

# Real-Time Medicine Packet Recognition System in Dispensing Medicines using YOLOV8

Chris Xavier Mathias<sup>1</sup>, Meng Lin<sup>2</sup>, Arsh Imtiyaz Assadi<sup>3</sup>, Bhavish B S<sup>4</sup>, Daksha Shenoy<sup>5</sup>,  
Rolston D'Souza<sup>6</sup> and Aravinda C V<sup>7</sup>

<sup>1, 3-7</sup> NMAM Institute of Technology, Department of Computer Science & Engineering, Karkala, Udupi, Karnataka, India  
Email: chrisxm15@gmail.com, arshassadi29@gmail.com, bhavish104@outlook.com, dakshashenoy19@gmail.com,  
aravinda.cv@nitte.edu.in

<sup>2</sup>Dept. of Electronic and Computer Engineering (The Graduate School of Science and Engineering), Ritsumeikan University  
1-1-1 Noji-higashi, Kusatsu Shiga, 525-8577 JAPAN

**Abstract**—Dispensing medicine to elderly residents is a crucial task in nursing homes, requiring specialized medical staff to prevent errors and manage inventory. With rapid advancements in deep learning technology, high-performance computers can replace some supervision tasks. In this paper, we have proposed a system consisting of a Convolutional Neural Network- YOLOV8 custom-trained image recognition model, Optical Character Recognition (OCR), and RoBERTa, a pre-trained Large Language Model (LLM) that uses the architecture of the transformer, and it has attained cutting-edge performance on a variety of natural language processing tasks. The RoBERTa model is particularly effective in handling unstructured text data, making it an ideal choice for this project. The model is hosted on a virtual machine instance on the Google Cloud Platform, a cloud computing service that provides a range of tools and services for building, deploying, and managing applications and services. The use of the Google Cloud Platform ensures that the system is scalable, reliable, and accessible from anywhere at any time, making it an ideal choice for a method for recognising pharmaceutical packets in real time.

**Index Terms**— Medical dispensing, inventory management, computer vision, optical character recognition, transfer learning, natural language processing and large language model.

## I. INTRODUCTION

Administering medications to elderly residents is a crucial responsibility in the daily operations of nursing homes. To ensure that medications are dispensed accurately, nursing homes must employ specialized medical personnel and monitoring staff to prevent medication errors. A Real-Time Medicine Packet Recognition System is an automated solution designed to assist in the dispensing of medicines and the management of medical inventory. The system uses advanced technologies such as machine learning algorithms, image processing techniques, and natural language processing to recognize and segment the characters on medicine packets. Once the system identifies the medicine packet, it stores the relevant information in an online database, enabling healthcare providers to access it in real time. This makes it easier to track inventory, ensure that the correct medication is administered to patients, and prevent errors in the dispensing process. While medicine packet recognition systems have been developed in the past, there is a need for a more accurate and efficient solution that can be easily integrated into existing healthcare systems. The main objective of this project is to create a

novel solution that can solve this issue and have a beneficial effect on the healthcare sector.

## II. RELATED WORK

Ref. [1], coauthored by Lin Meng, described a three-phase system. The characters are first divided into segments using morphological and binarization techniques, and their positions are then preserved. Additionally, a slight convolutional neural network (SCNN) is used to effectively and swiftly recognise the characters. In the end, the system organises the results and presents various data for checking and automatically avoiding medication dispensing errors. During the test, the computation time hit five frames per second, and the recognition accuracy reached 100%. The outcomes of the experiments further demonstrate the viability and efficacy of the suggested approach. Despite how attractive this might seem, the dataset used for the study was perfect and almost non-existent.

Ref. [2], coauthored by Mark Swainson, developed a camera-based automatic system that is capable of quickly and accurately identifying the expiration date information printed on various food packaging. Food containers with expired dates printed incorrectly will be collected. In order to produce much more food, this technology will allow for far greater control over the correctness and readability of crucial "use-by" dates and important traceability information in food and drink manufacturing activities. To find the expiration date region, a full convolutional network deep neural network approach is used. The extracted date character blobs will be determined by the detection findings. Characters will be categorised based on related traits that have been extracted. A method like this might improve the assurance of food safety and quality.

