MediSynthAI: A Multimodal AI Framework Using OCR, NLP, and CNN for Medical Report Interpretation, Medication Verification, and Patient Routine Management

[1]Lakshmipriya S, [2]Yeshmitha Sri V, [3]Sri Vani S, [4]Keerthana R, [5]Faahima K
[1]Assistant Professor, [2,3,4,5]Student
[1]Department of Artificial Intelligence and Data Science,
[1]Sri Manakula Vinayagar Engineering College, Puducherry, India

Abstract : Prescriptions being illegible and patients taking the wrong medications represent only two examples of how medication errors occur while providing care to patients. The purpose of this study is to develop an AI-based application to merge the process of checking prescriptions, identifying pills and creating reminders into one compact group, called MediSynthAI. Using Optical Character Recognition (OCR) and Natural Language Processing (NLP), the system will parse through the information related to a prescription (e.g. name, dosage, frequency) from handwritten or printed prescriptions, whilst at the same time check the uploaded photos of pharmacy bottles for the names of the medications located on the bottle or box in relation to the prescription history. Based on this information, the system will automatically create reminders that include both the name of the medication along with a picture of the particular pill to avoid confusion when it comes time to provide it to the patient. This medication adherence tracking capability keeps track of doses missed or taken by the patient and will generate a complete report on adherence. Furthermore, this application will be designed to work offline thus allowing patients who have limited Internet access the opportunity to utilize the application. Therefore, by creating this comprehensive and integrated system for medication safety and compliance, MediSynthAI intends to reduce the likelihood of medication errors and increase medication safety for patients.

***IndexTerms* - OCR, NLP, CNN, Transformer based OCR (TrOCR), Patient Reminder System.**

I. INTRODUCTION

Lack of adherence amongst patients regarding prescribed medications contributes substantially to adverse health outcomes as well as unnecessary hospitalizations. Approximately 50% of patients do not follow the directions provided on an Rx label and this is primarily attributable to misunderstanding or misinterpretation of the information on the Rx label, according to the World Health Organization (WHO). Unreadable prescriptions lead to further misunderstandings and increase the risk of medication errors. Prior research has explored these issues separately. Yaniv et al. introduced the Pill Image Recognition Challenge to identify the significance of identifying pills based on their shape and imprint, while Heo et al. developed a CNN based methodology for automatically identifying pills, achieving an accuracy level of over 95%, as discussed in their paper. Ali et al. developed a method to automatically extract medication details from handwritten prescriptions using multi-head-attention mechanisms. None of the above examples integrate prescription parsing, pill verification, and user-created reminders as part of one integrated system. MediSynthAI is designed to fill this gap in research by bringing together OCR-processed prescription interpretation, CNN-driven pill verification, and intelligent automatic reminders into a unified workflow. Unlike previous methodologies, the MediSynthAI framework is capable of providing patients with visual reminders displaying images of their prescribed pills with dosage information, which will ultimately improve patient compliance.

II. LITERATURE REVIEW

Yaniv et al. [1] developed the National Library of Medicine Pill Challenge dataset, which provides pill retrieval features like color, shape and imprint. Following that, Delgado et al. [5] published a paper on efficient methods to retrieve pills within real life scenarios. Madsen and Payne [6] presented automatic pill recognition from pillbox images, addressing robustness issues. Deep learning has been exploited with recent developments. Heo et al. [2] designed a pill recognition system with high accuracy on constrained datasets using CNN. Nguyen et al. [7] developed a multimodal fusion model based on Graph Neural Network (GNN) for multi-pill detection with high accuracy. Ashraf et al. [8] showed code-free deep learning for pill recognition that can be done without expertise. Analysis of prescriptions has also been a topic of interest. Ali et al. [3] used multi-head attention for handwritten prescriptions to enhance text extraction accuracy. Nguyen et al. [9] introduced PIKA, which is a prescription knowledge-driven model that integrates text and visual features. These works do not make use of the patient-facing reminder system, which is addressed by our proposed system.

