استخدام الدالة المميزة التربيعية في تمييز أنماط الأرقام العربية

خالد ضاري عباس الطائي* والمرحوم. محمد رمضان عتاب
*** مدير مركز الحاسبة، كلية الإدارة والاقتصاد، جامعة بغداد.**

الخلاصة

لتصميم منظومة الكترونية لتحليل انماط الارقام العربية، تطلب ذلك مناقشة تمييز الانماط وعلاقته بالتحليل االاحصائي. مع عرض لتحويل فورير واشتقاق للدالة المميزة التربيعية المستخدمة في عملية التمييز. واختيار المتغيرات الداخلة في الدالة كذلك المعايير المستخدمة لقياس اداء الدالة المميزة باستخدام تقدير خطأ التصنيف الذي استخدم لأول مرة باعتماد طريقة التعويض. ولأثبات ذلك تم اعتماد عينة من (150) شخص كتبوا الارقام العربية من (9–0) يدوياً وبذلك حصلنا على (1250) نمط.

اعطى الاسلوب الاحصائي للتمييز نتائج جيدة في تمييز الارقام العربية من حيث خطأ التصنيف اذ اعطى (15) متغيرا وهو افل عدد بين باقي الاساليب ووصل تقدير احتمال خطا التصنيف الى الصفر. من خلال برامج كتبت بلغة ++C.

1– 1 المُـقدمة

من الظواهر البارزة للتقدم العلمي محاولات الإنسان المستمرة لإسناد جزء كبير من نشاطاته المختلفة إلى الآلة لتمكنه من إنجاز هذا النشاط بشكل أسرع وأفضل وبجهد أقل. ويعد اختراع وتطور الحاسوب من أبرز هذه المحاولات وأكثرها مساندة للجهد البشري.

وتتضمن مهمة تمييز الأنماط بتصميم منظومة إلكترونية يمكن من خلالها تمييز الأنماط الجديدة الغير معروفة التميز. ويتم ذلك بأخذ مجموعة من الأنماط المعروفة التميز كعينة تعليمية حيث يؤخذ منها P من المتغيرات الخاصة بكل نمط بحيث تتوفر فيها افتراضات طريقة التصنيف المطلوبة بعدها تختزل هذه المتغيرات بأسلوب استخلاص الصفات بحيث تبقى q من المتغيرات التي تكون كافية لعملية التميز وهذه المتغيرات هي التي تكون الدالة المميزة الخاصة بالمنظومة والخطوة الأخيرة تتمثل باختبار كفاءة حالة التميز عن طريق تقدير احتمال خطأ التصنيف [3].

1–2 هدف البحث

إن الهدف من البحث هو تميز للأرقام العربية آلياً باستخدام الدالة التمييزية التربيعية وذلك من خلال بناء منظومة تمييز آلية.

2– الجانب النظري

2–1 تحويل فورير Fourier Transform

يعرف تحويل فورير F(k) للدالة f(t) بالشكل الآتي:

$$F(k) = \int_{-\infty}^{\infty} f(t)\, e^{-ikt}\, dt,\ i = \sqrt{-1} \ \ldots\ldots\ldots\ldots\ldots\ldots (2.1)$$

وتحويل فورير العكسي

$$f(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(k)\, e^{-ikt}\, dk \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.2)$$

ويتطلب تحويل فورير الشروط الآتية:

أ– f(t) دالة واحدة إلى واحد في المجال $-\infty < t < \infty$.

ب– f(t) قابلة للتكامل في المجال $-\infty < t < \infty$.

أي أن:

$$\int_{-\infty}^{\infty} f(t)\, dt,\ t < \infty$$

إن هذا التحويل بنقل الدالة f(t) من المجال الزمني إلى المجال الترددي. وإن عملية حساب تحويل فورير تتطلب إجراء عمليات حسابية كثيرة (n^2 من العمليات المركبة) لذلك جاءت الحاجة لاستخدام الطرق المركبة المطورة المقترحة من قبل (Cooley and Takeg, 1959) وهي تحويل فورير السريع التي تقلل العمليات الحسابية إلى n log(n) من العمليات المركبة[5].

هناك الكثير من الخوارزميات التي تعرف بأسم تحويل فورير السريع لذلك تم ذكر واحدة منها تسمى Radix-2FFT algorithm وتعرض كما يأتي [5]:

for even values:

$$\left.\begin{array}{l}\text{for even values:}\\ F(2k)=\sum_{t=0}^{\frac{n}{2}-1}\left[f(t)+f(t+\frac{n}{2})\right]W^{2+k},\ k=0,\ldots,\frac{n}{2}-1\\ \text{for odd values:}\\ F(2k+1)=\sum_{t=0}^{\frac{n}{2}-1}\left[f(t)-f(t+\frac{n}{2})\right]w^{t}w^{2ik},\ k=0,\ldots,\frac{n}{2}-1\\ i=\sqrt{-1},\ \ w=e^{-i2\pi/n}\end{array}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.3)$$

تستند هذه الخوارزمية على تقسيم سلسلة الأرقـــام $f(t)$ إلى جزأين بطول $n/2$ وهو أسلوب سهل التنفيذ ويختصـــر كثير من العمليات الحسابية ويمكن أيضاً اختصار حســابات تحويل فورير إلى النصف بسبب تناظر حدود النصف الأول والأخير من تحويل فورير، أي أن:

$$F(k)=a_k+ib_k \Rightarrow F(k+\frac{n}{2})=a_k-ib_k,\ k=0,\ldots,\frac{n}{2}-1 \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.4)$$

لذلك يقتصر الحساب للنصف الأول من السلسلة فقط.

2-2 الدالة المميزة

يتم حساب الدالة المميزة بطرائق متعددة منها اللامعلمية مثل طريقة فشر ومنها المعلمية مثل طريقة الإمكان الأعظم وطريقة قرار بيز. وتوفر كذلك أساليب أخـــرى للتصـــنيف مثل طريقة التحليل العنقودي الاتجاه الأقرب وطريقة النواه، الخ.

في هذا البحث تم استخدام إحدى الطرق المعلميـــة فـــي التميز وهي طريقة قرار بيز.

2-2-1 اشتقاق الدالة المميزة

فـــي حالـــة وجـــود مجمـــوعتين مـــن مجتمعتـــين توزيعهما $f_2(x), f_1(x)$ لهما معالم تقدير من خلال عينات تدريبية (*Training Sample*)، ولدينا كذلك مشاهدات غير معروفة الانتماء، لذلك الفرضية الخاصة باختبار انتماء كل مشاهدة من هذه المشاهدات تكون [9] [11].

$$H_0: \underline{X} \in f_2(\underline{x})$$
$$H_1: \underline{X} \in f_1(\underline{x})$$

والحل بطريقة قرار بيز بتقليل مخاطرة بيز باســتعمال نسبة الإمكان والاحتمالات الأولية P_1 وP_2 وكلف التصنيف الخاطئ C_{12} وC_{21} كما يأتي:

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} < \frac{C_{12}P_2}{C_{21}P_1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.5)$$

وفي حالة كون f_1 وf_2 هما توزيعان طبيعيان يكون بعد أخذ لوغاريتم الطرفين.

$$\mathrm{Ln}\frac{\left|\sum_1\right|}{\left|\sum_2\right|}+(\underline{x_{1i}}-\underline{\mu_1})'\textstyle\sum_1^{-1}(\underline{x_{1i}}-\underline{u_1})-$$
$$(\underline{x_{2i}}-\underline{u_2})'\textstyle\sum_2^{-1}(\underline{x_{2i}}-\underline{u_2})<-2\mathrm{Ln}\frac{C_{12}P_2}{C_{21}P_1} \quad \ldots (2.6)$$

وبعد التبسيط تكون الدالة المميزة كما يأتي:

$$h(x)=\underline{x}'A\underline{x}b'x+C\langle\ \rangle T(1) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.7)$$
$$A=\textstyle\sum_1^{-1}-\sum_2^{-1}$$
$$b=2\left(\textstyle\sum_2^{-1}\underline{\mu_2}-\sum_1^{-1}\underline{\mu_1}\right)$$
$$C=\left(\underline{\mu_1'}\textstyle\sum^{-1}\underline{\mu_1}-\underline{\mu_2'}\sum_1^{-1}\underline{\mu_2}+\mathrm{Ln}\frac{\left|\sum_1\right|}{\left|\sum_2\right|}\right)$$
$$T=-2\mathrm{Ln}\frac{C_{12}P_2}{C_{21}P_1}$$

حيث تصف المشاهدات x على أســاس أنهــا تنتمــي للمجتمع f_1 إذا كانـــت $h(\underline{x})\langle' T$ وتنتمـــي إلـــى f_2 فـــي حالة $h(x)\rangle T$ وتعرف بالدالة المميزة التربيعية.

أمـــا عنـــد تســـاوي مصـــفوفات التبـــاين والتبـــاين المشترك $\sum_i$ تصبح الدالة بالشكل الآتي:

$$h^*(\underline{x})=b'^*\underline{x}+c^*\langle\ \rangle T \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.8)$$
$$\underline{b}^*=2\textstyle\sum(\underline{u_2}-\underline{u_1}), C^*=\underset{-1}{\underline{u'}}\sum^{-1}\underline{u_1}-\underline{u_2}\sum^{-1}\underline{u_2}$$

وهي الدالة المميزة الخطية.

ويلاحظ بالنسبة للمجموعتين أنه إذا كانـــت كلـــف التصنيف الخاطئ تساوي واحد دائماً، فإن مخـــاطرة بيـــز تتحول إلى احتمال خطأ التصنيف، لذلك فإن الحل الذي يقلل مخاطرة بيز يقلل أيضاً احتمال خطأ التصنيف والحل يكون:

$$p_1f_1(\underline{x})<p_2f_2(\underline{x}) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.9)$$

---

(1) يعني الرمز $\langle\rangle$ أن تصنيف المشاهدة x على أنها تنتمي للمجتمع f_1 المميزة أقل من T وتنتمي إلى المجتمع f_2 إذا كانت قيمتها أكبر من T.