Ref. [3], coauthored by Tzu-Chin Yang proposes a deep learning-based intelligent recognition system dubbed ST-Med-Box. The approach attempts to assist chronic patients in managing a variety of medications, avoid medication mistakes, and avoid drug interactions that may result from taking the wrong prescriptions. It can also include a variety of pharmaceutical-related functions, such as chronic patient information management, drug information, and reminders for taking prescriptions on schedule. The ST-Med-Box is made up of a deep learning training server, an Android mobile app, an intelligent medicine recognition device, and a cloud-based management platform.

Ref. [4], coauthored by Dilbagh Singh, presents a network called DKNet, created to address issues including over-fitting and gradient vanishing during the recognition of Kuzushiji characters. A trilateral joint filter is used to reduce noise from training images, and contrast-limited adaptive histogram equalisation (CLAHE) is used to increase image visibility. The final layers of a MobileNet model are removed when using the DKNet architecture, and the input is flattened before a fully connected classification layer with rectified linear units (ReLU) and dropouts. Finally, recognition results are generated using the softmax activation function. These methods increase the precision of character segmentation and recognition.

Ref. [5], coauthored by V.B. Ashwin introduces an automated process for creating subtitles and creating semantic video summaries. There is a substantial demand for automatic video summarization in the age of big data. This method helps to preserve important content while facilitating efficient storage and simple browsing through huge video collections. Regardless of whether the video had subtitles when it was first created, the method uses text summarization algorithms to create video summaries with subtitles generated through speech recognition. The suggested method uses weight-based learning and intersection, two different ensemble methods, to improve the efficiency of creating subtitles and summarising videos.

Ref. [6] coauthored by Kenjiro Sugimoto presents Linear Manifold Colour Descriptor (LMCD), a colour-based approach for drug package recognition. Ref. [6] coauthored by Kenjiro Sugimoto describes a set of colour pixels from a coloured package image as an affine subspace in the colour space and identifies an anonymous package through linear manifold matching. Due to the limited dimensionality of colour spaces, LMCD can offer a more succinct description and faster computation than other description styles. The research also suggests distance-based dissimilarities for matching linear manifolds, with a focus on matching colour distributions.

Ref. [7], coauthored by Paula Trigueros proposes a mobile system with two distinct modes, learning and recognition. The former is responsible for profiling and storing pills. This mode begins by capturing an image that includes both the marker and the pill. Next, the pill's shape, size, and color are determined to create a pill profile, which is then stored in the system's database for future use. In contrast, the recognition mode is designed to identify pills based on their profile and characteristic filtering. This mode starts by capturing and resizing the image and then detecting the marker. The pill is then segmented, and filtering takes place, which involves sequentially removing groups of pills during feature detection. Shape is the first feature used to exclude pills, followed by size and color filtering. The process ends with one of three outcomes: no pill detected, one pill detected, or multiple pills detected.

Ref. [8], coauthored by Aravinda C V is a study which aims to enhance the accuracy of recognizing people suffering from apathy through Doppler radar images, which is a practical technique. To achieve this, the doppler radar image dataset based on the elderly person suffering apathy is used, and the pictures are pre-processed using binarization and division to extract the eigenvalues. For learning and prediction on the dataset, many classic classification models such as support vector machine, k-nearest neighbour, decision tree, random forest, and neural network are used. To discover the ideal solution for each model within the same binarization threshold, the parameters of all models are optimised, and a three-dimensional data graph is generated based on testing. Finally, by integrating these best models, an ensemble model is built, increasing the apathy categorization prediction accuracy from 75% in earlier research to 92.86%.

Ref. [9], coauthored by H. N. Prakash, uses a correlation approach to identify Kannada handwritten characters. It can be more difficult to recognise Kannada characters since they can be formed into compound shapes known as Kagunita. To solve this, a variety of preprocessing methods are applied to the digitised input image before a straightforward segmentation algorithm is used to separate the image's constituent characters. Then, a correlation between each segmented character and a template is performed. An editable version of the template with the highest correlation value is shown.

