ព្រះរាជាណាចក្រកម្ពុជា

ជាតិ សាសនា ព្រះមហាក្សត្រ

REPORT OF PROGRAMMING FOR DATA SCIENCE

GROUP: I3-AMS-B-Group3

Name of Students	ID of Students	Score
1. LENG DEVID	e20220888	……….
2. LY CHUNGHEANG	e20220980	……….
3. NGOUN LYHORNG	e20221532	……….
4. NANG CHETTRA	e20220468	……….
5. NHEN THEARY	e20220455	……….
6. NY CHANTHARITH	e20220472	……….

Lecturer : Mr. OL Say (Course)
Mr. KHEAN Vesal (TP)

Academic Year 2024-2025

Table of Content

I. Introduction

In financial management, processing and verifying transaction receipts manually is time-consuming and prone to error. Bank receipts often contain critical data, but they come in various formats and languages. To address this, our project automates the extraction and classification of transaction data from scanned bank receipts. We focus on two banks — ABA and ACLEDA — and support receipts in both Khmer and English. The goal is to improve processing speed, reduce manual errors, and build a scalable system for wider adoption.

II. Objectives

- Automatically classify receipts from supported banks.
- Preprocess images to improve OCR performance.
- Extract key transaction details including ID, date, amount, and currency.
- Store structured data into a MongoDB database.
- Provide an easy-to-use web interface for uploads, tracking, and downloads.

III. Methodology

1. Data Collection

We collected a diverse set of bank receipt images from Sand Bakery in Khmer and English formats.

2. Bank Classification Model Training

Images undergo preprocessing steps for enhanced clarity:

- Convert to grayscale
- Resize
- Invert
- Adjust contrast and brightness
- Sharpen
- Normalize
- For light backgrounds: apply thresholding

3. Preprocessing

We trained a CNN-based image classifier to identify whether a receipt belongs to ABA or ACLEDA bank. The model achieves high accuracy by learning logo and layout patterns.

4. Data Extraction Templates

For each classified bank, custom templates define where and how to extract transaction fields using OCR output patterns.

5. Validation

We verify the structure and format of each extracted value to ensure consistency and completeness.

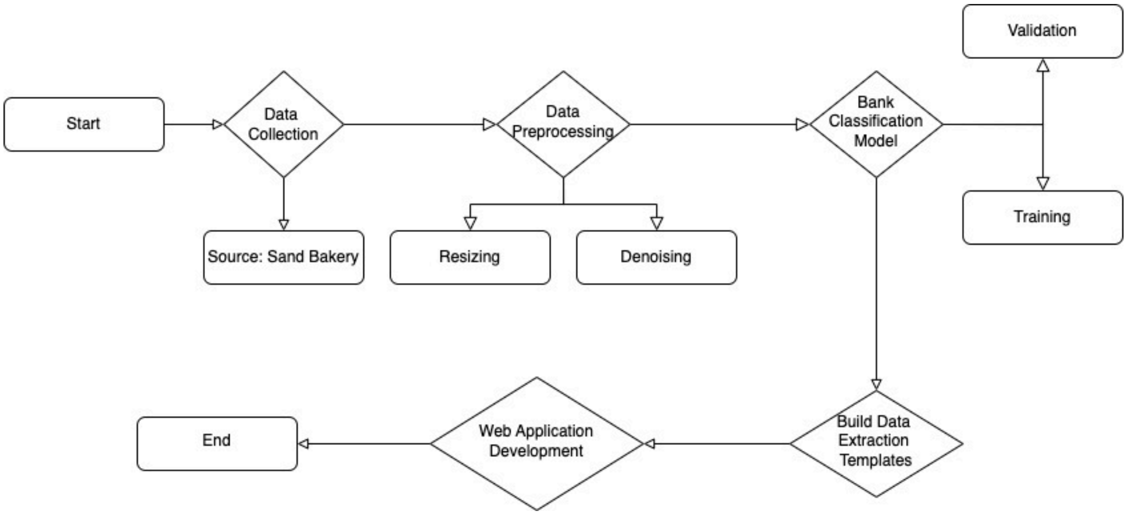
6. Web Application

Built using FastAPI (backend) and Next.js (frontend), the platform includes:

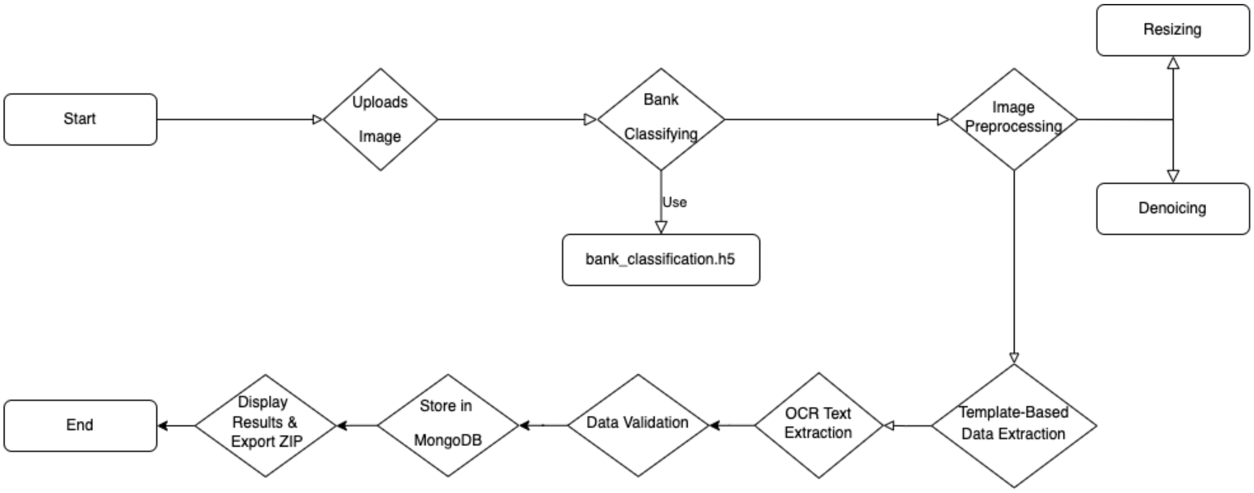
- Upload interface
- Viewable history log
- Downloadable extracted data
- About Us page

IV. Flow

1. Research Flowchart



2. System Workflow



V. Results

The system successfully extracts transaction information from ABA and ACLEDA receipts in both Khmer and English. Preprocessing significantly improved OCR performance. Key fields — transaction ID, date, amount, and currency — were extracted with high accuracy. Extracted data is structured and stored in MongoDB. Users can upload, track, and download results via a web app with three pages: Home, History, and About Us.

VI. Evaluation

The OCR and extraction pipelines were tested on numerous receipt samples. The classification model achieved consistent accuracy across test cases. Preprocessing proved essential, especially in cases with noisy or low-contrast images. Template-based extraction showed robustness in known layouts but highlighted limitations in generalizability.

VII. Challenges

- **Template rigidity**: Templates were manually built and rely on fixed OCR patterns.
- **Image cropping**: We couldn’t yet isolate data fields before extraction.
- **Storage constraints**: The system doesn’t store the original images due to limited space.

VIII. Conclusion

The system successfully classifies banks and extracts key transaction data from ABA and ACLEDA receipts in both Khmer and English. It improves OCR accuracy through advanced preprocessing and provides a user-friendly web interface for viewing and downloading structured results.

IX. Future Work

We aim to:

- Support full Khmer OCR and field-wise extraction.
- Integrate a new OCR engine capable of cropping text fields.
- Optimize preprocessing pipeline for even greater clarity.
- Extend support to more banks and formats.
- Add backend image storage support.

X. Reference

This our GitHub Link Repository: Bank-Transaction-Scanner