أما في حالة المجموعات العديدة فالمشكلة تتسع إلى ما يسمى اختبار الفرضيات المتعددة (*Multiple Hypothesis Test*) والفرضية تكون:

$$H_1 : \underline{x} \in w_i,\ \ i = 1,...,g$$

والحل في حالة الكلف تساوي واحد يكون باختبار المجموعة التي تحقق ما يأتي كمجموعة تنتمي إليها المشاهدات $\underline{x}$ [11].

$$p_k f_k(\underline{x}) > p_i f_i(\underline{x}), \text{for } i \neq k, i = 1,...,g \ ...... (2.10)$$

وبعبارة أخرى فإن لدينا (g) من الدوال التميزية $h_i(\underline{x})$ والحل هو اختبار الدالة التي تعطي أعظم قيمة، كما يأتي:

$$h_k(\underline{x}) = \max h_i(\underline{x}), h_i(\underline{x}) = p_i f_i(\underline{x}) \ .... (2.11)$$

وهناك نظرية بهذا الخصوص تنص على أن أي دالة متزايدة رتيبة $h_i(\underline{x})$ يمكن أن تحل بدلاً من $h_i(\underline{x})$

وفي حالة توزيع البيانات طبيعياً يكون [9]:

$$h_i(\underline{x}) = P_i N(\underline{x_i}\underline{u_1}, \textstyle\sum_i) \ .... (2.12)$$

حيث $N(\underline{x_i}\underline{u_i}, \sum_i)$ هي دالة التوزيع الطبيعي المتعدد، ولكن للسهولة وحسب النظرية السابقة تبسط هذه الدوال لتصبح بالشكل الآتي:

$$h_i(\underline{x}) = -(\underline{x} - \underline{u_i})' \textstyle\sum_i^{-1} (\underline{x} - \underline{u_i}) - \mathrm{Ln}\left|\sum_i\right| + 2\mathrm{Ln}P_i, i = 1,...,g \ .... (2.13)$$

وهي الدالة المميزة التربيعية المتعددة (*Multiple Quadratic Discriminant Function*) وفي حالة تساوي مصفوفات التباين والتباين المشترك تصبح الدوال بالشكل الآتي:

$$h_i^*(\underline{x}) = 2u_i^{\sim} \textstyle\sum^{-1} \underline{x} - \underline{u_i} \sum^{-1} \underline{u_i} + 2\mathrm{Ln}P_i, i = 1,...,g \ .... (2.4)$$

وهي الدالة المميزة الخطية.

3-2 تقدير معالم النموذج

يتم تقدير $\sum_i, \underline{U_i}$ $i = 1,...,g$ بطريقة الإمكان الأعظم لوجود خاصية الثبات (*Invariance*) في هذه الطريقة، حيث:

$$\hat{\mu}_i = \underline{\bar{x}_i} = \frac{1}{n_i} \sum_{j=i}^{n_i} \underline{x}_{ij} \ .... (2.15)$$

$$\hat{\textstyle\sum}_i = S_i^2 = \frac{1}{n_i} \sum_{j=i}^{n!} (\underline{x}_{ij} - \underline{\bar{x}_j})(\underline{x}_{ij} - \underline{\bar{x}_i})^2$$

ويمكن استبدال تقدير $\sum_i$ المتحيز بتقدير $\sum_i^{-1}$ غير المتحيز المشتق من توزيع (*Wisbert*) وهو [3]

$$\hat{\textstyle\sum}_i^{-1} = \left(\frac{n_i - P - 3}{n_i}\right) S_i^{-1} \ .... (2.16)$$

حيث S_i هو تقدير الإمكان الأعظم لـ $\sum_i^{-1}$، وكذلك بدلاً من $\sum_i$ يمكن استخدام التقدير غير المتميز لمسافة مهلنوبس [3] [6].

$$\hat{D}_{ij}^2 = \frac{n_i + n_j - p - 3}{n_i + n_j - p - 2} (\underline{\bar{x}}_j - \underline{\bar{x}}_i)' S^{-1} (\underline{\bar{x}}_j - \underline{\bar{x}}_i) - \frac{p(n_i - n_j)}{n_i n_j} \ .... (2.17)$$

وإذا كانت $\hat{D}_{ij}^2$ سالبة فتؤخذ صفراً.

أما بالنسبة لتقدير الاحتمالات الأولية فتقدر هنا بالطريقة المعروفة:

$$\hat{P}_i = P_i = \frac{n_i}{n} \ .... (2.18)$$

وذلك لكبر حجم العينة حيث أن P_i هو

$$P_i = \underset{n \to \infty}{\mathrm{Lin}} = \frac{n_i}{n} \ .... (2.19)$$

حيث n الحجم الكلي للعينة و n_i حجم عينة المجموعة i

ويمكن معرفة الاحتمالات أيضاً من معرفة طبيعة المشكلة، فهي هنا احتمالات متساوية وتساوي $1/g$ لذلك أن احتمال ظهور أي رقم متساوي أي أن ظهور الأرقام يتوزع توزيعاً منتظماً متقطعاً.

4-2 اختيار المتغيرات (استخلاص الصفات)

تتمثل المرحلة باختيار المتغيرات ذات التأثير المعنوي في عملية التميز، وذلك لإنجاز مهمة التميز بسهولة وسرعة ودقة، وكذلك إزالة المعلومات المكررة غير المؤثرة في عملية التميز والتي تؤدي إلى تقليل الدقة. فإذا كان لدينا (Z) من المتغيرات. المطلوب اختبار مجموعة جزئية من هذه المتغيرات التي تحقق أعلى تميز أي أقل احتمال خطأ تصنيف، وهناك عدد من الأساليب وقد تم استخدام أسلوبين

في البحث وأهملت الأساليب الأخرى، لكون نتائجها مطابقة إلى نتائج الأسلوبين المذكورين في أدناه.

2-4-1 أسلوب الاختيار الأمامي المتسلسل Sequential Forward Selection (SFS)

هذا الأسلوب يتم اختيار أفضل متغير وهو الذي يعظم معيار الاختيار بعدها يتم أخذ المتغير الثاني الذي يعطي مع المتغير الأول أعلى قيمة للمعيار وتستمر هذه الخطوات إلى أن تصل إلى العدد المطلوب.

ويأخذ هذا الأسلوب عدم الاستقلالية الموجودة بين المتغيرات بنظر الاعتبار، لكنه مع ذلك يحتوي بعض المساوئ منها أنه يمكن جذب المتغير الذي يضاف مستقبلاً والذي يؤدي إلى تقليل من معنوية ذلك المتغير بحيث يصبح ذو تأثير مهم، وهذا ما يسمى بظاهرة "**التكدس**"*Nesting*، أي تكون المجاميع الجزئية لمتداخلة للمتغيرات ذات تأثير متناقص [4] P.217.

2-4-2 أسلوب الاختيار التدرجي Stepwise Selection

وهو من أشهر أساليب اختبار المتغيرات، لأنه يعتمد في عملية التوقف على الاختبارات الإحصائية الخاصة بذلك، وكما يأتي:

أ- يؤخذ المتغير الأكثر معنوية إلى النموذج، وذلك بعد أن يختبر فيما إذا كان غير معنوياً، فإذا كان كذلك يتم التوقف.

ب- تختبر جميع المتغيرات في النموذج فإذا كان منها متغيرات غير معنوية نحذف من النموذج.

ج- عاد الخطوات أ و ب إلى أن يتم التوقف ويلاحظ في الخطوة أ أن اختيار المتغير الأكثر معنوية يكون آنياً وليس متفرداً.

2-5 معايير اختبار المتغيرات

إن أفضل معيار يستخدم لوصف معنوية متغيرات التصنيف هو احتمال خطأ التصنيف، ولكن لسوء الحظ صعوبة تقديره تحول دون استخدامه عملياً من قبل ولكن تم استخدامه في هذا البحث.وقد تم ابتكار العديد من المعايير البديلة من قبل الاحصائين و المعايير الخاصة باختيار المتغيرات المستخدمة في هذا البحث هي:

2-5-1 إحصاءه $\wedge$ Wilk's

وتدعى أحياناً بإحصائه (*U-Statistic*) وهي نسبة مجموع مربعات داخل المجموعات إلى المجموع الكلي للمربعات [7].

حيث:

W مصفوفة مجموع المربعات بين المجموعات

$$\wedge = \frac{|w|}{|T|} \quad \ldots\ldots (2.20)$$

T مصفوفة مجموع المربعات الكلية

$$T = \sum_{i=1}^{g}\sum_{j=1}^{ni}(x_{ij}-\bar{x})(x_{ij}-\bar{x})'$$

وأن قيمة ^ تتراوح بين 0 و 1 وكلما قلت ^ دل ذلك على زيادة انفصال المجموعات.

$$w = \sum_{i=1}^{g}\sum_{j=1}^{i}(x_{ij}-\bar{x})(x_{ij}-\bar{x})'$$

2-5-2 إحصاءه Rao's V

وتعرف كذلك بـ *Lowley-Hotelling Tree* وتعرف كما يلي [3] [8] :

$$V = (n-g)\sum_{i=1}^{p}\sum_{j=1}^{p}a_{ij}\sum_{k=1}^{g}n_k(\bar{x}_{ik}-\bar{x}_i)(\bar{x}_{jk}-\bar{x}_j) \quad \ldots\ldots (2.21)$$

حيث:

n_k عدد المشاهدات في المجموعة k

a_{ij} هو العنصر ij في معكوس مصفوفة مجموع مربعات الانحرافات بين المجموعات.

