أنظمة تمييز الأحرف لنقاط الخلفية الضوئية background pixels شدة إضاءة brightness قيمتها (صفر) أي أنها تمثل الخلفية بلون اسود وتعطي نقاط النص الضوئية text pixels شدة إضاءة brightness قيمتها (1) أي تمثل النص باللون الأبيض ويتم ذلك بالاعتماد على مقدار حد العتبة threshold المحدد لإعطاء الخلفية اللون الأسود وإعطاء النص اللون الأبيض, يستخدم هذا النوع من التمثيل لأنه عادة ما يسهل عملية تمييز الشوائب في النص للتخلص منها بعملية إزالة الشوائب noise removal [11]. وقد تم تنفيذ هذه العملية في البحث الحالي عن طريق استخدام الإيعاز (im2bw) بلغة MATLAB البرمجية المستخدمة لبرمجة النظام.

2. التخلص من الشوائب (Noise Removal):

تشمل هذه العملية تحديد النقاط الضوئية pixels التي تمثل الشوائب في الصورة والقيام بالتخلص منها من دون التأثير على النص الداخل [9], حيث يتم التخلص من المكونات المتصلة والتي تشمل اقل من 5 نقاط ضوئية pixels من الصورة الداخلة [12]. وقد تم تنفيذ هذه العملية في نظامنا عن طريق استخدام الإيعاز (bwareopen) بلغة MATLAB والذي يقوم بحذف المقاطع المتكونة من عدد محدد من النقاط الضوئية وفي نظامنا تم تحديد هذا العدد بـ 8 نقاط ضوئية.

3. تطبيع سمك خط الكتابة (Thickness normalization):

يتم تطبيع سمك الخط بخطوتين, الأولى, يتم تطبيق خوارزمية التثخيف (thinning algorithm) لتقليل سمك الخطوط (strokes) إلى سمك نقطة ضوئية واحدة (1 pixel). ومن ثم إجراء عملية إعادة تثخين re-thickening منتظمة يتم تطبيقها للحصول على سمك عدة نقاط ضوئية pixels في الخط stroke الواحد [12]. تتضمن عملية التثخيف thinning تحويل الصورة الخام البيضاء والسوداء إلى رسوم خطية متكونة من سمك موحد عن طريق مسح النقاط الضوئية pixels الزائدة بدون الأضرار بمظهر الحرف. هذه العملية المسماة أيضا بعملية الهيكل skeletonization يجب أن تحافظ على الهيكل الأساسي وترابط النموذج الأصلي. أما عملية إعادة التثخين re-thickening يتمثل عملها ببساطة بجعل قيمة شدة الإضاءة brightness مساوي للواحد (1) للنقاط الضوئية pixels السوداء الواقعة في المحيط المجاور neighbor pixels للنقاط الضوئية pixels البيضاء [12]. وقد تم تنفيذ هذه العملية

في نظامنا باستخدام إيعازي التثخيف thinning ومن ثم التوسع dilation بمقدار 3 نقاط ضوئية بلغة MATLAB.

4. تطبيع الميل العمودي (Slant Normalization):

إن عملية تصحيح الميل slant correction هي عملية تجريد الصورة الداخلة باتجاه موازي للمحور الأفقي لإجبار الميل العام لنموذج الحرف الداخل أن يكون عمودياً، يتم القيام بهذه العملية للحصول على هيكل عام لصورة الحرف الداخل بغض النظر عن مصدر الخط، بالإضافة إلى أن وجود الميل يقلل من نسبة التمييز لذا يجب تقليله أو محاولة التخلص منه بشكل كامل [11]. إن الخوارزمية المستخدمة لتصحيح الميل تعتمد بشكل أساسي على الملاحظة التالية وهي أن: صورة الحرف تمتلك أقل عرض عندما يكون الحرف بأقل درجة ميلان (أي عندما يكون عمودياً تقريباً) [10]، أي أن الحرف المائل بدرجة معينة يكون اعرض من حالته العمودية وعلى هذا الأساس يتم كشف درجة الميل ومحاولة تعديلها عن طريق تدوير النص بنفس درجة الميل وبعكس اتجاهه [12]. وقد تم تنفيذ هذه العملية في نظامنا باستخدام المعادلة الآتية:

$$\forall p(x, y) = 1, b(x, y) = 1$$

where:
$$\begin{cases} \bar{x} = x - y \operatorname{tg}(\alpha) \\ \quad y = y \end{cases}$$

5. التخلص من حافات الصورة (Image Edges removal):

تكون الصورة الداخلة حاوية على عدة نقاط بيانية من الممكن أن تعيق أداء التمييز، لهذا للتخلص من هذه النقاط يتم تأطير الصورة عن طريق التخلص من الحافات البيضاء المحيطة بالنص للتقليل من الذاكرة المطلوبة لخرن صورة النص والوقت المطلوب للمعالجة، بحيث نحصل على صورة للنص فقط [12].

2.2: مرحلة التقطيع: Segmentation Stage

تعتبر عمليات التقطيع إلى الأسطر line segmentation, الكلمات word segmentation والأحرف character segmentation من أهم العمليات في نظام تمييز الحروف وذلك لكونها تؤثر على نسبة التمييز. إن عملية التقطيع segmentation ليست مشكلة آلية فحسب، فالبشر يمتلكون أكفاً جهاز قراءة ضوئي ألا وهو "العينان"، إلا أنهم مع هذا يواجهون في بعض الأحيان صعوبة في تمييز بعض النصوص المكتوبة بالخط المتصل cursive script، فنحن نملك نسبة خطأ في التمييز تصل إلى 4% من عمليات القراءة التي تتم بغياب مضمون الجملة [11]. تشمل عملية التقطيع segmentation ثلاث مراحل ثانوية:

1) تقطيع الأسطر Line Segmentation:

عندما تصبح مصفوفة الصورة جاهزة للمعالجة، فإن أول خطوة تلي عملية الإعداد هي فصل كل سطر من النص text Line عن الوثيقة. نستخدم تقنية التحليل الأفقي لهذا الغرض horizontal projection profile technique، حيث يقوم البرنامج بمسح الوثيقة أفقياً لإيجاد أول وآخر نقطة ضوئية pixel بيضاء لكل سطر line. بمجرد إيجاد هذه النقاط الضوئية pixels فإن المساحة ما بين هذه النقاط الضوئية pixels تمثل المنطقة التي تحوي حرف أو أكثر وباستخدام التقنية المذكورة أعلاه يتم مسح الوثيقة بشكل كامل لاكتشاف كل سطر ومن ثم تخزينه في مصفوفة مؤقتة للمعالجة. [11]

2) تقطيع الكلمات Word Segmentation:

تكون عملية التقطيع سهلة نوعاً ما بالنسبة للوثائق المطبوعة آلياً، ماعدا الخطوط من نوع extremely tightly-set و micro – justified، حيث تكون الكلمات فيها متداخلة مع بعضها البعض. أما بالنسبة للوثائق المكتوبة بخط اليد، فإنها عادة ما تحوي فراغات كبيرة تفصل بين الكلمات [3]، لذا فإننا من الممكن أن نستخدم نفس الأسلوب المستخدم في تقطيع الأسطر Line Segmentation ولكن هنا سيتم مسح الوثيقة عمودياً لإيجاد أول وآخر نقطة ضوئية لكل حرف أو كلمة في كل سطر ومن ثم تخزين الكلمة أو الحرف أيضاً في مصفوفة مؤقتة ليتم إرسالها إلى برنامج تقطيع الأحرف ليتفحص المصفوفة المستلمة فيما إذا كانت مصفوفة كلمة أم حرف، وبالتأكيد يعتمد اتخاذ القرار على عرض المصفوفة المتمثل بعدد الأعمدة، حيث يقارن عرض المصفوفة المستلمة مع العرض التخميني للحرف لتحديد فيما إذا كانت المصفوفة حرفاً أم كلمة [10].

3) تقطيع الأحرف Character Segmentation:

إن أسلوب تقسيم الأحرف المقترح تم تنفيذه كعملية تكرارية recursive process، والتي تتفاعل مع مرحلة التمييز recognition. يتضمن الأسلوب المقترح لتقسيم أحرف الكلمة الخطوات التالية:

- 1- قراءة صورة الوثيقة المراد تمييزها.
- 2- إجراء عمليات الإعداد عليها.
- 3- حساب مجموع نقاط المقدمة الضوئية foreground pixels والمتمثلة بالنقاط الضوئية البيضاء لكل عمود column. العمود الذي يكون مجموع نقاطه الضوئية pixels ما بين [صفر، 3]، نقوم بتخزين رقم العمود في مصفوفة، تحوي أرقام الأعمدة المرشحة كأعمدة قطع (candidate segment columns CSC).

