DeCognize:

Prescription Digitization Using Knowledge Graphs.

Group Members:

Muhammad Sharjeel Akhtar (20P-0101)

Mahad Ashraf (20P-0563)

Project Supervisor

Sir Sohaib Muhammad Khan

Introduction

The project addresses the need for:

- Enhanced Medical Record Digitization
- Improved Healthcare Decision-Making
- Research Opportunities
- Efficiency And Time Savings
- Competitive Recognition
- Scalable Deployment
- Accessibility
- Contribution To Local Healthcare

Motivation

The project addresses critical healthcare challenges by:

- Improving data management.
- Preventing medication errors.
- Ensuring quick access to accurate patient information.
- Facilitating informed decision-making.
- Reducing manual data entry.
- Exploring healthcare technology research.
- Enhancing handwritten prescription handling.
- Leveraging Knowledge Graph technology.
- Ultimately, enhancing patient safety and healthcare quality.

Problem Statement

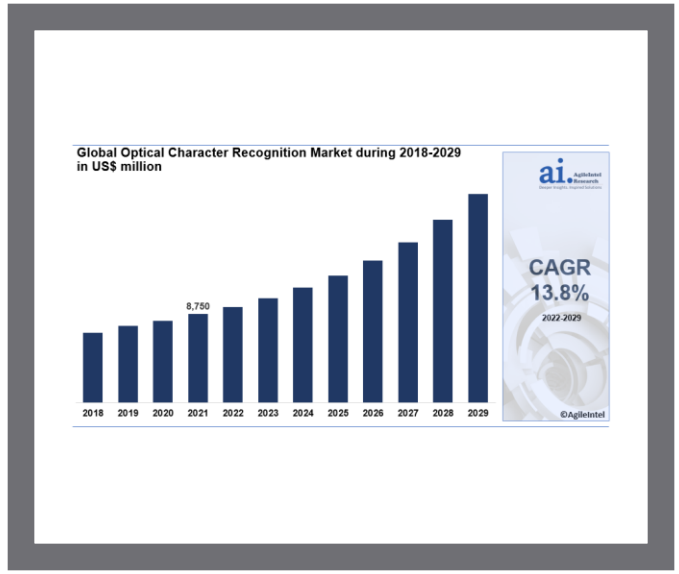
- **Problem**: Inefficient healthcare data management for prescriptions.
- **Challenge**: Illegible handwriting , medical jargon and Knowledge Graph
- **Consequence**: Errors in healthcare due to traditional OCR systems.
- **Goal**: Develop NLP-based system for accurate prescription transcription.

Solution

- **OCR Enhancement:** Improve scanned image quality and transcribe medical handwriting accurately using advanced models.
- **NLP-Based Enhancement:** Use of NLP with enhance our Project in overall readability.
- **Bio-NER:** Identify drugs, diseases, and procedures in prescriptions.
- **Bio-NEL (BERN):** Link entities to a biomedical knowledge graph, enriching data.
- **BERT Text Enhancement:** Improve readability by disambiguating text and providing definitions/synonyms.
- **Zero-Shot Relation Extraction:** Identify relationships like drug-dosage links.
- **Approach:** This approach combines OCR and NLP, with BERN adding specialized biomedical knowledge and context from a biomedical knowledge graph.
- **Knowledge Graph:** Collect medical data, construct a structured graph using standardized vocabularies, and enable efficient data retrieval.
- **Node Embedding for Context:** Use techniques to understand entity significance and relationships in prescriptions within the knowledge graph.

Scope of the project

- Global OCR in Healthcare: Booming market, 15.4% CAGR (2022-2030).
- Recent Projects: Automated doctor prescription by Nano Net Technologies Inc and Neurodata Group.
- OCR in Healthcare in Pakistan: Active research by Seerat Rani, Abd Ur Rehman, Beenish Yousaf, Hafiz Tayyab Rauf, Emad Abouel Nasr, and Seifedine Kadry.
- Summary: OCR enhancing healthcare in Pakistan through innovation and integration.



Goals

- Prescription Support
- Entity Accuracy
- Object Detection Integration
- Synthetic Dataset Creation
- Improve data with the help of NLP
- Accurate Text Mining
- Single-shot Processing
- Document AI Integration
- MLOps Implementation
- AWS Cloud Deployment
- User-friendly interface
- Documentation and Training
- Testing and Evaluation
- Contribution To Healthcare

Expected Result

The system will be capable of accurately and efficiently digitizing medical documents.

The key expected results include:

- Accurate Data Extraction
- Efficient Object Detection
- Comprehensive Bio-Medical Knowledge Graph
- User-friendly Interface
- Documentation and Training Resources
- Enhanced Healthcare Data Management
- Enhanced Decision-Making
- Improved Patient Safety
- Contribution To Healthcare
- Scalability

Objectives

SEP-OCT

STAGE 1: STARTING

- Initial Research+Literature Review
- Data Gathering
- Annotations
- Labelling
- Text Detection Techniques

NOV-DEC

STAGE 2: BASIC IMPLEMENTATION

- Detailed Literature Review
- Analysis
- Training Text Extraction Models
- Improving Text Detection Techniques
- Creating FrontEnd and BackEnd (Initial Stage)

JAN-MAR

STAGE 3: TRAINING AND TESTING

- Completing FRONT-END and BACK-END Web Application
- Improving Text Extraction Module
- Testing And Validation
- Integration Of Improved Detection Techniques
- User Testing And Feedback

APR-MAY

STAGE 4: FINALIZING

- Documentation
- Training And Support
- Performance Optimization
- Deployment
- Project Presentation And Final Evaluation

Team Work (Timeline)

SEP-OCT:

- Initial Research+Literature Review
- Data Gathering
- Annotations
- Labelling
- Text Detection Techniques

NOV-DEC:

- Detailed Literature Review
- Analysis
- Training Text Extraction Models
- Improving Text Detection Techniques
- Creating FrontEnd and BackEnd (Initial Stage)

JAN-MAR:

- Completing FRONT-END and BACK-END Web Application
- Improving Text Extraction Module
- Testing And Validation
- Integration Of Improved Detection Techniques
- User Testing And Feedback

APR-MAY:

- Documentation
- Training And Support
- Performance Optimization
- Deployment
- Project Presentation And Final Evaulation

References

- TowardsDataScience-BioMedical Graphs Using NLP
- ACM.di-OCR Based Image Features
- Ncbi.Nlma-Figure Text Extraction In Literature
- Citeseerx-Improving The OCR Performance
- ScienceDirect-AI Backed OCR
- JournalsKuwait-Comparitive Study On Text Detection