أن V يتوزع تقريباً حسب مربع كاي بدرجة حرية $p(g-1)$

2-5-3 مسافة مهلنوبس Mahalanobis Distance (D^2)

وهو قياس للمسافة بين مجموعتين ويعرف كما يلي:

$$D^2_{ij} = (\bar{x}_i-\bar{x}_j)'S^{-1}(\bar{x}_i-\bar{x}_j) \quad \ldots\ldots (2.22)$$

وكلما زادت مسافة مهلنوبس يعني ذلك زيادة في انفصالية المجموعتين وأن قيمة مهلنوبس تتناسب طردياً مع قيمة معامل التحديد في حالة تحليل البيانات بواسطة الانحدار المتعدد.

2-5-4 إحصاءه F لمسافة مهلنوبس (FD2)

يمكن اختبار الفرضية القائلة بتساوي وسطي مجتمعين بواسطة الإحصائه التالية:

$$F_0 = \frac{(n_1+n_2-p-1)n_1n_2}{p(n_1+n_2-2)(n_1+n_2)} D_{ij}^2 \sim F(p, n_1+n_2-p-1)$$
.... (2.23)

وتستخدم هنا لاختبار المتغيرات التميزية، فكلما زادت F_0 كلما دل ذلك على زيادة انفصالية المجاميع وكان ذلك المتغير مميزاً أفضل، ويلاحظ هنا أن نتائج استخدام هذا المعيار قد تختلف عن سابقتها التي تستخدم مسافة مهلنوبس.

وفي حالة استخدام أسلوب الاختبار المتدرج *Stepwise* يتم اختبار تأثير المتغير المضاف فقط عن طريق الإحصاءه التالية [7].

$$F_0 = \frac{(n_1+n_2-p-1)n_1n_2(D_{p+1}^2 - D_p^2)}{(n_1+n_2)(n_1+n_2-2)+n_1n_2D_p^2} \sim F(1, n_1+n_2-p-2$$
.... (2.24)

حيث D_p^2 و D_{p+1}^2 هما مسافة مهلنوبس لـ p و p+1 من المتغيرات على التوالي.

2-5-5 معيار التباعد Divergence Criterion (JD)

يستخدم لقياس التباعد بين مجموعتين [4][9].

وفي حالة توزيع البيانات طبعاً يكون:

$$JD = \frac{1}{2}(\underline{u}_2 - \underline{u}_1)'(\Sigma_2^{-1} + (\underline{u}_2 - \underline{u}_1) +$$
$$\frac{1}{2}\mathrm{tr}\left\{\Sigma_1^{-1}\Sigma_2 + \Sigma_2^{-1}\Sigma_1 - 2I\right\} \quad \text{.........} (2.25)$$

2-5-6 معيار JB Battacharyya

يعرف كما يلي [4][9].

$$JB = \frac{1}{8}(\underline{u}_2 - \underline{u}_1)'(\frac{1}{2}\Sigma_1 + \Sigma_2)]^{-1}(\underline{u}_2 - \underline{u}_1) +$$
$$\frac{1}{2}\mathrm{Ln}\frac{|\frac{1}{2}(\Sigma_1 + \Sigma_2)|}{[|\Sigma_1||\Sigma_2|]^{\frac{1}{2}}} \quad \text{......................} (2.26)$$

نلاحظ في حالة تساوي مصفوفات التباين والتباين المشترك أن:

$$JD = 8JB = D^2$$

أي أن هذه المعايير تعطي نفس النتيجة عند التحليل خطياً.

2-5-7 معيار Matusita (JT)

ويعرف كما يأتي: [2][4].

$$J_T = \left\{\int_{\forall \underline{x}} \left[\sqrt{f(\underline{x}|w_1)} - \sqrt{f(\underline{x}|w_2)}\right]^2 d\underline{x}\right\}^{\frac{1}{2}}$$
.... (2.27)

حيث تزداد قيمته بزيادة الانفصالية وفي حالة توزع البيانات طبيعياً يكون:

$$J_T = \sqrt{2[1-\exp(J_B)]} \quad \text{..............................} (2.28)$$

حيث J_B هو معيار باتاجاريا *Battacharyya*

2-5-8 معيار Patric Fisher

ويعرف كما يأتي [4]:

$$J_p = \left\{\int_{\forall \underline{x}} [f(\underline{x}|w_1)p_1 - f(\underline{x}|w_2)p_2]^2 \, d\underline{x}\right\}^{\frac{1}{2}}$$
.... (2.29)

إذ يأخذ بالحسبان الاحتمالات الأولية لظهور المجموعات p_1 وp_2 ويزداد هذا المعيار بزيادة الانفصالية، وفي حالة توزيع البيانات طبيعياً يكون:

$$J_p = \sqrt{p_1^2\left[(2\pi)^p|2\Sigma_1|\right]^{-\frac{1}{2}} + p_2^2\left[(2\pi)^p|2\Sigma_2|\right]^{-\frac{1}{2}}} \Big/$$
$$-2p_1p_2\left[(2\pi)^p|\Sigma_1+\Sigma_2|\right]^{\frac{1}{2}}\exp\left\{-\frac{1}{2}(u_2-u_1)'(\Sigma_1+\Sigma_2)^{-1}(\mu_2-\mu_1)\right\}$$
.... (2.30)

2-5-9 معيار احتمال خطأ التصنيف

احتمال خطأ التصنيف هو معيار مثالي لقياس أداء الدالة المميزة، وصعوبة تقديره تحول دون استخدامه عملياً [4] بالرغم من ذلك فإن التطور الحاصل في مجال الحاسبات أزال هذه العقبة وتم استعماله هنا من قبل الباحث وذلك عن طريق تقديره بطريقة إعادة التعويض يذكر أن احتمال خطأ التصنيف في حالة المجموعات العديدة هو من قبل الباحث

$$\mathrm{error} = \sum_{i=1}^{g}\int\sum_{i\neq k}^{g} p_i f_j(\underline{x})dx$$
$$= \int_{Rx} \min\left[\sum_{i\neq j}^{g} p_i fj(\underline{x})\right] d\underline{x} \quad \text{...............} (2.31)$$

حيث:

R_k مجال المجموعة k.

$R_{\underline{x}}$ مجال $\underline{x}$.
والمعيار لا يمتلك صفة الرتابة

أ- طريقة إعادة التعويض (Resubstitution)

يتم تقدير احتمال خطأ التصنيف بعدة طرائق منها (طريقة إعادة التعويض) ويتم حسابها بنسبة عدد المشاهدات التي صنفت

خطأ إلى العدد الكلي، وكما يأتي:

$$error = \frac{n_e}{n} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.32)$$

n_e عدد المشاهدات التي صنفت بشكلٍ خاطئ.

n العدد الكلي للمشاهدات

ويمكن حساب الأخطاء التفصيلية كما يأتي:

$$e_{ij} = \frac{n_{ij}}{n_i} \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.33)$$

حيث:

n_{ij} عدد المشاهدات في المجموعة i التي صنفت على أساس أنها المجموعة j.

n_i عدد المشاهدات للمجموعة i.

ب- طريقة (Rotation)

وإن هذه الطريقة لها تميز عالي وكذلك تباين عالي أيضاً [8].

3- يتم تقسيم العينة إلى $\frac{n}{v}$ من العينات حيث n تقبل القسمة على v و v عدد صحيح موجب. بعدها يتم حذف العينة الأولى وحساب الدالة المميزة للعينات الباقية باعتبارها عينة واحدة وتقدير احتمال خطأ التصنيف باستخدام العينة المحذوفة بطريقة إعادة التعويض ثم تكرر هذه العملية أكثر من العينة الثانية بدلاً من الأولى هكذا في كل العينات الجزئية بعدها يتم أخذ المعدل للتقديرات المحسوبة كتقدير لاحتمال خطأ التصنيف.

3- الجانب التطبيقي

في هذا الجزء من البحث ستطبق الدالة التربيعية لتمييز الأرقام العربية وذلك عن طريق بناء منظومة حاسوبية. تقوم بتحويل صور الأنماط إلى الحاسبة أولاً ومعالجتها بالتخلص من الضوضاء ثم تحديد إطار كل نمط وأخذ تحويل فورير وذلك بعد تحويل الإطار إلى أرقام كمية تمثل إحداثيات نقاط الإطار.

وطبقت بعد ذلك أساليب اختزال المتغيرات لاختبار المتغيرات التمييزية فقط، وذلك يعد اختبار البيانات الناتجة عن المرحلة السابقة.

ولتقييم الدالة المميزة الناتجة بكل أسلوب من أساليب اختزال المتغيرات ثم تقدير احتمال خطأ التصنيف بعدة طرق كذلك حساب وقت التنفيذ ووقت تمييز النمط الواحد والملحق رقم (1) يوضح هذه الخطوات بالتفصيل. ثم كتابة البرامج الحاسوبية للقيام بالمهام السابقة من قبل الباحث وذلك باستخدام لغة (C++) لما تتمتع به من مواصفات فريدة منها السرعة العالية وإمكانية البرمجة بطريقة الكيانات Objected Oriented Programming بالإضافة إلى المرونة التي تتمتع بها هذه اللغة وإمكانية الوصول إلى المستويات الواطئة من برامجيات الحاسوب.

3-1 عينة البحث

شملت البيانات في هذا البحث مجموعة من الأرقام العربية المكتوبة باليد من قبل عدد من الأشخاص الذين تم اختيارهم عشوائياً من مختلف المستويات الدراسية والثقافية حيث قام كل شخص بكتابة الأرقام من صفر إلى تسعة بسطر واحد. وأخذ بالحسبان عدم تقييد الشخص بالكتابة بخط واضح وإنما يكتب الأرقام المطلوبة بخطه العادي دون تركيز. كان عدد الأشخاص الذين تم أخذ خط يدهم كعينة (125) شخص إلى أن عدد الأنماط المأخوذة هو (1250) تم بعدها تحويل هذه الأنماط إلى جهاز الحاسوب عن طريق سحبها على شكل صورة بواسطة جهاز الماسح البصري Scanner حيث يقوم بأخذ كل نقطة من نقاط الصورة المأخوذة (صورة الأرقام) ويحولها إلى الرقم الذي يمثل لونها ويمكنه كذلك سحب الصورة على أساس أنها بالأبيض والأسود أو ما يسمى بالتدرج الرصاصي Gray Level.

