A Computer Vision-Based Pill Recognition Application: Bridging Gaps in Medication Understanding for the Elderly

Taif Alahmadi[1], Rana Alsaedi[2], Ameera Alfadli[3], Ohoud Alzubaidi[4], Afnan Aldhahri[5]
Dept. of Computer Science and Artificial Intelligence-College of Computers,
Umm Al-Qura University, Makkah 21955, Saudi Arabia[1,2,3,4]
Dept. of Software Engineering-College of Computers, Umm Al-Qura University, Makkah 21955, Saudi Arabia[5]

***Abstract*—Identifying prescribed medication accurately remains a challenge for many people, particularly older individuals who may experience medication errors due to impaired vision, lack of English proficiency, or other disabilities. This problem is more prevalent in healthcare settings where pills are often distributed in strips rather than in traditional packaging, increasing the risk of dangerous consequences. To address this issue, a mobile application has been developed using Computer Vision and Artificial Intelligence to accurately recognize pills and provide relevant information through text and speech formats. The approach integrates the GPT-4 API for imprint extraction and YOLOv8 for image detection, significantly enhancing the application's accuracy. The goal is to improve medication management for vulnerable populations facing unique accessibility challenges. The application has achieved an overall accuracy of 90.89%, demonstrating its effectiveness in assisting users to identify and manage their medication.**

Keywords*—Pill detection; seniors; computer vision; artificial intelligence

I. Introduction

In healthcare, pills are a ubiquitous form of medication, typically round or oval-shaped, designed for oral consumption [1]. Their widespread use is attributed to their convenience, precise dosage control, and ease of administration. Pills encompass a diverse array of medications, including pain relievers, vitamins, antibiotics, and various therapeutic agents. As healthcare systems increasingly prioritize patient needs, the importance of innovative solutions in medication management becomes evident. The critical problem is that medication management poses significant challenges for older individuals, who are particularly susceptible to medication errors. Medication errors, particularly among older individuals, underscore the need for tools that empower individuals to take control of their health. Ensuring accurate adherence to medication is crucial for effective treatment outcomes. However, the complexity of modern drug regimens and the variety of pill presentations often lead to confusion and errors [2]. Older individuals are particularly susceptible to medication errors due to impaired vision or a lack of proficiency in English, which can lead to difficulties in correctly identifying and administering prescribed medications [3]. Studies have shown that between 75% and 96% of older patients make mistakes with their medication. While some errors may have minor effects, others can result in severe health consequences, and in some cases, even death. The risk is further exacerbated when medications are provided in unconventional forms, such as strips instead of traditional packaging [4]. Current solutions to address this problem are inadequate, as most advancements focus on reducing medical errors from the healthcare professionals' perspective, neglecting patient involvement. There is a significant gap in research on pill detection and classification, which hampers progress in medication management and safety. Existing studies have primarily developed trained model systems rather than user-friendly mobile applications, creating accessibility challenges for older individuals.

In this study, we focus on introducing and evaluating the efficacy of a novel pill detection application. This application leverages advanced technologies such as computer vision and Artificial Intelligence (AI), specifically utilizing deep learning techniques like YOLO (You Only Look Once) and GPT-4 API. YOLOv8 is employed for image detection, identifying pills with high precision, while the GPT-4 API extracts and analyzes imprints on pills to provide detailed information. The application addresses the challenges of medication recognition and management by providing a user-friendly platform capable of accurately and efficiently identifying pills from images.

The user experience is at the core of our design, with features such as a simple interface, text-to-speech functionality, and multi-language support to assist older adults and non-English speakers. This ensures users receive accurate and tailored information, addressing challenges associated with various medication forms and securing a higher degree of safety. Aligned with the ideological focus on patient-centric care, the application empowers users, regardless of age or language barriers, to confidently manage their medications. This innovative solution aims to improve overall health outcomes and reduce risks associated with medication errors, addressing a critical need in the healthcare industry.

The rest of the study is organized as follows: Section II reviews previous work. In Section III, we introduce our methodology. Section IV presents the results of this study. Finally, Section V concludes the paper.

II. Related Work

In recent years, there has been increasing research in the field of pill detection. Based on the current state of the art, existing pill recognition systems can be categorized into two groups: pill detection using an image of the pill, and pill imprint detection.

A. Pill Detection

This section will provide an overview of several research papers developed to identify pills solely using an image of the pill. These studies have shown promising results in computer vision and image recognition.

A research study conducted by Caban, Jesus j et al. [5] developed a model for automatically identifying prescription drugs. They used Shape Distribution Models to capture the shape, imprint, and color characteristics of pills. The model samples equally spaced points along pill boundaries from the centroid and calculates distances to describe the shape. It also estimates imprints through edge points analysis. The model uses HSV color space conversion to analyze boundary pixel values. When tested on 568 common US drugs, the model achieved an accuracy of 91.13%. This approach laid a foundation for pill identification in real-world applications where accurate recognition of diverse pill appearances is essential.

