MULTI-MODAL AI FOR INTELLIGENT CONTENT SUMMARIZATION: IMAGES AND TEXT

Khushboo Bhoolwani
Dept. of Computer Science and Engineering,
Terna Engineering College,
Nerul, Navi Mumbai, India
bhoolwanikhushboo@gmail.com

Anay Bhave
Dept. of Computer Science and Engineering,
Terna Engineering College,
Nerul, Navi Mumbai, India
anaybhave@gmail.com

Sahil Saini
Dept. of Computer Science and Engineering,
Terna Engineering College,
Nerul, Navi Mumbai, India
saini.sahil6766@gmail.com

Prof. Umesh B. Mantale
Dept. of Computer Science and Engineering,
Terna Engineering College,
Nerul, Navi Mumbai, India
umeshmantale@ternaengg.ac.in

***Abstract*— This project presents a unified system that harnesses the power of artificial intelligence to address two critical tasks: text summarization and pill recognition. The text summarization component allows users to input large blocks of text and receive concise and coherent summaries, aiding in information digestion. The pills recognition module accepts images of pills as input, identifying them and providing relevant information, such as their names and uses, enhancing medication management and safety.**

I. INTRODUCTION

The project, "Multi-Modal AI for Intelligent Content Summarization: Images (Pills) and Texts," aims to provide users with a versatile platform for both text summarization and pill summarization. In today's information-rich world, users often face information overload, and our project addresses this issue by offering efficient content summarization solutions.
By combining text summarization and image-based pill summarization, we provide users with a comprehensive toolset for managing and extracting valuable insights from textual and visual information.

In the era of information explosion, the need for comprehensive content summarization has become increasingly crucial. This research addresses the challenges of traditional summarization methods by integrating advancements in both text and image modalities.

Identifying pills based on visual cues can be error-prone and time-consuming, which poses potential health risks, especially for individuals with multiple medications. Users struggle with information overload, making it challenging to extract essential insights from lengthy texts, leading to time inefficiency and cognitive strain. Our project addresses these problems by offering a user-friendly interface for text summarization and pill recognition, improving information handling and medication management processes.

The primary objectives include developing a robust text summarization model, pioneering intelligent pill recognition, and innovatively integrating both modalities into a hybrid model.

The research holds significance in healthcare for improving medication adherence and in content management for transforming how information is comprehended.

The scope of this project encompasses the development of a versatile and user-friendly MultiModal AI system that offers intelligent content summarization for textual data and

recognition for images of pills. The project aims to address information overload by providing concise text summaries while enhancing medication management by identifying pills from images and providing relevant information. Key objectives include developing AI models, creating an intuitive user interface, ensuring data privacy and security, and promoting ethical AI practices, all while offering scalability and user support.

The study is imperative in response to the evolving landscape of information processing, where the integration of text summarization and pill recognition addresses critical gaps. Existing methods often specialize in one modality, hindering a comprehensive understanding of diverse content types. The fusion of text and image modalities contributes to a more nuanced approach to content summarization and holds significant implications for healthcare, particularly in medication management, where accurate pill recognition is paramount. The study responds to the pressing need for holistic content understanding in domains ranging from healthcare to educational materials. Additionally, it contributes to the emerging field of AI applications, providing insights into the integration of diverse information sources and promising enhanced user experiences through context-aware summaries. Furthermore, the potential cross-domain applications of the proposed models underscore the study's relevance in addressing the challenges of modern information processing and advancing the capabilities of intelligent summarization models.

II. LITERATURE REVIEW

The Paper [1] proposes a trainable text summarization framework using diverse features, comparing its performance to non-trainable baselines with Naive Bayes and C4.5 decision tree classifiers. Naive Bayes notably surpassed the baselines, emphasizing the classifier's influence on the trainable summarizer's effectiveness. Future work aims to design a tailored classification algorithm for text summarization. [2] The proposal presents a deep learning system for visually impaired patients, utilizing image analysis to classify pills. It includes a mobile app and wearable smart glasses, aiming to enhance medication safety and reduce issues for chronic visually impaired patients. Swapnil Acharya at [3] created a web-scraped Wikipedia text summarization system with preprocessing and feature extraction. LSA excelled over LDA, and future work involves RNN-based topic modeling for improved summaries. The paper [4] uses a database using SQLite store medicine names and images. Preprocessing sharpens images with an unsharp filter using OpenCV. The SIFT algorithm detects local features in the image and matches them with the database. If a match is found, the medicine name is converted to audio via Google Text to Speech for blind identification.

