# Decognize: PrescriptionDigitization Using Knowledge Graphs


**Group Members:**

Muhammad SharjeelAkhtar(P20-0101)

Mahad Ashraf(P20-0563)

**Supervisor:**

**Sir.Muhammad Shoaib Khan**

# Table ofcontents

# OVERVIEW SECTION:

# 1. Problem Statement

# 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 Knowledge-Graph-based system for accurate prescription transcription

# 2. Project Objective

# Project Objective

- **Creating** a User-friendly system to digitize any prescription document.
- **Increasing** the readability of prescription for illegible-handwriting.
- **Storing** the Medicines and Doses in the form of Knowledge-Graphs .
- **Helps** the END-USER in better decision-making by the outcome.

# DIAGRAMS SECTION:

# 3.Architecture Design

# System Diagram

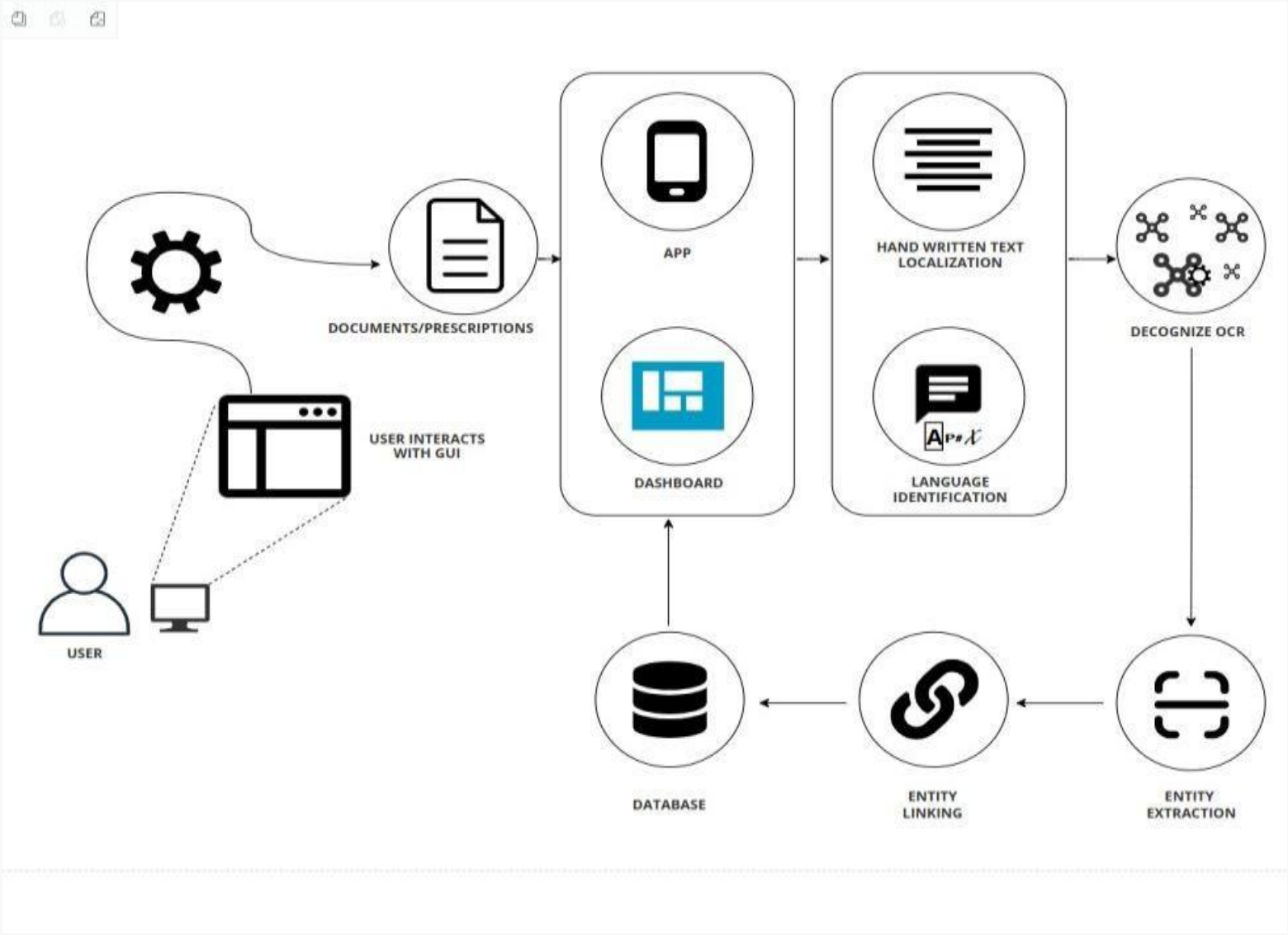



**Figure:** Architecture- Diagram of DeCognize

# 4. Activity Diagram

# ActivityDiagram

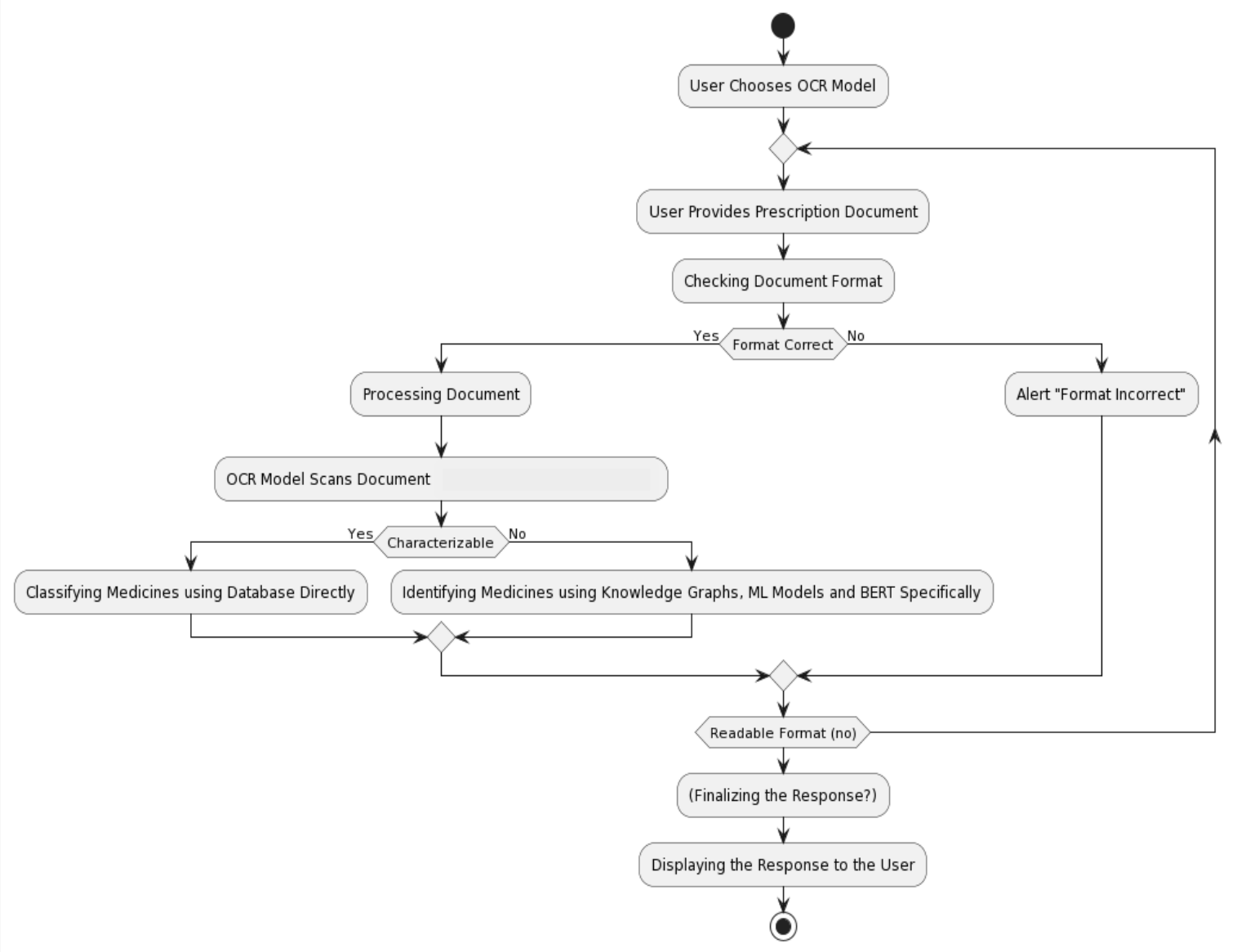



