

LỜI CẢM ƠN

Trong bối cảnh các cơ quan, doanh nghiệp đẩy mạnh chuyển đổi số để tối ưu quy trình vận hành, nhu cầu số hóa và tự động hóa xử lý giấy tờ ngày càng cấp thiết. Công việc nhận diện và phân loại các loại văn bản—như CMND/CCCD, bằng lái, hoá đơn, hợp đồng, phiếu thu/chi, công văn—nếu làm thủ công thường tốn nhiều thời gian, dễ sai sót và khó truy xuất. Nhiều đơn vị hiện vẫn dựa vào nhập liệu tay hoặc tra cứu thủ công, dẫn đến độ trễ xử lý, thiếu minh bạch về nguồn dữ liệu và khó khăn trong việc kiểm soát phiên bản, lịch sử chỉnh sửa.

Xuất phát từ thực tế đó, đề tài “Xây dựng hệ thống nhận diện và phân loại giấy tờ” được thực hiện nhằm phát triển một giải pháp số hóa trọn vẹn vòng đời tài liệu: từ khâu tiếp nhận ảnh/chứng từ, tiền xử lý (chỉnh nghiêng, khử nhiễu, tăng tương phản), nhận dạng vùng văn bản, OCR trích xuất nội dung, đến phân loại tự động theo loại giấy tờ và trích xuất trường thông tin quan trọng (tên, số định danh, ngày cấp, số tiền...). Hệ thống hướng tới mục tiêu rút ngắn thời gian nhập liệu, nâng cao độ chính xác, tiêu chuẩn hóa cấu trúc dữ liệu đầu ra và hỗ trợ tra cứu—lưu trữ tập trung, phục vụ kiểm toán cũng như khai thác dữ liệu về sau.

Mặc dù đã nỗ lực để bảo đảm tính ứng dụng và độ tin cậy của mô hình, do hạn chế về thời gian và kinh nghiệm, đề tài khó tránh khỏi những thiếu sót—đặc biệt ở các trường hợp ảnh đầu vào chất lượng kém hoặc bố cục tài liệu không chuẩn. Em rất mong nhận được những ý kiến đóng góp của thầy cô để hoàn thiện hệ thống, nâng cao độ chính xác nhận dạng–phân loại và tối ưu hiệu năng triển khai trong môi trường thực tế.