يتمثل ناتج الخطوة الثالثة بعدد من أعمدة القطع المرشحة وهذا العدد يكون عادة أكبر من العدد الحقيقي لأعمدة القطع الفعلية، لذا فقد قمنا بتحديد مقدار عتبة threshold ليمثل العرض التقريبي للحرف، وقد تم تحديده بالتجربة، وعلى أساسه يتم تحديد الأحرف وتقطيعها [12].

3.2: النظام المناعي الحيوي: Biological Immune System

يتكون النظام المناعي للفقريات من أنواع مختلفة من الجزيئات molecules, الخلايا cells, والأعضاء organs المنتشرة في جميع أنحاء الجسم. لا يمتلك النظام المناعي عضو مركزي يسيطر على عمله، بل يمتلك عدة مكونات منتشرة في أماكن مختلفة تؤدي عملها وفق قوانين متناسقة. حيث أن الوظيفة الأساسية للنظام المناعي هي الحفاظ على حياة العضو من خلال التمييز ما بين الخلايا الضارة وخلايا ذلك العضو، مثل (خلايا الورم والخلايا السرطانية)، والعناصر الخارجية المسببة للأمراض مثل (البكتريا والفيروسات). [1] [2]

تسمى العناصر التي يميزها النظام المناعي بالمستضدات Antigens أو بالمستضدات الغير-ذاتية non-self antigens, في حين أن الخلايا التي تعود لأجسامنا تسمى بالمستضدات الذاتية self antigens أو الأجسام المضادة antibodies, لذلك يجب أن يمتلك النظام المناعي القابلية على التمييز ما بين المستضدات الذاتية self والغير ذاتية non-self. ويتم ذلك بشكل أساسي عن طريق سلسلة من فعاليات تمييز الأنماط pattern recognition. انطلاقاً من مفهوم تمييز الأنماط pattern recognition فإن أهم خاصية يمتلكها النظام المناعي هو امتلاكه لجزيئات الاستقبال receptor molecules المتواجدة على سطح الخلايا المناعية، والقادرة على تمييز عدد غير محدود تقريباً من الأنماط المعادية للجين Antigenic patterns. يمتلك النظام المناعي مجموعتين رئيسيتين من الخلايا المناعية وهي خلايا B-cells و خلايا T-cells, تتشابه هذه الخلايا مع بعضها البعض كثيراً، إلا أنها تختلف في كيفية تمييزها للمستضدات antigens وفي قوانين عملها، فخلايا B-cells قادرة على تمييز المستضدات antigens بشكل مباشر مثلاً في مجرى الدم، بينما تحتاج خلايا T-cells إلى خلايا مساعدة accessory cells كي تتمكن من تمييز المستضدات antigens. [1] [2]

4.2: النظام المناعي الاصطناعي: Artificial Immune System

انبثقت الأنظمة الاصطناعية AIS, مؤخراً على أنها نماذج للحسابات الخفيفة soft computation المبتكرة، باحثة عن الوحي في الأنظمة المناعية الحيوية لتطوير نماذج حسابية

لحل المشاكل المعقدة. تصنف هذه النماذج بشكل عام إلى ثلاث مجاميع: نماذج الشبكات المناعية immune network models [2], خوارزميات الاختيار النسيلى clonal selection algorithms [3] وخوارزميات الاختيار السلبي negative selection algorithms [4].

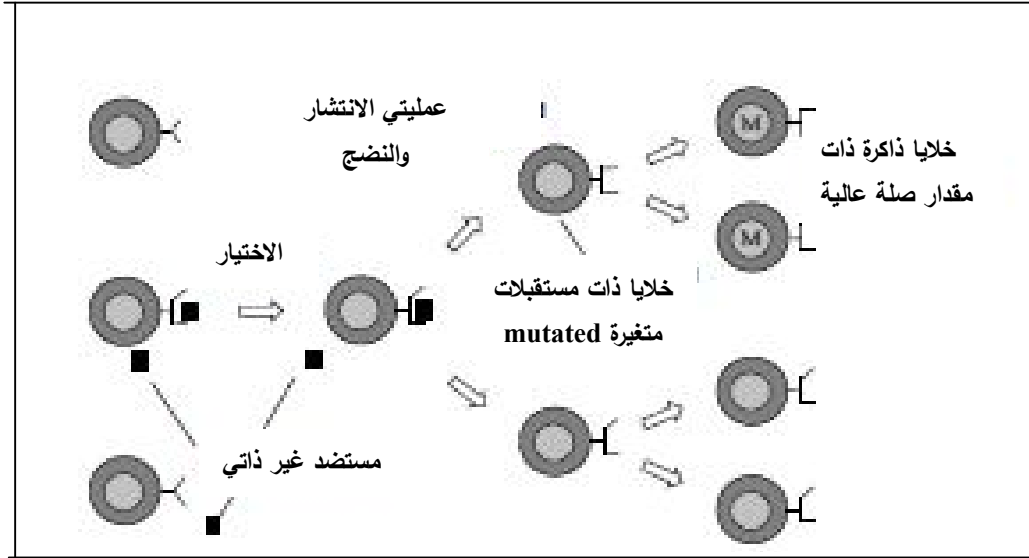
5.2: خوارزمية الاختيار النسيلى: العملية الحيوية:

وفقا لنظرية Burnet 1959 للاختيار النسيلى clonal selection, تنفذ هذه الخوارزمية ميكانيكية اختيار خلال فترة حياة الفرد, حيث تنص هذه النظرية على أنه عندما يحدث تماسك مع مستضد antigen مناسب فسوف تحدث عملية تفعيل للخلايا للمفاوية. وبمجرد حدوث عملية التفعيل فإنه يتم إنتاج مجموعة من النسخ clones للخلية للمفاوية ذات مستقبلات مماثلة للخلية للمفاوية الأصلية التي وجدت المستضد, وبالتالي يحدث لهذه الخلية عملية توسع نسيلى clonal expansion. وهذا يدل على أنه فقط الخلايا للمفاوية التي تفعل من قبل المستضد يتم استنساخها لعدة مرات. [14]

ذكرت نظرية الاختيار النسيلى clonal selection بأن أي خلية لمفاوية تملك مستقبل مشابه لجزيئات الشخص الذاتية يجب أن يتم حذفها خلال عملية بناء الخلايا للمفاوية. وهذا يؤكد بأنه فقط المستضد الذي ينتمي لمجموعة المستضدات المسببة للأمراض هو الذي سوف يتسبب بحدوث التوسع النسيلى للخلية للمفاوية التي سوف تحفز استجابة مناعية تدميرية للمستضد الداخل. ومن هذا المنطلق بالإمكان أن نعتبر أن خوارزمية الاستنساخ النسيلى clonal selection هي خوارزمية تصنيف للمستضد لتحديد كونه مستضد ذاتي أم غير ذاتي, على فرض أن المستضد الغير ذاتي هو مستضد مسبب للأمراض ويجب أن يتم إزالته من الجسم.

خلال عملية التوسع لخلايا B-cells فقط (وليست خلايا T-cells), يزداد معدل الصلة ما بين الجسم المضاد antibody والمستضد antigen المسبب للتوسع النسيلى clonal expansion, حيث يطلق على هذه الظاهرة اسم نضوج الصلة affinity maturation, وهذه الظاهرة هي المسؤولة عن حقيقة أن التعرض المتكرر للمستضد antigen يزيد من الفعالية المناعية للأجسام المضادة antibodies وذلك بسبب امتلاك هذه الأجسام المضادة صلة أكبر بالمستضد. تنتج عملية نضج الصلة عن طريق الطفرة الجسدية somatic hypermutation وميكانيكيات الاختيار التي تحدث خلال التوسع النسيلى لخلايا B-cells. حيث تغير الطفرة الجسدية من الأجسام المضادة antibodies عن طريق إحداث تغييرات عشوائية في الجينات

المكونة لها [2] [3] [14]. وفيما يلي الشكل رقم (1) يوضح خوارزمية الاختيار النسيلى :clonal selection



الشكل (1): الاستنساخ النسيلى, الانتشار (التوسع), نضوج الصلة, والحصول على خلايا الذاكرة.

الخوارزمية:

طور كل من Von Zaber و De Castro واحدة من أشهر الأنظمة المناعية الاصطناعية المعتمدة على الاختيار النسيلى وأكثرها استخداما والمسماة بـ CLONALG, التي استخدمت لتنفيذ مهام مطابقة النماذج ودوال تحسين النماذج المتعددة. عند استخدامها في عملية مطابقة النماذج pattern matching, يتم اعتبار مجموعة من النماذج S المراد مطابقتها على أنها المستضدات antigens, وبهذا تكون مهمة CLONALG تكوين مجموعة من الأجسام المضادة antibodies, والمتمثلة بالمجموعة M, والتي تطابق العناصر في S [12] [14]. يتم ذلك باستخدام الخوارزمية التالية:

الإدخال: S = مجموعة النماذج المراد تمييزها, n = عدد الأجسام المضادة التي تملك أقل مقدار من قيم الصلة. الإخراج: M = مجموعة من كاشفات ذاكرة قادرة على تصنيف النماذج الجديدة.

