www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 04, April 2025, pp : 2693-2696

e-ISSN : 2583-1062
Impact Factor : 7.001

INTELLIGENT PRESCRIPTION ANALYSIS SYSTEM USING NLP MODELS

Himanshu Kumar[*1], Ashish[*2], (Prof) Dr. Pankaj Aggarwal[*3]

[*1,2]Department Of Computer Science And Engineering, K.R. Mangalam University, Gurugram, Haryana, India.

[*3]Prof., Department Of Computer Science And Engineering, K.R. Mangalam University, Gurugram, Haryana, India.

DOI: https://www.doi.org/10.58257/IJPREMS40198

ABSTRACT

To effectively analyze research papers and provide precise answers to user inquiries, we present a novel natural language processing approach. We can allow cutting-edge frameworks for document preparation, analysis, and question answering with this architecture. Text summarization systems, named entity recognition systems, topic modeling, generative language models, organization gain task generation text classification frameworks, and many more are examples of methodologies. The model's ability to employ complex doctor's prescription queries to extract essential terms, identify the type of question posed, and generate useful responses demonstrates its promise for knowledge discovery.

Keywords: Text Summarization, Prescription, Topic Modeling, NLP Model.

1. INTRODUCTION

In the era of information overloading, efficient document analysis and question answering systems have become indispensable tools. This paper delves into the development of a comprehensive NLP model capable of analyzing handwritten prescription and providing accurate informative answers to queries. The model aims to streamline the process of extraction of information, comprehension and retrieval. The NLP model will function similarly to how a user would read, analyse, and respond to questions posed by a prescription and it will respond according to the behaviour of the illness and give answer to the queries accordingly. In essence, the model will cut down on the time and effort needed to read those handwritten prescription which may lead to incorrect treatment of the patient to get carefully considered responses to the queries that raise doubts.

A. The need for text preprocessing in NLP model

The text pre-processing is a important and significant step in processing the model which gives key knowledge about each and every part of the text given in a dataset or created by the model based on asked questions by the user. They are different ways to preprocess text which includes tokenization, stop-word removal, stemming etc. in which the key aspect is to create a NLP pipeline which includes all the important key elements from creating to training the model and training with billions of parameter in which regress training and accuracy is needed for both the pre-trained model and self-reliant models. They are different ways for text -preprocessing which includes stop-word removal, tokenization and stemming.

B. Why are stop words removed ?

The main motive of stop word removal is to reduce the dimension of text data and improve accuracy and efficiency of the model.

1. Reduce Noise : Stop words does not add any insightful information about analysis so by removing them the model can focus on more relevant and meaningful words reducing unnecessary complexity and computational load.

2. Efficiency : Extracting stop word reduces the number of tokens used in the model and number of tokens the distinct words which needs to process which leads to faster computation it makes easier to work with large datasets specially in NLP models.

www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 04, April 2025, pp : 2693-2696

e-ISSN : 2583-1062
Impact Factor : 7.001

2. OBJECTIVES OF THE STUDY

A. Problem Statement

Analysing the text written in medical prescription like different handwriting styles and complex medical terminology. Many times the pharmacists misinterpret the prescribed drug leading to wrong information under various brand names. Users became anxious and are unable to understand the prescriptions.

3. DATA DESCRIPTION

A. For this study, we used a dataset that we obtained from Kaggle. The dataset's link is provided below:

https://www.kaggle.com/datasets/mehaksingal/illegible-medical-prescription-images-dataset/data.

The dataset contains 129 images which contains different medical prescriptions written by different doctors in different style and format

Each image is uploaded into Google's API and JSON database is extracted from each image and then a separate database is trained for the model.

