حوسبة الحرف العربي بين الواقع و المأمول

أ/ سليمة يحياوي

قسم اللغة العربية– جامعة أبي بكر بلقايد– تلمسان

توطئة:

اتسمت السنوات الأخيرة بزيادة اهتمام الباحثين بالحوسبة العربية و تطبيقاتها المختلفة، وقد كان سبب التزايد الملحوظ في وتيرة هذه البحوث ناتج عن عوامل عديدة منها:" الاهتمام المتزايد بتطبيقات الحاسوب في مجالات جديدة مثل تسيير المكاتب آليا و مكاتب الخدمات البريدية و كذا التقدم التقني الذي ينزل بكثير من معدات الحاسوب المادية إلى متناول القطاعات و بالإضافة إلى التطبيقات الجارية بالنسبة إلى اللغات الأخرى غير العربية "[36] و يلاحظ هذا الاهتمام في العديد من تطبيقات الحوسبة العربية و منها التعرف الآلي (أوما يعرف بالتعرف الضوئي) على الكتابة العربية.

إن التعرف الآلي (أو التعرف الضوئي) على الكتابة العربية اليدوية هو عملية تحويلها إلى نصّ رقمي اعتمادا على صورة الكلمة المكتوبة فقط؛ بمعنى تحويل اللّغة المكتوبة من شكلها الرّسومي المعروف إملائيا إلى شكلها الرّقمي المعروف لدى أجهزة الحاسوب.

و قد قد حدّد أستاذي الدكتور عمر ديدوح من جامعة تلمسان في بحثه الموسوم :توطين تقنات تعليم الحاسوب تمييز الخط العربي اليدوي بعدين رئيسيين متكاملين لهذا المشروع و هما البعد السياسي و البعد الحضاري، وأكد أن ترميز الحرف العربي و توصيفه للحاسوب يعد امتدادا طبيعيا للدعامتين السالفتين لتتواصل حلقات الخط العربي ماضيها بحاضرها و مستقبلها و يكمل البعد السياسي البعد الحضاري للغة العربية بشقيها المنطوق و المخطوط[37].

الخط العربي مطبوعا كان أم يدويا يبقى محل بحث و اهتمام من قبل الباحثين و اللغويين خاصة في عصرنا الحاضر عصر التكنولوجيات الحديثة والاتصالات، "و قد تم خلال العقود الأربعة

---

[1]- نبوي عاصم عبد الفتاح،محمود صبري عبد الله، الحمودي خالد بن عبد الرحمن: تمييز حروف اللغة العربية المكتوبة آليا باستخدام الشبكات العصبية ذات الاتشار الرجوعي- مجلة جامعة الملك سعود– المجلد9– الرياض 1417هـ/1997م– ص: 1

[2]- ينظر.عمر ديدوح: مستقبل اللغة العربية في سوق اللغات-منشورات المجلس الأعلى للغة العربية – الجزائر-2009-ص 76،ص 77

الأخيرة تطوير نظم تعرف آلية في اللغات اللاتينية و اليابنية و الصينية،لكن تبقى البحوث في هذا المجال بالنسبة للغة العربية محدودة نسبيا"[38]، وذلك لما تتميز به الكتابة العربية من خصائص، وكثرة أنواع أساليب الخط المستعمل، "حيث يقارب عددها 450 نوعا،إذ يمكننا أن نصل في بعض التجارب إلى نسب عالية من التعرف أين يكون الخط المستعمل من نوعية واحدة و بمعطيات غير شاملة،لكن تتدنى هذه النسبة إذا كان النص يحتوي على عدة نوعيات خطية و إذا أخذ بعين الاعتبار جودة الوثيقة و سلامتها"[39].

يستعمل هذا المشروع في "المعالجة الآلية للملفات الإدارية و الاستبيانات و العناوين البريدية و الشيكات البنكية و ميدان ترقيم التراث الثقافي العربي القديم و المحافظة عليه"[40]، كما أن هذا المجال سيخدم اللغات التي تستخدم الحروف العربية كالفارسية والأوردية والكردية والأمازيغية و الماليزية والأندنيسية [41].