بداية

- كون بشكل عشوائي مجموعة أولية من الأجسام المضادة وضعتها في المجموعة A
- لجميع النماذج في S, أفعال:
- حدد مقدار الصلة مع كل جسم مضاد في A.
- كون نسخ لمجموعة جزئية من الأجسام المضادة في A والتي تملك أكبر قيم صلة, عدد النسخ يتناسب طرديا مع مقدار الصلة.
- غير mutate صفات هذه النسخ للمجموعة A (مقدار التغيير يتناسب عكسيا مع مقدار الصلة), وضع نسخة الجسم المضاد التي تملك أكبر مقدار صلة في مجموعة الذاكرة M.
- قم باستبدال n من الأجسام المضادة في A والتي تملك أقل مقدار من قيم الصلة affinity بمجموعة الأجسام المضادة الجديدة المتكونة.
- كرر إلى أن يتحقق شرط التوقف

نهاية [14]

ملاحظة: هناك العديد من خوارزميات الاختيار النسيلي التي يمكن اعتمادها حسب نوع المشكلة المراد حلها, للحصول على معلومات أكثر بالإمكان الاستفادة من المصدر [12].

6.2: مرحلة التمييز: Recognition Stage

كما ذكرنا بأنه تم تنفيذ مرحلة التمييز باستخدام خوارزمية الاختيار النسيلي المناعية والتي تم تهجينها مع الخوارزمية الجينية. ليكن ag يمثل المستضد وهو الحرف المراد تمييزه, يمتلك كل ag صفتين الأولى هي الصنف class والتي تكون بالشكل التالي:

عدد الأحرف الإنكليزية $ag.c \in C = \{c_1, c_2, c_3, \dots, c_n\}$; $n = 26$.

والصفة الثانية هي متجه الخصائص feature vector. ولتكن AB هي الأجسام المضادة التي ستقوم بالتمييز بحيث $AB = \{ab_1, ab_2, ab_3, \dots, ab_n\}$ حيث أن كل ab_i هي خلية مناعية وبالنسبة لبحثنا فقد تم تكوينها بشكل هياكل structures تحوي الخلايا المناعية كل خلية تمثل حرف وسوف نوضح تفاصيل تكوينها لاحقاً. يمتلك كل ab_i صفتين مشابهة لصفات المستضدات المفردة, وتتمثل هاتين الصفتين بما يلي: $ab_i.c \in C = \{c_1, c_2, c_3, \dots, c_n\}$ وهي معلومات الصنف class و $ab_i.f$ وهي متجه الخصائص feature vector. قبل البدء بعملية التمييز يجب تحويل صورة الحرف المراد تمييزه إلى صورة أحادية البعد vector وهذا يدل على أن الصورة نفسها سوف تستخدم كخارطة خصائص feature map. إن التطابق ما بين مثل هذه المتجهات $S(F_1, F_2)$ هو عبارة عن ناتج لثلاث مراحل, تتمثل المرحلة الأولى والثالثة بقياس معامل الارتباط الذاتي autocorrelation coefficient ما بين F_1 و F_2 وكما هو معرف في العلاقة التالية:

$$S(F_1, F_2) = \frac{1}{2} - \frac{S_{10} S_{01} - S_{00} S_{11}}{2\sqrt{(S_{11} + S_{10})(S_{01} + S_{00})(S_{11} + S_{01})(S_{10} + S_{00})}} \quad \text{..... (1) المعادلة}$$

حيث أن:

S_{00} = عدد المطابقات ما بين النقاط الضوئية pixels ذات القيم صفر (zeros matches).

S_{11} = عدد المطابقات ما بين النقاط الضوئية pixels ذات القيم واحد (ones matches).

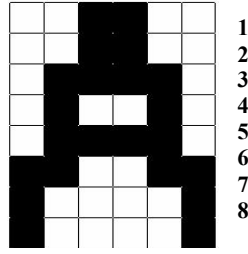
S_{01} = عدد عدم المطابقات التي تحدث للنقاط الضوئية pixels ذات القيمة صفر (zero mismatches).

S_{10} = عدد عدم المطابقات التي تحدث للنقاط الضوئية pixels ذات القيمة واحد (ones mismatches).

تعطي S قيمة تقع في المدى $[0,1]$, حيث تمثل القيمة (1) أعلى مقدار من التطابق (الارتباط) ما بين النماذج و قيمة (صفر) تمثل أقل قيمة تطابق (أي أنها تدل على عدم وجود أي تطابق ما بين النماذج). أما المرحلة الثانية فتتمثل بقياس مقدار التطابق في عدد معابر الصفوف row

crossings ما بين الخلايا المناعية الناتجة من المرحلة الأولى, والصورة الداخلة التي تمثل المستضد Ag حيث يتم حساب معابر الصفوف لكلا الطرفين كالآتي:

يمثل مقدار معابر الصفوف row crossings عدد المرات التي ينقطع بها الخط الأفقي المستقيم في كل صف. يوضح الشكل ذو الرقم (2) صورة الحرف A, وعلى سبيل المثال أن أبعاد الصورة هي 6*8, حيث أن عدد الصفوف 8 صفوف, فإذا أخذنا الصف الأول مثلا يكون مقدار معابر الصفوف row crossings فيه (واحد), بينما مقدار معابر الصف الرابع هو (2) أي أن الخط الأبيض والذي يمثل خلفية الصورة قد تم قطعه مرتين بالنقاط السوداء التي تمثل نقاط النص, وهكذا يتم حساب المعابر لجميع الصفوف ولكل صورة داخلة ومقارنتها مع معابر صفوف row crossings الصور المخزنة مسبقا في قاعدة البيانات, يمثل عدد معابر الصفوف row crossing بشكل (كود) code للصورة, كما وأنه يمكن حسابه للأعمدة أيضا ولكن وجدنا من التجارب التي أجريت بأن معابر الصفوف row crossings لها دقة أكبر في التمييز من معابر الأعمدة column crossing. بعد أن يتم حساب عدد معابر الصفوف row crossings للصورة الداخلة يحسب مقدار التطابق بينها وبين عدد معابر صفوف صور قاعدة البيانات المكونة مسبقا, ليتم اختيار الخلايا العشرة التي تملك أعلى مقدار من الترابط وإرسالها إلى آخر مرحلة والمتمثلة باختيار الخلايا الخمسة التي تملك أعلى مقدار من معامل الارتباط الذاتي autocorrelation.



الشكل ذو الرقم (2): صورة الحرف A لتوضيح كيفية حساب معابر الصفوف

من المهم أن نذكر هنا أن التطابق يمثل مقدار الصلة affinity حيث استعمل معامل الارتباط الذاتي ومقدار التطابق في عدد معابر الصفوف row crossings لقياس التطابق/الصلة خلال التفاعلات ما بين المستضد Ag- الجسم المضاد Ab. [12] تتضمن عملية التمييز المراحل التالية:

1- التهيئة : Initialization

تتضمن هذه المرحلة اختيار الخلايا المناعية التي ستقوم بعملية التمييز والتي تمثل الأجسام المضادة AB. وفي حالة بحثنا فقد تم اختيار 20 خلية من الخلايا المناعية التي تم خزنها مسبقا لتكوين الأجسام المضادة التي ستقوم بالتمييز, ويجب أن نشير هنا إلى أن عدد الخلايا المناعية المختارة في مرحلة التهيئة يؤثر على أداء النظام [12] [14].

2- تكوين النسخ : Clones generation

يتم في هذه المرحلة اختيار الأحرف الأقرب لصورة الحرف الداخل، وبالنسبة لبحثنا فقد تم اختيار أقرب خمسة أحرف (والتي تمثل أقرب خمسة أجسام مضادة antibodies)، ويكون اختيار الأقرب معتمد على قيمة معامل الارتباط الذاتي autocorrelation ومقدار التطابق في عدد معابر الصفوف row-crossings لكل فرد من الأجسام المضادة antibodies التي تم تهيئتها في المرحلة الأولى مع خلية المستضد Ag (والذي يمثل هنا صورة الحرف المراد تمييزه) فالخلية التي تملك أكبر مقدار ارتباط ذاتي وأكبر عدد من التطابقات في عدد معابر الصفوف row crossings تكون هي الحرف الأقرب والخلية المرشحة لتمييز صورة الحرف الداخلة. تدخل بعدها هذه الخلايا الخمسة في عملية الانتشار proliferation والمتمثلة بتكوين النسخ لهذه الخلايا، بحيث أن الخلية التي تملك أعلى مقدار صلة affinity وهي الخلية الأقرب سوف تملك أكبر عدد من النسخ، وهكذا يتم تكوين النسخ لكل خلية حسب مقدار صلتها وبالتدريج، فمثلا في بحثنا الحالي تم تكوين النسخ بالشكل التالي:

الخلية الأولى (تمتلك أكبر مقدار صلة affinity) لها 5 نسخ، والخلية الثانية 4 نسخ، والخلية الثالثة 3 نسخ، والرابعة نسختين والأخيرة تمتلك نسخة واحدة فقط، وبذلك يصبح مجموع النسخ 15 نسخة.