وكذلك يمكن أن تكون الصورة الناتجة ثنائية اللون Binary Image ثم خزنت الصور على شكل ملف من نوع BMP ومثلت البيانات الخام والتي تتضمن جميع التفاصيل وكبرت عدة مرات لنقل جميع تفاصيل الصورة بدقة. وبعدما تم حذف بعض الأنماط المشوهة أثناء عملية

سحب الصورة ونتج عن ذلك (1129) نمط جيد في ملف الصورة.

3-2 المعالجة الأولية

تمت المعالجة الأولية بإزالة الضوضاء أي النقاط السوداء المعزولة والحواف السوداء. وذلك بواسطة برنامج يقوم بإزالة الضوضاء عن طريق حساب حجم نقطة الضوضاء بحيث تحذف إذا كانت أكبر من (Pixels 4).

3-3 تقطيع الصورة وتحديد إطار الأنماط

بعد أن تهيأت لنا صورة نقية لهذه المرحلة التي تتكون من تقطيع الصورة وتحديد الإطار.

3-4 إدخال متغيرات الاستجابة

تتطلب الدالة المميزة وجود عينة من المتغيرات معروفة الاستجابة أي أن كل مشاهدة منها تنتمي إلى مجموعة معروفة وإن هذا الانتماء يرمز بشكل متغير متقطع يمثل كل رقم فيه رقم المجموعة التي تنتمي إليها المشاهدة الخاصة بذلك الرقم ويدعى هذا المتغير بمتغير الاستجابة وتم في هذه المرحلة إدخال تصنيف أو استجابة كل شخص في ملف يسمى ملف الاستجابة.

3-5 التطبيع وإزالة نقطة البداية

إن تعدد أساليب الكتابة اليدوية من حيث ميلان وانحناء الكتابة وحجمها إضافة إلى عوامل أخرى. تؤدي إلى اختلاف النقطة التي تبدأ منها عملية تتبع مسار إطار النمط في مرحلة اخذ إطار النمط مما يؤدي إلى إزاحة سلسلة أرقام إطار النمط بمقدار هو t_s وهذا يؤدي إلى زيادة الاختلافات بين الأنماط من نفس المجموعة ومن ثم زيادة المتغيرات اللازمة لتصنيفها وكذلك زيادة احتمال تصنيفها الخاطئ. لذلك جاءت هذه المرحلة لمعالجة هذه المشكلة. وتبدأ أولاً بتوحيد عدد الأرقام التي يتكون منها إطار النمط ويتم ذلك باستخدام خوارزمية التطبع (Normalization) حيث تقوم أولاً برسم إطار النمط ومن ثم تقطيعه من جديد إلى عدد ثابت هو (128) خط نأخذ رقم اتجاهها وتخزن في إطار الملف الجديد. وجاء استخدام الرقم (128) لأسباب عديدة لا مجال لذكرها الآن.

أما الخطوة الثانية في هذه المرحلة تتجسد في اختيار أحد الأنماط كنمط قياسي وذلك لكل مجموعة على حدة ثم تحسب t_s التي تعطي أعلى r كما في المعادلة [9]

$$r = \max_{t_s} \sum_{i=1}^{128} d_i^* d_i + t_s \quad \ldots\ldots (3.1)$$

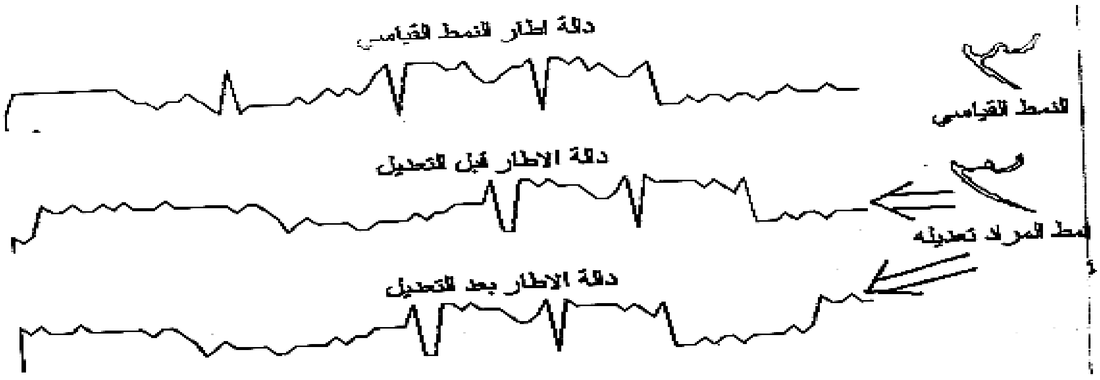
حيث:

d_i^* أرقام قطع النمط القياسي

$d_i + t_s$ أرقام قطع النمط المطلوب معالجته بعد إزاحته بمقدار t_s وباعتبار di كسلسلة زمنية فإن t_s هو مقدار الإزاحة الذي يعطي أعلى ضرب متقاطع cross-product ويتم تعديل النمط بواسطة إزاحته عكسياً بمقدار t_s كما يأتي:

$$d_i = \begin{cases} d_i + t_s, & \text{if } i + t_s \le 128 \\ d_i + t_s - 128, & \text{if } i + t_s > 128 \end{cases} \quad \ldots\ldots (3.2)$$

وتخزن النتائج في ملف الإطار الجديد شكل (10) يوضح هذه المرحلة حيث رسمت القطع الخاصة بالنمط القياسي والنمط المعالج على شكل دالة تعرف بدالة الإطار Contour Function إذ يمثل المحور السيني فيها برقم القطعة والصادي برقم اتجاه القطعة.



شكل (10).

3-6 أخذ القياسات

إن أحد الفروض للتحليل التمييزي هو توزيع المتغيرات توزيعاً طبيعياً متعدداً، ولتحقيق هذا الفرض وجميع الفروض الأخرى الخاصة بالتحليل التمييزي. استعملنا لهذا الغرض تحويل فورير السريع للحصول على القياسات المطلوبة، ولكن بدلاً من أخذ التحويل للأرقام الخاصة بإطار النمط أي دالة الإطار استخدمنا قيم الإحداثيات الناتجة عنها ويتم ذلك برسم إطار النمط تدريجياً قطعة قطعة حيث تؤخذ إحداثيات نقطة البداية لكل قطعة ذلك لأن النقطة هي نفسها للقطعة هي نفسها لضغطة البداية للقطعة التي تليها.

وينتج من هذه العملية مجموعتان من الأرقام الأولى تمثل الإحداثي السيني لنقطة بداية القطع وهي (128) والأخرى تمثل الإحداثي الصادي وهي (128) رقم أيضاً. ويكون مجموعها (256) رقم.

إن سبب هذا الإجراء يكمن في تحويل البيانات النوعية المتمثلة بأرقام اتجاهات القطع إلى بيانات كمية يمكن أن يؤخذ بها تحويل فورير كذلك فإن خطأ التصنيف يكون أقل عند أخذ هذا الإجراء إضافة إلى إمكانية الحصول على متغيرات تمييزية أفضل.

3-7 اختبار المتغيرات التمييزية

إن مرحلة اختيار الصفات تتلخص باختيار مجموعة المتغيرات المعنوية (التمييزية) والتي تؤخذ بحيث تجعل احتمال خطأ التصنيف أقل ما يمكن. واستخدم في هذه المرحلة أسلوب SFS وStepwise أما باقي الأساليب استبعدت أما لردائتها أو بسبب عدم إمكانية تنفيذها مثل أسلوب SBS.

أما بالنسبة لمعايير اختيار المتغيرات فقد استخدمت جميع المعايير المذكورة في الجانب النظري. والجدول من (1) إلى (13) يعرض النتائج حسب خطأ التصنيف بطريقة إعادة التعويض لكل خطوة لبيان تأثير إدخال المتغيرات الجديدة عليه.

جدول (1)

استخدام معيار باتاجاريا JB لاختيار المتغيرات بأسلوب SFS.

رقم الخطوة	رقم المتغير	معيار باتاجاريا JB	خطأ التصنيف
1	3	2.277202	0.528787
2	132	3.660289	0.345538
3	4	4.899622	0.216121
4	130	6.36998	0.115146
5	5	7.625295	0.058459
6	133	8.662451	0.039858
7	8	9.576944	0.030115
8	7	10.54771	0.026572
9	131	11.46625	0.024801
10	135	12.39233	0.020372
11	134	13.26674	0.017715
12	1	14.12649	0.012400
13	143	14.93798	0.007972
14	140	15.72945	0.007972
15	139	16.53152	0.007972
16	136	17.29593	0.005314
17	6	18.03422	0.007086
18	10	18.78597	0.003543
19	11	19.41228	0.002657
20	137	20.03324	0.005314
21	138	20.71119	0.003543
22	141	21.48489	0.002657
23	9	22.13233	0.001772
24	13	22.77441	0.001772
25	17	23.36195	0.001772
26	12	23.95972	0.001772
27	14	24.61053	0.001772
28	15	25.23784	0.001772
29	16	25.80959	0.001772
30	19	26.30163	0.001772

جدول (2)

استخدام معيار ماتوسيتا JT لاختيار المتغيرات بأسلوب SFS.

رقم الخطوة	رقم المتغير	معيار ماتوسيتا JT	خطأ التصنيف
1	3	0.99459	0.528787
2	132	1.220627	0.345439
3	5	1.317149	0.186005
4	4	1.360937	0.131975
5	130	1.388724	0.058459
6	1	1.399136	0.031001
7	8	1.405097	0.026572
8	2	1.408223	0.023915
9	133	1.410013	0.018601
10	7	1.411284	0.017715
11	6	1.412201	0.011515
12	131	1.412794	0.012400
13	238	1.413172	0.006200
14	10	1.413449	0.006200
15	134	1.413657	0.003543
16	43	1.413796	0.003543
17	13	1.413910	0.002657
18	11	1.413992	0.003543
19	27	1.414055	0.002657
20	59	1.414102	0.001772
21	21	1.414135	0.001772
22	46	1.414158	0.001772
23	83	1.414174	0.000886
24	136	1.414185	0.000000

جدول (3)

استخدام معيار التباعد JD لاختيار.