**Figure:** Activity-Diagram of DeCognize

# 5. Sequence Diagram

# Sequence Diagram

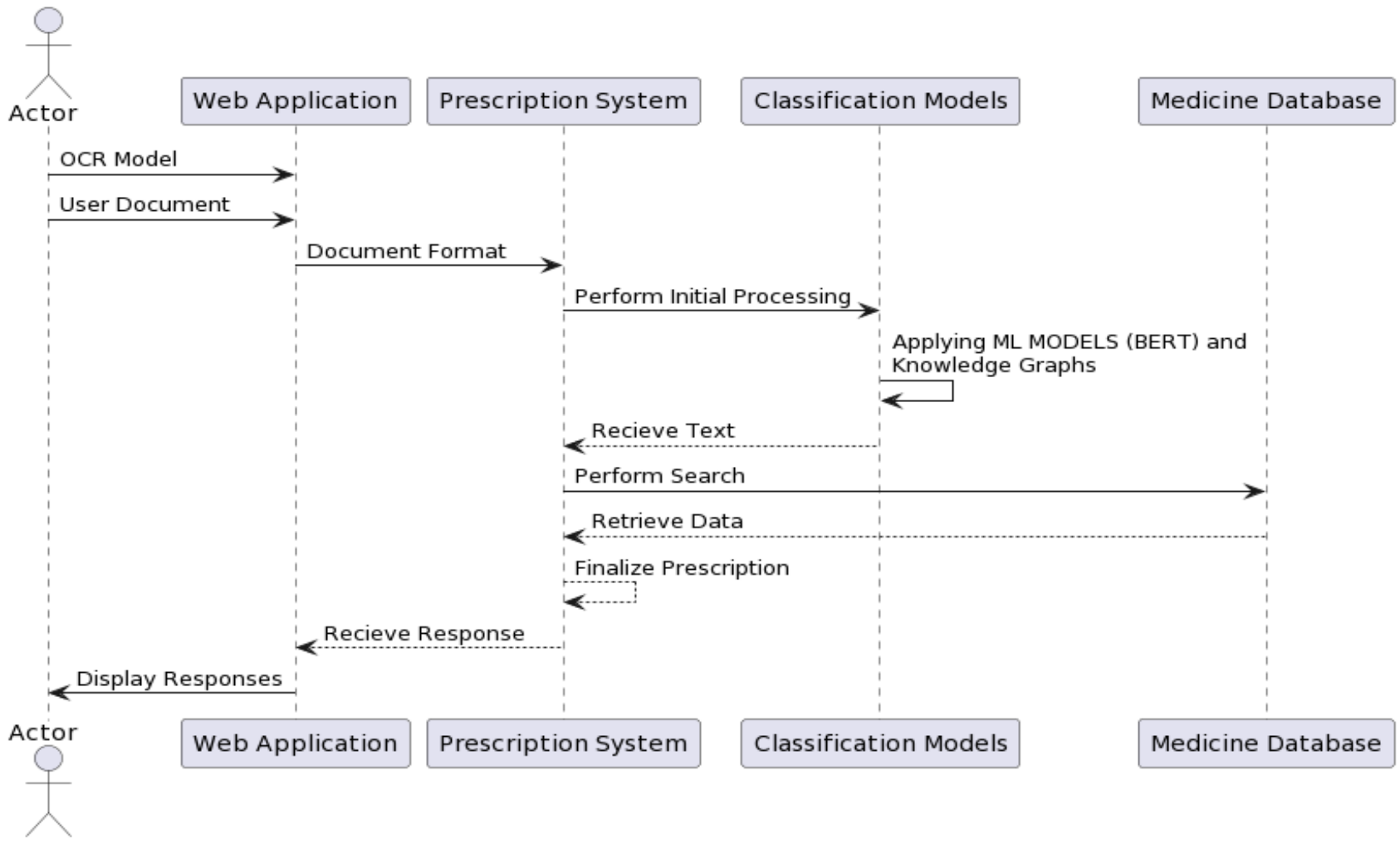



**Figure:** Swimlane- Diagram of DeCognize

# 6. Layered Diagram

# Layered Diagram

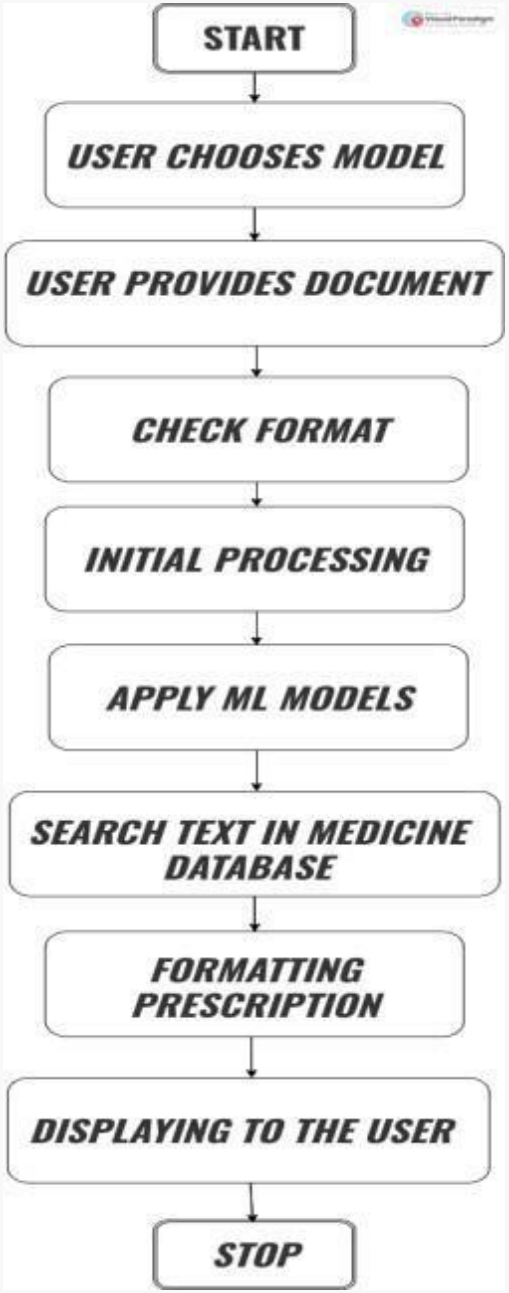



**Figure:** Layered-Diagram of DeCognize

# 7. Flow Diagram

# Flow Diagram

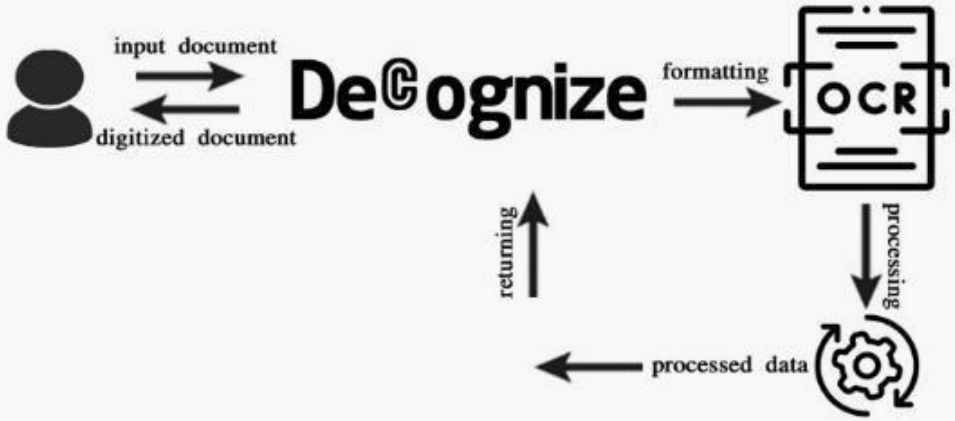