تلي عملية تكوين النسخ، عملية التغيير أو ما تسمى بالطفرة mutation. حيث ترسل جميع النسخ المتكونة إلى برنامج الطفرة ليتم تغييرها بنسب معينة، بحيث يكون مقدار التغيير متناسب عكسيا مع مقدار الصلة (والتي تمثل هنا قيمة معامل الارتباط الذاتي والتطابق بعدد معابر الصفوف row crossings)، بعد التغيير يتم اختيار الفرد الذي يملك أكبر قيمة ارتباط ذاتي autocorrelation على أنه الحرف الداخل. [12]

تحديث الذاكرة المناعية: Update immune memory

بعد أن يتم تمييز الحرف الداخل، تحسب قيم عزومه السبعة seven invariant moments ليتم الاحتفاظ بها في الذاكرة المناعية مع تسلسل الحرف والذي يأخذ إحدى القيم ما بين 1-26 (عدد أحرف اللغة الإنكليزية). وبهذا عند إدخال هذا الحرف مرة ثانية، سوف يتم فحص الذاكرة المناعية باستخدام الخوارزمية الجينية Genetic algorithm أولا، وبما أن النظام قد تعرض لهذا النموذج مسبقا فسوف يتم تمييزه مباشرة من قبل الخوارزمية الجينية GA دون الحاجة إلى استدعاء خوارزمية الاختيار النسيلي clonal selection، أما في حالة كون النظام يتعرض لهذا الحرف لأول مرة فسوف يتم تمييزه باستخدام خوارزمية الاختيار النسيلي clonal selection وبعد الانتهاء من تمييزه يتم خزن قيم عزومه السبعة وتسلسله index في الذاكرة المناعية ليكتسب النظام المناعة immunity ضد هذا الحرف. [12]

Recognition Analysis

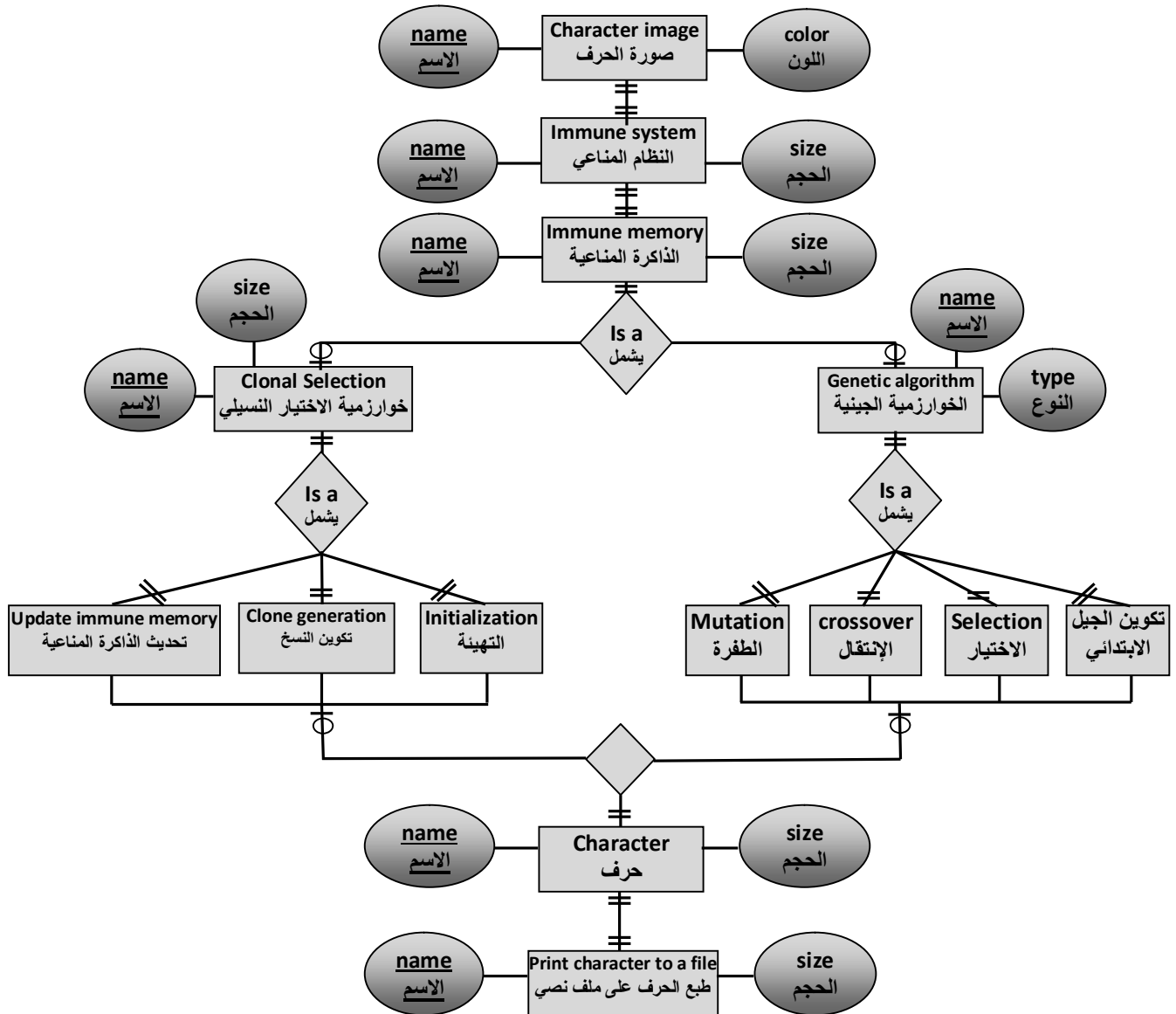
7.2: تحليل مرحلة التمييز:

تشمل هذه المرحلة تكوين ثلاث نماذج، وهي النموذج البياني data model، النموذج الوظيفي functional model والنموذج السلوكي behavioral model بالإضافة إلى تكوين وثيقة وصف specification لكل مخطط (نموذج) لتوضيح مواصفاته [13].

(Recognition Data Model)

1.7.2: النموذج البياني لمرحلة التمييز:

يمثل المخطط البياني باستخدام نموذج "الكائن-علاقة" E-R model، لتوضيح المكونات الرئيسية التي تتكون منها مرحلة التمييز، وكما هو موضح في المخطط ذي الرقم (3):