Björne et al. [6] conducted a study on drug name recognition and drug-drug interaction extraction using the Turku Event Extraction System (TEES) combined with machine learning techniques, specifically Support Vector Machine (SVM) classification and domain knowledge. Their methodology involved integrating several external resources, such as DrugBank and MetaMap, and syntactic features derived from deep parsing. The study achieved F-scores of 60% for drug name recognition and 59% for interaction extraction. This showcases the potential of combining domain-specific knowledge and machine learning to improve medication information retrieval. The work emphasizes the importance of integrating structured external data with advanced computational methods to enhance medication management systems.

A study conducted by Cunha et al. [7] introduced HelpmePills, a tool created to help elderly individuals identify pills using image processing techniques. The system distinguishes between different pill images based on their shape, dimensions, and colors. It operates in two main steps: learning, where pill information is gathered and stored in a local database by a caregiver, and recognition, where pill properties are compared against the database using a decision tree. When tested on a Samsung Galaxy Note II, the tool accurately identified all the learned pills. This research highlights the importance of user-friendly mobile solutions, especially for older adults who may have difficulties with managing their medication.

With little labeled data, Wang et al. [8] created a CNN-based system to identify medicines from mobile phone photos. Using a dataset from the National Institutes of Health (NIH) with 1000 distinct pill classes, they assessed their approach by applying data augmentation techniques to create fake pill visuals. With a Mean Average Precision (MAP) score of 0.328, they recommended eliminating bias in the dataset, shifting the domain, and investigating other deep-learning architectures and methodologies.

Zeng et al. [9] created MobileDeepPill, a mobile-based system for identifying pills using advanced AI techniques. The system utilizes a multi-Convolutional Neural Network (CNN) architecture along with Knowledge Distillation to enhance performance on mobile devices. Furthermore, a triplet loss function improves the model's ability to differentiate between pills. When evaluated with the NIH NLM dataset, MobileDeepPill achieved a Top-1 accuracy of 73.7% and a Top-5 accuracy of 95.6% for recognizing both sides of a pill. This study highlights the potential of integrating sophisticated AI models into mobile platforms to enhance pill identification. However, it also notes limitations such as the inability to recognize multiple pills in a single image, indicating areas for future improvement.

In response to the issue of pill misidentification, Wong et al. [10] developed a Fine-Grained Pill Identification Algorithm utilizing a Deep Convolutional Network (DCN). The DCN model surpassed standard techniques, achieving a mean accuracy rate of 95.35% at the Top-1 return. Future research could focus on expanding the dataset and enhancing the model's robustness to ensure its applicability in real-world healthcare settings.

A study conducted by Ou et al. [11] developed a computer system to help identify and classify different types of drug pills. The system has two stages: detection and classification. In the first stage, a deep CNN is used to determine the location of the pills, while in the second stage, another CNN is used to classify the pills based on their type. Scientists created a database containing 131 categories of drug pills to train the models. They used popular deep learning frameworks such as TensorFlow or PyTorch, and likely ran experiments on GPUs to speed up the process. The system achieved a top-1 accuracy rate of 79.4%, with top-3 and top-5 accuracies of 88.3% and 91.8%, respectively. Future improvements could include expanding the drug pill database and integrating the system with mobile or handheld devices for easy access and convenience.

As the third most common cause of death in the United States, medical errors are addressed by Delgado et al. [12], with a major focus on drug errors. The main issue is the challenging identification of prescription medications. To enhance accuracy, they employed deep learning with CNN architectures (ResNet50, SqueezeNet, MobileNet, 37 InceptionV3), fine-tuning them using an Adam optimizer with a decreasing learning rate strategy to improve the performance and convergence of the models, and pill localization to accurately identify the location of pills within the images by using a blob-detection neural network and post-processing techniques which were vital for precise identification. The training involved the NIH Pill Image Recognition Challenge dataset and synthetic images. The application achieved a 94% accuracy in identifying prescription medication from images, recognizing the correct pill within the top five results.

In a paper by Ou et al. [13] they discussed drug pill detection challenges in medication safety. using a two-stage architecture with EFPN for drug localization and Inception-ResNet v2 for classification. They developed the Drug Pills Image Database,

which included 612 drug categories, and analyzed the system, attaining over 96% accuracy in localization as well as Top-1, Top-3, and Top-5 accuracy levels of 82.1%, 92.4%, and 94.7%. The study used high-quality images captured with a DSLR camera, two NVIDIA 1080Ti GPUs, and the Adam optimizer. Future improvements include enhancing accuracy, simplifying the model, and expanding the database for broader pharmaceutical applications.

Tan et al. [14] examined three object detection algorithms, RetinaNet, SSD, and YOLOv3, for real-time pill and hard sample detection. Each algorithm was trained on a pill image dataset and their performance was analyzed. RetinaNet had the highest mean average precision (MAP), but its frames per second (FPS) was only a third of that of YOLOv3, making it difficult to achieve real-time performance. SSD did not perform as well on MAP or FPS. YOLOv3 featured a little lower MAP but significantly faster detection speed and harder sample detection. The authors concluded that YOLOv3 is more suitable for deployment in hospital equipment.