**Figure:** Flow-Diagram of DeCognize

# 8. Use Case Diagram

# Structured Use-Case-Diagram

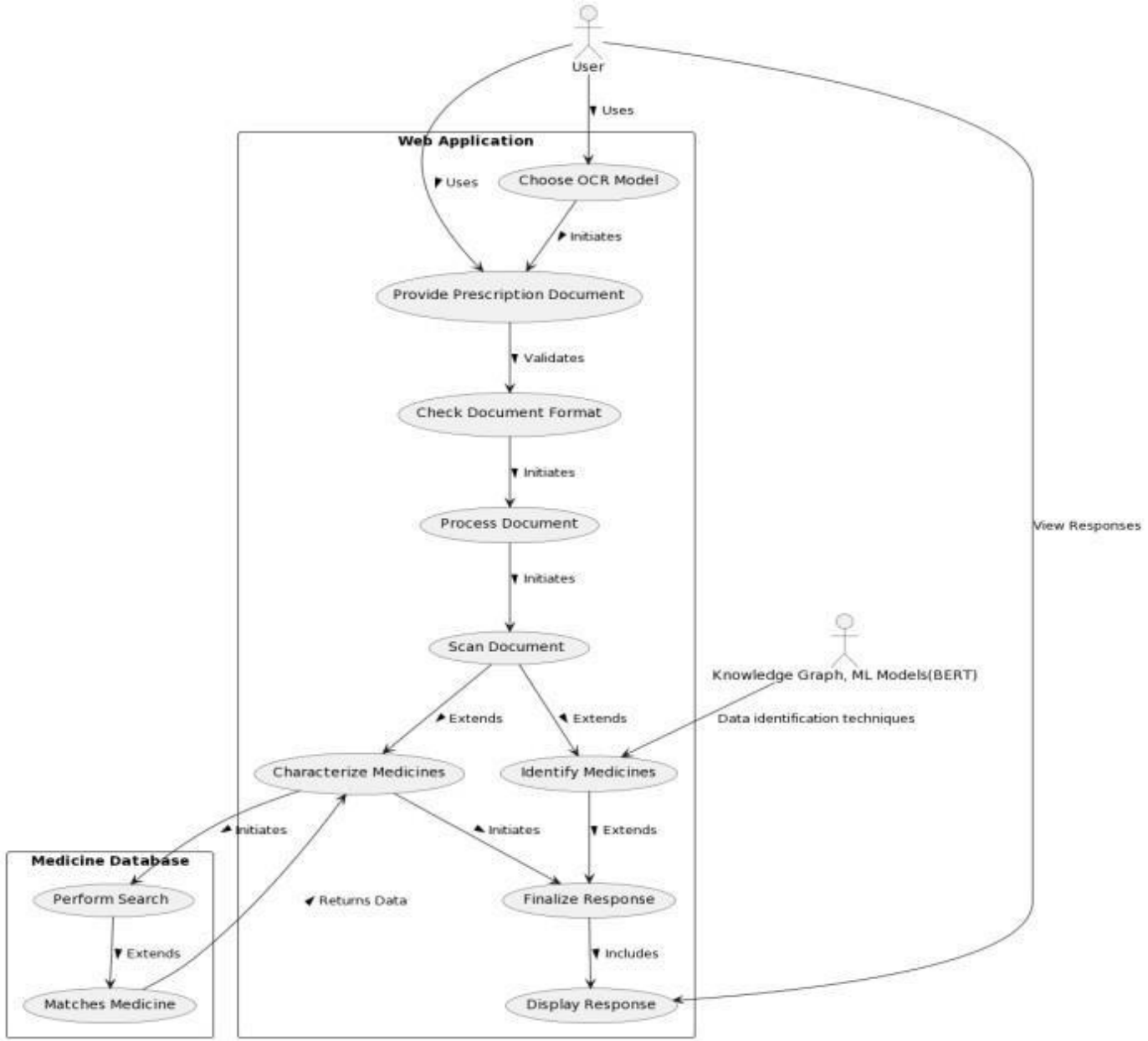



**Figure:** Structured-Use-Case-Diagram of DeCognize

# 9. State Machine Diagram

# State Machine Diagram

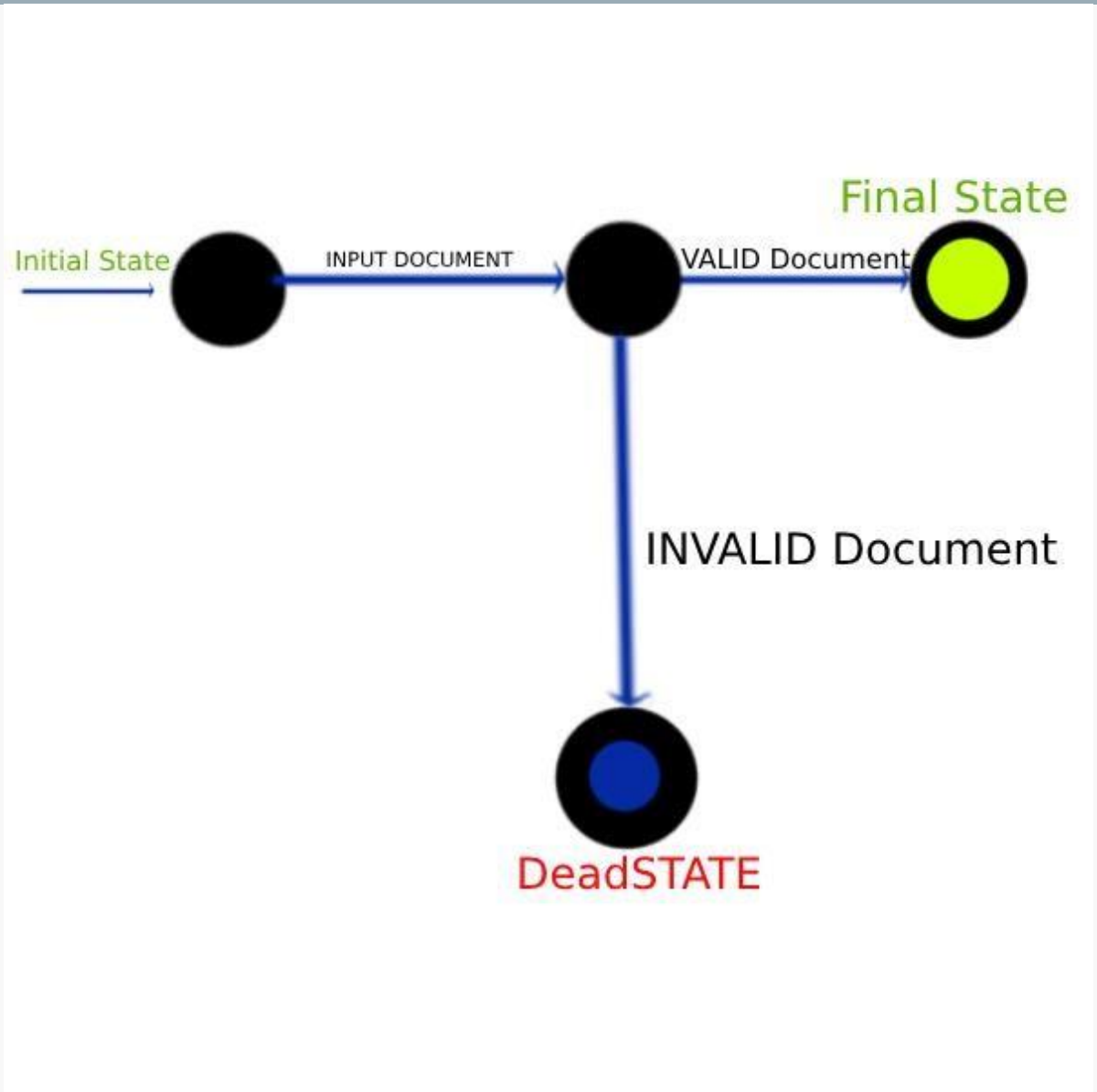



**Figure:** State-Machine-Diagram of DeCognize

# 10. Data Flow Diagram

# Data Flow Diagram

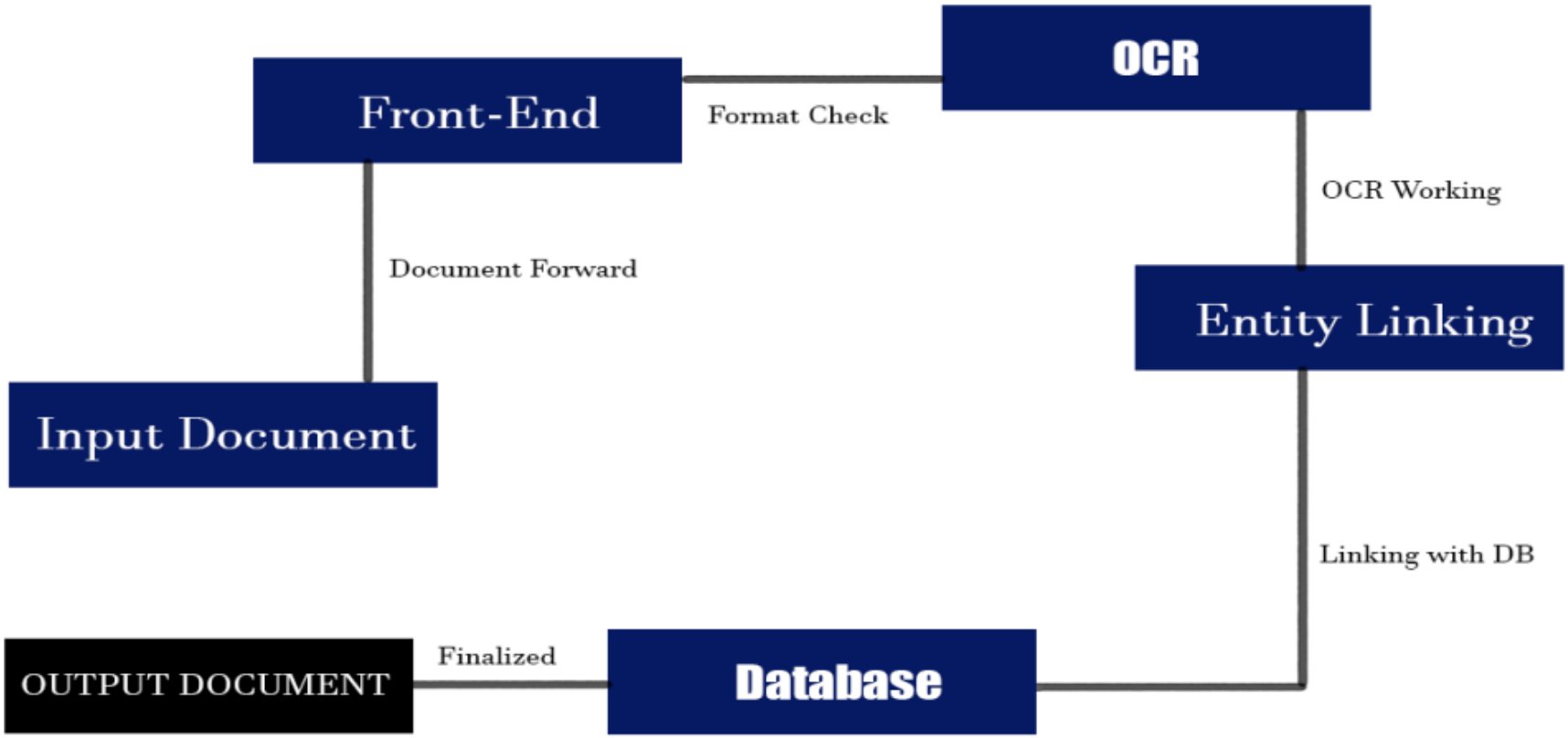



**Figure:** Data-Flow-Diagram of DeCognize

# 11. Class Diagram

# Class Diagram

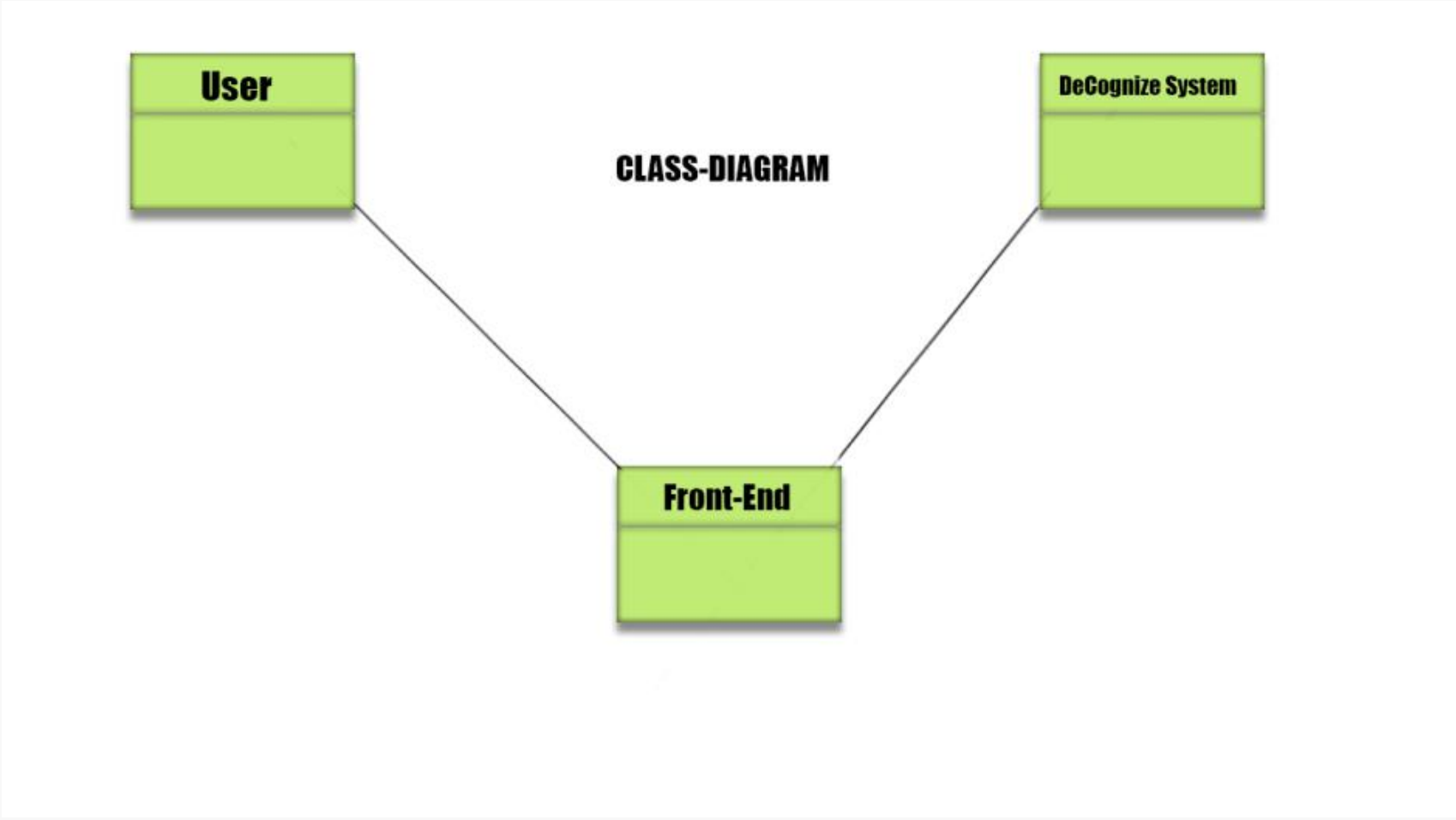