III. METHODOLOGY

The methodology employed in this research integrates rigorous techniques from both text summarization and pill recognition domains to achieve a comprehensive multi-modal summarization framework. The study adheres to ethical guidelines, ensuring the responsible handling of data and the privacy of participants, where applicable.

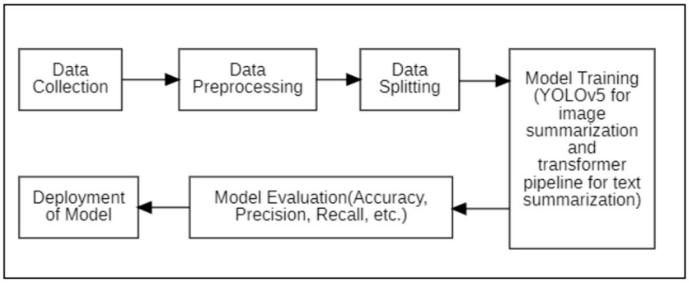

Fig 1 Architecture of Proposed System

Phase 1: Data Collection and Splitting
Data Collection: We have collected a dataset from Robolflow for the Pill Summarization Model.

➢ Characteristics:

Total images: 2202
Training set images: 1353
Validation set images: 575
Testing set images: 274
Data Splitting: Splitting dataset of images into training, testing and validation.

Phase 2: Model Development
Text Summarization Model: Implement a Transformer pipeline for the text summarization
model.
Pills Recognition Model: Training model to identify pills in images. Using Pytorchs
framework YOLOv5, Identifying the pills and providing its information from Wikipedia.

Phase 3: Integration and Cross-Modal Functionality

Cross-Modal Integration: Integrate both the text summarization and pills recognition modules
into the Multi-Modal AI system.
Allow users to input text summaries related to recognized pills or vice versa for cross-modal
content analysis.

Phase 4: User Interface and Back-End Development
User Interface Design: Create an intuitive and user-friendly web-based interface for user
interaction.
The system built is a user-friendly web interface that enables seamless interaction with the system. This interface is developed using the flask framework, providing a robust foundation for web application building. The web interface design follows a minimalist approach, ensuring a simple user experience.
Back-End Development: Build the back-end infrastructure, including the web server and AI model server. Implement the logic for processing user input, invoking the AI models, and managing user data.

Phase 5: Testing, Evaluation, and Optimization
Testing and Evaluation: Evaluate the system's performance and accuracy using metrics.
Collect user feedback on the quality of text summaries and recognized pill information.
Optimization: Continuously optimize the AI models for better accuracy and efficiency based
on user feedback and usage data.

Phase 6: Deployment
Deployment: Deploy the Multi-Modal AI system to a production-ready environment, ensuring high availability and scalability.

IV. IMPLEMENTATION AND RESULT

The result of this project will be a versatile and user-friendly Multi-Modal AI system that empowers users with two key functionalities: intelligent content summarization for textual data and pills recognition from images. Users will be able to obtain concise text summaries for efficient information consumption and access detailed information about pills by simply uploading images. This system will enhance information management, facilitate medication identification, and contribute to improved user experiences, making it a helpful tool for widespread of applications in healthcare, education, and content management.

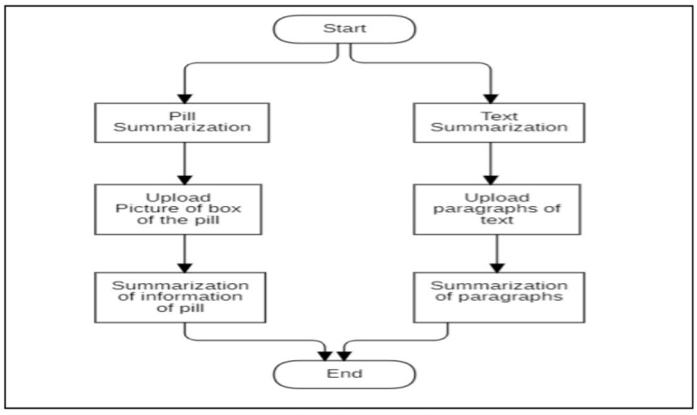

Fig 2. Flowchart of webApp

The result of the summarization experiment clearly demonstrates that our proposed system outperforms the state-of-the-art baselines in precision and F1-Score. Figs 3 and 4 show the results.