Kwon et al. [15] proposed a deep learning system to improve pill detection even with limited training data. Normally, the algorithm learns from various pill images when individual pills are detected. But, as the number of different pill types to identify rises the 38 combinations of pills in an image grow significantly. The algorithm follows a two-step structure that includes single-class pill area detection learning and the optimization of area dilation for multi-class pill detection. They trained the algorithm with single pill images, utilized the Mask Region-Based Convolutional Neural Network (RCNN) model, and incorporated post-processing techniques for improving pill detection. Despite the limited image and dataset sizes, the method outperformed previous algorithms. It achieved better detection performance in terms of pill identification.

In a study by Heo et al. [16], an accurate deep learning-based system for automatic pill recognition was developed, identifying tablets automatically. The system consists of two main steps - pill recognition and pill retrieval, both of which use deep learning models to train pill images and imprinted characters. The authors compiled a pill database from both South Korea and the United States. The system obtained top-1 candidate accuracy ratings of 85.6% (South Korea) and 74.5% (United States) for pill kinds that had not been trained on two databases. For future enhancements, the authors propose incorporating transfer learning approaches such as multitasking learning or adapters.

In their study, Al-Hussaeni et al. [17] aimed to improve the accuracy and efficiency of identifying pills through image retrieval. They suggested using CNNs instead of traditional methods to prevent medication errors. The authors offered three distinct CNN architectures, two of which were hybrid networks combined with classification methods (CNN + Support Vector Machine and CNN + KNearest Neighbors), and the third was a ResNet-50 network. The researchers employed a real-life dataset from the National Library of Medicine database (NLM) and achieved an accuracy of 90.8% in pill image retrieval. However, the CNN + KNN architecture showed better retrieval accuracy by 10% compared to other models. The study could be improved further by exploring advanced classification methods and refining the CNN architecture.

B. Pill Detection that Focuses on Imprint Information

This section will provide an overview of several research papers developed to automatically identify pills from images, focusing on detecting imprint information.

Lee et al. [18] conducted a study on identifying illicit drugs using a feature extraction method based on edge-based characteristics and invariant moments. Their approach effectively accounted for the variability in pill images caused by different lighting conditions and viewpoints. By creating multiple templates during edge detection to improve resilience against these variations, the study achieved a remarkable 76.74% rank-1 matching accuracy using a comprehensive dataset of 822 illicit drug pill images and 1,294 legal pill images. This pioneering use of edge localization for imprint extraction represents a significant advancement in pill identification accuracy, demonstrating impressive capability in handling variations in image quality and environmental conditions. It provides a valuable reference for enhancing the precision of pill recognition in various settings, cementing its relevance for imprint extraction techniques.

Chen et al. [19] introduced an automated methodology for identifying pills by using imprint information. The text was obtained through a modified stroke width transform (MSWT) and characterized using the weight shape context (WSC). By employing this approach, the researchers achieved a classification accuracy of as high as 93.03% when categorizing over 10 thousand query pill images into approximately 2000 distinct groups.

Yu et al. [20] developed a high-accuracy automatic pill recognition system that employs imprint information as the primary differentiation between different pills. It utilized algorithms for imprint extractions, which adopt a modified stroke width transform for imprint extractions and uses Loopy belief propagation for image segmentation of printed imprint pills. The results were promising, with up to 97.16% accuracy in identifying 12,500 pill images into 2,500 categories. The authors suggested accelerating the algorithm and improving the accuracy for lower-quality images.

Chupawa et al. [21] developed a pill identification system for pharmacists using a detailed three-stage approach. Firstly, the preprocessing stage enhances image quality by removing background noise, cropping, and applying filters to improve clarity. Secondly, during feature extraction, the system isolates and analyzes imprint characteristics, such as shape and texture, which are crucial for accurate identification. Finally, the classification stage employs a neural network to categorize pills based on the extracted features. This system achieved an impressive accuracy rate of 94.4% in identifying six different types of pills, demonstrating the effectiveness of deep learning techniques in enhancing pill identification and supporting precise medication management.

Suntronsuk et al. [22] described a method for automatically identifying text from pill impressions. The method relied on a set of predetermined rules for recognizing imprint places, as well as a methodology for removing noise from binary images.

Initially, the photos were treated to normalization and enhancement techniques to improve contrast. Then, the imprint area was identified using different criteria. The selected area was subsequently trimmed and converted to binary representations via either Otsu's thresholding approach with noise reduction or K-means clustering. Finally, the binary result was fed into a trained Tesseract model, which extracted the text. The study found that Otsu's thresholding method surpassed K-means clustering, with precision and recall rates above 57%. Pill Image Binarization to Detect Text Imprints.

III. METHODOLOGY

The following section provides a detailed overview of the methods utilized in this study to combine the YOLOv8 model with GPT-4 for optical character recognition (OCR) capabilities, with the goal of improving the pill detection application. This integration is crucial for accurately capturing and interpreting text imprints on pills, which is essential for proper medication identification. The methodology is broken down into key areas such as research design, procedure, data acquisition, and preparation, all customized to address specific challenges and objectives identified in the initial stages of the project.