أولا: خصوصية الكتابة العربية

ممّا لاشّك فيه أنّ اللّغة العربية لغة مميّزة، وحروفها على ذلك المستوى من التمييز، وهذه الصّفات المميّزة للحروف العربية تعطي لعملية الكتابة بها بعض الخصوصية، لكنّها تضيف بعض الصعوبات بالنّسبة لعملية التعرّف عليها بواسطة الحاسوب،كاتجاه الكتابة و عملية اتّصال الحروف العربية عند كتابتها الذي يؤدّي إلى تداخل حدود الحروف فيما بينها، و ضرورة وجود بعض المجموعات الّتي لها الشكل العام نفسه مع بعض الاختلافات البسيطة؛ كعدد النّقاط وموقعها،

---

[3]- د فريال حاج حسن: التعرف الآلي على الأرقام الهندية –قسم تقنية المعلومات-كلية علوم الحاسب و المعلومات- جامعة الملك سعود- المملكة العربية السعودية.

[4]– سامي بن موسى ،منصف الشرفي،محمد عادل العليمي: نظام للتعرف على نوعية الخط العربي باستعمال الهندسة الكسرية و العصبيات الاصطناعية بحث المؤتمر الدولي لممارسات علوم الحاسب باللغة العربية1-4 أبريل 2008 الدوحة– قطر.

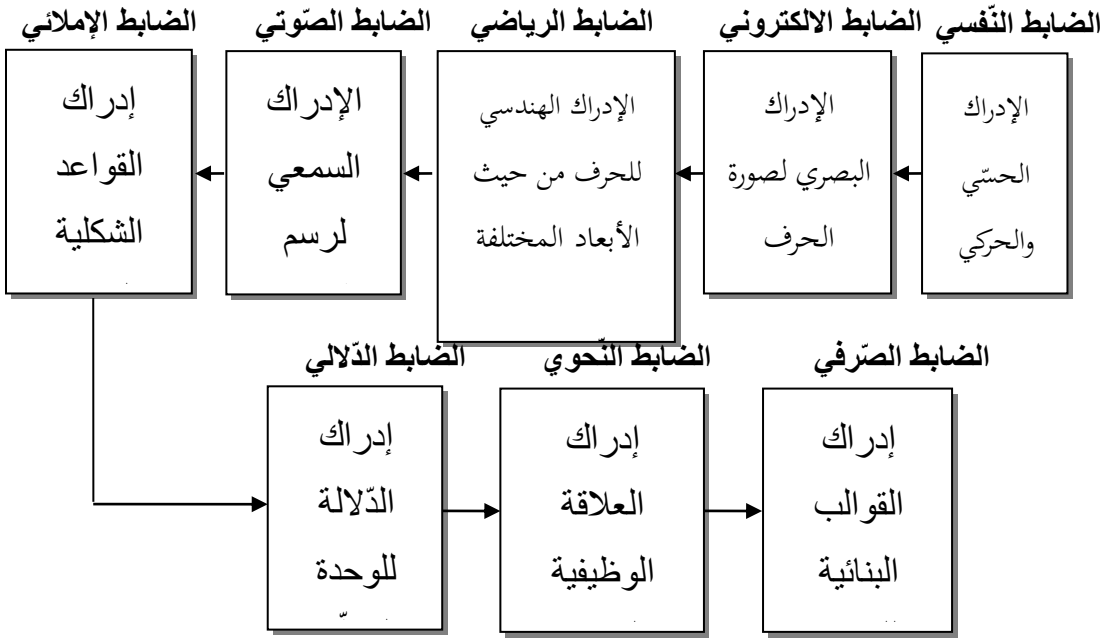
[5]– عبد الكريم البعتي ،منجي خير الله ،عبد اللطيف الناجي ،محمد عادل العليمي : استرجاع التسلسل الزمني للكتابة المنفصلة باستعمال الخوارزميات الجينية بحث المؤتمر الدولي لممارسات علوم الحاسب باللغة العربية1-4 أبريل 2008 الدوحة قطر.

[41]-.عمر ديدوح: المرجع السابق- ص: 79 -

بالإضافة إلى اختلاف شكل الحرف اعتمادا على موقعه بالنّسبة إلى الكلمة (كونه منفصلا أو واقعا في بداية الكلمة أو وسطها أو نهايتها)[42].