III. LITERATURE REVIEW

3.1Preprocessing Steps

To enhance OCR accuracy, every prescription image will undergo preprocessing. All input is converted to grayscale to decrease the effect of color noise. Their images are denoised using gaussian and median filters, which smooth the image while maintaining many of their important edges. Adaptive thresholding is applied for binarization purposes to help improve the segmentation of text in conditions of uneven illumination. Morphological operations, such as dilation and erosion, will separate characters that may have merged due to imperfections in handwriting. Final steps include resizing and normalizing the data, so they are the same format when processed through the OCR model. All these steps significantly enhance the quality of prescription images and improve OCR accuracy.

3.2Models Used in Each Module

3.2.1 Prescription Text Extraction Module

The prescription text extraction module is in charge of reading printed and handwritten prescriptions. Handwritten samples present a challenge for OCR (TrOCR), and printed samples will use Tesseract OCR. Raw unstructured output will be the result of an OCR; tokenization and filtering will create structured, filtered textual output from this Ocr. This Module represents an attempt to build an automated solution for capturing the essential information regarding medical entries contained within prescription images, while overcoming potential limitations associated with image noise, image distortion, and poor-quality handwriting.

3.2.2 Natural Language Processing (NLP) Module

The text from the prescription of a patient is converted into actionable medical instructions through Natural Language Processing (NLP). Information such as medication name, dosage, and frequency is extracted from text using SpaCy or a dedicated BERT NLP model. For example, the input "Amoxicillin 500mg bid for 5 days" will be processed to produce output that clearly defines medication name, dosage, and frequency for accurate scheduling and verifying each prescribed medication by subsequent modules.

3.2.3 Pill Verification Module

The Pill Verification Module is a verification module that provides confirmation prior to ingesting any medication based on what your doctor prescribed you — that is, confirming that the drug you received is actually the same medication that was prescribed. Instead of using a pill image dataset, as was required in some of the earlier versions of this module, Optical Character Recognition (OCR) is now integrated into the Pill Verification Module to directly extract information from uploaded pill packaging and labels. The OCR module processes the text that it extracts from the image (an example is "Paracetamol 650 mg") and matches it to the prescription information processed by the Natural Language Processing (NLP) module. If both match, the Pill Verification Module verifies the pill as being valid. The Pill Verification System's dynamic approach enables it to accommodate both branded and generic medications, without requiring an extensive pill image database.

3.2.4 Reminder Scheduling Module

The reminder scheduling module automatically creates reminders according to the organized prescription information. With the dosing frequency processed by the NLP module, the scheduler schedules alerts at suitable times of the day. The reminder contains the name of the medicine and the uploaded pill image by the user for visual verification. For instance, a reminder can show: "Time to take Amoxicillin 500 mg" with the uploaded photo of the pill. The module allows for the setup of reminder notifications using one of the following reminder notifications types; alarm, push and voice.

3.2.5 Tracking and History Module

This module has an entire drug history, marking when a dose was taken, missed, or delayed. Storing adherence data and timestamps in a local SQLite database allows the system to create a timeline that can be referenced by patients or caregivers later. The module also has offline capabilities, where logs and reminders remain operable even without the internet.

3.3 Real-Time Deployment

The MediSynthAI model is intended for real-time deployment in a mobile setting. Prescription and pill images are imaged directly using the smartphone camera, followed by on-device preprocessing and OCR inference. For supporting deployment in rural or resource-constrained environments, the models are optimized through quantization and pruning methods to deploy efficiently on edge hardware. SQLite is employed in local data storage to facilitate offline capability so that reminders and logs are still active without ongoing internet access. The architecture supports real-time feedback, and patients can instantly verify if the pill they uploaded corresponds to their prescription.

3.4 Data Annotation and Ground Truth

Manual annotations of medicine name, dosage, and frequency were employed as ground truth for assessing accuracy in OCR and NLP for the prescription dataset. In testing, prescription information extracted was matched to the annotated ground truth to calculate recognition accuracy. In pill verification, the correctness of the OCR extracted label from the pill image was checked against the prescription text. In this way, verification is made dynamic and does not rely on a static pill set, so the system can be made adaptive to a broad range of branded and generic drugs.