A. Study Design and Procedure

This study assessed the effectiveness of integrating YOLO (You Only Look Once) object detection models with GPT-4 for OCR in developing the pill detection application. YOLOv8 was selected due to its proven efficacy in object detection [23] and GPT-4's advanced capabilities for text recognition [24]. This combination is pivotal for accurately extracting and interpreting text imprints on pills, crucial for medication identification. The study followed a comprehensive procedure encompassing the training of detection models and the integration of OCR capabilities. Initially, YOLOv8 was fine-tuned on a specifically curated dataset containing images of various pills captured under different environmental settings to simulate real-world usage. This was followed by applying the GPT-4 API to perform OCR on the detected pills, focusing on the imprints containing essential medication information. The algorithmic flow of these processes is depicted in Fig. 1, which illustrates the program flow from model training to pill detection and information display in the application. First, the model will undergo training using a labeled dataset until it reaches a good level of accuracy. Next, it will be connected to our Android application. When the user opens the application and uploads or takes a picture, the model will analyze the image to identify any recognizable pills. If it detects a pill, the image will be sent to the GPT-4 API for imprint extraction. If the imprint and detected pill match up, the relevant information will be displayed for the user.

B. Data Acquisition and Preparation

The dataset was meticulously constructed with a focus on medications commonly prescribed for prevalent chronic diseases in Saudi Arabia, encompassing hypertension, diabetes, and heart diseases. Medications included hypertension pills such as Tabuva, Amlor, and Tenoryl; diabetes medications including Glucare and Jardiance; and heart disease treatments like Aspirin, Cardicor, Diusemide, and Isobide. The selection of these medications was informed by consultations with several pharmacists to ensure the dataset reflects real-world medical needs. It was essential to choose a dataset with a wide variety of medications to represent different conditions and therapeutic classes, as well as a diverse range of pill shapes, colors, and sizes, as diversity is crucial for confirming the model's ability to be applied across different medical situations.

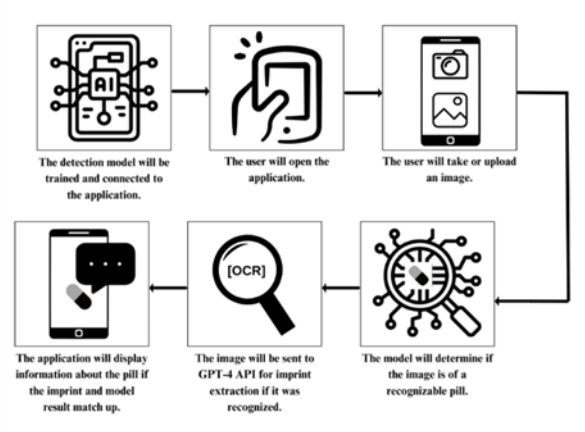

Fig. 1. Pill detection application flow.

To further enhance the dataset's relevance, specific choices were made to include pills without imprints, such as Aspirin, and pills with only engraved imprints and not inked, like Cardicor. Additionally, pills that appeared very similar except for their imprints, such as Jardiance and Glucare, were included to test the model's ability to distinguish between subtle differences.

Fig. 2. Dataset capturing conditions.

To simulate realistic usage scenarios, the dataset creation involved a detailed and structured image collection process shown in Fig. 2. Pills were photographed under varied lighting conditions to reflect different environmental settings patients might encounter. Specifically, 200 images were equally distributed across four lighting categories: bright, dim, natural, and artificial light. To further enhance the model's ability to perform under diverse backgrounds, 100 images were captured against both plain and complex backgrounds. The dataset also included images taken from multiple angles and orientations—top view, side view, angled view, and random orientations—to ensure comprehensive coverage of how pills might be presented to the application by users. In addition to these variations, 50 images focused on scale and proximity, with close-up shots and standard-distance shots, and another 50 showcased pills on different surface types, split between glossy and matte finishes. This comprehensive application to data collection ensures that the model is well-prepared for effective deployment in diverse and challenging settings.

Following collection, the images were annotated using RoboFlow, which facilitates precise and efficient object labeling within images. The images will be split into three parts for training, testing, and validation, with 70%, 15%, and 15% respectively. This division is based on research that suggests that for datasets between 100 and one million, this is the most ideal splitting method [25]. To enhance the model's capacity for generalization and improve its detection accuracy, the dataset was augmented by introducing an 'unknown' class comprising pills that were either similar to or different from the target medications but commonly encountered in the region. Additionally, negative samples, consisting of images devoid of any pills, were incorporated to train the model to recognize scenarios absent of relevant objects, a critical step for minimizing false positives in real-world applications.

C. Testing and Evaluation Methods

The pill detection application was thoroughly evaluated to test the effectiveness of the YOLO models and the GPT-4 OCR integration. The evaluation involved using various quantitative metrics to gain valuable insights into the system's performance. The confusion matrix was utilized to examine true positive, false positive, false negative, and true negative predictions for each type of detected pill, providing a comprehensive overview of classification performance as well as a visual depiction of the model's accuracy and misclassifications. Accuracy was calculated to determine the proportion of correct predictions out of all predictions made, which is crucial for evaluating the overall effectiveness of the detection system.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

Precision, which measures the exactness of the detection, ensures that the model doesn't produce a high number of false positives. It measures the proportion of true positive (TP) instances among those that the model predicts as positive.

