

ملخص

في سياق علم الأدوية والرعاية الصحية، تحتوي الملصقات الطبية على معلومات أساسية مثل أرقام الدُفعات، تواريخ التصنيع والانتها، تعليمات الجرعات، والأسعار. ومع ذلك، وبسبب غياب تنسيق موحد للملصقات الطبية معتمد من السلطات الصحية الجزائرية، يظل تطوير الأنظمة الآلية للتعرف على الملصقات الطبية تحديًا كبيرًا.

يقترح هذا البحث نظامًا قائمًا على الذكاء الاصطناعي (AI) للتعرف تلقائيًا واستخراج النصوص من الملصقات الطبية. وللتعامل مع هذه المشكلة، قمنا بإنشاء مجموعة بيانات مخصصة تحتوي على أكثر من 2500 صورة ملصق طبي مُعلّمة. ومن أجل تقييم الحلول، تم اقتراح ثلاثة محركات OCR: Tesseract، EasyOCR، و TrOCR واختبارها على مجموعة البيانات المذكورة.

بالإضافة إلى ذلك، توسعت الدراسة عبر نماذج محسّنة ومعايرة دقيقة، إلى جانب تحسين المعايير (HPO) لتعزيز أداء النماذج في سياق طبي محدد. تم تطبيق طريقتين لتحسين المعايير: Optuna Configurator والخوارزميات الجينية (GA) لتحسين أداء نموذج OCR المقترح.

أظهرت النتائج أن الخوارزميات الجينية تفوقت باستمرار على تقنية Optuna عبر جميع نماذج OCR الثلاثة. تُشير هذه النتائج إلى أن نموذج TrOCR، عند اقترانه بالتحسين التطوري، هو النهج الأكثر كفاءة للتعرف الآلي على النصوص في الملصقات الطبية الجزائرية.

يقدم هذا المشروع حلًا قويًا لاستخراج النصوص من الملصقات الطبية، مما يمهّد الطريق أمام أنظمة ذكية يمكنها دعم تصنيف الأدوية وتحسين إدارة البيانات الصيدلانية في الجزائر.

الكلمات المفتاحية:

الذكاء الاصطناعي، الملصقات الطبية، التعرف على النصوص، الاكتشاف، التجزئة، التصنيف، OCR، EasyOCR، TrOCR، Tesseract، تحسين المعايير، الخوارزميات الجينية، Optuna.