IV. PROPOSED SYSTEM

The envisioned system, MediSynthAI, incorporates prescription text extraction, natural language processing, pill identification, reminder scheduling, and adherence monitoring into a single framework. All modules are constructed to execute sequentially but independently for keeping the system robust and modular. The system receives two main inputs from the user: (1) a prescription image that is handwritten or printed and (2) an image of the medicine (pill packaging or label). The architecture then processes these inputs to authenticate the medicine and generate reminders automatically.

4.1 Data Annotation and Ground Truth

The prescription text extraction module handles the uploaded prescription image Ip. Handwritten prescriptions are scanned through a Transformer-based OCR (TrOCR) model, and printed prescriptions are run through Tesseract OCR. The OCR gives raw text output Traw.

$$T_{raw} = f_{OCR}(I_p)$$

To enhance recognition quality, preprocessing methods like grayscale normalization, adaptive thresholding, and morphological processing are used prior to OCR. This allows even noisy and partially illegible prescriptions to be read reliably.

4.2 Natural Language Processing (NLP) Module

Raw OCR output is frequently cluttered with extraneous or ambiguous text. The NLP module cleanses this output through named entity recognition (NER) and token classification to yield structured medical information: medicine name (M), dosage (D), and frequency (F).

$$P = f_{NLP}(T_{raw}) \Rightarrow \{M, D, F\}$$

In this case, P is the structured prescription output. For instance, "Amoxicillin 500 mg twice daily" is parsed into:
M = Amoxicillin
D = 500mg
F = 2 times / day.

4.3 Pill Verification Module

The pill authentication module verifies that the patient is taking the right medicine by examining the uploaded pill packaging or label image I_m. OCR is used once more, resulting in extracted text M_{img}. The system performs a similar string comparison between the extracted name M_{img} and the prescription medicine name M.

$$Verify = \begin{Bmatrix} 1\,, & if\ M_{img} \equiv M \\ 0\,, & otherwise \end{Bmatrix}$$

4.4 Reminder Scheduling Module

The pill authentication module verifies that the patient is taking the right medicine by examining the uploaded pill packaging or label image I_m. OCR is used once more, resulting in extracted text M_{img}. The system performs a similar string comparison between the extracted name M_{img} and the prescription medicine name M.

$$Verify = \begin{Bmatrix} 1\,, & if\ M_{img} \equiv M \\ 0\,, & otherwise \end{Bmatrix}$$

4.5 Tracking and History Module

The pill authentication module verifies that the patient is taking the right medicine by examining the uploaded pill packaging or label image I_m. OCR is used once more, resulting in extracted text M_{img}. The system performs a similar string comparison between the extracted name M_{img} and the prescription medicine name M.

$$Verify = \begin{Bmatrix} 1\,, & if\ M_{img} \equiv M \\ 0\,, & otherwise \end{Bmatrix}$$

4.6 Real-Time Deployment

The entire system is designed for mobile use. It is designed to utilize the OCR and NLP components as efficient, quantifiable and pruned so that they will run as on device offline. In order to ensure that users are still able to receive reminders and logs while not connected to the Internet, SQLite stores all the reminder and log data locally. By utilizing Lean models, MediSynthAI is guaranteed to be usable in both urban and rural health care settings.

4.7 System Workflow and Integration

After user uploads both a pill package and a prescription image to the system, many actions will occur; however, the first step in processing the uploaded image file is to use Optical Character Recognition (OCR) to extract text from the prescription. The extracted text will then go through several preprocessing tasks, including spurious character removal and word spacing normalization. Once the preprocessing is complete, the Natural Language Processing (NLP) capabilities of MediSynthAI will be used to convert the extracted text from the prescription into structured data, including medicine's name, dosage amount and frequency of that medicine. At this point of the process MediSynthAI has transformed the originally unstructured text from the prescription into a structured data format that will facilitate the scheduling process of the reminders and pill verification. The uploaded package pill photo will also be processed using either the OCR Model or the CNN image classifier to read the label on it with respect to the parsed details from the prescription. If both of these entries match, then the pill will have been identified by the system as verified and, therefore, both the verified photo of the pill and the associated dosage amount will be used to set the reminder so that the user can visually confirm to him or herself that they are taking their medication.