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

Recall, also known as sensitivity, evaluates the model's ability to properly identify every relevant instance. This metric is particularly crucial when it's important to capture as many true positives as possible, such as in medical applications where missing a relevant pill type could have significant consequences.

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

The F1-score, which balances precision and recall, is particularly useful in scenarios with uneven class distribution and provides insights into the model's robustness when dealing with various pill types.

$$F1 = \frac{2 \times (precision \times recall)}{(precision + recall)} \quad (4)$$

The mean average precision (mAP) was utilized to examine the precision of the bounding boxes generated by the YOLO models across different types of tablets. This metric offers a comprehensive perspective on the model's performance in locating and accurately identifying pills in an image. This set of metrics provides a comprehensive evaluation framework that measures individual aspects of model performance and ensures the system's reliability and effectiveness in a real-world application setting, confirming the system's suitability for practical deployment in the healthcare sector.

IV. Results And Discussion

A. Results

The evaluation process in this study is conducted in two stages: first for the YOLOv8 detection model, and then for the entire application with YOLOv8 and GPT-4 for imprint extraction.

Evaluate the YOLOv8 detection model: to evaluate our pill detection model, we employed several standard YOLO evaluation metrics. The main metric we used was the mAP, which assesses both precision and recall across various classes by averaging the mAP scores. We utilized a confusion matrix, as depicted in Fig. 3, to compute the accuracy, precision, recall, and F1 score. The confusion matrix enables us to determine the accuracy, precision, recall, and F1 score for each class, and the corresponding values are presented in Table I.

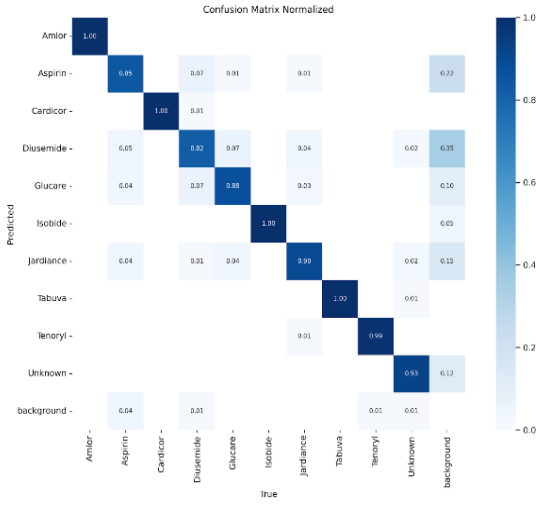

Fig. 3. YOLOv8 normalized confusion matrix.

TABLE I. ACCURACY, PRECISION, RECALL, AND F1-SCORE FOR EACH CLASS

	Accuracy	Precision	Recall	F1-score
Amlor	1	1.0	1.0	1.0
Aspirin	0.956	0.73	0.83	0.78
Cardicor	0.999	0.99	1.0	0.99
Diusemide	0.936	0.61	0.83	0.70
Glucare	0.967	0.79	0.88	0.83
Isobide	0.995	0.95	1.0	0.97
Jardiance	0.972	0.78	0.90	0.84
Tabuva	0.999	0.99	1.0	0.99
Tenoryl	0.998	0.99	0.99	0.99
Unknown	0.988	0.89	0.99	0.94

The efficacy of the YOLOv8 model is evaluated and illustrated across a variety of performance metrics, each represented in comprehensive graphical form. Fig. 4 showcases the Precision-Recall Curve, highlighting the model's exceptional capability to achieve a balance between precision and recall. As shown in Table I. Amlor stands out with perfect scores of 1.0 in both metrics, exemplifying flawless detection capabilities. Similarly, Cardicor, Isobide, Tabuva, and Tenoryl demonstrate near-perfect performances, affirming their high detection accuracy with precision and recall rates of 0.99 or higher.

The analysis continues with the Precision-Confidence Curve and F1-Confidence Curve, depicted in Fig. 5 and Fig. 6 respectively. These graphs reveal precision and F1-scores across different classes, highlighting areas for improvement and strengths. Aspirin, for example, shows a precision of 0.73 and an F1-score of 0.78, suggesting a need to minimize false positives. Conversely, Glucare and Jardiance perform robustly, with precision scores of 0.79 and 0.78 and F1-scores of 0.83 and 0.84, showcasing reliable detection and classification at varied confidence levels.

Fig. 7 shows the Recall-Confidence Curve, further illustrating the model's effectiveness in identifying true positives. High recall rates are maintained across most classes, with notable achievements from Glucare and Jardiance, who reach recalls of 0.88 and 0.90, ensuring comprehensive detection of relevant objects.

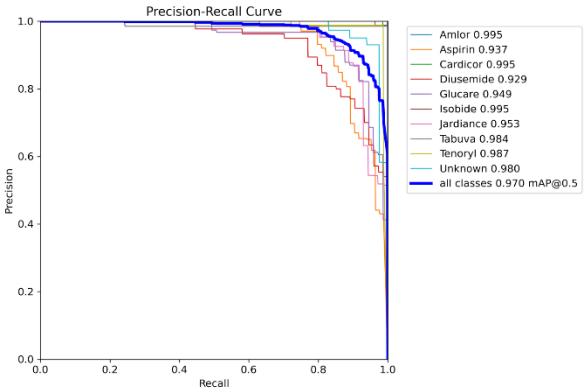

Fig. 4. YOLOv8 model precision-recall curve.