رقم الخطوة	رقم المتغير	معيار التباعد JD	خطأ التصنيف
1	3	20.80572	0.528787
2	132	38.86807	0.345439
3	130	63.77779	0.223206
4	133	122.1737	0.140833
5	141	182.5742	0.118689
6	134	306.1321	0.109832
7	137	446.7687	0.093888
8	138	703.8744	0.085917
9	136	934.8292	0.079717
10	140	2242.342	0.067316
11	144	4910.781	0.065545
12	152	7003.883	0.063773
13	156	14537.37	0.057573
14	148	34662.39	0.046944
15	160	50056.27	0.041630
16	146	97415.92	0.040744
17	164	111594.8	0.038087
18	203	131426.5	0.032772
19	142	178847.3	0.031001
20	149	310586.1	0.032772

جدول (4)

استخدام معيار باترك فيشر JP لاختيار المتغيرات بأسلوب SFS.

رقم الخطوة	رقم المتغير	معيار باترك فيشر JP	خطأ التصنيف
1	253	0.001849	0.644818
2	255	0.001161	0.594331
3	251	0.000856	0.587245
4	247	0.000669	0.553587
5	243	0.000740	0.521701
6	239	0.000784	0.502214
7	249	0.000661	0.460585
8	245	0.000947	0.417183
9	235	0.000904	0.378211
10	231	0.001177	0.368468
11	227	0.001719	0.352524
12	223	0.002323	0.325066
13	241	0.003075	0.308237
14	237	0.005029	0.284322
15	233	0.010150	0.266608
16	229	0.037075	0.255093
17	225	0.229449	0.235607
18	221	0.541012	0.224978
19	250	0.262213	0.199291
20	246	0.616365	0.180691

جدول (5)

استخدام مسافة مهلنوبس D^2 لاختيار المتغيرات بأسلوب SFS.

رقم الخطوة	رقم المتغير	معيار مسافة مهلنوبس D^2	خطأ التصنيف
1	3	13.30035	0.528787
2	132	20.88645	0.345439
3	4	28.03035	0.216121
4	5	35.22627	0.131975
5	130	42.43944	0.058459
6	8	47.27641	0.040744
7	7	50.78545	0.035430
8	1	53.44953	0.021258
9	134	56.08984	0.019486
10	133	58.03013	0.017715
11	135	60.22570	0.014172
12	12	61.81015	0.010629
13	10	63.62260	0.009743
14	13	64.93796	0.007972
15	136	66.11231	0.005314
16	11	67.19295	0.005314
17	6	68.47606	0.004429
18	52	69.55452	0.003543
19	236	71.07785	0.002657
20	140	71.92889	0.003543
21	24	72.70078	0.002657
22	2	73.55698	0.001772
23	15	74.43018	0.000886
24	137	75.23367	0.000886
25	139	75.92458	0.000886
26	16	76.57693	0.000000

جدول (6)

استخدام مجموع التباين غير المشروح D^2r لاختيار المتغيرات بأسلوب SFS.

رقم الخطوة	رقم المتغير	معيار التباين غير المشروح D^2r	خطأ التصنيف
1	3	13.50857	0.528787
2	5	4.020176	0.313552
3	130	1.773403	0.158547
4	1	0.863325	0.079717
5	132	0.505016	0.036315
6	131	0.397322	0.035430
7	8	0.321979	0.023915
8	6	0.265765	0.021258
9	4	0.219467	0.017715
10	13	0.193349	0.015058
11	133	0.170750	0.015058
12	10	0.151483	0.011515
13	7	0.135368	0.010629
14	12	0.125437	0.011515
15	134	0.116253	0.008857
16	11	0.109752	0.008857
17	27	0.103838	0.005314
18	137	0.098848	0.003543
19	25	0.094033	0.001772
20	135	0.090020	0.000000

جدول (7)

استخدام معيار F لمسافة مهلنوبس FD^2 لاختيار المتغيرات بأسلوب SFS.

رقم الخطوة	رقم المتغير	معيار F لمسافة مهلنوبس FD^2	خطأ التصنيف
1	3	760.3151	0.528787
2	132	591.8251	0.345439
3	4	525.5814	0.216121
4	5	494.4899	0.131975
5	130	475.8990	0.58459
6	8	436.8750	0.040744
7	7	400.9141	0.035430
8	1	367.2565	0.021258
9	134	341.4457	0.019486
10	133	316.6206	0.017715
11	135	297.4232	0.014172
12	12	278.6415	0.010629
13	10	263.7025	0.009743
14	13	248.8945	0.007972
15	136	235.4404	0.005314
16	11	223.3331	0.005314
17	6	213.0660	0.004429
18	52	203.4757	0.003543
19	236	196.1195	0.002657
20	140	187.5714	0.003543
21	24	179.6535	0.002657
22	2	172.7274	0.001772
23	15	166.3990	0.000886
24	137	160.3967	0.000886
25	139	154.6097	0.000886
26	16	149.2262	0.000000

جدول (8)

استخدام إحصاءة Rao's V لاختيار المتغيرات بأسلوب SFS.

رقم الخطوة	رقم المتغير	إحصاءة Rao's V	خطأ التصنيف
1	3	6748.791	0.528787
2	132	10598.09	0.345439
3	4	34223.00	0.216121
4	5	17874.32	0.131975
5	130	21561.79	0.058459
6	8	23988.73	0.040744
7	7	25769.27	0.035430
8	1	27121.06	0.021258
9	134	28460.81	0.019486
10	133	29445.32	0.017715
11	135	30559.38	0.014172
12	12	31363.36	0.010629
13	10	32283.04	0.009743
14	13	32950.46	0.007972
15	136	33546.35	0.005314
16	11	34094.68	0.005314
17	6	34745.74	0.004429
18	52	35292.98	0.003543
19	236	36065.91	0.002657
20	140	36497.76	0.003543
21	24	36889.46	0.002657
22	2	37323.89	0.001772
23	15	37766.95	0.000886
24	137	38174.66	0.000886
25	139	38525.25	0.000886
26	16	38856.26	0.000000

جدول (9)

استخدام إحصاءة Wilks' Λ لاختيار المتغيرات بأسلوب SFS.

رقم الخطوة	رقم المتغير	معيار مسافة مهلنوبس D^2	خطأ التصنيف
1	3	0.142012	0.528787
2	132	0.034194	0.345439
3	5	0.009568	0.186005
4	4	0.003242	0.131975
5	7	0.001467	0.120461
6	8	0.000749	0.064659
7	130	0.000384	0.035430
8	134	0.000253	0.032772
9	1	0.000178	0.019486
10	133	0.000127	0.017715
11	135	9.65E-05	0.014172
12	10	7.55E-05	0.012400
13	12	5.83E-05	0,009743
14	13	4.87E-05	0.007972
15	6	0.000041	0.006200
16	129	3.47E-05	0.006200
17	131	2.95E-05	0.003543
18	11	2.54E-05	0.002657
19	57	2.24E-05	0.003543
20	140	1.99E-05	0.002657
21	137	0.000018	0.001772
22	55	1.63E-05	0.001772
23	9	1.49E-05	0.000886
24	20	1.37E-05	0.002657
25	27	1.26E-05	0.001772
26	144	1.16E-05	0.000886
27	21	1.07E-05	0.000000

جدول (10)

استخدام تقدير احتمال خطأ التصنيف بطريقة Resubstitution لاختيار المتغيرات بأسلوب SFS.

رقم الخطوة	رقم المتغير	خطأ التصنيف
1	3	0.528787
2	5	0.313552
3	130	0.158547
4	1	0.079717
5	132	0.036315
6	206	0.025686
7	24	0.019486
8	39	0.014172
9	4	0.010629
10	9	0.008857
11	6	0.006200
12	20	0.003543
13	48	0.000886
14	13	0.000886
15	2	0.000000

جدول (11)

استخدام إحصاءة Wilks' ∧ لاختيار المتغيرات بأسلوب Stepwise.

رقم الخطوة	نوع الخطوة	رقم المتغير	DF_1	DF_2	F	خطأ التصنيف
1	enter	3	9	1119	743.0139	0.528787
2	enter	132	9	1118	387.1244	0.345439
3	enter	5	9	1117	315.5554	0.186005
4	enter	4	9	1116	238.4997	0.131975
5	enter	7	9	1115	147.5833	0.120461
6	enter	8	9	1114	116.3703	0.064659
7	enter	130	9	1113	115.3150	0.035430
8	enter	134	9	1112	62.1760	0.032772
9	enter	1	9	1111	50.7983	0.019486
10	enter	133	9	1110	47.9115	0.017715
11	enter	135	9	1109	37.1902	0.014172
12	enter	10	9	1108	32.9736	0.012400
13	enter	12	9	1107	34.9506	0,009743
14	enter	13	9	1106	22.7567	0.007972
15	enter	6	9	1105	21.7805	0.006200
16	enter	129	9	1104	20.9354	0.006200
17	enter	131	9	1103	20.0902	0.003543
18	enter	11	9	1102	18.5253	0.002657
19	enter	57	9	1101	15.2566	0.003543
20	enter	140	9	1100	13.9008	0.002657
21	enter	137	9	1099	12.3722	0.001772
22	enter	55	9	1098	11.0117	0.001772
23	enter	9	9	1097	10.2758	0.000886
24	enter	20	9	1096	9.7604	0.002657
25	enter	27	9	1095	9.4273	0.001772
26	enter	144	9	1094	9.4866	0.000886
27	enter	21	9	1093	8.3884	0.000000
28	enter	143	9	1092	8.1979	0.000000
29	enter	16	9	1091	8.0395	0.000000
30	enter	14	9	1090	6.9047	0.000000

جدول (12)

استخدام مسافة مهلنوبس لاختيار المتغيرات بأسلوب Stepwise.

