Received: 23 February 2024 | Revised: 2 August 2024 | Accepted: 31 October 2024 | Published online: 9 December 2024

Artificial Intelligence and Applications
2024, Vol. 00(00) 1–12
DOI: 10.47852/bonviewAIA42022666

RESEARCH ARTICLE

# A Machine Learning-Based Approach for the Detection of Drugs in Drug Self

**Md. Nazmul Sakib[1] and Mahfuzulhoq Chowdhury[1,*]**

[1]*CSE Department, Chittagong University of Engineering and Technology, Bangladesh*

**Abstract:** Accurate detection of specific named medicines on drugstore shelves is critical for pharmaceutical inventory management. Existing research did not properly use machine learning technology for the accurate detection of medicine from drug samples. The purpose of this paper is to create a machine learning-based medicine detection system capable of automatically recognizing and localizing medicine boxes on shelves. The system uses Faster region-based convolutional neural network (R-CNN) and YOLOv5 to recognize and extract medicine boxes from images. Text recognition techniques, such as Tesseract OCR, are used to extract medicine names from boxes. This work includes collecting and annotating a dataset, training and evaluating models, and implementing text recognition algorithms. The simulation results show that the proposed faster R-CNN and Tesseract-based system is more accurate at detecting medicine boxes and extracting text than the existing Yolov5 and Tesseract-based systems.

**Keywords:** drug detection, object detection, machine learning, Faster R-CNN, YOLO, Tesseract, accuracy performance, pharmaceutical inventory management

## 1. Introduction

In Bangladesh, clinically inappropriate and ineffective medication use leads to serious health issues [1]. It is widely believed that more than half of all used drugs are delivered or sold incorrectly to customers. Inadequate access to medications and incorrect medication use result in significant morbidity and death. The World Health Organization provided appropriate guidance for each country's design of national drug use policies. As a result, despite all obstacles, Bangladesh is constantly working to reach the level. The government is now attempting to ensure proper monitoring of the pricing, safety, quality, and access issues of medicine for both pharmaceutical companies and customers. Some new medications are constantly being developed, while others, such as rofecoxib, are being phased out. Monitoring medication use, mandatory interdisciplinary activities, and adequate government funding are all critical to success. Koumpagioti et al. [2] found that the total dispensing error rate, prescribing error rate, and administration error rate (all within the total medication error) were 0.065, 0.342, and 0.316, respectively.

Gates et al. [3] discussed the disadvantages of medication error. Some of the errors are the result of the pharmacist providing the incorrect medicine. Illiterate people are unable to tell whether the medicine provided by the pharmacist is incorrect or correct. Pharmacists occasionally provide medicine as part of their business. Mekonnen et al. [4] found that thirty-three of fifty-one hospital cases in an African hospital were due to medication errors, while fifteen were due to adverse drug consumption issues. They also reported that the major reason for medication errors was due to inefficient use of medicine by practitioners (i.e., due to a lack of knowledge or skills), as well as environmental factors (i.e., unfavorable workplace and heavy workload). Tariq et al. [5] stated that at least 6,000 to 9,000 people die each year in the United States as a result of inefficient drug use. They also reported that at least one hundred thousand people had health problems (i.e., unreported cases) as a result of high dosage usage or a medication-related error. The associated cost of treating patients as a result of medication errors can be as much as USD 40 billion per year. The preceding discussion demonstrates that pharmaceutical errors (i.e., pharmacist errors) and the use of incorrect medication can cause serious harm to people's health and personal lives. Sometimes it happens because there aren't enough pharmacists in many pharmacies.

Several countries, including Japan, now rely on robots to run pharmacies and perform tasks such as medicine detection from the drug itself. This type of robot-based pharmacy requires an automated medicine detection system. An algorithm to detect drugs or medicine shelves in pharmacies is therefore required for an automated drug delivery and purchase process. To assist users with medicine box detection and name identification, a machine learning (ML)-based prediction system is required. To locate a specific medicine box on a bookshelf, the researchers must first identify all of the medicine boxes on the shelf before recognizing the names on the boxes.

### 1.1. Motivation

Manual inventory management in pharmacies and healthcare facilities can be time-consuming and error-prone, resulting in inefficiencies and patient safety risks. A highly accurate automated system is required to improve medicine detection and

*Corresponding author: Mahfuzulhoq Chowdhury, CSE Department, Chittagong University of Engineering and Technology, Bangladesh. Email: mahfuz@cuet.ac.bd

text recognition processes on medicine shelves in today's world. This work aims to streamline inventory management, reduce errors, and increase overall efficiency by developing an automated system that can accurately detect and recognize medicine boxes [6]. Furthermore, this research is motivated by the potential to apply ML and computer vision techniques to real-world healthcare challenges, thereby improving patient safety and healthcare operations. To address the aforementioned issues, this paper proposes a ML-based system for detecting medicines on medical shelves. The ultimate goal is to identify specific named medicines by analyzing images of medicine shelves. The automated system can be used to control quality in pharmaceutical manufacturing. By verifying the presence and correctness of medicine labels, the system can help identify packaging errors, such as incorrect labels or missing information, ensuring regulatory compliance and product quality. The automated drug detection feature of the drug self-system can aid in prescription verification by comparing extracted text from medicine boxes to the prescribed medication. This can help detect discrepancies, ensuring patients receive the correct medication based on their prescriptions, and lowering the risk of medication errors [7]. Counterfeit medicines pose a serious threat to public health.

By accurately detecting and recognizing medicine boxes, the system can assist in identifying counterfeit products by comparing label information to authorized databases. This can help to prevent the spread of counterfeit drugs and protect patients. The automated drug detection work can be used to build a robotic medicine dispensary shop [8].