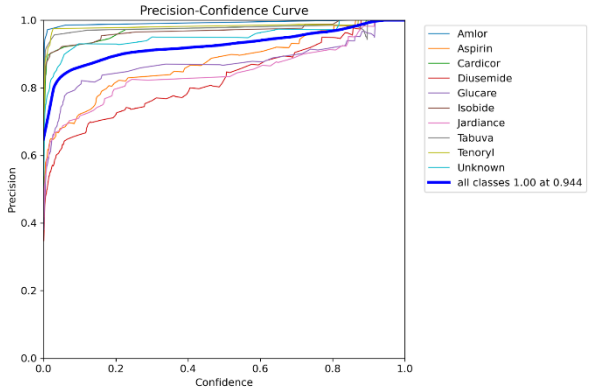

Fig. 5. YOLOv8 model precision-confidence curve.

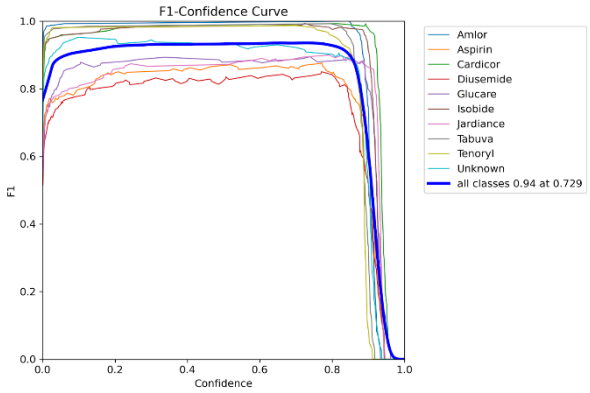

Fig. 6. YOLOv8 model f1-confidence curve.

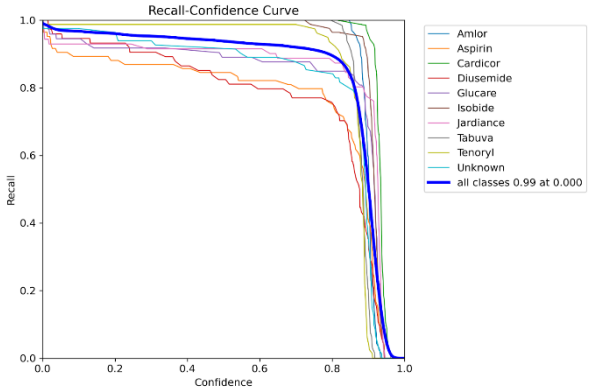

Fig. 7. YOLOv8 model recall-confidence curve.

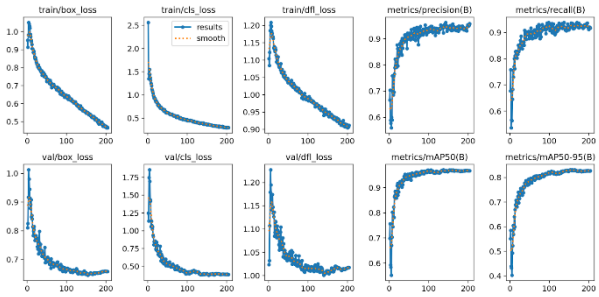

Fig. 8. Training and validation losses of YOLOv8 model.

Lastly, Fig. 8 presents the Training and Validation Loss Curves, providing insights into the model's learning dynamics over time. These curves demonstrate a consistent decrease in training and validation losses, including Train/Box_Loss, Train/Cls_Loss, and Train/Dfl_Loss, as well as Val/Box_Loss, Val/Cls_Loss, and Val/Dfl_Loss. This indicates an improvement in the model's ability to predict bounding boxes, classify objects correctly, and estimate attributes such as distance and focal length, confirming effective learning and generalization to new, unseen data.

Furthermore, we measure the mAP on both the validation and test datasets. We attained a mAP of 97% on the validation dataset at a threshold of 0.5, which is a remarkable achievement. Also, we measured the mAP of the test dataset, we attained a mAP of 96.3% with a threshold of 0.5 on the test dataset. This result is compared to a related study [8] that focused on detecting pills using YOLOv5 whereas the achieved accuracy was 85.6%. Our pill detection model demonstrates a significant performance difference, highlighting its robustness and precision compared to the previous study.

It is possible for the mAP percentage to decrease on the test dataset compared to the validation dataset. This is because the model was never exposed to the test dataset during training. According to the predictions in Fig. 9 yolov8 model can identify a single pill with a confidence ranging from 87% to 94%.

Fig. 9. YOLOv8 model predictions (confidence).