**Figure:** Class- Diagram of DeCognize

# 12. Component Diagram

# Component Diagram

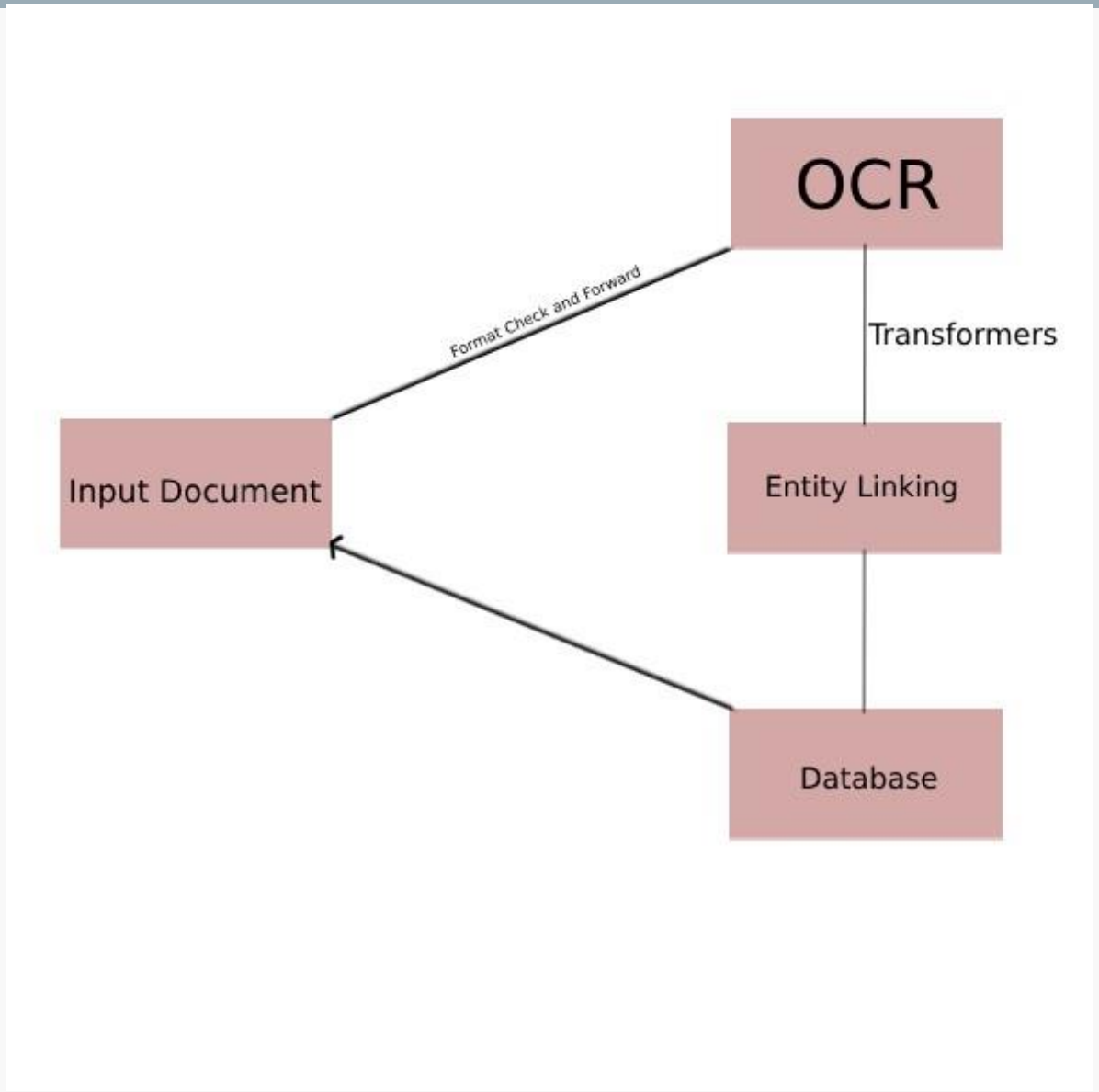



**Figure:** Component-Diagram of DeCognize

# Technologies, Scope, Poster Section:

# Technologies Used





# 13. Business Opportunity

# Business Opportunity

1. **Efficient Data Digitization:** Streamline conversion of handwritten medical notes to digital records for time savings and reduced errors.