Ref. [10], coauthored by Yucong Duan, explores the feasibility of utilizing doppler radar imaging for the diagnosis of elderly apathy. To construct a dataset for the research, a total of 178 participants who classify as elderly were enrolled, who were required to complete a questionnaire and walk while being subjected to doppler radar imaging. The choice of walking as an activity of interest was motivated by its ubiquity and potential for providing crucial health information. In order to protect the privacy of participants, doppler radar imaging was preferred over an RGB camera. On the basis of the walking doppler radar images, seven machine learning models—including a suggested neural network—were used to categorise indifference. Prior to classification, a simple image pre-processing technique was employed to extract features. Specifically, each image was divided into four pieces on both the axes, vertical and horizontal. The resulting four-by-four grid was then binarized, and the number of feature points in each of the resulting sixteen sub-regions were tallied to create eight features. The optimized binarization threshold was determined through an experimental sliding of the threshold. Using the suggested neural network, they have achieved an apathy classification accuracy of over 75%. Although this accuracy is lower than other current methods for classifying objects, this study provides evidence for the potential of utilizing doppler radar images for apathy classification in elderly populations. Future research will focus on improving the accuracy of the approach.

Ref. [11], coauthored by Keerthi Shetty aims to improve the weighted K-means clustering with the help of Firefly Optimization algorithms for an effective data clustering system. Data analysis applications, including data mining, pattern recognition, and picture analysis, frequently employ weighted K-means clustering. Although it is effective, scalable, and simple to implement, its primary drawback is randomly selecting initial centroids, which can result in local optimums. Global optima are obtained by combining weighted K-means clustering and the firefly optimisation algorithm, an optimisation technique inspired by nature, to address this problem. The efficiency of sharing and searching for information was found to increase by 0.02-0.4 (label value) when this technique was applied to various medical datasets, including liver disease, heart disease (original), heart disease (stat-log), and Indian liver patients.

### III. PROPOSED SYSTEM

The proposed system employs advanced technologies such as Convolutional Neural Network (CNN) and Optical Character Recognition (OCR) to automatically identify and segment the medicines in an inventory. Users can take a picture of their inventory using a single application, which will be sent to the server's Application Programming Interface (API). Upon receiving the image, the server searches for the region of interest using the highly accurate CNN-YOLOv8 algorithm. The OCR engine then detects and segments the region of interest to identify the text string, which is pre-processed using a Natural Language Processing (NLP) Large Language Learning Model - Robustly Optimized Bidirectional Encoder Representations from Transformer (BERT). If the initial search does not return any result, the system will magnify the region of interest by 30% to improve character recognition. The model is hosted on a cloud server using Google Cloud Platform (GCP), and the Virtual Machine instance returns a JSON file of the predicted image data. On receiving the data, the application displays vital information such as expiry date, and manufacturing date of the medicine.

#### IV. METHODOLOGY

##### A. Workflow for the Proposed System

The process of our proposed system is depicted in Fig. 1. The arrow in the figure indicates the direction of the workflow, illustrating the path that the workflow follows within the diagram.