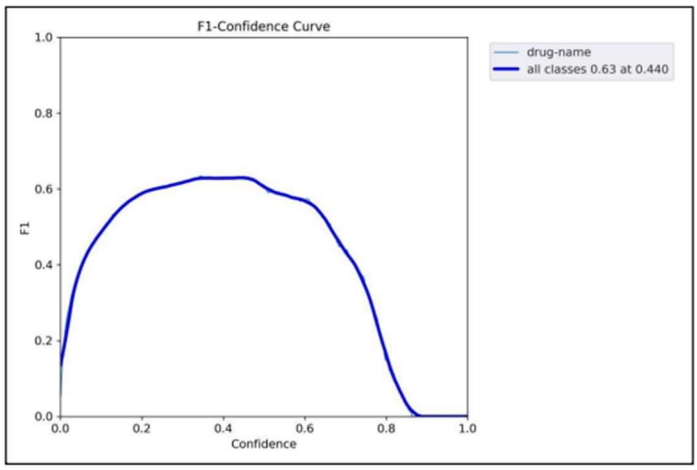

Fig 3 F1-Score

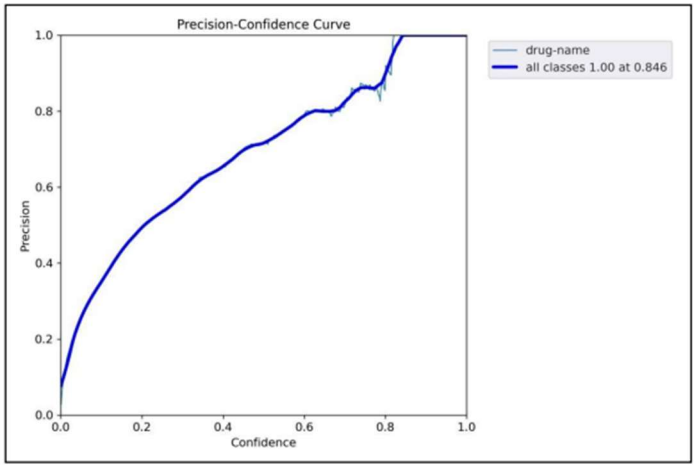

Fig 4 Precision Curve

This model was deployed on a Progressive Web App with a user-friendly interface . The Screenshots of which are given below:

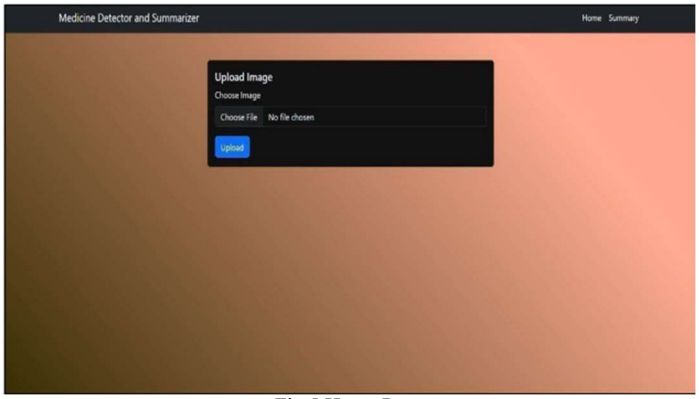

Fig 5 Home Page

Fig 6 Result of Pill Summarizer

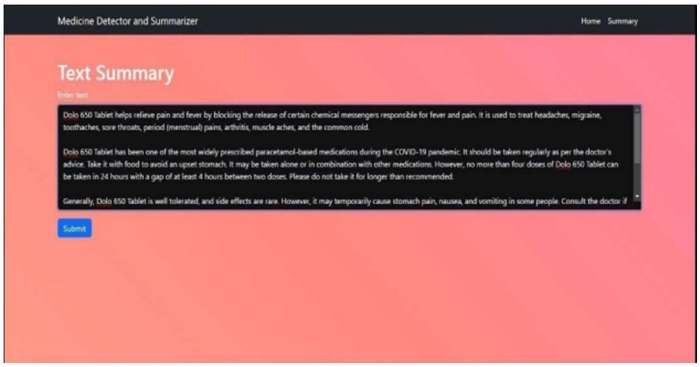

Fig 7 Text Summarization Model

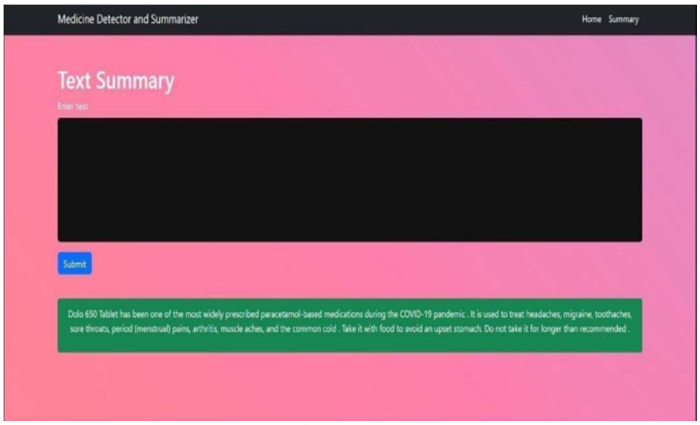

Fig 8 Result of Text Summarizer

The above screenshots show to working of our model by providing input to pill summarization model and text Summarization model respectively.