المخطط ذو الرقم (3): مخطط الكائن-علاقة لمرحلة التمييز

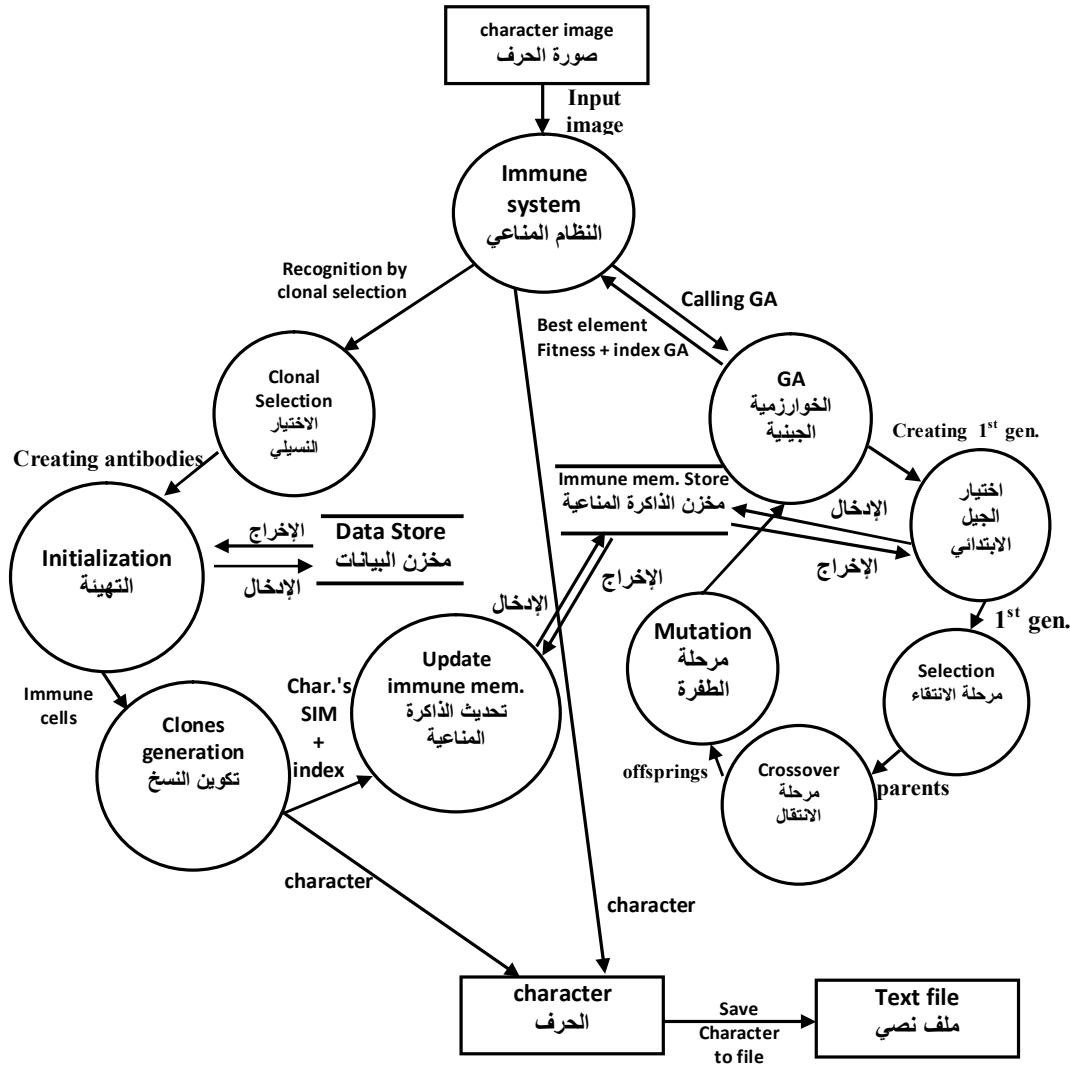
2.6.2: وصف المخطط البياني: (Recognition data model specification)

كما هو موضح في المخطط (3) تحتوي مرحلة التمييز recognition أربعة عشر كائن ممثلين بالأشكال المستطيلة ومرتبطين مع بعضها البعض بعلاقات ممثلة بالأشكال الماسية التي تحوي أسماء العلاقات، ولكل كائن صفات متصلة به وممثلة في المخطط ذي الرقم (3) بالأشكال البيضوية.

3.7.2: النموذج الوظيفي وتدفق المعلومات لمرحلة التمييز:

(Recognition functional model and information flow)

تم تمثيل المخطط الوظيفي باستخدام "مخطط تدفق البيانات"، والذي يرمز له بالرمز DFD اختصاراً لـ Data Flow Diagram وكما هو موضح في المخطط ذي الرقم (4):



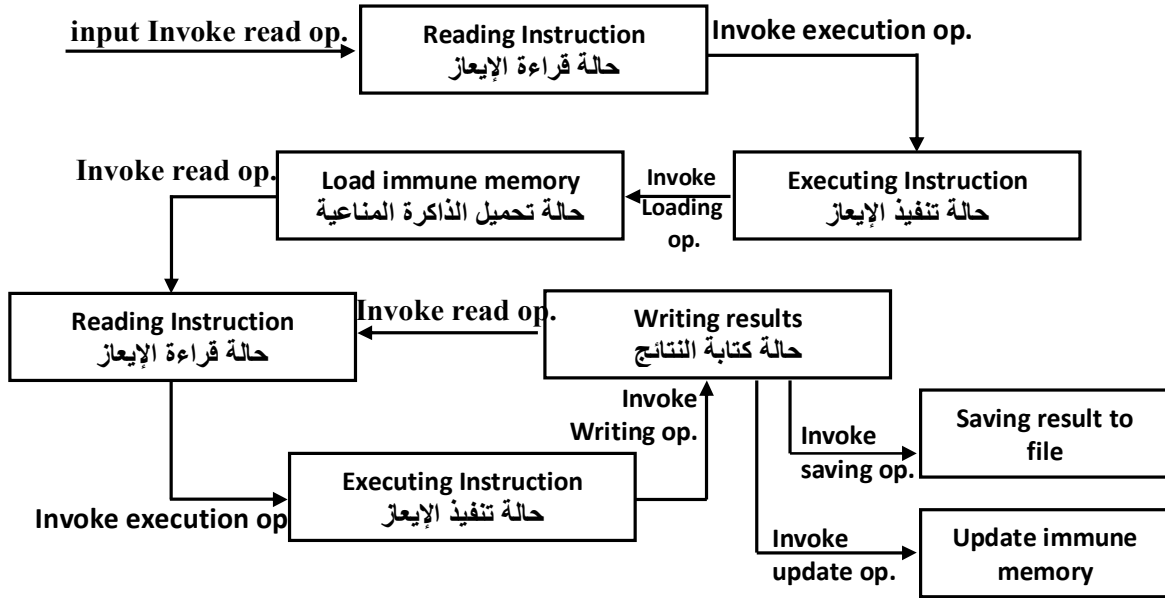
المخطط نو الرقم (4): مخطط تدفق البيانات DFD لمرحلة التمييز

4.7.2: وصف مخطط تدفق البيانات: (Data Flow Diagram Specification)

يوضح المخطط ذو الرقم (4) كيفية تدفق البيانات في مرحلة التمييز recognition وما هي العمليات التي ستمر بها هذه البيانات، حيث تم تهجين نظامنا المناعي مع الخوارزمية الجينية GA، عن طريق استخدام الخوارزمية الجينية GA لفحص الذاكرة المناعية عند بدء مرحلة التمييز recognition. فبعد وصول صورة الحرف المراد تمييزه إلى النظام المناعي immune system يتم استدعاء الخوارزمية الجينية GA لفحص الذاكرة المناعية immune memory، تبدأ الخوارزمية الجينية GA بتنفيذ جميع مراحلها بدءاً بتكوين الجيل الابتدائي وذلك بعد تحميل محتويات الذاكرة المناعية immune memory، مروراً بالانتقاء selection، الانتقال crossover وأخيراً الطفرة mutation ولعدد محدد من الأجيال وفي بحثنا الحالي يبلغ عدد الأجيال بالتحديد 20 جيلاً. بعد انتهاء الخوارزمية الجينية GA من تنفيذ جميع مراحلها سوف تعيد قيمة لياقة fitness أفضل فرد ناتج من الجيل الأخير والمتمثل بالفرد الذي يمتلك أقل قيمة لياقة fitness وذلك لكون دالة اللياقة عبارة عن دالة لحساب المسافة الاقليدية Euclidean distance، بالإضافة إلى تسلسله index، إلى النظام المناعي immune system. يقوم النظام المناعي بفحص قيمة اللياقة fitness value فإذا كانت القيمة واقعة ما بين [0 - 0.1] فسوف يتم تمييز الحرف مباشرة وباستخدام تسلسله index، أما إذا كانت القيمة لا تقع في هذا المدى فسوف يتم استدعاء خوارزمية الاختيار النسيلى clonal selection لتميزه. تبدأ خوارزمية الاختيار النسيلى clonal selection بتكوين جيل من الأجسام المضادة antibodies التي ستقوم بعملية التمييز من مجموعة الأحرف المخزنة مسبقاً، بعد الانتهاء من مرحلة التهيئة initialization سوف تتكون 20 خلية مناعية يتم اختيار أفضل خمسة خلايا منها فقط لتدخل إلى المراحل التالية وهي تكوين النسخ clones generation، الانتشار proliferation، الطفرة mutation وأخيراً يتم اختيار أفضل فرد ناتج بعد انتهاء جميع مراحل الخوارزمية، يمثل هذا الفرد الحرف المطلوب، بالإضافة إلى أنه يتم تحديث الذاكرة المناعية immune memory بإضافة قيم العزوم السبعة (SIM) seven invariant moments إلى الذاكرة المناعية ليكتسب النظام المناعة ضد هذا الحرف، وأخيراً يتم تخزين الحرف الناتج من مرحلة التمييز recognition في ملف نصي text file.