Fig. 1 Architecture of the model

B. Methodologies

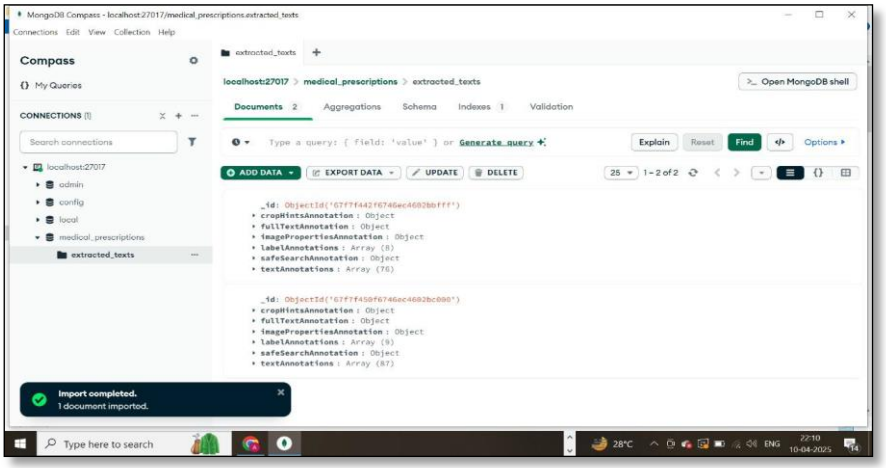
In this paper, we have used three api's google free vision api for creating Json dataset extracting text from image and annotations and after that created a dataset and ocr-based model for extracting text from image and annotations which is also running in backend. MongoDB is also used in backend for extracting text in the backend where json file is uploaded and text is extracted.

Step 1. Choosing a dataset

The dataset we have chosen is from Kaggle which contains 129 images of different medical prescription.

Step 2. Text Preprocessing

In data preprocessing, firstly we try to extract the text from the images using python's pytesseract, ocr based technology which extracts text and after that the text was pre-processed and trained on the given dataset.



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
www.ijprems.com
editor@ijprems.com
Vol. 05, Issue 04, April 2025, pp : 2693-2696
e-ISSN : 2583-1062
Impact Factor : 7.001

Step 3. Image annotation

The text is extracted from image and inspite of doing manual annotations we have used google vision api which auto-generates Json files that contains the extracted the text from the image.

Step 4. Backend using MongoDB and Fast Api

MongoDB is fast and easy to accessible for the image to text extraction because it stores the Json dataset which comprises of both structured and semi-structured and this kind of extracted data is stored easily without any changes in the dataset.

Fast Api is helpful in image to text extraction because it allows a dynamic structure for the json data to be used for training and validation inputs which gives a better understanding of text extracted

It creates an end-points which is quite helpful in processing the text from the image which is running in the backend.

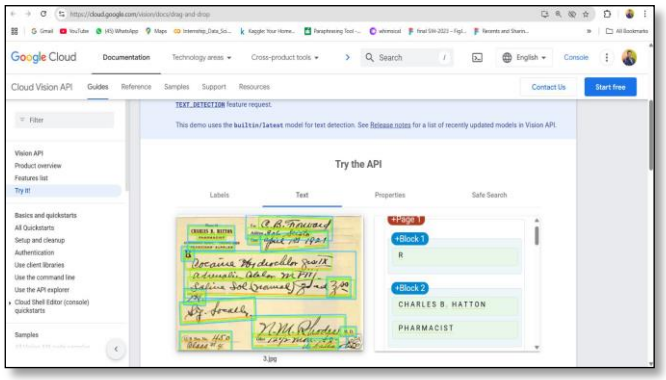


Fig. 2 Text extracting using Google Vision Api

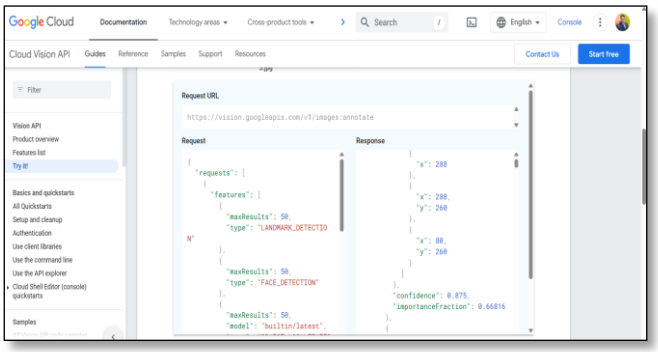


Fig. 3 Creating annotations and Json dataset

4. RESULT

The model is trained on system prompt to behave like a doctor so that it can answer queries to the prescription according to user's needs and requirements and the model is given knowledge and it is trained like that to extract the text from the given document in any format and the data is extracted in the form of json files in the form of key-value pairs which will answer and give suggestions accordingly.