V. CONCLUSION

In conclusion, the development of the Multi-Modal AI system for intelligent content summarization and pills summarization represents a significant leap forward in user-centric technology. By seamlessly integrating natural language processing and image recognition capabilities, this project addresses the challenges of information overload and medication management. The system's user friendly interface ensure a holistic solution. This endeavor not only transforms how users interact with information but also demonstrates the potential of AI in enhancing everyday tasks. In the rapidly evolving landscape of technology, this project stands as a testinomy to innovation, usability, and ethical AI implementation.

REFERNCES

1] Naoto Usuyama, Natalia Larios Delgado, Amanda K. Hall Microsoft healthcare "ePillID Dataset: A Low-Shot Fine-Grained Benchmark for Pill Identification"arxiv:2005:14288v1 [cs.CV] 28 May 2020

2] Sudarshan Borude, Sakshi Patil, Roshni Bhoirkar, Isheeta Shirsat "Drug Pill Recognition System Using Deep Learning" ISO 9001:2008 Certified Journal ,Volume: 09 Issue: 11 | Nov 2022

3] R Shashidhar1□, V Sahana2, Sudeshna Chakraborty3, S B Puneeth4, M Roopa5Recognition of Tablet using Blister Strip forVisually Impaired using SIFT Algorithm, Indian Journal of Science and Technology 2021;14(23):1953–1960

4] Mudasir Mohd, Nowsheena, Mohsin Altaf Wani, Hilal Ahmad Khanday, Umar Bashir Mir, Sheikh Nasrullah, Zahid Maqbool, Abid Hussain Wani, "Semantic-Summarizer: Semantics-based text summarizer for English language text",Software Impacts,Volume 18,2023,100582,ISSN 2665-9638

5] K Veningston, P V Venkateswara Rao, M Ronalda, "Personalized Multi-document Text Summarization using Deep Learning Techniques",Procedia Computer Science,Volume 218,2023,Pages 1220-1228,ISSN 1877-0509

6] Asad Abdi, Shafaatunnur Hasan, Siti Mariyam Shamsuddin, Norisma Idris, Jalil Piran,"A hybrid deep learning architecture for opinion-oriented multi-document summarization based on multi-feature fusion,Knowledge-Based Systems",Volume 213,2021,106658,ISSN 0950-7051

7] Mudasir Mohd, Nowsheena, Mohsin Altaf Wani, Hilal Ahmad Khanday, Umar Bashir Mir, Sheikh Nasrullah, Zahid Maqbool, Abid Hussain Wani, Semantic-Summarizer: Semantics-based text summarizer for English language text, Software Impacts, Volume 18, 2023, 100582, ISSN 2665-9638, https://doi.org/10.1016/j.simpa.2023.100582. (https://www.sciencedirect.com/science/article/pii/S2665963823001197)

8] Shashidhar R, Sahana V, Chakraborty S, Puneeth SB, Roopa M. (2021) Recognition of Tablet using Blister Strip for Visually Impaired using SIFT Algorithm. *Indian Journal of Science and Technology*. 14(23): 1953-1960

9] Mudasir Mohd, Mohsin Altaf Wani, Hilal Ahmad Khanday, Umar Bashir Mir, Sheikh Nasrullah, Zahid Maqbool, Abid Hussain Wani, *" Semantic-Summarizer: Semantics-based text summarizer for English language text,"* Software Impacts. Elsevier, ISSN : 2665-9638. IF: 2.1.

10] P.A. CUNDALL,
Computer Simulations of Dense Sphere Assemblies,
Editor(s): Masao Satake, James T. Jenkins,Studies in Applied Mechanics,Elsevier,Volume 20,1988,Pages 113-123,ISSN 0922-5382,ISBN 9780444705235,
https://doi.org/10.1016/B978-0-444-70523-5.50021-7.
(https://www.sciencedirect.com/science/article/pii/B9780444705235500217)

11] K Veningston, P V Venkateswara Rao, M Ronalda, Personalized Multi-document Text Summarization using Deep Learning Techniques, Procedia Computer Science, Volume 218, 2023, Pages 1220-1228, ISSN 1877-0509,
https://doi.org/10.1016/j.procs.2023.01.100.
(https://www.sciencedirect.com/science/article/pii/S187705092300100X)