رقم الخطوة	نوع الخطوة	رقم المتغير	DF_1	DF_2	F	خطأ التصنيف
1	enter	3	1	223	746.5153	0.528787
2	enter	132	1	222	400.0998	0.345439
3	enter	4	1	221	363.4503	0.216121
4	enter	5	1	220	354.1005	0.131975
5	enter	130	1	219	346.0904	0.058459
6	enter	8	1	218	220.5540	0.040744
7	enter	7	1	217	158.2258	0.035430
8	enter	1	1	216	118.0433	0.021258
9	enter	134	1	215	115.3289	0.019486
10	enter	133	1	214	83.5619	0.017715
11	enter	135	1	213	93.4659	0.014172
12	enter	12	1	212	66.6140	0.010629
13	enter	10	1	211	75.4195	0.009743
14	enter	13	1	210	54.1315	0.007972
15	enter	136	1	209	47.8791	0.005314
16	enter	11	1	208	43.6695	0.005314
17	enter	6	1	207	51.4106	0.004429
18	enter	52	1	206	42.8131	0.003543
19	enter	236	1	205	59.9587	0.002657
20	enter	140	1	204	33.1612	0.003543
21	enter	24	1	203	29.8435	0.002657
22	enter	2	1	202	32.8545	0.001772
23	enter	15	1	201	33.2445	0.000886
24	enter	137	1	200	30.3487	0.000886
25	enter	139	1	199	25.8966	0.000886
26	enter	16	1	198	24.2717	0.000000
27	enter	19	1	197	24.0755	0.001772
28	enter	143	1	196	21.7162	0.000886
29	enter	9	1	195	19.8030	0.000886
30	enter	57	1	194	21.0847	0.000886

جدول (13)
استخدام إحصاءة Rao's V لاختيار المتغيرات بأسلوب Stepwise.

رقم الخطوة	نوع الخطوة	رقم المتغير	Rao's V	خطأ التصنيف
1	enter	3	6685.831	0.528787
2	enter	132	3815.170	0.345439
3	enter	4	3590.586	0.216121
4	enter	5	3620.364	0.131975
5	enter	130	3654.342	0.058459
6	enter	8	2404.869	0.040744
7	enter	7	1764.797	0.035430
8	enter	1	1339.775	0.021258
9	enter	134	1327.947	0.019486
10	enter	133	975.8164	0.017715
11	enter	135	1104.328	0.014172
12	enter	12	796.1934	0.010629
13	enter	10	912.1680	0.009743
14	enter	13	661.6367	0.007972
15	enter	136	590.3359	0.005314
16	enter	11	543.6836	0.005314
17	enter	6	645.4297	0.004429
18	enter	52	542.0508	0.003543
19	enter	236	765.6250	0.002657
20	enter	140	427.8711	0.003543
21	enter	24	387.9453	0.002657
22	enter	2	430.5664	0.001772
23	enter	15	439.1172	0.000886
24	enter	137	404.0781	0.000886
25	enter	139	347.2500	0.000886
26	enter	16	328.1328	0.000000
27	enter	19	327.8008	0.001772
28	enter	143	297.8594	0.000886
29	enter	9	273.6602	0.000886
30	enter	57	293.1914	0.000886

3-8 اختبار البيانات

لأجل التأكد من تحقق شروط التحليل وجب إجراء الاختبارات الآتية:

أ-اختبار توزيع البيانات: حيث اختبرت البيانات كلاً على حدة بواسطة اختبار حسن المطابقة بمربع كاي حيث وجدت تتوزع طبيعياً ويمكن التنبؤ بذلك من ملاحظة كيفية استخراج هذه المتغيرات والذي تم بأخذ المجموع لمجموعة من الحدود (تحويل فورير) الذي يؤدي حسب نظرية الغاية المركزية إلى توزع متعدد المتغيرات الطبيعية. اختبرت كذلك المتغيرات الناتجة عن مرحلة اختيار المتغيرات التمييزية ووجدت بأنها تتوزع على وفق التوزيع الطبيعي المتعدد بواسطة اختبار التفرطح المتعدد حيث وجد معنوياً لجميع الأساليب والجدول (14) التالي يوضح النتائج.

جدول (14)

اختبار توزيع البيانات بواسطة التفرطح المتعدد.

group 4			group 3			group 2			group 1			group 0			الأسلوب
level	B	kurt	level	B	kurt	level	B	kurt	level	B	kurt	level	B	kurt	
0.055	-1.919	929.72	0.035	-2.109	927.85	0.065	-1.85	925.31	0.013	-2.497	924.15	0.094	-1.674	923.11	JB
0.137	1.487	623.49	0.104	1.629	624.31	0.119	-1.56	600.97	0.091	-1.69	602.79	0.166	-1.385	599.76	JT
0.202	1.277	439.79	0.186	-1.323	425.82	0.234	1.192	424.46	0.003	3.002	448.99	0.295	1.049	436.82	JD
0.188	-1.318	426.02	0.204	1.271	439.7	0.213	1.246	438.78	0.003	3.051	449.25	0.275	1.093	437.09	JP
0.1	-1.647	705.2	0.072	-1.799	703.88	0.105	1.621	726.11	0.045	-2.005	702	0.138	1.484	723.83	D^2
0.208	-1.259	426.33	0.134	-1.498	424.89	0.166	1.386	439.6	0.121	-1.551	424.33	0.223	1.219	437.89	D^2r
0.1	-1.647	705.2	0.072	-1.799	703.88	0.105	1.621	726.11	0.045	-2.005	702	0.138	1.484	723.83	FD^2
0.1	-1.647	705.2	0.072	-1.799	703.88	0.105	1.621	726.11	0.045	-2.005	702	0.138	1.484	723.83	RAOsV
0.088	1.705	782.64	0.057	1.903	783.95	0.087	1.712	781.22	0.043	-2.022	755.45	0.122	-1.546	752.63	Wilks' Λ
0.432	-0.786	189.13	0.315	-1.005	188.28	0.337	0.961	194.97	0.46	-0.739	189.11	0.39	-0.86	187.07	ERROR
0.06	-1.88	930.02	0.037	2.087	961.01	0.065	1.847	957.38	0.024	2.257	962.19	0.093	1.68	954.27	Stepwise Λ
0.057	1.903	959.68	0.045	-2.01	928.63	0.07	1.811	957.07	0.026	-2.236	926.25	0.096	1.665	954.13	StepwiseD^2
0.057	1.903	959.68	0.045	-2.01	928.63	0.07	1.811	957.07	0.026	-2.236	926.25	0.096	1.665	954.13	Stepwise V

group 9			group 8			group 7			group 6			group 5			الأسلوب
level	B	kurt	level	B	kurt	level	B	kurt	level	B	kurt	level	B	kurt	
0.046	2.002	960.21	0.045	2.004	959.82	0.046	-2.001	926.66	0.026	2.234	962.22	0.068	-1.826	924.16	JB
0.123	1.543	623.68	0.135	1.494	623.08	0.139	1.479	622.89	0.098	-1.654	603.52	0.141	-1.473	600.7	JT
0.215	1.239	439.47	0.209	-1.256	425.47	0.187	1.32	439.65	0.122	1.545	441.19	0.241	1.172	438.08	JD
0.211	-1.251	426.04	0.228	-1.206	425.74	0.212	1.25	439.26	0.124	-1.537	424.77	0.244	-1.164	424.01	JP
0.084	1.729	728.06	0.1	1.645	727.16	0.102	-1.637	703.48	0.066	1.839	728.96	0.109	-1.604	700.71	D^2
0.167	-1.383	425.33	0.187	1.32	439.72	0.195	1.298	439.52	0.13	-1.514	424.89	0.187	-1.32	423.07	D^2r
0.084	1.729	728.06	0.1	1.645	727.16	0.102	-1.637	703.48	0.066	1.839	728.96	0.109	-1.604	700.71	FD^2
0.084	1.729	728.06	0.1	1.645	727.16	0.102	-1.637	703.48	0.066	1.839	728.96	0.109	-1.604	700.71	Rao's V
0.076	-1.778	757.37	0.087	-1.712	756.87	0.09	1.699	781.91	0.047	-1.988	756.34	0.103	1.63	780.12	Wilks' Λ
0.408	-0.827	188.83	0.431	0.788	194.54	0.429	0.79	194.52	0.354	0.927	195.17	0.362	0.912	194.68	ERROR
0.051	-1.954	928.69	0.052	1.947	959.36	0.059	-1.891	927.57	0.031	-2.153	927.69	0.074	-1.785	924.52	Stepwise Λ
0.051	1.95	959.8	0.056	1.91	959.06	0.063	1.862	958.51	0.032	2.141	961.49	0.068	1.823	956.63	Stepwise D^2
0.051	1.95	959.8	0.056	1.91	959.06	0.063	1.862	958.51	0.032	2.141	961.49	0.068	1.823	956.63	Stepwise V

ب- اختبار معنوية الدالة المميزة: وهو اختبار اختلاف أوساط المجاميع وتم باستخدام اختبار F المستند على إحصاءة ∧ المقترح من قبل Rao. فقد أظهرت جميع الأساليب وجود اختلاف كبير في أوساط المجاميع الذي يعني إمكانية استخدام الدالة المميزة في التمييز بين المجتمعات موضوعة البحث، الجدول (15) يلخص نتائج الاختبار.

ج- اختبار تجانس مصفوفات التباين والتباين المشترك: حيث تم اختبار التجانس ووجد بأن مصفوفات التباين والتباين المشترك تختلف فيما بينها لذلك توجب استخدام الدالة المميزة التربيعية في التحليل. الجدول (16) يوضح نتائج الاختبار.

جدول (15)

اختبار معنوية الدالة المميزة لجميع الأساليب.