Fig.4 Extracting text in backend using MongoDB

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
www.ijprems.com
editor@ijprems.com
Vol. 05, Issue 04, April 2025, pp : 2693-2696
e-ISSN : 2583-1062
Impact Factor : 7.001

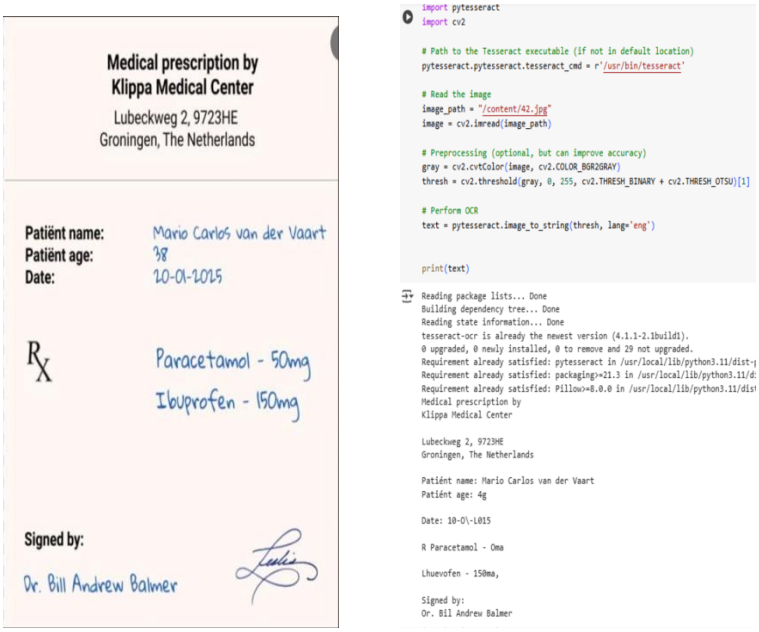


Fig 4. Extracting text in backend using OCR model

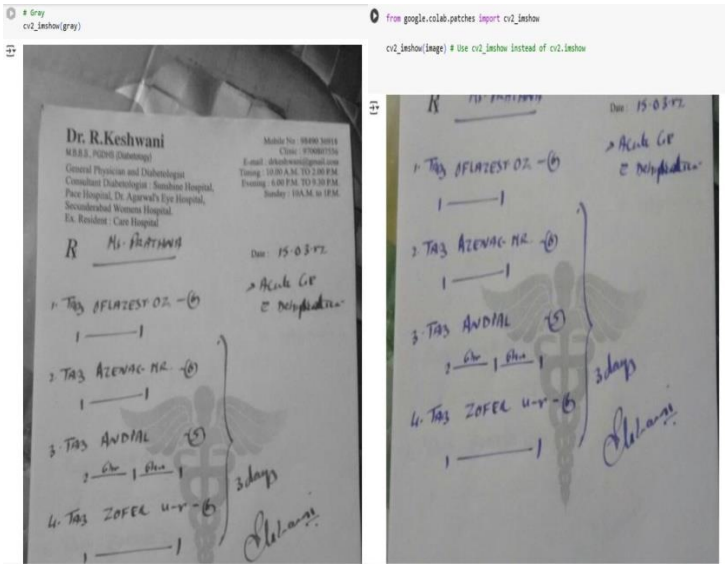


Fig 5. Taking different formats of prescription for extracting text

5. CONCLUSION

This prescription analysis system will be an advancement in the text to image extraction as it will help to understand the names of the medicines and medical terminologies like RxNorm which will help users in understanding the medical terms and use the medicine of same brand as prescribed by the doctor. It will reduce human-error and improve patient safety.

6. REFERENCES

[1] https://www.kaggle.com/datasets/mehaksingal/illegible-medical-prescription-images-dataset/data.

[2] https://whimsical.com/engineering-home-DywDqtWujh6AVhN8wHo45o