و إذا كان جمال الخطّ العربي يمثّل منزلة يسعى إليها كلّ من أراد الإبداع، فإنّ الكتابة السّليمة والصّحيحة هي الرّكيزة الأساسية لهذا الجمال والإبداع، من هنا اهتمّ العلماء العرب في القديم والحديث بالخطّ العربي وربطوا بينه وبين الكثير من الفروع كـ: الأصوات والنحو والصّرف والدّلالة والمعاجم وعلم النفس وما إليها.

و المخطط التالي يفسر ظاهرة تظافر الضوابط اللغوية وغير اللغوية لاستنتاج الوظيفة الخطية[43]:



ثانيا: طرق التعرف الآلي على الخط العربي اليدوي

إن إنّ أهمّ طرائق التعرّف على الحروف ضوئيا هي[44]:

---

7)- ينظر: كمال جنبي- استخدام الحاسوب في التعرّف على الكلمات العربية المكتوبة بخطّ اليد- سجل بحوث المؤتمر الهندسي السّعودي الرّابع- 12-15 جمادى الأخيرة 1416هـ الموافق لـ05-08 نوفمبر 1995م- ص: 69.

8)- سهام موساوي: الضّوابط اللّغوية للتعرّف الآلي على الخطّ اليدوي العربي (رسم الهمزة أنموذجا)- رسالة ماجستير- قسم اللغة العربية- كلية الآداب والعلوم الإنسانية والاجتماعية- جامعة أبي بكر بلقايد- تلمسان- 2006-2007- ص: 20.

9)- عبد الملك السلمان:التعرف الضوئي للكتابة العربية كلية علوم الحاسب و المعلومات جامعة الملك سعود المملكة العربية السعودية: WWW.CCIS.EDU.SA

– التعرف على الحروف بطريقة الاتّصال المباشر بالحاسوب (on-line) في أثناء الكتابة أو ما يعرف بالنظام المتزامن؛ وهو خاص بتلك الحروف المفردة غير الموصولة والتي تكتب باليد بقلم خاص موصول بالحاسوب، وتتمّ معالجتها في أثناء الكتابة، الأمر الّذي لا يزوّد نظام التعرّف بصورة الحرف فقط بل بمعلومات عن توقيت كتابة كلّ جزء من الحرف أيضا من الأمثلة في ذلك "**القلم العربي**" من شركة الأقصى (الكويت) و "**القلم الضوئي العربي**" من شركة دار حوسبة النص العربي (الأردن) .

– التعرّف على الحروف المكتوبة باليد(off-line) أو ما يعرف بالنظام غير المتزامن ، من أمثلة البرامج التجارية "**القارئ الآلي**" صخر.

ثالثا: النظام المقترح للتعرف الآلي على الحروف اليدوية المعزولة

في بحثنا هذا، و لكي نتمكن مستقبلا من التعرّف على الكلمة العربية المكتوبة يدويا يستوجب علينا التعرّف أولا على الحرف اليدوي في حالته المعزولة، هذا الأخير تتمّ دراسته لغرض تحديد الصّفات المميّزة له، ثم يتم استخراج قاعدة المعطيات الخاصّة به،و من ثم التعرف عليه داخل الكلمة.

و ينقسم العمل التطبيقي إلى مرحلتين الأولى خاصة **بمعالجة الصورة** و المرحلة الثانية متعلقة **بتعليم الشبكات العصبية الإلكترونية**.

المرحلة الأولى: معالجة الصورة

إنّ عملية المعالجة الآلية للصّورة لها طرق متنوّعة ومراحل متعدّدة وفيما يخصّ الطّرق المتّبعة في هذا البحث فقد تمّ إجراؤها باستعمال لغة البرمجة (matlab)، وهي لغة اصطناعية كغيرها من لغات البرمجة، تحتوي على علبة حسابية لكلّ مرحلة عملية إجرائية، يتوقّف اختيار طريقة المعالجة التي تسمح بالمرور إلى المراحل الموالية على نوعية العيّنة المختارة للدّراسة.

معالجة الصّورة هي أحد فروع علوم الحاسب تهتمّ بإجراء عمليات على الصّورة (Image Processing) بهدف تحسينها طبقا لمعايير محدّدة أو استخلاص بعض المعلومات منها[45].