2. **Enhanced Decision-Making:** Improve data accuracy to facilitate quicker, informed healthcare decisions, ultimately enhancing patient care.

3. **Research and Efficiency:** Utilize biomedical knowledge graphs to enable advanced research and analytics, driving innovation in the medical field.

4. **Cost Savings Automate:** Manual data entry, leading to significant operational cost reductions for healthcare institutions.

# 14. Project Scope

# Project Scope

- **Global Market Growth:** The global OCR market is experiencing rapid growth, showcasing its significance in addressing diverse industry needs. OCR market was valued at USD 8.93 billion A Compound Annual Growth Rate (CAGR) of 15.4% is anticipatedbetween 2022 and 2030

- **Recent Projects:** Automated doctor prescription by Nano Net Technologies Incand 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.

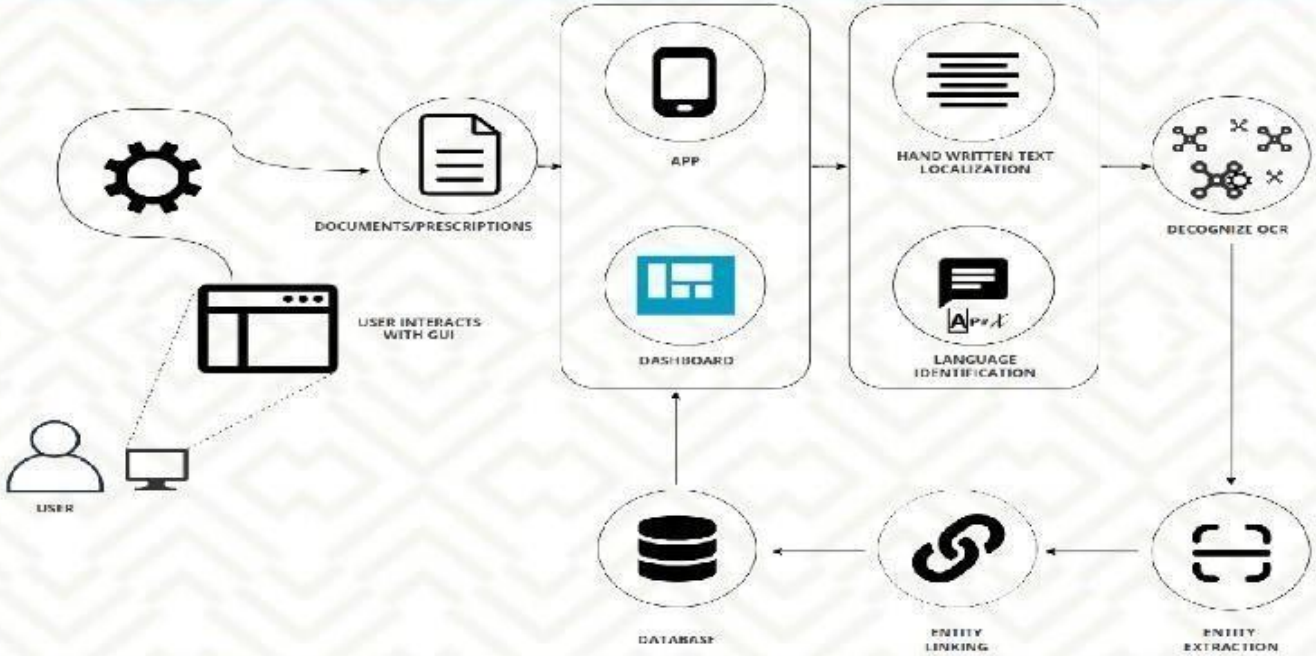
# 15. Poster

# DeCognize

Prescription Digitization
Using Knowledge Graphs

OUR METHODOLOGY INCLUDES EXTRACTING TEXT FROM DOCUMENTS AND THEN PARSING CRUCIAL ENTITIES TO ORGANIZE INFORMATION IN A STRUCTURED MANNER. THIS STRUCTURED FORMAT EMPOWERS ORGANIZATIONS TO ANALYZE THE DATA EFFECTIVELY, FACILITATING DATA-DRIVEN DECISIONS BASED ON THE DOCUMENT INFORMATION.

