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 challenges in healthcare data management and decision making. It includes:

- Improved healthcare data management.
- Prevent errors in medication administration and patient care.
- Ensure quick access to accurate patient information, especially in emergencies.
- Enable healthcare professionals to make well-informed decisions.
- Optimize resource allocation by reducing the need for manual data entry.
- Seize research opportunities in healthcare technology, including OCR.
- Enhance the handling of handwritten prescriptions.
- Leverage Knowledge Graph technology for improved medical data processing.
- Ultimately, improved patient safety and healthcare quality.

Solution

- **1. OCR Enhancement**
- **Preprocessing**: Improve scanned image quality with noise reduction, contrast adjustment, and binarization.
- **Handwriting Recognition**: Employ advanced models, fine-tuned with medical handwriting data, for accurate text transcription.
- **2. NLP-Based Enhancement**
- **Bio-NER**: Use a Bio-NER model to identify drugs, diseases, and procedures in prescriptions.
- **Bio-NEL**: Apply BERN for linking entities to a biomedical knowledge graph, enriching data.
- **BERT Text Enhancement**: Enhance readability by disambiguating text and providing definitions/synonyms.
- **Zero-Shot Relation Extraction**: Identify relationships like drug-dosage links.
- This approach combines OCR and NLP, with BERN adding specialized biomedical knowledge.
- **Biomedical Knowledge Graph & Context**
- **3. Knowledge Graph**
- **Data Collection**: Gather medical data from trusted sources (e.g., drug databases, clinical guidelines).
- **Graph Construction**: Create a structured graph using standardized vocabularies (e.g., SNOMED CT, UMLS).
- **Efficient Querying**: Enable fast data retrieval for NLP processing.
- **4. Node Embedding for Context**
- **Node Embedding**: Use techniques to understand entity significance and relationships in prescriptions.

Problem Statement

- Inefficient and error-prone process of healthcare data management, particularly related to handwritten and printed medical prescriptions

  The problem at hand is the digitization and improvement of scanned doctor prescriptions. These handwritten prescriptions are often challenging to decipher due to illegible handwriting, medical jargon, and various abbreviations. Traditional OCR (Optical Character Recognition) systems struggle with accurately extracting text from such documents, leading to errors that can have serious consequences in healthcare settings. The goal is to develop a robust system that can not only transcribe these prescriptions but also enhance their readability and context using NLP techniques and a biomedical knowledge graph.

Scope of the project

- **Global Scope of OCR in Healthcare**
  - The OCR market in healthcare is booming, with an estimated CAGR of 15.4% between 2022 and 2030. It's driven by the need for efficient data management, enhanced patient care, and digitization of medical records. Countries worldwide, including the United States, Canada, Germany, France, the United Kingdom, Russia, China, and Australia, are actively adopting OCR for healthcare, emphasizing its transformative role in data management and patient safety.

- **Recent Projects in Healthcare OCR**

- Prominent organizations and research groups, including Nano Net Technologies Inc. and Neurodata Group, have made significant contributions to healthcare OCR. In 2021, Nano Net Technologies developed an automated system for detailed healthcare data extraction. Furthermore, Neurodata Group has furthered this field by creating an automated system in 2023,

- **Scope of OCR in Healthcare in Pakistan**

- In Pakistan, OCR in healthcare is gaining recognition with active research by Seerat Rani, Abd Ur Rehman, Beenish Yousaf, Hafiz Tayyab Rauf, Emad Abouel Nasr, and Seifedine Kadry . The scope includes:

- Efficient Data Management

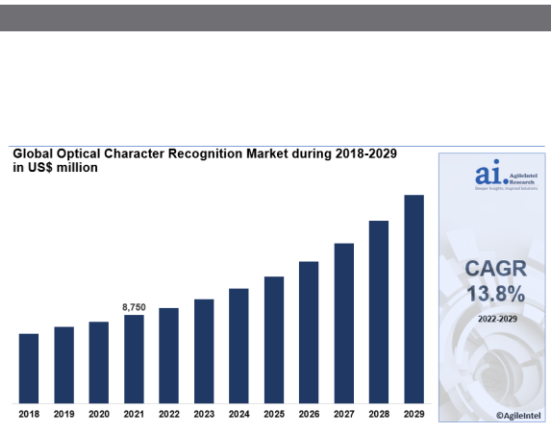
- Medication Safety

- Accessibility

- Research and Innovation

- Integration with Healthcare Systems

- In summary, OCR in healthcare is gaining traction in Pakistan, improving data management, safety, innovation, and system integration for better healthcare services.ancing detailed healthcare data extraction.



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 Evaulation

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