---

10)-vue : alain DEITERLEN : Traitement d’image- Cours laboratoire MIPS

نظام معالجة الصورة الذي إعتمدناه يتألّف من ستّة مراحل متتالية وهي على الترتيب:

1- استقبال الحروف العربية وتخزينها في ذاكرة الحاسوب:

بعد الحصول على عيّنات خطّية مختلفة للحروف العربية اليدوية المعزولة ، نجري عليها عملية المسح الضّوئي، لأجل الحصول على صورة رقمية قابلة للتمثيل في مصفوفة بقيم رقمية لها شدّة ضوئية وكثافة موزّعة بنقط مختلفة على مساحة الصّورة، حيث كلّ نقطة ضوئية من الصّورة تحتوي على بعدين بقيمة رقمية (س، ع)، وتسمّى عناصر الصورة أو الوحدات الضوئية (pixel).

2- مرحلة قلب مصفوفات الصورة الضوئية (Inversion des pixels):

تتمثّل هذه المرحلة في قلب مصفوفات الصورة الضوئية وذلك بتحويل النّقط البيضاء إلى نقط سوداء والعكس، تسهيلا منّا لنقلها إلى مرحلة ما قبل المعالجة.

3- مرحلة الترشيح (Filtrage):

تهدف هذه المرحلة إلى تهيئة صورة الهاء لتسهيل معالجتها داخل الحاسوب، وذلك بإزالة جميع الشوائب الطّارئة، والتي هي عبارة عن معلومات زائدة مزاحمة للمعلومات الأصلية الموجودة في الصورة، فقد تكون عبارة عن نقط سوداء وخطوط عشوائية ليس لها وظيفة في شكل الحرف ويمكن تعديل هذه المعلومات وإزالة شوائبها بعدّة طرق وذلك بحسب درجتها المتواجدة في الصّورة.

4- التقطيع (Segmentation):

عملية تقطيع أو تجزئة مناطق الصّورة إلى مناطق متساوية لمقارنتها مع بعضها البعض ثمّ تحديد جميع الخصائص العامة لصورة الحرف من حيث الشّكل والمحيط واستخراج مكوّناتها الأساسية باستعمال عدّة مؤثّرات (Opérateurs) على الصّورة.

والمؤثّر (Deriche canny) الأكثر استعمالا في معالجة الصورة، و الأدق حساسية للشوائب ، والأكثر دقّة في تحديد محيط وحجم ومساحة الحرف لتسهيل عملية التقطيع وإنجاحها.

5- استخراج الخصائص:

بالرّغم من وجود العديد من الصّفات المميّزة والتي يتمّ استخدامها في عملية التعرّف إلاّ أنّ الصّفات التالية هي التي قد تم استعمالها في هذه المرحلة:

- حساب المسافة بين قطع المحيط
- حساب المسافة بين قطع المحيط و مركز الثقل
- حساب الزوايا:

الزاويا Angle polaire

الزاويا Angle des vertexe

المرحلة الثانية: تعليم الشبكات العصبية الإلكترونية

إنّ الشبكات العصبية الاصطناعية هي: "تركيبات للمعالجة المتوازية الموزّعة، تعتمد أساسا على عنصر المعالجة القادر على العمل كذاكرة محلّية مع إجراء عمليات المعالجة المختلفة، تعتمد على القيم المدخلة وكذلك القيم المخزونة للذّاكرة المحلّية لهذه العناصر"[46].

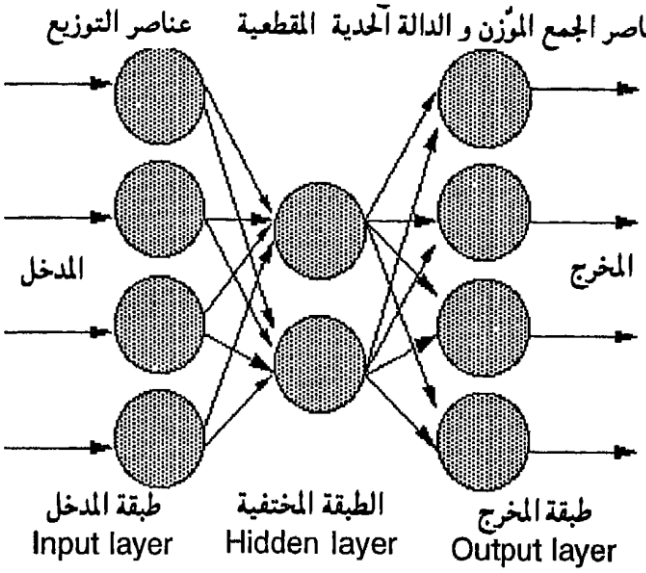
نقترح في هذا البحث استخدام الشّبكة العصبية ذات التركيب البنيوي ويسمّى بالإدراك المتعدّد الطّبقات (Multi layer perception) بعد القيام بمعالجة الصّورة الضّوئية واستخلاص خصائصها.