In addition, an individual's entire history of interaction with the system, called the tracking module, will be contained within it and the module will provide the capability for individuals to store this history and access it when they have no internet connectivity. The flexible, step-by-step structure of the workflow allows for modules to function independently of one another; however, the workflow's flexibility permits the modules to continue to communicate with one another, thus enhancing the strength and scalability of this system when used in real-world healthcare environments.

V. SYSTEM ARCHITECTURE

The proposed model has 4 layers. The User Input Layer is where patients or caregivers input either handwritten or printed images of their prescriptions or images of pills via their cell phones. These will be the source of untapped data. The Processing Layer utilizes Optical Character Recognition (OCR) and Natural Language Processing (NLP) to convert these images into a model that can identify medicine name(s), dosage, and frequency for activity. A Convolutional Neural Network (CNN) and You Only Look Once (YOLO) model will use these two layers to process additional information about the pill through its shape or imprint or color for verification against the prescription. The Application Layer provides an interface for medication reminders based on the information collected from both the User Input Layer and Processing Layers. Monitoring Modules will also allow tracking patient compliance. The final, Output Layer of this system provides alert capabilities to users through notifications, alarms, or via voice, and a display that graphs the complete medication history, compliance, and future reminders. These four layers of the proposed model will be integrated to create a robust solution that enables accurate verification of prescribed medications and optimally manage patient adherence issues.

