MONASH University

MONASH INFORMATION TECHNOLOGY

DATA SCIENCE

FINAL YEAR PROJECT

Eunice Lee Wen Jing, Foo Kai Yan, Alicia Quek Chik Wen, Jesse Yow San Gene
Supervised by Dr Muhammad Fermi Pasha

Automated Health Information System (AHIS)

INTRODUCTION

- Traditional Health Information Systems (HIS) reduce paper records but still rely heavily on manual data entry, which is time-consuming and error-prone.
- **Objective:** To develop a smart, web-based system that uses automated handwriting recognition to accurately extract and digitize handwritten patient data, reduce administrative workload, improve data accuracy and ensure secure, scalable data management.
- AHIS was designed to address these issues by transforming handwritten registration forms into digital records automatically.
- AHIS supports various handwriting styles (cursive and non-cursive) that improves operational efficiency and strengthens patient data confidentiality.

METHODS AND MATERIALS

System Architecture

- Web-Based Platform
  - Developed using MEAN stack framework:
    - MongoDB: Scalable database storing patient, appointment, medical records.
    - Express.js & Node.js: Backend server, handles API routing and model integration.
    - Angular: Frontend framework, provides responsive and user-friendly interface.
- Security
  - User authentication using bcrypt for encrypted password storage.
  - Role-based access control to restrict sensitive data access to authorized users only.
  - Validation mechanisms on input fields (e.g. login, registration).

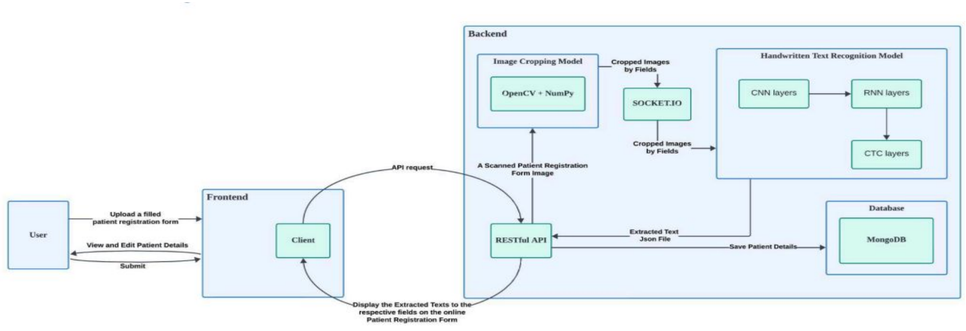

Figure 1. Project Final System Design

Handwriting Recognition Pipeline

- Image Cropping Model (**ICROP**)
  - Designed to crop uploaded form images into individual fields based on predefined pixel coordinates.
  - Cropping ensures that only targeted handwritten sections are sent to the recognition model.
  - Preprocessing includes image resizing (standardized to 1242×1755 pixels) and grayscale conversion.
- Handwritten Text Recognition Model (**HTR**)
  - Combines CNN for feature extraction, RNN for sequence learning and CTC for sequence decoding without explicit alignment.
  - Training Dataset: IAM Handwriting Dataset — 1,539 pages from 657 writers, covering various handwriting styles (cursive and block letters) which includes uppercase, lowercase and punctuation.
  - Output is in JSON format, mapping each cropped image field to its recognized text.

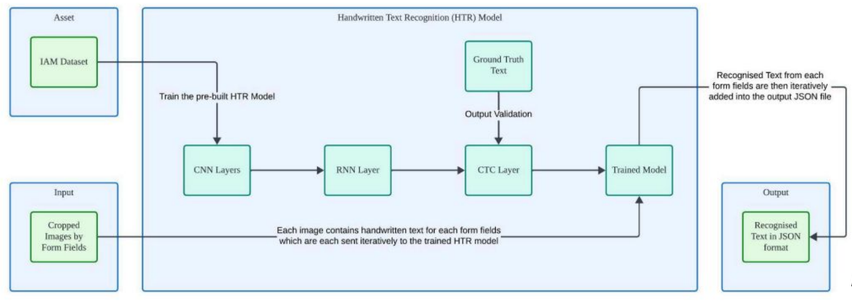

Figure 2. HTR Model Design

RESULTS AND DISCUSSION

Web Application Outcomes

- Core features include:
  - Smart patient registration (with auto-filled data from HTR).
  - Doctor appointment scheduling, diagnosis, prescription, and medication management.
  - Profile management for patients and physicians.
- Strong security implementation:
  - Bcrypt used for password hashing to prevent data leaks.
  - Role-based access control ensures only authorized users access sensitive data.
- Scalable MongoDB database:
  - Designed using a reference-based model for efficient data organization and easy future expansion.

HTR Model Performance

- Achieved **~75% accuracy** in recognizing handwritten text from patient forms.
- High accuracy on standard fields: name, address, date of birth.
- Minor errors with special characters (e.g. @, /, #).
- Outputs text as structured JSON, auto-filled into web forms for review.
- User feedback highlighted faster registration and improved data accuracy.
- Accuracy depends on image quality (clear, shadow-free scans perform best).
- Works best with consistent form layouts to align fields correctly.

CONCLUSION

AHIS successfully revolutionizes patient data management through a web-based system that automates handwritten data entry using HTR technology. We achieved key goals of creating a scalable, accurate, and efficient solution for patient registration, while future improvements include increasing the model's accuracy and reducing its loading time. We extend our sincere thanks to Dr. Fermi Pasha for his invaluable support and guidance.