F_0	df_2	df_1	Wilks' ∧	λ	m	S	الأسلوب
96.186	9441.615	270	0.000011	67	1108	8.642	JB
110.439	9290.782	216	0.000021	53.5	1111	8.459	JT
56.573	9093.002	180	0.002032	44.5	1113	8.250	JD
10.284	9093.002	180	0.216840	44.5	1113	8.250	JP
107.915	9355.397	234	0.000014	58	1110	8.533	D^2
134..061	9093.002	180	0.000023	44.5	1113	8.250	D^2r
107.915	9355.397	234	0.000014	58	1110	8.533	FD^2
107.915	9355.397	234	0.000014	58	1110	8.533	Rao's V
108.230	9381.630	234	0.000011	60.25	1109.5	8.564	Wilks' ∧
133.635	8286.957	117	0.000362	28.75	1116.5	7.474	Error
99.695	9441.615	270	0.000009	67	1108	8.642	Stepwise ∧
97.536	9441.615	270	0.000010	67	1108	8.642	Stepwise D^2
97.536	9441.615	270	0.000010	67	1108	8.642	Stepwise V

جدول (16)

اختبار تجانس مصفوفات التباين والتباين المشترك للمجموعات لجميع الأساليب.

الأسلوب	M	C	χ^2_{cal}	d.f.	χ^2_{tab}
JB	32012.48	0.899794	28804.63	4185	4336.955
JT	18694.14	0.919553	17190.25	2700	2822.25
JD	14896.97	0.932734	13894.91	1890	1992.445
JP	11847.13	0.932734	11050.22	1890	1992.445
D^2	23402.53	0.912965	21365.7	3159	3291.152
D^2r	17310.09	0.932734	16145.71	1890	1992.445
FD^2	23402.53	0.912965	21365.7	3159	3291.152
Rao's V	23402.53	0.912965	21365.7	3159	3291.152
Wilks' $\wedge$	24219.52	0.909672	22031.82	3402	3539.102
Error	7184.525	0.955834	6867.216	819	886.779
Stepwise $\wedge$	27616.05	0.899794	24848.75	4185	4336.955
Stepwise D^2	27932.29	0.899794	25133.3	4185	4336.955
Stepwise V	27932.29	0.899794	25133.3	4185	4336.955

3-9 تقدير معالم النموذج وخطأ التصنيف

في البداية تم تقدير معالم النموذج $(\sum_i, \underline{\mu_i})$ بطريقة الإمكان الأعظم وحسب منها الدوال التمييزية التي استخدمت في عملية اختيار المتغيرات التمييزية حيث أعطت النتائج المشار إليها في المباحث السابقة بعدها تم استخدام التقديرات غير المتحيزة لـ $D^2, \sum^{-1}$ (المعادلتين 2.16، 2.17) على التوالي إلا أنها أعطت نفس الإمكان الأعظم ويرجع سبب ذلك إلى كبر العينات المأخوذة وتقارب حجومها لاحظ الجدول (17).

جدول (17)

أحجام العينات للمجموعات والاحتمالات الأولية.

المجموعة	حجم العينة	الاحتمالات الأولية Pi
0	89	0.7880
1	120	0.1063
2	102	0.0903
3	123	0.1089
4	125	0.1107
5	97	0.0859
6	124	0.1098
7	113	0.1001
8	115	0.1019
9	121	0.1072
المجموع	**1129**	**1.0000**

بعد ذلك نتائج طريقة تم تقدير احتمال خطأ التصنيف للدوال الناتجة عن أساليب اختيار المتغيرات المختلفة بطرائق التقدير المذكورة سابقاً وهي طريقة إعادة التعويض Resubstitution، Rotation، Holdout وتم استبعاد طريقة Leave One Out لأنها تتطلب وقتاً طويلاً للحساب بالإضافة إلى أن طريقة Rotation تكون بديلاً لها عادة ويظهر الجدول (18) نتائج عملية التقدير.

فقد قسمت العينة إلى خمس عينات في طريقة Rotation وذلك لتجاوز التحيز من جهة وتقليل زمن الحساب من جهة أخرى كذلك فإن زيادة عدد العينات يؤدي إلى زيادة قليلة جداً لا تذكر في دقة تقدير احتمال خطأ التصنيف. إن طريقة إعادة التعويض متفائلة جداً وتأتي بعدها طريقة Holdout بعبارة أخرى فإن تحيز كل من الطريقتين السابقتين يكون بالاتجاه التفاؤلي ويكون كبيراً كذلك فإن طريقة Holdout تتأثر بطريقة تقسيم العينة بشكل كبير جداً بحيث أن التقديرات الناتجة تختلف بشكل كبير يصعب معه الاعتماد عليها. أما طريقة Leave One Out فإن تحيزها قليل لكنها تتطلب كماً كبيراً جداً من الحساب لأنها تحتاج إلى تقدير الدوال المميزة في كل مرة يتم فيها حذف إحدى المشاهدات فهي تتطلب هنا بحدود عشرين ساعة من الوقت في حاسبة متوسطة السرعة تقريباً[1] لذلك استخدم بدلاً عنها طريقة Rotation المتوسطة التحيز.

[1] حاسبة متوسطة السرعة ( ).

أما بالنسبة لتباين تقديرات هذه الطرائق فتختلف باختلاف الطريقة المستعملة فطريقة إعادة التعويض تمتلك تباين عالي بعدها تأتي طريقة Holdout وبعدها طريقة Rotation متوسطة التباين ثم طريقة Leave One Out قليلة التباين.

جدول (18)

تقدير احتمال خطأ التصنيف بعدة طرق.

ت	الأسلوب	Resubstitution Error	Holdout Error $n_1 = 546$ $n_2 = 565$	Rotation Error $n_1 \ldots n_4 = 226$ $n_5 = 225$	عدد المتغيرات
1	JB	0.017715	0.072642	0.046020	30
2	JT	0.000000	0.042478	0.016842	24
3	JD	0.032772	0.283152	0.101069	20
4	JP	0.180691	0.493805	0.640411	20
5	D^2	0.000000	0.076106	0.026605	26
6	D^2R	0.000000	0.047788	0.030987	20
7	FD^2	0.000000	0.051327	0.025654	26
8	Rao's V	0.000000	0.056637	0.027462	26
9	Wilks' $\wedge$	0.000000	0.083180	0.023907	27
10	Error	0.000000	0.038938	0.036413	15
11	Stepwise $\wedge$	0.000000	0.069027	0.032734	30
12	Stepwise D^2	0.000886	0.115044	0.029630	30
13	Stepwise V	0.000886	0.060177	0.033591	30

3-10 تحليل النتائج

من ملاحظة الجداول (1) إلى (13) يلاحظ بأن أسلوب SFS باستخدام تقدير خطأ التصنيف بطريقة إعادة التعويض أعطى (15) متغير وهو أقل عدد بين باقي الأساليب، إذ وصل تقدير احتمال خطأ التصنيف إلى الصفر ويلاحظ في الخطوات (13 و14) من جدول (10) تكرار نفس قيمة الخطأ وهذا لا يعني عدم وجود معلومات إضافية في المتغير المضاف الجديد في الخطوة (14) وهو المتغير X_{13} وإنما عدم كفاية المعلومات الموجودة فيه لتقليل خطأ ونرى في الخطوة (15) إضافة المتغير X_2 أدت إلى تقليل الخطأ إلى الصفر وذلك لا يعني أن المتغير X_2 أكثر معنوية من المتغير X_{13} ولكن المعلومات في X_2 قد جمعت مع المعلومات في X_{13} والمتغيرات السابقة أدت إلى تقليل الخطأ إلى الصفر ونلاحظ هذه الحالة في أغلب الأساليب.

يأتي بعده أسلوب SFS باستخدام مجموع التباين غير المشروح المحسوب من مسافة مهلنوبس (D^2r) الذي أعطى (20) متغير ووصل الخطأ إلى الصفر أيضاً، كما يلاحظ هنا أن الخطوة (14) في الجدول (6) وهي إضافة المتغير X_{12} إلى النموذج أدى إلى زيادة تقدير خطأ التصنيف وفي الوقت نفسه تقليل المعيار المستخدم وهو مجموع التباين غير المشروح والسبب يعود هنا إلى عدم وجود خاصية الرتابة في احتمال خطأ التصنيف من جهة وإلى الاختلاف بين المعيارين من جهة أخرى والكلام نفسه ينطبق على الحالة نفسها في باقي الأساليب بمختلف المعايير.

أما بالنسبة إلى أسلوب SFS باستعمال معيار Matusita (JT) فقد أعطى (24) متغير ووصل الخطأ إلى الصفر وأعطى أسلوب SFS باستخدام مسافة مهلنوبس و F لمسفة مهلنوبس FD^2 وإحصاءة Rao's V نفس النتيجة وهي (26) متغير لحين وصول الخطأ إلى الصفر ويدل ذلك على الترابط العالي بين هذه المعايير.

أعطى أسلوب SFS باستخدام إحصاءه Wilks' $\wedge$ (27) متغير ووصل الخطأ إلى الصفر وباستخدام معيار JB باتارجاريا (30) متغير ووصل الخطأ إلى (0.0017715)

حيث تم التوقف لصعوبة وعدم جدوى الاستمرار بحساب المعيار وطول الوقت المستغرق لحسابه أنظر جدول (1).

أما المعايير التي أعطت نتائج سيئة باستخدام أسلوب SFS فهي معيار التباعد JD ومعيار باترك فيشر JP أنظر الجدولين (3 و4) حيث لم يؤديان بإضافتهما للمتغيرات الجديدة إلى تقليل كبير في تقدير احتمال خطأ التصنيف كما هو الحال في باقي الطرق لذلك تم التوقف في الخطوة (20) لعدم جدوى الاستمرار فيهما ويلاحظ أن المتغيرات الناتجة تختلف في عمومها عن المتغيرات الناتجة في باقي المعايير.

