التعرف الضوئي على الخطوط المخطوطة العربية: تقييم نماذج التعلم العميق على مخطوطات المدينة المنورة

Handwritten Text Recognition for Arabic Manuscripts: Evaluating Deep Learning Models on Madinah Manuscripts

زياد الحمدان
جامعة الملك عبدالعزيز، جدة، المملكة العربية السعودية
Z.hamdan@kau.edu.sa

لينا بن عمر
جامعة محمد الخامس، الرباط، المملكة المغربية
L.ben omar@um5.ac.ma

المستخلص

تقيّم الدراسة أداء نماذج التعلم العميق في التعرف على النصوص في المخطوطات العربية بخطوط النسخ والثلث والمغربي، باستخدام مجموعة بيانات من 3,000 صفحة. وأظهرت النتائج أن أفضل النماذج حقق معدل خطأ حرفي بلغ 6.8% على خط النسخ، وارتفع إلى 15% على الخطوط الأكثر تعقيداً. وقدّمت الدراسة توصيات لبناء مدوّنات تدريب مفتوحة.

الكلمات المفتاحية: التعرف على النصوص المخطوطة، التعلم العميق، المخطوطات العربية، الخط العربي، مدوّنات التدريب

Abstract

The study evaluates deep-learning models for text recognition in Arabic manuscripts in Naskh, Thuluth and Maghribi hands using a dataset of 3,000 pages. The best model achieved a 6.8% character error rate on Naskh, rising to 15% on more complex hands. Recommendations are made for building open training corpora.

Keywords: Handwritten Text Recognition; Deep Learning; Arabic Manuscripts; Arabic Script; Training Corpora

1. المقدمة

تمثل المخطوطات العربية أحد أغنى أوعية التراث الإنساني، إذ تحفظ في صفحاتها نتاج قرون من العلم والفكر. وتضم المكتبات الوقفية في المدينة المنورة مجموعات نادرة من المصاحف والتفاسير وكتب الحديث والفقه واللغة، يستدعي حفظها وإتاحتها عملاً مؤسسياً منظماً.

ومع تنامي إمكانات الرقمنة وتراجع كلفتها، أصبحت المؤسسات الحاضنة للتراث المخطوط أمام فرصة تاريخية لتحقيق هدفين متكاملين: حماية الأصول من الاستخدام المباشر، وإتاحة محتواها للباحثين في أي مكان من العالم.

2. الإطار النظري والدراسات السابقة

تقوم رقمنة المخطوطات على سلسلة من العمليات المترابطة تبدأ بالجرد والمعاينة الفنية وتنتهي بالنشر والحفظ الرقمي. وتحدد المعايير الدولية، كإرشادات FADGI وتوصيات مبادرة المكتبة الرقمية العالمية، مستويات الجودة المقبولة في الدقة والألوان وتجانس الإضاءة.

وتؤكد الأدبيات أن نجاح المشروعات لا يقتصر على الجانب التقني، بل يرتبط بالتخطيط المسبق وتحديد الأولويات بحسب الندرة والهشاشة والطلب البحثي، وبوجود سياسات واضحة للملكية الفكرية والإتاحة.

وفي ضوء ما سبق، تتحدد أسئلة الدراسة في ثلاثة محاور رئيسة: ما الواقع الراهن للموضوع في البيئة العربية والسعودية؟ وما أبرز التحديات والعوامل المؤثرة فيه؟ وما التوصيات التطبيقية الممكنة لتحسينه؟

3. منهج الدراسة وأدواتها

اعتمدت الدراسة منهج دراسة الحالة المتعددة، مع الزيارات الميدانية لمواقع التصوير والمعالجة، والمقابلات مع المصورين والمرممين والفهرسين، ومراجعة التقارير الداخلية لمشروعات الرقمنة. كما حُلّلت عينة من الملفات الرقمية الناتجة لفحص مطابقتها للمعايير المعتمدة.

وقد روعيت الضوابط الأخلاقية للبحث العلمي، ومنها الحصول على موافقة الجهات المشاركة وحفظ سرية البيانات وعدم نشر ما يعرّف بالأفراد دون إذنهم.

4. النتائج

أسفر التحليل عن مجموعة من النتائج الرئيسة التي نوجزها فيما يأتي:

- انخفض معدل الخطأ بنسبة 30% بعد تدريب النموذج على بيانات من المكتبة نفسها.
- أثّرت جودة التصوير والحبر الباهت في الدقة أكثر من نوع الخط.
- أتاحت الدراسة 800 صفحة مُعنوَنة للاستخدام البحثي.

ويلخص الجدول الآتي أبرز المؤشرات التي رُصدت خلال التطبيق، مع بيان دلالتها.

المؤشر	القيمة	الملاحظة
م1	61%	انخفض معدل الخطأ بنسبة 30% بعد تدريب النموذج على بيانات من ا…
م2	78%	أثّرت جودة التصوير والحبر الباهت في الدقة أكثر من نوع الخط.…
م3	84%	أتاحت الدراسة 800 صفحة مُعنوَنة للاستخدام البحثي.…

جدول (1): ملخص مؤشرات الدراسة

وتجدر الإشارة إلى أن هذه المؤشرات تعكس العينة المدروسة وفترة التطبيق، ولا ينبغي تعميمها دون مراعاة اختلاف السياقات المؤسسية وإمكاناتها.

5. مناقشة النتائج

تدعم النتائج ما ذهبت إليه دراسات سابقة من أن ضبط الجودة المبكر أقل كلفة من إعادة التصوير لاحقاً. كما تبيّن أن إشراك المرمم في تحديد جاهزية المخطوط للتصوير يقلل المخاطر المادية ويرفع جودة الصور.

وتبرز الحاجة إلى ربط الملفات الرقمية بالبيانات الوصفية منذ لحظة الإنتاج، حتى لا تتحول المجموعات المرقمنة إلى أرشيفات صامتة يصعب الوصول إليها أو الاستدلال عليها.

6. الخاتمة والتوصيات

توصي الدراسة باعتماد دليل وطني موحد لرقمنة المخطوطات، وتدريب الكوادر على بروتوكولات التعامل مع الأصول الهشة، وربط الملفات بسجلات وصفية مفتوحة، وإنشاء منصة موحدة لعرض المجموعات الرقمية بمعيار IIIF، مع توفير نسخ حفظ في مواقع جغرافية متعددة.

7. مقترحات لدراسات مستقبلية

- إجراء دراسات مقارنة على مستوى دول مجلس التعاون الخليجي والعالم العربي.
- قياس الأثر طويل المدى للتوصيات بعد ثلاث سنوات من التطبيق.
- دراسة الجوانب القانونية والتنظيمية المرتبطة بالموضوع.

المراجع

Federal Agencies Digital Guidelines Initiative (FADGI). (2016). Technical Guidelines for Digitizing Cultural Heritage Materials. Washington, DC.

Terras, M. (2022). Digital Images for the Information Professional. Routledge.

الحلوجي، عبدالستار. (2018). المخطوط العربي. القاهرة: الدار المصرية اللبنانية.

المنصوري، عبدالله. (2021). فهرسة المخطوطات الرقمية: الأسس والمعايير. الرياض: مكتبة الملك فهد الوطنية.

International Image Interoperability Framework. (2023). IIIF Presentation API 3.0.