INSTITUTE OF TECHNOLOGY OF CAMBODIA

DEPARTMENT OF APPLIED MATHEMATICS AND STATISTICS

RESEARCH AND DATA ANALYTICS LABORATORY

Department of Applied Mathematics and Statistics

BANK TRANSACTION SCANNER

Automating Financial Data Extraction From Bank Statements

Under the supervision of professor **KHEAN Vesal**

LENG Devid, LY Chungheang, NANG Chettra, NGOUN Lyhorng, NHEN Theary, NY Chantharith

INTRODUCTION

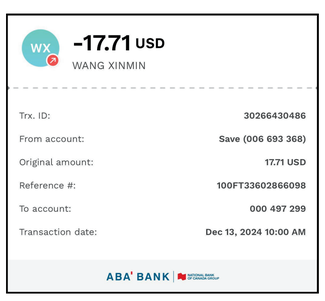
In today's fast-paced financial environment, businesses and auditors frequently handle large volumes of bank statements in scanned PDFs or image formats. Manual extraction of financial data is time-consuming, error-prone, and inefficient. Our project, Bank Transaction Scanner, addresses this challenge by automating the extraction process, enabling efficient and accurate handling of bank statement data through a user-friendly web platform.

OBJECTIVE

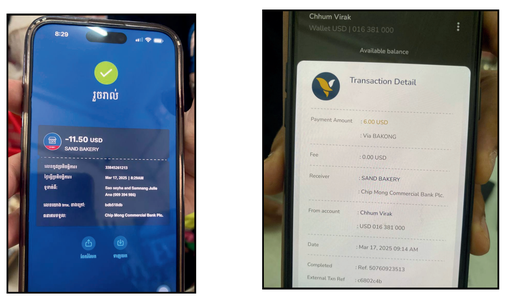
The Bank Transaction Scanner automates financial data management by capturing and converting bank document photos into text, extracting and calculating transaction details like dates and amounts, and organizing the data on an easy-to-use website for viewing or downloading, saving time and effort.

DATA COLLECTION

Collecting Sand Bakery's transactions via screenshots and camera-captured photos for spending analysis.



Screenshots



Camera-Captured Photos

RESEARCH WORKFLOW

Data Collection

- Gather a diverse set of transaction images from different banks to use for training and testing.

Bank Classification Model

- Model Training: Train a deep learning model to classify images by bank.
- Model Validation: Evaluate the model's accuracy using a validation dataset.

Data Preprocessing

- Denoising: Remove visual noise to improve OCR accuracy.
- Resizing: Standardize image dimensions for consistent processing.

Build Templates

- Define Template Rules per Bank: Create custom extraction patterns based on each bank's statement layout.

VALIDATION

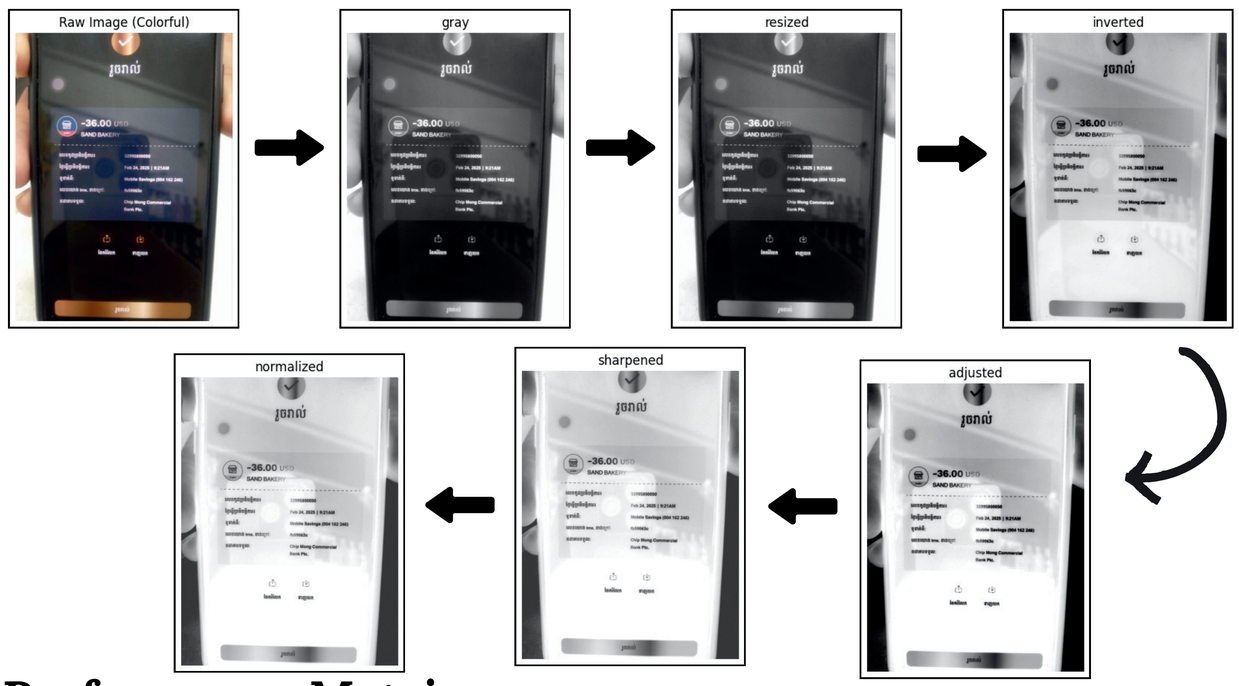
- Check Accuracy and format :Verity correctness of extracted fields and detect missing or malformed information.

Web Application Development

- User Interface + Backend Integration: Design a user-friendly web interface with backend logic to automate the workflow.

RESULT

We tested the full pipeline on various bank transaction images. Preprocessing enhanced OCR accuracy, and template-based extraction successfully captured key details like amounts and dates. The system validated and stored the data, confirming its reliability for automated processing.



Performance Metrics

Metric	Value
OCR Text Extraction Accuracy	**89.50%**
Complete Record Extraction Rate	**92%**
Missing Field Detection Accuracy	**100%**
Average Processing Time	**~2.1 seconds/image**

CONCLUSION

Integrated with deep learning, OCR, and web technologies, the project developed a comprehensive and automated image-based transaction data extraction system. The whole process integrates image cleaning, classification, template-driven extraction, and validation, accomplishing high precision and efficiency. Users can effortlessly get obtained transactional data organized in a precise format through an intuitive web application that was successfully implemented.