يتكوّن النّموذج العام للشبكات متعدّدة الطّبقات من ثلاث طبقات:

1- طبقة المدخل.

2- الطّبقة المختفية.

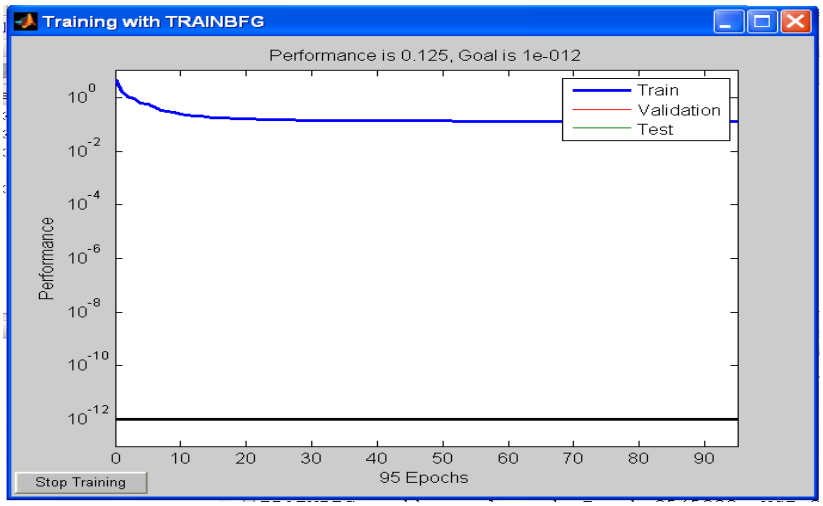
3- طبقة المخرج



[46]-.محمد علي الشرقاوي: الذّكاء الاصطناعي والشبكات العصبية- سلسلة علوم وتكنولوجيا حاسبات المستقبل- ص:- ص: 259-260.

نموذج: لقد وقع اختيارنا على حرف الهاء العربية مع انتقاء خطّ النّسخ كنموذج للدّراسة بوصفه الخطّ الأقرب إلى الخطّ المرجعي المطبعي، قمنا بتعليم الشّبكات العصبية على مرحلتين:

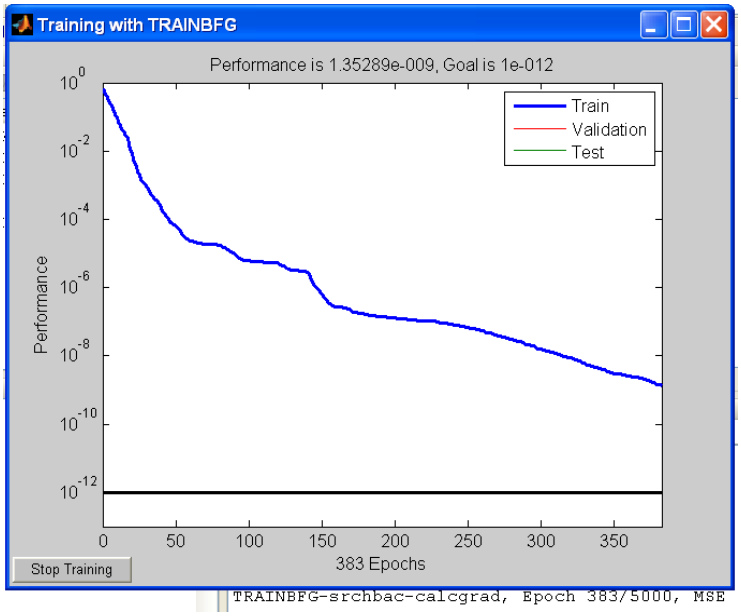
المرحلة الأولى: تعليم الشبكات ثلاث عيّنات خطّية فقط (هـ، هـ، هـ)، فكان عدد الحالات المتعرّف عليها 12 حالة من مجموع 26 بدالّة خطأ مسموح 10^{-1}



حقيقي / مفترض	هـ	هـ	هـ
هـ	1	4	2
هـ	1	2	3
هـ	8	4	1
المجموع	10	10	6

المرحلة الثانية:

تمّ تعليم الشّبكات العصبية أربع عيّنات خطّية (هـ، هـ، ه، هـ) فكان عدد الحالات المتعرّف عليها 18 الحالة من مجموع 38 بدالّة خطأ مسموح: 10^{-11}.