5.7.2: النموذج السلوكي لمرحلة التمييز: Recognition Behavioral model

يمثل النموذج السلوكي باستخدام "مخطط انتقال الحالة"، والذي يرمز له بـ STD اختصاراً لـ State Transition Diagram، وكما هو موضح في المخطط ذو الرقم (5): [13]



المخطط ذو الرقم (5): مخطط انتقال الحالة لمرحلة التمييز

6.7.2: وصف مخطط انتقال الحالة: State Transition Diagram Specification

إن مخطط انتقال الحالة STD يمثل سلوك النظام عن طريق توضيح حالاته states والأحداث التي تتسبب بتغيير حالته من حالة state إلى أخرى وكما هو موضح في المخطط ذي الرقم (5)، تضم مرحلة التمييز recognition ثمانية حالات states رئيسية وهي حالات القراءة، التنفيذ، التحميل، ومن ثم حالات القراءة، التنفيذ والكتابة بحيث يستمر النظام بتنفيذ هذه الحالات الثلاثة إلى حين انتهاء إيعازات النظام المناعي ويتم تمييز الحرف لتحديث بعدها حالتين أخيرتين وهي خزن الحرف الناتج وتحديث الذاكرة المناعية.

8.2: تصميم مرحلة التمييز: Recognition Design

يركز التصميم في مجال هندسة البرمجيات على أربع مناطق رئيسية: البيانات، المعمارية، الواجهة والمكونات. ولذلك تشمل هذه المرحلة تكوين النماذج التصميمية التالية: الهيكل البياني، معمارية النظام وتفاصيل مستوى المكونات، أما بالنسبة لتصميم الواجهة فلن يتم إدراجه لضيق المجال المتاح في هذا البحث. [13]

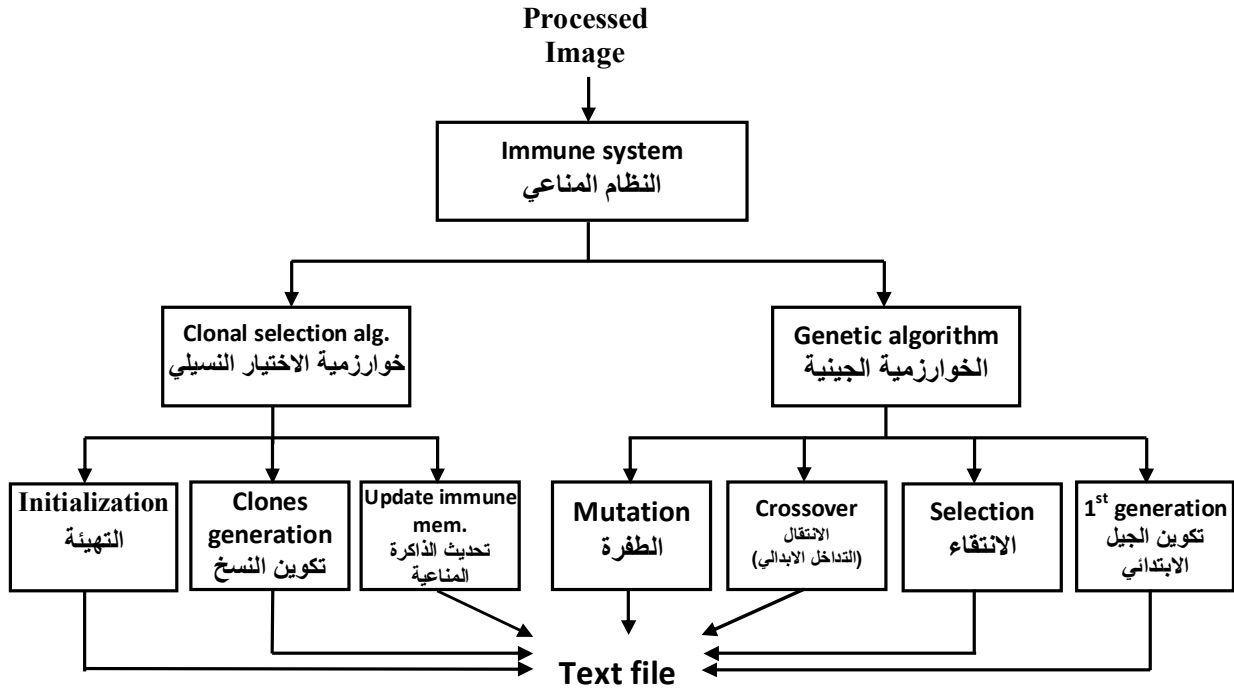
1.8.2: الهيكل البياني لمرحلة التمييز: (Recognition Data Structure)

يمثل الهيكل البياني data structure العلاقات المنطقية بين عناصر البيانات الفردية وكيفية تمثيل هذه العناصر من قبل النظام أثناء انتقالها بين مراحل البرنامج المختلفة [13]. وبالنسبة لمرحلة التمييز recognition يتمثل العنصر البياني الأساسي data item فيها بصورة

الحرف المراد تمييزه, حيث تتعامل لغة MATLAB البرمجية المستخدمة في برمجة البحث مع بيانات الصورة المراد تمييزها كمصفوفة ثنائية الأبعاد two dimensional array, وبذلك ستمثل, ويتم التعامل معها من قبل النظام بشكل مصفوفة ثنائية الأبعاد.

2.8.2: التصميم المعماري لمرحلة التمييز: (Recognition Architectural design)

يستخدم التصميم المعماري لتمثيل هيكل البيانات ومكونات البرنامج المطلوبة لبناء النظام, وهناك عدة أساليب (نماذج) للتصميم المعماري كل أسلوب (نموذج) يستخدم حسب معمارية البرنامج المراد تصميمه وحسب متطلباته, وبالنسبة لمرحلة التمييز recognition فان أفضل أسلوب لتمثيل تصميمه المعماري هو أسلوب "معمارية الاستدعاء والعودة" Call & return architecture وبالتحديد من نوع معمارية "البرنامج الرئيسي/البرنامج الفرعي" main program/sub program architecture والذي يمثل المكونات والهيكل البياني بشكل هرم وكما هو موضح في المخطط ذي الرقم (6):



المخطط ذو الرقم (6): التصميم المعماري لمرحلة

3.8.2: تصميم مستوى-المكونات لمرحلة التمييز: Component – Level Design

هناك العديد من الطرق المستخدمة لتصميم مستوى المكونات مثل المخططات الانسيابية وأسلوب التصميم الجدولي واستخدام (الكود) المزيف [13] pseudo code, أما بالنسبة للبحث

الحالي فسوف يتم تمثيل تصميم مستوى - المكونات component – level design باستخدام لغة برمجة التصميم Programming Design Language والتي يرمز لها اختصارا بـ PDL وهذه الطريقة تتدرج تحت طريقة استخدام (الكود) مزيف للتصميم. وبالنسبة لتصميم مستوى- المكونات لمرحلة التمييز فسوف يتم إدراج جزء صغير منه، وعلى النحو الآتي:

```

PROCEDURE recognition
INTERFACE RETURN character;
TYPE image IS ARRAY TYPE;
% immune system procedure include genetic & clonal selection algorithms
PRODEDURE immune-system
PROCEDURE RETURN char;
[ fitness index] = CALL GA PROCEDURE;
IF fitness >= 0 & fitness <= 0.1
THEN char = CALL rec PROCEDURE FOR index;
ELSE
THEN char = CALL clonal-selection PROCEDURE;
ENDIF
END immune-system PROCEDURE;
    
```

9.2: عملية برمجة مرحلة التمييز : Recognition Coding

تضم هذه المرحلة (كود) البرنامج، وسوف يتم إدراج نموذج من برامج هذه المرحلة وذلك بسبب كبر حجم (الكود) حيث وصلت عدد صفحات (الكود) إلى أكثر من 30 صفحة، وفيما يلي جزء من (الكود) المنفذ بلغة MATLAB:

- Immune system code:

```

function [char]=immune_system(letter)
% checking immune memory for the letter
[index_letter min]=GA(letter);
if min>=0 && min<=0.1
    char=rec(index_letter);
else
    char=IMS(letter);
end
    
```

- Genetic Algorithm code:

```

function [index_letter,minv]=GA(letter)
format short g;
[fname,pname]=uigetfile('*.bmp','please select the ...
                        image');
sh1=fullfile(pname,fname);
sh=imread(sh1);
sh2=im2bw(sh);
letter=imresize(sh2,[50 50]);
    
```

```

a=load ('immunememory.dat');
b=simoments(letter);
[r c]=size(a);
    generation=a;
    for genr=1:20 %no. of generations
        for j=1:r
            sum1=0;
            for i=1:c
                var =sqrt(abs((a(j,i)^2-abs(b(1,i).^2)));
                sum1=sum1+abs(var);
            end
            fatarray(j)=sum1;
        end
        index=1;
        m=fatarray(1);
        for i=1:r
            if(fatarray(i)>m)
                m=fatarray(i);
                index=i;
            end
        end
        pip1=a(index ,1:7);
        randarray=randperm(26);
        index2=randarray(1);
        while index2==index
            randarray=randperm(r);
            index2=randarray(1);
        end
        pip2=a(index2:index2 ,1:7);
        [ch1 ch2]=Arith_Crossover(pip1,pip2);
        [ch1 ch2]=mutation_add(ch1,ch2);
        for ex=1:r
            if(ex==index)
                for j=1:7
                    gen(ex,j)=ch1(j);
                end
            elseif (ex==index2)
                for j=1:7
                    gen(ex,j)=ch2(j);
                end
            else
                for j=1:7
                    gen(ex,j)=generation(ex,j);
                end
            end
        end
        generation=gen;
    end
    fatarray
    index_letter=1;
    minv=fatarray(1);