## ARCHITECTURE




## FLOW

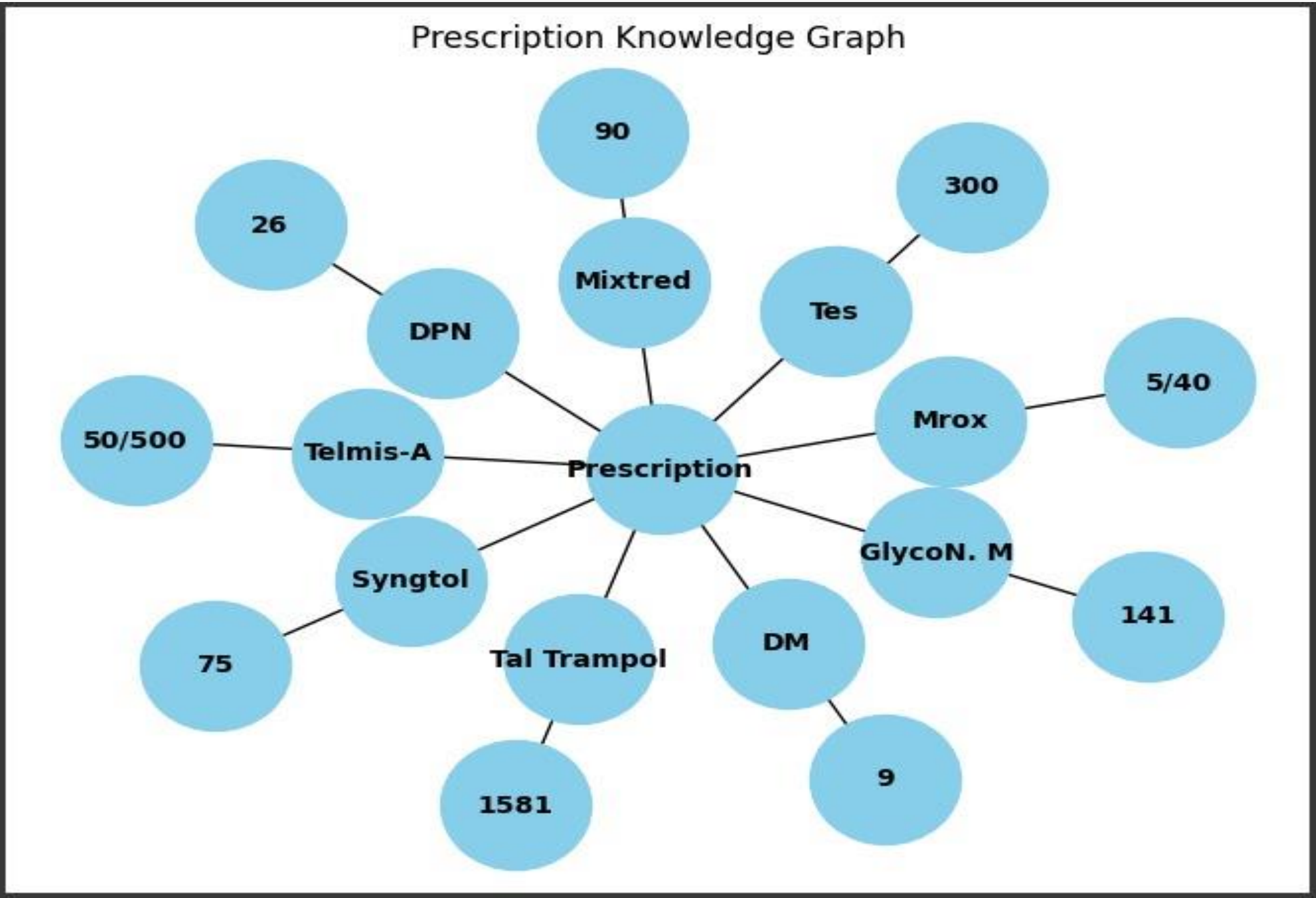

## GOAL

"DECOGNIZE" AIMS TO SIMPLIFY HEALTHCARE DATA MANAGEMENT BY USING ADVANCED TECHNOLOGIES TO STREAMLINE THE DIGITIZATION OF PRESCRIPTIONS, ENSURING ENHANCED EFFICIENCY AND IMPROVED PATIENT SAFETY.


| GROUP MEMBERS: | SUPERVISOR | TOOLS |
|---|---|---|
| MUHAMMAD SHARJEEL AKHTAR (20P-0101)<br>MAHAD ASHRAF (20P-0563) | SIR.SOHAIB MUHAMMAD KHAN | neo4j docker TensorFlow K OpenCV |

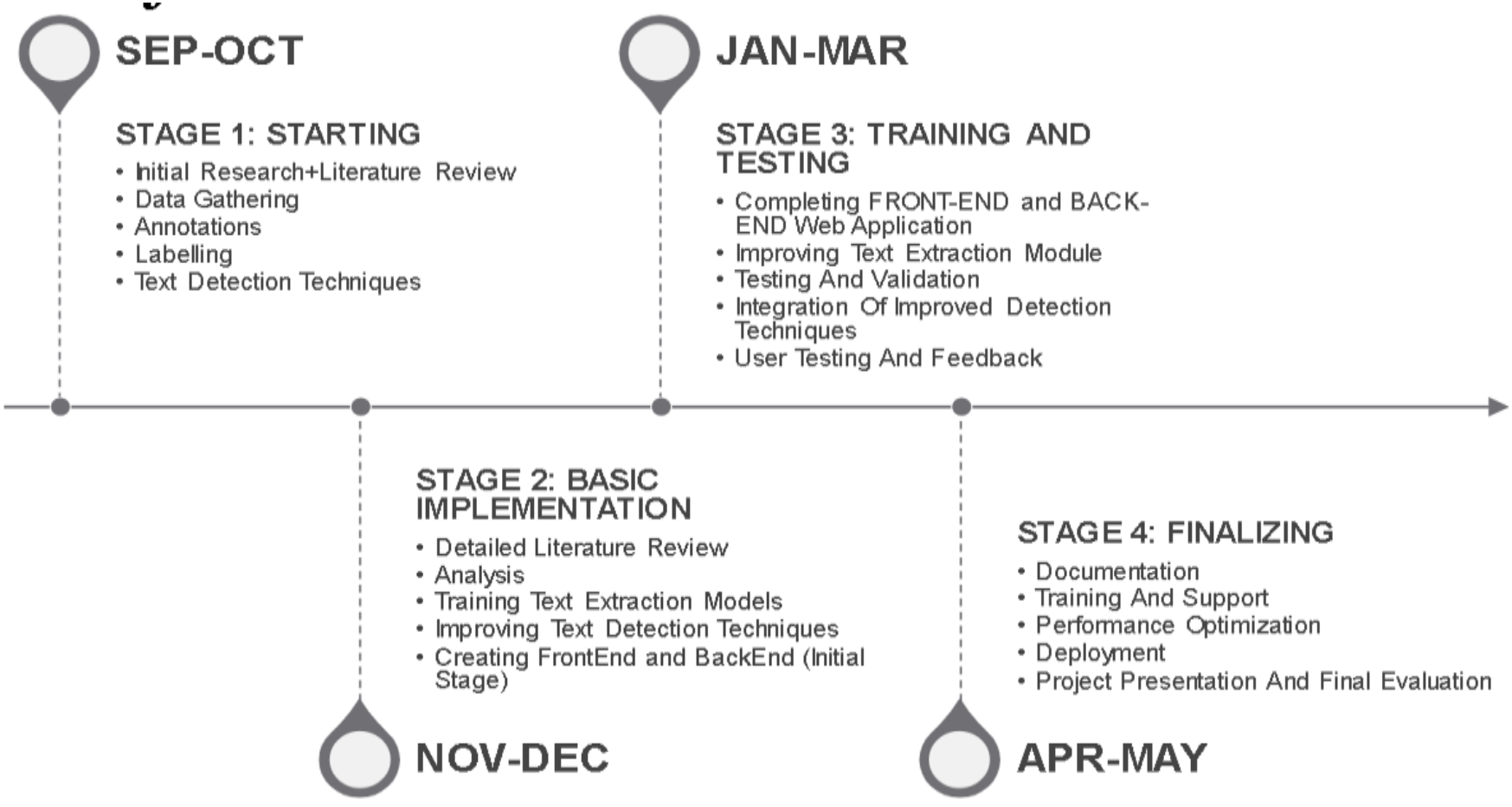
# REPRESENTATION, Timeline, Literature-Review Section:

# 16. Representation Technology


Prescription Knowledge Graph
90
300
26
Mixtred
Tes
DPN
5/40
Mrox
50/500
Telmis-A
Prescription
GlycoN. M
Syngtol
141
75
Tal Trampol
DM
9
1581