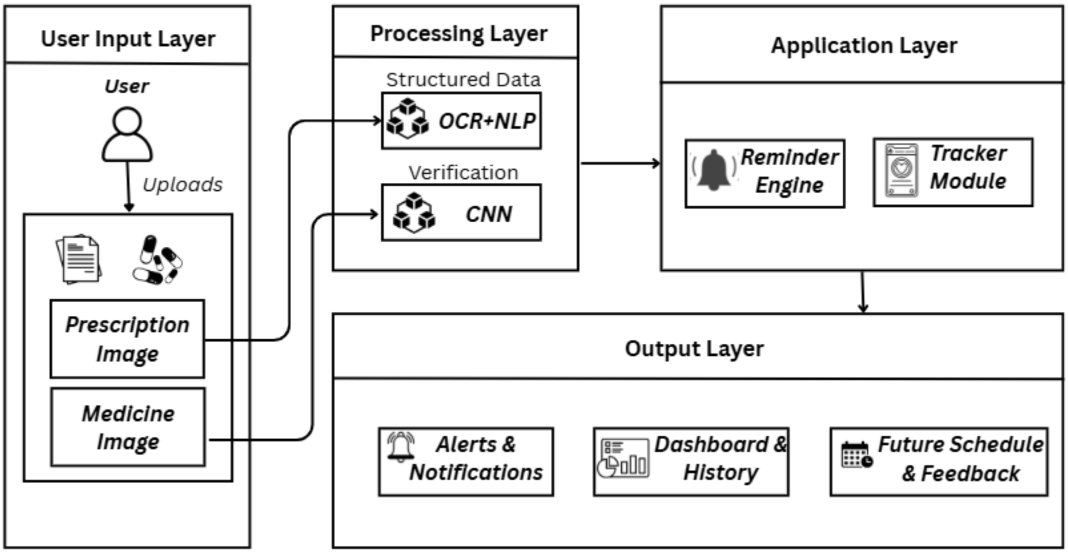

Fig. 1. Overall Architecture of MediSynthAIss

VI. PREDICTIVE MODEL DEPLOYMENT

In order to build this system, we used a combination of mobile application development technologies and some machine learning libraries. We employed Tesseract OCR to extract printed text from prescriptions, and TrOCR (a Transformer-based OCR library) for handwriting recognition on prescriptions. The raw text we extracted is then sent to a Natural Language Processing (NLP) model such as SpaCy or BERT for entity recognition, structuring the data into a format for use in creating the reminder engine. We applied OCR directly to photographs of packaged pills uploaded by users to identify the pill name. Next, the correlating pill name identified, and the user's prescription information was cross-referenced; both groups of information (i.e., structured data provided by the reminder engine and the pill name(s) obtained via Optical Character Recognition) were consolidated into a single SQLite database for offline storage and Firebase for synchronizing between different devices and the cloud. To develop the mobile application, the user interface was created with Flutter, while the back-end service was developed with Python using FastAPI. Moreover, Tesseract OCR and SpaCy NLP models had been further optimized using TensorFlow Lite, allowing for on-device real-time inference. As a result, this solution provides seamless and direct operation both online and offline, making it ideal for rural and low-resource environments.

VII. RESULTS

MediSynthAI's implementation is an ongoing process. So far, we have done preliminary testing of the performance of MEDISYNTHAI's three main components— OCR, NLP, and Pill Verification. The first experiments included building out a data preprocessing pipeline and validating the integration between its OCR and NLP capabilities to create structured prescriptions from their textual formats, e.g., an image of a handwritten prescription. The early findings indicate that the process as developed significantly increases the precision of OCR text extraction from handwritten prescriptions by improving the clarity of the input images for the OCR model. This was accomplished using several data preprocessing techniques: converting images to grayscale, removing noise, and applying adaptive thresholding to the input images prior to sending them to the OCR model. Follow-up experiments will compare its performance to established accuracy metrics (Accuracy, Precision, Recall, and F1 Score). The OCR module will be tested using the Illegible Medical Prescription Images dataset, while the NLP and Pill Verification modules will be validated using text samples taken from a large body of annotated text and an identified set of actual pill labelling images, respectively. In addition, we will evaluate the performance of the reminder-generation and tracking components with respect to reliability in scheduling the delivery of reminders and usability in terms of user experience. Overall, this aim is to create a robust, multimodal, AI platform that can improve patient compliance by providing real-time automated interpretation of prescriptions and verifying medication use through an automated reminder system. Follow-up testing will numerically demonstrate the extent to which these objectives can be achieved and therefore validate the functional applicability of the completed system.

VIII. NOVELTY AND CONTRIBUTION

MediSynthAI has many innovative features that increase patient trust, usability and accuracy compared with other healthcare-based AI technologies. The system brings together Optical Character Recognition (OCR), Natural Language Processing (NLP), and Convolutional Neural Networks (CNN) into a single multimodal pipeline for prescription parsing, pill verification, and automated reminder generation. Traditional systems are limited by using fixed datasets of pill images; MediSynthAI uses Optical Character Recognition (OCR) to create dynamic verification, which requires no extensive training when verifying branded or generic medications, since the label image is uploaded directly to the system. Due to Model Quantization and Pruning, the Framework also has Offline capabilities, thereby allowing on-device inference and reminder scheduling to occur regardless of internet connectivity. The built-in visual reminder pairing images of medications with the names also increases patient clarity in identifying medications and helps patients follow their schedules. MediSynthAI maintains an ongoing medication history log, enabling caregivers and patients to track dosage patterns and adherence trends. The combination of these features is an important part of our efforts to create a highly intelligent, dependable and patient-focused medication management solution.

IX. CHALLENGES AND REAL-WORLD CONSIDERATIONS

While MediSynthAI has experienced success, many problems affect its ability to maintain steady performance in a much larger, real-world healthcare setting. Doctors write in many different styles with many common abbreviations that can make it difficult for OCRs to accurately read handwritten prescriptions, which may be impossible to read at all. The use of pill image recognition software has a number of factors that impede its ability to correctly identify medications through digital photos of their labels: poor lighting, glare and low-quality photos taken on mobile devices. The way users engage with the app also affects the application's ability to accurately monitor medication adherence; this includes uploading non-optimal (blurred) photos when scanning medication, as well as ignoring the app's requests to upload images of medications. In addition, there are some technical hurdles involved in operating the application offline; though offline optimizing the application through methods such as Quantization and Pruning minimizes the amount of computational requirements to run the application, this can create an increase in inaccuracy when identifying pills correctly. Data privacy and data security are other major issues for patients; therefore, adherence logs and prescription records must be protected per regulations (i.e., HIPAA and GDPR) to build patient trust. Finally, a variety of prescription formats, languages, and drugs in different geographic regions limit the ability to generalise the results from this system to other areas. Addressing these challenges requires the continual expansion of the dataset, consideration of user feedback, and strict adherence to all standards for use of medical data.