```

```

for i=1:r
    if (fatarray(i)<minv)
        minv=fatarray(i);
        index_letter=i;
    end
end
index_letter
minv

```

- **Clonal Selection Algorithm code:**

```

mem=[ ]; mem=xlsread('immune_memory');
function [char]=IMS(letter); letter_feature=[ ];
format('shortg'); sh2=letter;
sh3=imresize(sh2,[50 50]); se = strel('arbitrary',3);
IM2 = imclose(~sh2,se); inletter1=slant1(IM2);
inletter1 = bwmorph(inletter1, 'bridge', inf);
inletter1=imgedgescrop(~inletter1);
inletter1= bwmorph(inletter1, 'thin', inf);
inletter1 = bwmorph(inletter1,'dilate');
inletter1=imresize(inletter1,[50 50]);
imagn=create_vector(inletter1);
[row top fin]=gen_initial_pop(imagn);
% row is initial population of 20 members
% fin is their affinity values
[top_row1 fin2 features]= top_five(sh3,row,top,fin);
%top_row1 is top5 elements' array chosen according to
their %affinity values,fin2 is top 5 elements' affinity
values
clones=zeros(1,15);
%creating 5 clones of best element
for j=1:5
    clones_seq(j)= top_row1(1);
    clones(j)=fin2(1);
end
%creating 4 clones of second best element
for j=6:9
    clones_seq(j)= top_row1(2);
    clones(j)=fin2(2);
end
%creating 3 clone of 3rd best element
for j=10:12
    clones_seq(j)= top_row1(3);
    clones(j)=fin2(3);
end
%creating 2 clone of fourth best element
for j=13:14
    clones_seq(j)= top_row1(4);
    clones(j)=fin2(4);
end
%creating 1 clone of last best element
clones_seq(15)= top_row1(5);

```

برنامج لتحويل الصورة إلى متجه

```
clones(15)=fin2(5);
% mutating all clones and finding the best clone
b=0;mu_clones=zeros(1,15);
for kk=1:15
    mu_clones(1, kk)=mutate(clones(kk),kk);
    if mu_clones(1, kk)> b
        b=mu_clones(1, kk);
    end
end
for p=1:15
    if b== mu_clones(1,p)
        sequence=clones_seq(p);
        if p>=1 && p<=5
            immune_mem=features(1,:);
        elseif p>5 && p<=9
            immune_mem=features(2,:);
        elseif p>9 && p<=12
            immune_mem=features(3,:);
        elseif p>12 && p<=14
            immune_mem=features(4,:);
        else
            immune_mem=features(5,:);
        end
    end
end
[rm cm]=size(mem);
mem(rm+1,1)=sequence;
for i=2:8
    mem(rm+1,i)=letter_feature(1,i-1);
end
char=rec(sequence);
xlswrite('immune_memory.xls',mem);
```

10.2: مرحلة اختبار برامج مرحلة التمييز : Recognition programs testing

يتم في هذه المرحلة تصميم سلسلة من حالات الاختبار test – cases التي تملك احتمالية عالية في إيجاد الأخطاء. ويتم ذلك عن طريق استخدام تقنيات اختبار البرنامج software testing techniques والتي تكون على نوعين: [13]

(1) نوع يختبر المنطق الداخلي لمكونات البرنامج، ويسمى اختبار الصندوق الأبيض – white box testing .

(2) والنوع الثاني يفحص الإدخال والإخراج للكشف عن أخطاء البرنامج الوظيفية، السلوكية ومشاكل الأداء، ويسمى باختبار الصندوق الأسود [13].black – box testing

1.10.2: فحص الصندوق الأبيض: White – box testing

تتضمن هذه المرحلة كما ذكرنا، تصميم حالات اختبار test cases، عن طريق استخدام هيكل السيطرة للتصميم الإجرائي procedural design، لاشتقاق حالات الاختبار، وفيما يلي خطوات تنفيذها. [13]

1.1.10.2: اختبار طريق القاعدة: (Basis - Path Testing)

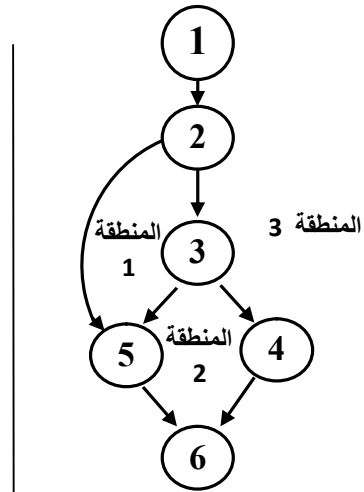
تبدأ هذه التقنية باشتقاق مقياس منطقي للصعوبة أو التعقيد logical complexity measure للتصميم الإجرائي procedural design، لكي يستخدم كدليل لتعريف مجموعة القاعدة basis set لمسالك التنفيذ، والتي عليها أن تضمن تنفيذ كل جملة statement في البرنامج على الأقل مرة واحدة خلال الفحص [13]. يبدأ اختبار طريقة القاعدة basis path testing باستخدام (الكود) أو تصميم مستوى - المكونات component - level design وهو مخطط PDL كقاعدة ليتم تصميم الرسم البياني للتدفق Flow Graph وبما أن مرحلة التمييز تتضمن عدة برامج لذلك سوف نقوم بإجراء عملية الاختبار لبرنامج واحد فقط لتوضيح كيفية القيام بعملية الاختبار. [13]

يبدأ اختبار طريقة القاعدة basis path testing بتكوين الرسم البياني للتدفق flow graph وبالاعتماد على تصميم مستوى - المكونات الخاص بكل برنامج لتحديد العقد nodes والحافات edges، وكما هو موضح في خطوات اختبار البرنامج الآتي:

1) Immune system procedure:

```

PRODEDURE immune-system
PROCEDURE RETURN char;
[ fitness index] = CALL GA PROCEDURE; } 1
    2          3
IF fitness >= 0.0001 & 0.1
THEN char = CALL rec PROCEDURE FOR index; } 4
ELSE char = CALL clonal-selection PROCEDURE; } 5
ENDIF
END immune-system PROCEDURE; } 6
    
```



المخطط ذو الرقم (7): الرسم البياني

لبرنامج النظام المناعي flow graph

سري: _____

(Cyclomatic Complexity)

تتضمن الخطوة الثانية من اختبار طريقة القاعدة basis path testing حساب قيمة التعقيد الدوري Cyclomatic Complexity للبرنامج، وكما يلي:

1- مقدار التعقيد الدوري الذي يرمز له بالرمز $V(G)$ مساوي لعدد المناطق المغلقة regions في الرسم البياني لمسالك البرنامج flow graph، وبالنسبة للمخطط ذي الرقم (7) الذي يمثل flow graph تبلغ عدد المناطق المغلقة 3 وذلك بحساب المنطقة الخارجية أيضا.

2- $V(G) = E - N + 2$ ، حيث أن E = تمثل عدد الحافات الممثلة بالأسهم، و N = عدد العقد .

وبالنسبة للمخطط (7) فإن التعقيد الدوري $V(G) = 7 - 6 + 2 = 3$.

3- $V(G) = \text{predicate} + 1$ ، حيث أن predicate تمثل عدد العقد التي ينطلق منها أكثر من مسلك. وبالنسبة لبرنامجنا فإن التعقيد الدوري $V(G) = 2 + 1 = 3$.

بالنتيجة فإن مقدرا التعقيد الدوري cyclomatic complexity لبرنامج النظام المناعي immune memory هو 3، أي انه هناك 3 مسالك مستقلة independent paths هي التي سوف تشكل مجموعة القاعدة basis set، لذلك تتضمن الخطوة الثالثة من اختبار طريقة القاعدة basis path testing تصميم 3 اختبارات tests لضمان تغطية جميع جمل statements برنامج النظام المناعي immune memory. [13]

وبالنسبة للمسالك التي ستشكل مجموعة القاعدة basis set للبرنامج فهي تشمل (مشتقة من المخطط ذي الرقم 7):

المسلك الأول: 1 - 2 - 3 - 4 - 6 .