أما أسلوب الاختيار التدريجي Stepwise الجداول (11) (12) (13) فلم تعطي النتائج المرجوة فقد استمر في إضافة المتغيرات بالرغم من وصول الخطأ إلى الصفر وذلك لجميع المعايير المستخدمة وهي إحصاءة ∧ Wilks'، D^2 وإحصاءة Rao's V وتم التوقف في الخطوة (30) لعدم جدوى الاستمرار وصعوبة وبطؤ الحساب للخطوات اللاحقة يذكر هنا أن نتائج أسلوب الاختيار التدريجي أعطى نتائج مشابهة تماماً لأسلوب SFS حين استعمل المعيار نفسه ويعود السبب إلى عدم حذف أي متغير بأسلوب Stepwise بعد إدخاله.

أما عند مقارنة المعايير المستخدمة من ناحية التعقيد وسرعة الحساب فإن تقدير احتمال خطأ التصنيف أكثر المعايير تعقيداً وبطيئ في الحساب ويأتي بعده معيار ∧ Wilks' وبعدها المعايير JP، JD، JT، JB على التوالي فهي متقاربة في السرعة والتعقيد وتعتبر بطيئة نوعاً ما.

أما المعايير Rao's V، D^2، D^2r، FD^2 فتعتبر سريعة الحساب وبسيطة. على الرغم من ذلك فإن تعقيد أحد المعايير لا يحول دون استخدامه وذلك في بعض الأساليب كذلك فإن هنالك طريقة الحساب التكراري لبعض المعايير [15] تؤدي إلى الإسراع في حسابها كما أن عملية اختيار المتغيرات تجرى لمرة واحدة يتم فيها اختيار المتغيرات المعنوية وبعدها يتم بناء النموذج المطلوب الذي يستعمل لبناء منظومة تمييز الأنماط.

من خلال النتائج السابقة نستنتج أن الهيكل العام لمنظومة تمييز الأنماط وفق الأسلوب الإحصائي باستخدام تحويل فورير لأخذ القياسات والدالة المميزة التربيعية للتصنيف يتكون من الخطوات الآتية:

أ- سحب صورة الأرقام ومعالجتها أولياً لإزالة الضوضاء والحواف السوداء.

ب- أخذ القياسات بواسطة حساب مقياس الاستدارة لإطار الأنماط وأخذ تحويل فورير لإحداثيات نقاط إطار كل نمط منها وذلك للمؤشرات التي تمثل المتغيرات التمييزية الناتجة عن استخدام تقدير خطأ التصنيف بأسلوب SFS (جدول 10) وهي:

$Rx_2, Rx_3, Ry_2, Rm, Ry_3, Ry_{40}, Ix_{12},$
$Rx_{20}, Ix_2, Rx_5, Ix_3, Ix_{10}, Ix_{24}, Rx_7, Rx_1$

ج- تدخل النتائج الدالة المميزة التربيعية المحسوبة مسبقاً لهذه المتغيرات لينتج عنها تصنيف الأنماط الجديدة.

4- الاستنتاجات والتوصيات

نود أن ننوه إلى أن موضوع البحث كان واسعاً ومتشعباً، ولكن حاول الباحث من خلال هدف البحث شمول كافة الجوانب المتعلقة بالتحليل المميز وكذلك استخدام الأساليب الأكثر شهرة في اختيار المتغيرات فضلاً عن استخدام تحويل فورير للحصول على القياسات المطلوبة واستعمال الأساليب المرنة في عملية تقطيع الصورة والتطبيع.

4-1 الاستنتاجات

من ملاحظة النتائج العملية في الفصل الثالث نستنتج ما يأتي:

1- أن معيار احتمال خطأ التصنيف هو معيار يمكن استخدامه في اختيار المتغيرات التمييزية وذلك بتقديره بطريقة إعادة التعويض.

2- أعطى معيار خطأ التصنيف أفضل مجموعة من المتغيرات بين باقي المعايير لذلك فإن التحيز الناتج عن تقديره لم يؤثر على أمثليته في اختيار المتغيرات.

3- إن مجموع التباين غير المشروح المحسوب من مسافة مهلنوبس هو معيار بسيط الحساب وسريع التنفيذ ويعطي نتائج جيدة ويمكن أن يستخدم لاختيار المتغيرات في حالة تعذر استخدام احتمال خطأ التصنيف.

4- ضرورة استخدام أسلوب البحث الأمثل لاختيار المجموعة المثلى من المتغيرات في حالة تحقق خاصية الرتابة في المعيار المستخدم وألا يتم استخدام البحث الشامل (أخذ جميع التوافيق الممكنة) هذا إذا كان عدد المتغيرات قليل أما إذا زاد عدد المتغيرات فيجب استخدام أسلوب البرمجة الديناميكية (Z_l, Z_r) وذلك باختيار r, l المناسبتين حسب تعقيد المسألة وكما ذكر سابقاً فإن اختيار r, l يمكن أن يؤدي إلى اختزال هذا الأسلوب إلى أسلوب أبسط.

5- أعطى الأسلوب الإحصائي نتائج جيدة في تمييز الأرقام العربية من حيث احتمال خطأ التصنيف وكذلك من ناحية وقت التمييز فكان يساوي (0.0102) ثانية للنمط الواحد وذلك على الحاسبة Pentium II 200MHz وهو وقت قليل ويمكن تقليله أكثر بإجراء بعض التعديلات التقنية على برنامج التمييز عند استخدامه عملياً.

6- ضرورة استخدام أسلوب أخذ القياسات الذي يستند على جميع نقاط النمط والذي يؤدي إلى أخذ جميع المعلومات الخاصة بالنمط وذلك في حالة استخدام الأسلوب الإحصائي في التمييز.

4-2 التوصيات

1- يوصي الباحث بضرورة البحث عن أساليب جديدة لأخذ القياسات تعطي قياسات متجانسة التباين بحيث يمكن أن تحلل بواسطة الدالة المميزة الخطية.

2- ضرورة دراسة استعمال الأساليب الأخرى في التمييز ومنها تمييز اللوجستك Logistic Discrimination وأسلوب اتخاذ القرارات المتسلسل Sequential Decision Making كما يمكن استخدام الأساليب اللامعلمية مثل طريقة النواة Kernel وطريقة الجار الأقرب K (Nearest Neighbor K) التي ثبت بأنها تعطي احتمال خطأ تصنيف يكون بقدر نصف خطأ التصنيف للدالة المميزة وذلك في حالة اقتراب حجم العينة من اللانهاية. كذلك يمكن استخدام الدالة المميزة الخطية المقطعة Piecewise Linear Discriminate Function السريعة التصنيف. كما يمكن اعتبار دالة الإطار كسلسلة زمنية تجرى عليها عملية التمييز بواسطة أساليب التصنيف الخاصة بالسلاسل الزمنية.

3- يوصي الباحث باستخدام تقدير احتمال خطأ التصنيف بطريقة إعادة التعويض كمعيار لاختيار المتغيرات.

4- كما يوصي باستخدام طرق جديدة لتقدير احتمال خطأ التصنيف واستخدامها في اختيار المتغيرات.

5- كذلك يوصي بالبحث عن معايير جديدة تترابط بصورة وثيقة مع احتمال خطأ التصنيف وتسهل عملية اختيار المتغيرات.

6- ضرورة دراسة الأساليب التكرارية في حساب المعايير لما لها الأثر الأكبر في تسهيل وتبسيط حساب معايير اختيار المتعيرات.

المصادر العربية

[1] عتاب، محمد رمضان (1999) " أستخدام الدالة المميزة التربيعية في تميز الارفام العربية. رسالة ماجستير في الاحصاء/ كلية الآدارة والآقتصاد/ جامعة بغداد.

المصادر الآجنبية

[2]Chen, C.H., (1977) "On the Use of Distance and Information measures in Pattern Recognition and Applications" in Fu, K.S &Whinston, A.B., "Pattern Recognition Theory and Application" Noordhoff–International Publishing. pp 45-60.

[3]Constanza, M.C. & Afifi, A.A., (1979) "Comparison of Stopping Rules in Forward Stepwise Discriminant Analysis" JASA. Vol. 74, No.368, pp 777 – 785.

[4]Devijver, P.A., &Kittler, J., (1982) "Pattern Recognition: A Statistical Approach" Prentice Hall International, Inc., London.

[5]Lanchenburch, P.A. & Mickey, M.R, (1968) "Estimation of Error Rate in Discriminant Analysis" Technometrics. Vol. 10, No. 1, pp 1- 12.

[6]Nussbaumer, H.J. (1982) "Fast Fourier Transform and Convolulion Algorithms" Springer – Verlag Berlin Heidelberg.

[7]Rao, C.R., (1973) "Linear Statistical Inference and Its Applications" John wiley &Sons Inc. Canada.

[8]Snapinn, S.M & Knoke, J.D., (1989) "Estimation of error rates in discriminant

analysis with selection of variables" Biometrics Vol. 45, No.1, pp 289-300 .
[9]Therrien, C.W., (1989) "Decision Estimation and classification" John wiley & Sons New York.
[10] Wasson, W.D., (1980) "A Pre – Processor for Hand printed Character Recognition" IEEE Pattern Recognition. Vol. 1, pp 703 – 777.
[11] Young, T.Y., (1974) "Classification, Estimation and Pattern Recognition" American Elsevier Publishing company Inc. New York.

Abstract

The objective is to design and determine electronic system for Arabic number pattern recognition.

To achieve this pattern recognition and It's relation to statistical analysis were discussed and recognized with outline for fourier transformation and deriving quadratic discriminate function used for recognition.

The process also include choosing input Variables in the function of criteria used to measure the performance discriminates function by using classification error estimation which used for the first time using substitution method.

To prove this sample of (150) person was implemented by writing Arabic numbers from (0-9) manually so we have (1250) patter dis Given the statistical methods used to distinguish good results in the crimination of Arabic numerals in terms of classification error as given (15) Lovell, a variable number of the rest of the methods and arrived in estimating the probability of classification error to zero. Through programs written in C++.