X. CONCLUSION

MediSynthAI's revolutionary multimodal solution incorporates pill verification with prescription parsing, reminder generation, etc. Whereas previous research in this area has applied the Individual Technologies separately, MediSynthAI will create a single App incorporating all technology types (OCR, CNN and NLP). Going forward, the company plans to expand its offering to include larger datasets of handwritten prescriptions and to factor in user feedback for adaptive reminder messages.

XI. FUTURE ENHANCEMENTS

Integration with Electronic Health Records (EHR) systems to allow real-time synchronization of prescriptions. Providing voice reminders and chatbot support for elderly patients. Making multilingual prescription comprehension available for greater access to medications. Developing cloud analytics to enable healthcare providers to monitor patient compliance with medications. Develop an AI-based model to predict missed doses and recommend remedial measures.

REFERENCES

[1] Yaniv, Z., et al.: The National Library of Medicine Pill Image Recognition Challenge. In: Proc. Int. Conf. Med. Image Comput. Comput. Assist. Interv. (2017)

[2] Heo, J., et al.: An Accurate Deep Learning–Based System for Automatic Pill Recognition. JMIR Med. Inform. 11(3) (2023).

[3] Ali, S., et al.: Leveraging Deep Learning with Multi-Head Attention for Accurate Extraction of Medicine from Handwritten Prescriptions. arXiv preprint arXiv:2412.18199 (2024)

[4] Singal, M.: Illegible Medical Prescription Images Dataset. Kaggle (2023). Available online: https://www.kaggle.com/datasets/mehaksingal/illegible-medical-prescription-images dataset

[5] Delgado, J., et al.: Fast and Accurate Medication Identification. IEEE Access (2019)

[6] Madsen, D., Payne, P.: Automatic Pill Identification from Pillbox Images. In: Proc. AMIA Annu. Symp. (2013)

[7] Nguyen, A.D., et al.: High-Accurate Multi-Pill Detection with Graph Neural Network Assisted Fusion. arXiv preprint arXiv:2303.09782 (2023)

[8] Ashraf, M., et al.: Leveraging Code-Free Deep Learning for Pill Recognition in Healthcare. Appl. Sci. (MDPI) (2024)
[9] Nguyen, A.D., et al.: PIKA: Prescription Knowledge + Visual Features for Pill Identification. In: Proc. Int. Conf. Med. Inform. (2022)
[10] A Computer Vision-Based Pill Recognition Application. In: Proc. Int. Conf. Computer Vision and Pattern Recognition Workshops (CVPRW) (2024)
[11] TechScience Group: A Comprehensive Review of Pill Image Recognition. TechScience Press 20(5), 45–67 (2024)
[12] Various Authors: Small CNN Prototypes & Educational Proof-of-Concept Projects for Pill Recognition. Educational ML Project Reports (2022–2024)
[13] Judge, C.S., Krewer, F., O'Donnell, M.J., Kiely, L., Sexton, D., Taylor, G.W., Skorburg, J.A.: Multimodal Artificial Intelligence in Medicine. Kidney360 5(11), 1771–1779 (2024). https://doi.org/10.34067/KID.0000000000000556
[14] Al-Hussaeni, K., Khan, I., Ajjarapu, R.M.: CNN-Based Pill Image Recognition for Retrieval Systems. Appl. Sci. 13(8), 5050 (2023). https://doi.org/10.3390/app13085050
[15] K. Al-Hussaeni, I. Khan, and R. M. Ajjarapu, "CNN-based pill image recognition for retrieval systems," Appl. Sci., vol. 13, no. 8, p. 5050, 2023. DOI: 10.3390/app13085050.