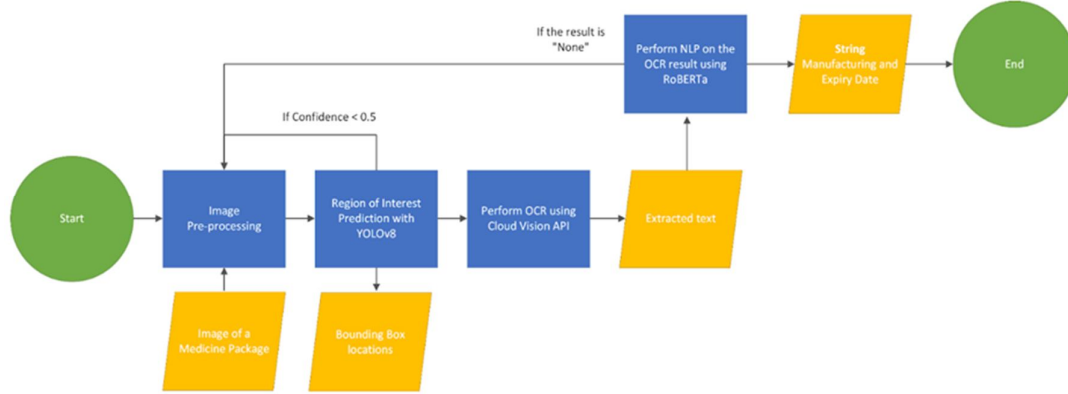


Figure 1. Workflow for the Proposed System

The given system contains three modules:

##### 1. Region of Interest Detection Module

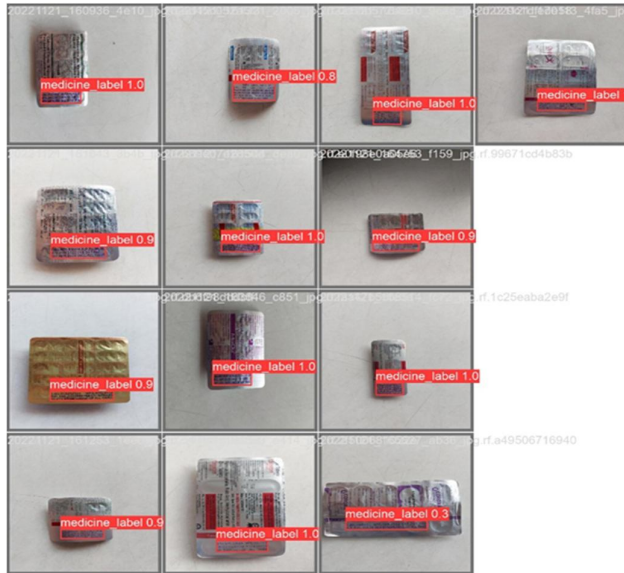


Figure 2. Region of Interest Detection



Figure 3. (a) Mis-predicted Data (b) Optimized Threshold using Rotational Approach

Our initial approach for detecting a specific pattern in the images was to label all the images and YOLOv8 to train the model to localize at one point. YOLOv8, short for You Only Look Once version 8, is an advanced object detection model used to detect objects accurately and quickly in images. YOLOv8 has been used in various applications, such as object detection in real-time video streams and the detection of defects in manufacturing processes. Its speed and accuracy make it an attractive option for developers and researchers who are engaged in the area of computer vision. The model's main strengths are its high speed and accuracy, making it an ideal choice for use in autonomous vehicles, robotics, and security systems. YOLOv8 is a deep neural network trained on large datasets of annotated images, such as COCO or Open Images, which allows it to detect

objects of various sizes, shapes, and colors with great precision. It can be used for various applications such as object detection and face recognition. This approach worked well when the image was properly aligned and the pattern was within the desired region of interest as shown in Fig. 2. However, we soon realized that this method was not ideal for images that were messy or when the desired pattern was in an opposing or different position, which can be noticed in Fig. 3(a).

To overcome this challenge, we employed a rotation approach that involves rotating the image until the bounding box prediction receives a threshold greater than 90%, indicating that it is in the correct location. This method proved to be effective in accurately detecting the desired pattern, even in images that were not properly aligned or had the pattern located in unexpected positions, as seen in Fig. 3(b).

## 2. Character Recognition Module

In order to achieve our main objective, the subsequent step is the vital process of Character Recognition and Segmentation. Fortunately, this step is relatively straightforward since we are already utilizing the Cloud Vision OCR. Cloud Vision OCR is a service provided by Google that allows users to extract and convert text from images into machine-readable text. It uses Optical Character Recognition (OCR) technology to analyze the image and identify the text. Our main objective in this step is to extract all the necessary information from the tablet packet, which is then returned to us in the form of a string. However, this seemingly simple step may not be without its complications. The extracted text may be noisy in nature, and therefore, it requires further processing before it can be used. This processing can be done using various methods like regular expression, or other artificial intelligence technologies to ensure that the extracted text is usable and accurate.