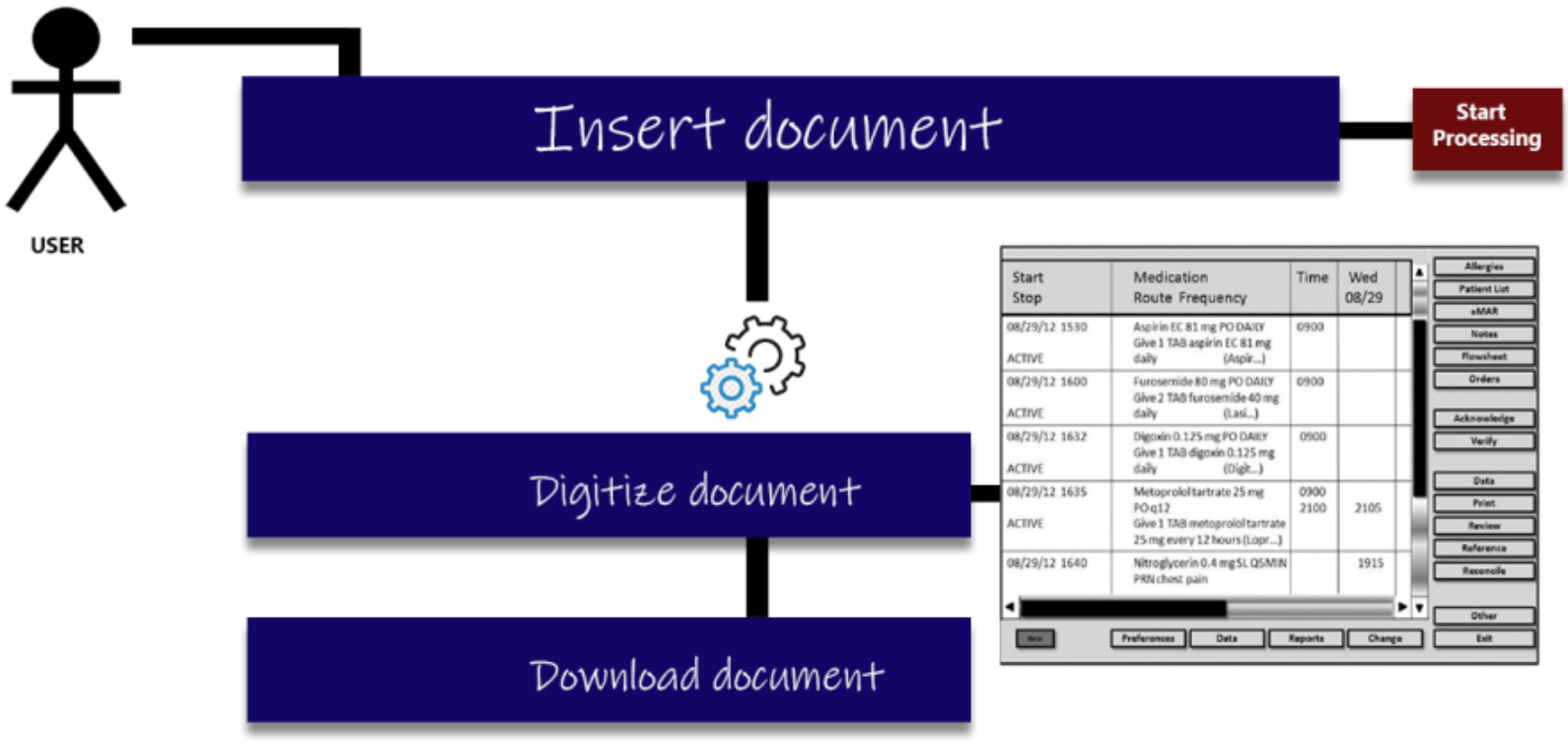
# 17. Timeline/Work-Distribution

# Timeline/Work-Distribution



# 18. User-Manual

# User-Manual via IMAGE



# 19. Literature Review

# Literature Review

| Sr. no | Year | Basic Idea | Methodologies | Results | Limitations |
|---|---|---|---|---|---|
| [1] | 2023 | OCR with Open CV and tesseract | Implemented Tesseract OCR with OpenCV in python. Focusing on image pre-processing for optimal results integrated text detection and recognition components | Achieved satisfactory OCR accuracy with well-preprocessed images. However, Tesseract struggled with complex backgrounds and artifacts, yielding suboptimal outputs. | Tesseracts accuracy is hindered by poor image quality. Requiring meticulous preprocessing. Challenges arise in handling artifacts handwriting and diverse languages. |
| [2] | 2021 | Optical Character Recognition Using TensorFlow | Implemented OCR with TensorFlow Enhanced model robustness with data augmentation technique. Implemented a custom ResNet architecture for OCR | These results showcase the effectiveness of the OCR model, particularly in accurately recognizing characters within the test set, demonstrating its robustness and suitability for the specified task. | Our Model can fail if the image is complex. E.g cursive writing images or images with Continous Characters Currently our model is trained only on digits and English language |
| [3] | 2021 | Construct a Bio Medical Knowledge Graph with NLP | Extracted text from biomedical document using OCR and applied BERN and utilized zero relation extractor. | Successfully established a Neo4j knowledge graph, showcasing versatility through demonstrated applications such as search engine, co-occurrence analysis and author expertise inspection. While emphasizing its utility for diverse biomedical machine learning applications. | Limitations include persistent NER challenges with BERN, potential inaccuracies in the zero-shot relation extractor and the need for expert validation with external database enrichment reliant on data consistency |
| [4] | 2018 | Build a Handwritten Text Recognition System using TensorFlow | Implemented HTR using TensorFlow, with NN trained on IAM word-images, including CNN, RNN and CTC layers. Preprocessed data with resizing normalization and potential augmentation. Utilized RMSProp for training and explored enhancements like data augmentation, input size adjustments and decoding strategies. | Implemented successful HTR on IAM word-images, enabling flexible NN customization and identifying areas for accuracy improvements. | Limited Diversity due to reliance on IAM dataset Potential recognition errors especially for non-dictionary words CPU based training may be slower : GPU recommended |
| [5] | 2022 | Doctor Handwritten Prescription recognition system in multilanguage using deep learning | Implemented a system employing machine learning techniques such as CNNs,RNNs,LSTMs for recognizing and translating handwritten prescription notes in diverse language | Successful recognition and translation of handwritten prescriptions in various languages. Demonstrated the efficiency of CNNs, RNNs, and LSTMS in multilingual handwritten text processing. | Sensitivity to variations in handwriting styles. Reliance on quality and diversity of training data for optimal performance |
| [6] | 2022 | A Comparison of various Machine learning Algorithms for recognizing Text on Medical prescriptions | Proposed approach involves image scanning pre processing and CNN-based feature extraction for recognizing handwritten medical prescriptions. Results are compared with drug name database using OCR for medicinal name identification | Successful implementation of CNN-based recognition for medical prescription. Need for further investigation into alternative machine learning algorithms for comprehensive comparision. | Limited Exploration of alternative machine learning algorithms<br>Identification challenges with low accuracy medical names in OCR |
| [7] | 2020 | Online Cursive Handwritten Medical Words Recognition System | Implemented an online cursive handwritten medical word recognition system using a bidirectional LSTM network. Employed data augmentation techniques to enhance recognition efficiency. | Successful Utilization of bidirectional LSTM for cursive medical word recognition<br>Recognition efficiency improvements achieved Through data augmentation | The system is restricted to providing output only for the trained data<br>Inability to generate output for the new unseen data due to lack of adaptability |
| [8] | 2021 | Medical Prescription Recognition Using Machine Learning | Developed a Medical Prescription Recognition System employing image processing techniques and machine learning algorithms to identify handwritten medicine names from prescription note images. | Successful integration of image processing and machine learning for medical prescription recognition . Acknowledged limitations include reliance on small dataset and lower accuracy | Limited dataset usage in the system<br>The system exhibits low accuracy levels |
| [9] | 2021 | Medical Prescription Identification Solution | Implemented a Medical Prescription Identification Solution employing a neural network for character recognition and knowledge-based maching for accurate results. | Utilized neural network approach and knowledge based matching for effective prescription identification | Restricted to reading only one line at a time |

# References

**1.** Filip Zelic and Anuj Sable. A review on on OCR with Tesseract OpenCV and Python. Nanonets, 2023.

**2.** Kamlesh Solanki . A review on optical character recognition using tensor flow. Medium, 10:39154–39176, 2021.

**3.** Tomaz Bratanic , D. Kim *et al*., "A Neural Named Entity Recognition and Multi-Type Normalization Tool for Biomedical Text Mining," in *IEEE Access*,Medium vol. 7, pp. 73729–73740, 2019, doi: 10.1109/ACCESS.2019.2920708. , 2021