المسلك الثاني: 1 - 2 - 3 - 5 - 6 .

المسلك الثالث: 1 - 2 - 5 - 6 .

3.1.10.2: اشتقاق حالات الاختبار: Deriving Test Cases

وهي الخطوة الأخير من اختبار طريقة القاعدة basis path testing وتتضمن تحضير حالات اختبار Test Cases تضمن تنفيذ كل مسلك في مجموعة القاعدة basis set:

حالة اختبار المسار الأول:

- تصف هذه الحالة كون صورة الحرف المراد تمييزه موجودة في الذاكرة المناعية immune memory فيتم تمييزه مباشرة عن طريق الخوارزمية الجينية GA.
- النتيجة المتوقعة: طباعة الحرف المميز على ملف نصي.

حالة اختبار المسار الثاني:

- يصف المسلك الثاني حالة عدم إيجاد الحرف المراد تمييزه في الذاكرة المناعية immune memory لهذا يتم استدعاء خوارزمية الاختيار النسيلي clonal selection لتمييزه.

- النتيجة المتوقعة: طباعة الحرف المميز على ملف نصي.

حالة اختبار المسار الثالث:

- يصف هذا المسلك حالة تحقق الجزء الأول من شرط الذاكرة المناعية وعدم تحقق الجزء الثاني مما يؤدي إلى استدعاء الاختيار النسيلي clonal selection ليتم تمييز صورة الحرف.

- النتيجة المتوقعة : أيضا طباعة الحرف المميز على ملف نصي.
وهكذا يتم تطبيق خطوات هذا الاختبار على بقية برامج مرحلة التمييز.

2.10.2: فحص الصندوق الأسود: Black – Box testing

أما بالنسبة لاختبار الصندوق الأسود فكما ذكرنا بأنه يركز على المتطلبات الوظيفية functional requirements للبرنامج ويحاول إيجاد الأصناف التالية من الأخطاء:

(1) انعدام أو نقصان بعض الوظائف.

(2) أخطاء في الأداء والسلوك.

(3) أخطاء في القيم البدائية والنتائج النهائية.[13]

وقد تم التأكد من خلو برامج مرحلة التمييز recognition من جميع هذه الأخطاء خلال مراحل تنفيذ هذه البرامج, وتم الحصول على النتائج المطلوبة.

Conclusions

الاستنتاجات:

1- تم تحقيق نسب التمييز الآتية:

نوع الإدخال	نسبة التمييز
في حالة تمييز صور الأحرف المتصلة cursive والمكتوبة بخط اليد التي دخلت في تكوين قاعدة البيانات	بلغت النسبة % 100
في حالة تمييز صور لأحرف متصلة cursive ومكتوبة بخط اليد لم تدخل في تكوين قاعدة البيانات	بلغ معدل النسبة % 94
في حالة تمييز صورة وثيقة تضم نصاً متكوناً من كلمات مكتوبة بحروف متصلة وبخط اليد	تراوحت النسبة بين % 85 - 80

ملاحظة: تم حساب نسبة التمييز لكل نموذج اختبار على حدة وذلك عن طريق حساب عدد أحرف وثيقة الاختبار الداخلة لبرنامج التمييز, وعدد الأحرف التي يتم تمييزها بشكل صحيح ومن ثم تم حساب النسبة.

2- إن استخدام خوارزمية الاختيار النسيلى وتهجينها بالخوارزمية الجينية GA قد ساعد كثيرا في تحسين نسبة التمييز وزيادة سرعة عملية التمييز, فقد تم تجربة تهجين الاختيار النسيلى بالشبكة العصبية back propagation عن طريق استخلاص العزوم السبعة للحرف SIM وتدريب الشبكة عليها وكانت النتائج كما هي موضحة في الجدول الآتى:

التهجين بالخوارزمية الجينية GA	التهجين بالشبكة العصبية
- لا تحتاج إلى التدريب, ولكن تحتاج إلى وقت معين للتمييز.	- بلغت مدة تدريب الشبكة أكثر من عشرين ساعة.
- حققت نسب التمييز المذكورة أعلاه باستخدام مئة نموذج فقط لكل حرف في بناء قاعدة البيانات.	- حققت نسبة تمييز لا تتجاوز الـ 10% بعد تدريب الشبكة على خمسة وأربعين نموذجا لكل حرف, وقد أشارت المصادر إلى أنها تحتاج إلى اثني عشر ألف نموذج [14] لكي تحقق النسبة التي تم تحقيقها في البحث الحالى, والمذكورة أعلاه.

Recommendations

التوصيات:

- من أجل تحسين نسبة تمييز الوثائق, بالإمكان تنفيذ عملية تقطيع الأحرف باستخدام الشبكات العصبية بعد تدريبها جيدا لتقليل نسبة الخطأ في عملية التقطيع (خصوصا في حالات الخط المتصل).
- بالإمكان زيادة كفاءة النظام المناعي عن طريق وضع حقول في واجهة المستخدم تمكنه من تصحيح الأحرف التي تم تمييزها بشكل خاطئ, ليكتسب النظام المناعة ضد مثل هذه الأحرف ويقوم بتمييزها بشكل صحيح في المرات التالية التي يدخل فيها للنظام صور مشابهة لصور هذه الأحرف.
- بالإمكان تجربة استخدام خوارزميات مناعية أخرى لتنفيذ عملية التمييز ومقارنتها مع نتائج بحثنا.
- تجربة تهجين النظام بتقنيات ذكائية أخرى مثل المنطق المضرب, أو خوارزميات مناعية أخرى مثل خوارزمية الاختيار السلبي negative selection ونظريات الشبكات المناعية immune network theories.

References

المصادر:

- 1) D. Dasgupta, Z. Ji, and F. Gonzalez, F, "Artificial immune system (AIS) research in the last five years", in Congress on Evolutionary Computation (CEC'03), 2003, Volume: 1, pp. 123- 130.
- 2) Zheng Tang, Koichi Tashima, and Qi P. Cao, "Pattern recognition system using a clonal selection-based immune network," Systems and Computers in Japan, Volume 34, Issue 12, pp. 56 - 63, 2003.
- 3) Z. Ji and D. Dasgupta, 2004, "Real-valued negative selection algorithm with variablesized detectors," in LNCS 3102, Proceedings of GECCO, pages 287–298.
- 4) L. N. d. Castro and F. J. V. Zuben, 2002, "Learning and Optimization Using the Clonal Selection Principle," *IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems*, vol. 6, pp. 239-251.
- 5) Horst Bunke and Tam as Varga, "O_-line Roman Cursive Handwriting Recognition", Swiss National Science Foundation program \Interactive Multimodal Information Management (IM)2 "in the Individual Project\Scene Analysis", part NCCR.
- 6) C. de Stefano, A. Della Cioppa, and A. Marcelli, 1999, "Handwritten Numeral Recognition by Means of Evolutionary Algorithms", in Proc. of the 5th Int. Conf. on Document Analysis and Recognition (ICDAR), Bangalore, India, page: 804- 808.
- 7) J. H. Carter, 2000, "The Immune System as a model for Pattern Recognition and classification", Journal of the American Medical Informatics Association. Vol. 7, no. 3, pp.28-41.
- 8) *Rafael Palacios and **Amar Gupta, "A system for processing handwritten bank checks automatically", *Universidad Pontificia comillas, Madrid, Spain; **Massachusetts institute of technology, Cambridge, MA, U.S.A.
- 9) Victor Lavrenko, Toni M. Rath and R. Manmatha, "Holistic Word Recognition for Handwritten Historical Documents", Center for Intelligent Information Retrieval University of Massachusetts Amherst Amherst, MA 01003.
- 10) "L. Lam, S.W. Lee, and C.Y. Suen, "Thinning Methodology – A Comprehensive Survy", in IEEE transactions on Pattern Analysis and Machine Intelligence, 14(9):869-885, 1992.
- 11) "United States Check Study", 2001, Result of 1999 Research. Edited by Green Sheet, Inc. <http://www.greensheet.com/CheckStudy/index99.html>, Oct 25.
- 12) Levent Burak Kara, "AUTOMATIC PARSING AND RECOGNITION OF HAND-DRAWN SKETCHES FOR PEN-BASED COMPUTER INTERFACES", doctor of philosophy, mechanical engineering, carnegie mellon university.
- 13) Roger S. Pressman, 2010, Ph.D., "Software engineering, a practitioner's approach", 7th edition, McGraw-Hill company .
- 14) *Utpal Garain, *Mangal P. Chakraborty and **Dipankar Dasgupta, "Recognition of Handwritten Indic Script using Clonal Selection Algorithm", *Indian Statistical Institute, 203, B.T. Road, Kolkata 700108, India **The University of Memphis, Memphis, TN 38152.