## 3. Large Language Model (LLM) Module

The Robustly Optimized BERT Approach (RoBERTa) module plays a critical role in our project post the character recognition and extraction step. RoBERTa is a pre-trained natural language processing (NLP) model created by Facebook AI and is an improved version of the original BERT (Bidirectional Encoder Representations from Transformers) model. On a variety of NLP tasks, including text categorization, question-answering, and language synthesis, it has demonstrated cutting-edge performance. One of RoBERTa's key strengths is its ability to generate high-quality datasets for specific tasks by learning the underlying patterns and relationships in the data. This approach can be particularly useful in situations where collecting or annotating data is difficult or expensive. Additionally, RoBERTa can be fine-tuned on a custom dataset using transfer learning, which improves performance on specific tasks. Prompt engineering is one such technique that fine-tunes RoBERTa to generate more accurate and relevant responses to specific prompts.

Despite having the OCR in place, the text string obtained is often messy, making it challenging to locate the desired text pattern, specifically the manufacturing and expiry date. Moreover, the packaging of medicine packets varies, with different formats, fonts, and characters, adding to the complexity. However, with the aid of this large language model, we can generate a dataset for prompt engineering by asking relevant questions like "when is the expiry," and the model can interpret the OCR's string, recognize the context, and produce the expiry date regardless of the font or format. To enhance the model's output based on the context given, we employ Transfer learning, further improving the accuracy of the system.

## B. System Implementation

The system works by providing a front-end platform for the model. This front end is connected to an API running on a virtual machine instance on the Google Cloud platform service. Users can either capture a new image in real-time or load an existing image from their library. Once the image is sent to the API, the detection process begins. The model searches for the region of interest in the image, which in this case, is the expiry date and the manufacturing date.

If the threshold received is less from the given input, the model performs a rotation to check for the threshold in every rotational angle and alignment. This allows it to find the region of interest quickly. Once the region of interest is found, it is cropped and sent to the OCR engine, which performs character recognition and extraction on the image. The OCR engine extracts all the text from the image in the form of a string and returns it back to the model.

However, the returned string may not be of perfect quality and may be noisy in nature depending on the pattern of the data given. To overcome this issue, the system employs the use of the NLP approach of large language models, specifically the RoBERTa model. This model is trained via prompt engineering to give the output for specific questions such as the expiry or manufacturing date, regardless of the format of the font of the text.

In some cases, the RoBERTa model fails to send back output and instead sends back "none". This is usually because the OCR model was not able to properly detect the text pattern in the image due to various reasons. In

such cases, the system loops back to the second stage where the recognition of the text is done using the YOLO model. It magnifies the image by 30% and sends it to the next character recognition stage, where it again tries to recognize the character pattern.

Once the output is received from the RoBERTa model, it is sent from the API to its application front and point, which is used to display the given output. The application can also be used for inventory management to keep a history of the medicine and to save the name of the medicines received because the model itself does not do the recognition of the medicine name.

## V. EXPERIMENTAL RESULT

The application displays the result on the front-end application. The model performs remarkably almost 80% of the time giving us the right output, but there are times where the model fails to get the result which could be because the image was too noisy or bad print of the text pattern on the medicine packet. Fig. 4 depicts the front-end application's main page where we have to provide an image of a medicine packet. Fig. 5 depicts the result page, which displays the picture with the position of the medicine label highlighted in the bounding box, the manufacturer, and the expiration date from the previously provided input image.