Evaluating the Integration of YOLOv8 and GPT-4: In efforts to improve the precision of the pill identification application, the YOLOv8 model was combined with the GPT-4 API. This integration significantly enhanced the reliability of detection outcomes. The models were tested separately under varying conditions such as different lighting, overlapping pills, and unclear imprints. The YOLOv8 model performed well under these diverse conditions, accurately identifying pills despite these challenges. However, to further improve accuracy and minimize false positives—which are particularly dangerous—the integrated performance of both YOLOv8 and GPT-4 was evaluated.

The GPT-4 API did not perform as well under different lighting conditions and struggled to provide predictions when the imprint or part of it was invisible. To assess the success of this integration, 450 tests were conducted with the application across various pill types in the dataset. The results demonstrated an impressive accuracy rate of 90.89%, indicating a substantial improvement in the application's ability to identify pills accurately. This comprehensive evaluation confirmed the integration's effectiveness in enhancing the accuracy and reliability of the pill identification application, especially under challenging conditions.

B. Discussion

The YOLOv8 model showed excellent performance, with high accuracy, precision, recall, and F1-scores across most classes. The mAP scores on validation and test datasets confirm its effectiveness. Classes like Amlor, Cardicor, Isobide, Tabuva, and Tenoryl achieved near-perfect detection, though Aspirin and Diusemide had higher false positives.

The precision-recall curves, confidence metrics, and loss curves highlight the model's strengths and areas for improvement. The consistent decrease in training and validation losses indicates effective learning and generalization.

Integrating YOLOv8 with the GPT-4 API improved overall accuracy, combining robust detection with enhanced imprint extraction. However, the GPT-4 API's performance depends on image quality and orientation, affecting reliability with poor images.

Compared to existing pill recognition systems, our approach is unique in integrating YOLOv8 and GPT-4. While most systems rely on either advanced image detection models or OCR technologies independently, our combination leverages the strengths of both. This integration allows our system to handle a broader range of identification challenges, enhancing overall accuracy and reliability, especially in distinguishing pills with similar appearances but different imprints.

A notable limitation of this study is the dependency of the GPT-4 API on image clarity. For the GPT-4 API to function properly, the images need to be clear and well-oriented. Poor quality images can lead to unreliable predictions, affecting the overall accuracy of the system.

V. Conclusion and Future Work

As the aging population grows, the likelihood of medication errors increases, particularly among older individuals who often rely on multiple medications for chronic conditions. Recognizing this challenge, we developed a novel application designed to mitigate the risk of such errors by facilitating accurate pill identification through imaging technology. This application leverages a YOLOv8 model trained on a meticulously created dataset, in conjunction with GPT-4 for enhanced text extraction capabilities. Insights were garnered from interviews with 11 pharmacists and 15 older individuals, highlighting the difficulties pharmacists face in identifying pills based solely on their appearance, given the necessity to recall each pill's specific codes, shapes, colors, and sizes. This task is compounded by the fact that identical medications produced by different manufacturers may vary significantly in appearance,

underscoring the need for a robust and precise pill classification system. The effectiveness of this system was demonstrated by achieving a mAP of 90.89%, validating the application's capability in accurately detecting and classifying pills. While the results are promising, there is still potential for further improvements.

Future enhancements include extending language options beyond English and Arabic to increase global accessibility, enriching the dataset with a more varied collection of pill images to boost the model's robustness, refining the text extraction feature to handle challenging imaging conditions, and expanding availability to iOS and Windows platforms. These advancements will further the development of our application, making it a more versatile and reliable tool for preventing medication errors among the elderly. Additionally, future work should focus on better image preprocessing and expanding the dataset to include more pill types for improved generalization. The integration of YOLOv8 and GPT-4 has shown significant potential for enhancing medication management, especially for older individuals and those with limited English proficiency. Continued refinement of these technologies can further reduce medication errors and improve patient outcomes. The planned improvements and ongoing testing will continue to refine the system, aiming for broader adoption and increased efficacy in the real world.

ACKNOWLEDGMENT

The authors would like to express their appreciation to the pharmacists and elderly individuals for taking part in the interviews. The valuable knowledge and insights contributed by these participants were crucial to the success of the project. Special thanks are also extended to a pharmacist who served as a primary source of information and aided in gathering the pills for the dataset.

REFERENCES

[1] "Medical definition of pill," *RxList*, Mar. 29, 2021. https://www.rxlist.com/pill/definition.htm

[2] B. C. Wimmer, J. S. Bell, J. Fastbom, M. D. Wiese, and K. Johnell, "Medication regimen complexity and polypharmacy as factors associated with All-Cause mortality in older people," Annals of Pharmacotherapy/˜the œAnnals of Pharmacotherapy, vol. 50, no. 2, pp. 89–95, Dec. 2015, doi: 10.1177/1060028015621071.

[3] "5. Vulnerable populations | ATrain Education." https://www.atrainceu.com/content/5-vulnerable-populations

[4] J. J. Mira, "Medication errors in the older people population," *Expert Review of Clinical Pharmacology*, vol. 12, no. 6, pp. 491–494, May 2019, doi: 10.1080/17512433.2019.1615442.

[5] J. J. Caban, A. Rosebrock, and T. S. Yoo, "Automatic identification of prescription drugs using shape distribution models," IEEE, Sep. 2012, doi: 10.1109/icip.2012.6467032.