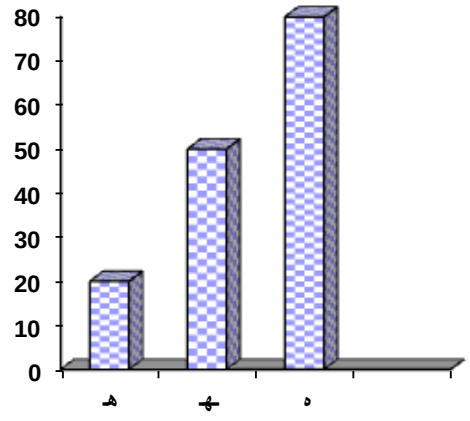


حقيقي / مفترض	ه	هـ	هـ	هـ
ه	9	2	0	2
هـ	0	1	0	0
هـ	0	0	0	0
هـ	3	3	10	8

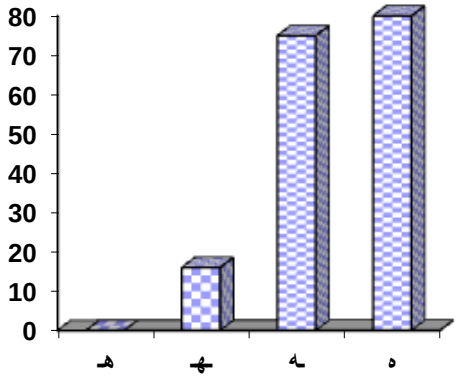
المجموع	12	6	10	10

تعليق:

بعد استقراء نتائج المرحلة الأولى والمرحلة الثانية نستنتج أنّه قد تمّ التعرّف على حرف الهاء في آخر الكلمة (ﻫ) بـ: 80% و(ﻪ) بـ: 75% وهي نسب إيجابية مقارنة مع نسب التعرّف على حرف الهاء في أوّل الكلمة ووسطها (هـ، ﻬ)؛وهذا يعني أنّ الصّفات التّمييزية لشكليّ الهاء في آخر الكلمة هي التي ساعدت الحاسوب على التعرّف عليه.



نتائج المرحلة الأولى



نتائج المرحلة الثانية

الخاتمة:

هذا البحث يعدّ محاولة في مجال حوسبة الحرف العربي حاولنا من خلاله المزاوجة بين الدّرسين اللّغوي والتّقني، وخلاصة ما انتهى إليه البحث:

❁ إدماج تقنيات الذّكاء الاصطناعي عن طريق نظام الشّبكات العصبية الإلكترونية في حوسبة الحرف العربي، باعتبارها نمذجة آلية للمخّ البشري، وبما يتّصف به هذا النّظام من البساطة حيث لا يوجد هناك معادلات رياضية معقّدة يلزم إيجاد الحلول لها ولا صفات تمييزية معقّدة يلزم التعامل معها.

❁ تقترح وكامتداد لهذا العمل البحثي استخدام النّظام نفسه للتعرّف على حرف الهاء العربية داخل الكلمة مع الأخذ بعين الاعتبار الزيادة في قيم المداخل وعدد العيّنات الخطّية الموضوعة للدّراسة الممسوحة ضوئيا بماسح ضوئي أكثر قوّة، وذلك لتمكين الشّبكات العصبية الالكترونية التعرّف على أكبر عدد ممكن من الخصائص التمييزية للصّورة الضّوئية.

خلاصة ما يمكن أن يقال أنّنا في حاجة ماسّة أكثر من أيّ وقت مضى إلى الحثّ على تداول اللّسان العربي عل جميع المستويات والأصعدة مستفيدينا من التطبيقات المعلوماتية المتطوّرة.