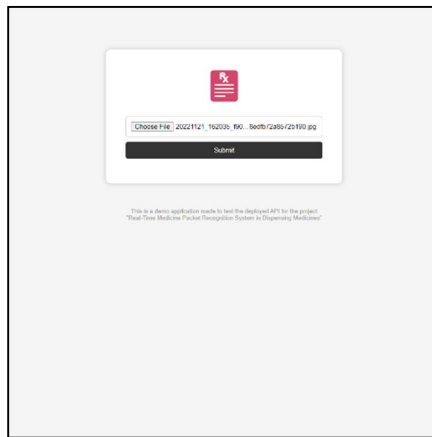


Figure 4. Home page

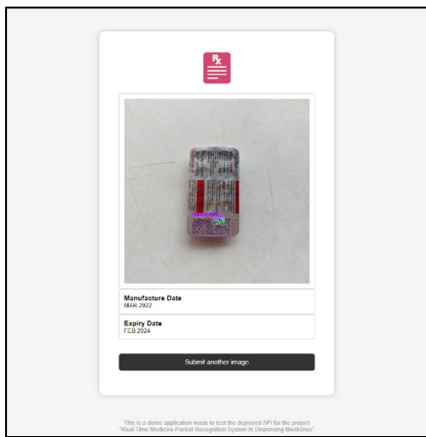


Figure 5. Result page

## VI. CONCLUSION

In conclusion, the system designed to detect the expiry and manufacturing date of tablets using deep learning, computer vision, and natural language processing has proved to be effective in locating dates with a success rate of at least 80%. Further improvement can be made by incorporating the extraction of tablet names into the model.

## REFERENCES

- [1] B. Lyu, Z.C. Wang, H.Y. Li, A. Tanaka, K. Furumoto, L. Meng, "Deep learning-based medicine packaging information recognition for medication use in the elderly," *2020 Int. Conf. On Identification, Information and Knowledge in the Internet of Things*, 2020.
- [2] L. Gong, M. Yu, W. Duan, X. Ye, K. Gudmundsson and M. Swainson, "A Novel Camera Based Approach for Automatic Expiry Date Detection and Recognition on Food Packages," *Artificial Intelligence Applications and Innovation. IFIP Advances in Information and Communication Technology*; Springer: Cham, Switzerland, 2018; pp. 133–142.
- [3] W. -J. Chang, L. -B. Chen, C. -H. Hsu, C. -P. Lin and T. -C. Yang, "A deep learning-based intelligent medicine recognition system for chronic patients," *IEEE Access*, 2019; 7:44441–58.
- [4] D. Singh, A. CV, M. Kaur, L. Meng, J. Shetty et al., "DKNet: Deep Kuzushiji Characters Recognition Network," *IEEE Access*, 2022, pp. 1–12.
- [5] A. VB, M. Javed, P. Parihar, A. K, D. CR, et al., "NLP Driven Ensemble Based Automatic Subtitle Generation and Semantic Video Summarization Technique," *Advances in Artificial Intelligence and Data Engineering*, Volume 1133, 2021.
- [6] K. Sugimoto, K. Inoue, Y. Kuroki, S. Kamata, "A Linear Manifold Color Descriptor for Medicine Package Recognition," *IEICE Trans. Info Syst.*, Volume E95-D (No.5), 2012.

- [7] A. Cunha, T. Adão, and P. Trigueiros, "HelpmePills: a mobile pill recognition tool for elderly person," *Procedia Technology* 16, pp. 1523–1532, 2014.
- [8] X. Yue, H. Li, K. Saho, K. Uemura, A. CV et al., "Machine Learning based Apathy Classification on Doppler Radar Image for Elderly Person," *2020 Int. Conf. On Identification, Information and Knowledge in the Internet of Things*, 2020.
- [9] C. V. Aravinda and H. N. Prakash, "Template Matching Method For Kannada Handwritten Recognition Based On Correlation Analysis," *2014 International Conference on Contemporary Computing Informatics (IC31)*, IEEE, 2014.
- [10] N. Nojiri, Z. Meng, K. Saho, Y. Duan, K. Uemura, A. CV, et al. "Apathy Classification Based on Doppler Radar Image for the Elderly Person," *Front Bioeng. Biotechnology*, Sec. Computational Genomics, Volume 8 – 2020.
- [11] K. Shetty and C. Aravinda, "An Effective Data Clustering System using Weighted K-Means and Firefly Optimization Algorithms," *2021 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, pp. 270–275, 2021.