[6] J. Björne, S. Kaewphan, and T. Salakoski, "UTURKu: Drug Named Entity Recognition and Drug-Drug Interaction Extraction using SVM classification and Domain knowledge," *Joint Conference on Lexical and Computational Semantics*, pp. 651–659, Jun. 2013, [Online]. Available: https://www.aclweb.org/anthology/S13-2108.pdf

[7] A. Cunha, T. Adão, and P. Trigueiros, "HelpMePills: a mobile pill recognition tool for elderly persons," Procedia Technology, vol. 16, pp. 1523–1532, Jan. 2014, doi: 10.1016/j.protcy.2014.10.174.

[8] Y. Wang, J. Ribera, C. Liu, S. K. Yarlagadda, and F. Zhu, "Pill Recognition Using Minimal Labeled Data," 2017 IEEE Third International Conference on Multimedia Big Data, Apr. 2017, doi: 10.1109/bigmm.2017.61.

[9] X. Zeng, K. Cao, and M. Zhang, MobileDeepPill: a Small-Footprint mobile deep learning system for recognizing unconstrained pill images. 2017. doi: 10.1145/3081333.3081336.

[10] Y. F. Wong, H. T. Ng, K. Y. Leung, K. Y. Chan, S. Y. Chan, and C. C. Loy, "Development of fine-grained pill identification algorithm using deep convolutional network," Journal of Biomedical Informatics, vol. 74, pp. 130–136, Oct. 2017, doi: 10.1016/j.jbi.2017.09.005.

[11] Y.-Y. Ou, A.-C. Tsai, J.-F. Wang, and J. Lin, "Automatic Drug Pills Detection based on Convolution Neural Network," IEEE, Oct. 2018, doi: 10.1109/icot.2018.8705849.

[12] N. L. Delgado et al., "Fast and accurate medication identification," Npj Digital Medicine, vol. 2, no. 1, Feb. 2019, doi: 10.1038/s41746-019-0086-0.

[13] Y. Ou, A. Tsai, X. Zhou, and J. Wang, "Automatic drug pills detection based on enhanced feature pyramid network and convolution neural networks," *IET Computer Vision*, vol. 14, no. 1, pp. 9–17, Jan. 2020, doi: 10.1049/iet-cvi.2019.0171.

[14] L. Tan, T. Huangfu, L. Wu, and W. Chen, "Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, Nov. 2021, doi: 10.1186/s12911-021-01691-8.

[15] H. Kwon, H.-G. Kim, and S.-H. Lee, "Pill detection model for medicine inspection based on deep learning," Chemosensors, vol. 10, no. 1, p. 4, Dec. 2021, doi: 10.3390/chemosensors10010004.

[16] J.-Y. Heo, Y.-J. Kang, S. Lee, D. Jeong, and K.-M. Kim, "An accurate Deep Learning–Based System for Automatic pill identification: model development and validation," Journal of Medical Internet Research, vol. 25, p. e41043, Jan. 2023, doi: 10.2196/41043.

[17] K. Al-Hussaeni, I. Karamitsos, E. A. Adewumi, and R. M. Amawi, "CNN-Based pill image recognition for retrieval systems," Applicationlied Sciences, vol. 13, no. 8, p. 5050, Apr. 2023, doi: 10.3390/application13085050.

[18] Y.-B. Lee, U. Park, and A. K. Jain, "PILL-ID: Matching and retrieval of drug pill imprint images," Istanbul, Turkey, Aug. 23, 2010. doi: 10.1109/icpr.2010.645.

[19] Z. Chen and S. Kamata, "A new accurate pill recognition system using imprint information," Proceedings of SPIE, Dec. 2013, doi: 10.1117/12.2051168.

[20] J. Yu, Z. Chen, S. Kamata, and J. Yang, "Accurate system for automatic pill recognition using imprint information," Iet Image Processing, vol. 9, no. 12, pp. 1039–1047, Dec. 2015, doi: 10.1049/ietipr.2014.1007.

[21] P. Chupawa, "Pill Identification with Imprints Using a Neural Network: doi: 10.14456/mijet.2015.7," 2015. https://ph02.tci-thaijo.org/index.php/mijet/article/view/10.14456.mijet.2015.7

[22] S. Suntronsuk and S. Ratanotayanon, "Automatic text imprint analysis from pill images," in Proceedings of the 2017 9th International Conference on Knowledge and Smart Technology (KST), Chonburi, Thailand, pp. 288-293, Feb. 1-4, 2017. doi: 10.1109/KST.2017.7886081.

[23] "YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness," IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/10533619

[24] M. Hajiali, "OCR post-processing using large language models," Digital Scholarship@UNLV. https://digitalscholarship.unlv.edu/thesesdissertations/4811/

[25] I. O. Muraina, "IDEAL DATASET SPLITTING RATIOS IN MACHINE LEARNING ALGORITHMS: GENERAL CONCERNS FOR DATA SCIENTISTS AND...," ResearchGate, Feb. 1262022,[Online].Available:https://www.researchgate.net/publication/358284895_IDEAL_DATASET_SPLITTING_RATIOS