The automated medicine recognition system enables a robotic arm or automated system to precisely locate and retrieve the requested medications based on the user's input. This can increase the efficiency, speed, and accuracy of medication dispensing in pharmacies, reducing the risk of human error and improving customer service. Data gathered during the detection and recognition processes can be used for research and analytics [9]. Furthermore, analyzing trends, medication usage patterns, and identifying common medication errors can provide useful information to healthcare providers, researchers, and policymakers. As a result of the preceding discussion, it is clear that a ML-based automated medicine box recognition system with high accuracy is extremely advantageous to a country like Bangladesh in terms of reducing human error and time spent searching for medicine boxes.

### 1.2. Limitation of the existing works and contribution of our work

There are several algorithms available for detecting objects like people such as Faster region-based convolutional neural network (R-CNN), YOLOv5, and SSD. To recognize text (e.g., name in the box), there are several models such as EAST Zhou et al. [10], Argman, CRAFT [11], CRNN scheme [12], Bgshih, FOTS [13], scene text detection scheme [14], MhLiao, and TPS-ResNet-BiLSTM-based framework [15]. For text extraction, Tesseract, Tesseract-Ocr has been utilized. However, most of the aforementioned work is limited to a small amount of dataset, with lower accuracy, recall, and precision values. Existing research has not proposed a suitable scheme for combining object detection and text recognition. The majority of these existing techniques were not proposed for use in medical name recognition or drug box detection. Furthermore, the proposed system's accuracy is extremely low. ML-based drug detection from drug self faces several challenges, including dataset collection, text extraction from medical boxes, dataset preprocessing, training and testing, and developing a combined system that detects and recognizes drugs.

To address current limitations, the primary goal of this work is to create a framework that detects and recognizes medicine boxes with high accuracy using ML technology. This work began with the collection of a large dataset of images of medical shelves. These images were meticulously annotated, with bounding boxes surrounding the medicine boxes and accompanying text. The dataset provided the foundation for creating and testing the object detection and text recognition models. The collected dataset was meticulously prepared, including several preprocessing steps. The dataset was divided into training and testing sets, and the images and annotations were checked for consistency and suitability. Data augmentation techniques increased the model's adaptability to real-world scenarios. The annotated dataset was used to train multiple advanced object detection models, including Faster R-CNN, SSD, and YOLOv5. The models used transfer learning to refine pre-trained weights from large datasets for the medicine detection task. This method produced accurate and efficient models for detecting medicine boxes with remarkable precision. Tesseract is a popular open-source OCR engine for extracting text from images. It supports a variety of languages and has accurate text recognition capabilities. Tesseract allows us to extract textual information from medicine boxes, such as names, dosages, and other details. The addition of Tesseract to our methodology enables us to accurately recognize and extract text from detected medicine boxes. This text contains useful information for future processing and analysis.

The framework's performance was thoroughly evaluated, with rigorous tests conducted to determine its accuracy, speed, and robustness. The model's performance was accurately assessed using evaluation metrics such as precision, recall, and mean average precision.

The key contributions of the proposed scheme are detailed below:

1) This paper collects a wide range of images of medicine shelves to ensure a representative sample of different shelf layouts, lighting conditions, and medicine box variations. This paper also meticulously annotated the images, marking the bounding boxes around each medicine box and labeling the boxes with the appropriate text. This paper provides a solid foundation for training and evaluating models by creating a comprehensive dataset.
2) This paper implements both Faster R-CNN and YOLOv5 models for object or medicine box detection work, using popular deep learning frameworks such as TensorFlow and PyTorch. This work also fine-tuned these models on the annotated dataset, optimizing their performance for accurate medicine box detection.
3) Following a thorough evaluation, this paper compares the performance, speed, and accuracy of these (Faster R-CNN, YOLOv5) models, analyzing their strengths and limitations in the context of medicine detection.
4) After comparing the accuracy values, this paper employs Faster R-CNN for medicine box detection and Tesseract OCR for medicine name recognition (text recognition) in the medicine box. This paper also includes performance comparisons with existing schemes.

The literature review will then be presented in Section 2. The proposed system is illustrated in Section 3. Section 4 provides the evaluation results. Section 5 includes the conclusion and future work.

## 2. Literature Review

Currently, some works in the literature focus on recognizing medicine. Ting et al. [16] proposed a method for detecting medicine names using deep learning. This model outperformed traditional computer vision systems at recognizing pharmaceuticals. The accuracy of their back-side detection model is 95.99%, while the front-side detection model is 93.72%. Their proposed model can only detect medicines. Their model simply assists the pharmacist in detecting the medicine correctly. However, it cannot detect medicines when they are present in large quantities.

Chen et al. [17] used a variety of image features (including color and shape) to distinguish between drug images. In their tests, the system can identify the top ten comparable medications from a picture. This system enables a user to locate a similar drug based on an image of a particular drug. Rupa et al. [18] proposed a system for recognizing medicines. Their proposed method is a mobile app that extracts text from medications using an optical character recognition (OCR) algorithm and the device's native camera. The extracted text is further analyzed using natural language processing techniques. They did not specify the dataset value or the accuracy value.

They used the DNR database to identify and categorize medication names. The app then generates an augmented reality model to explain why the medicine is used. This system works exclusively with 3D images from their database. It also recognized one medicine at a time. Some works in literature concentrate on detecting books from bookshelves. Sawaki et al. [19] proposed using multiple dictionaries to detect low-quality characters in scene photos. They began by estimating an input image's environmental conditions using an initial dictionary. The estimation is then used to automatically select and use an appropriate dictionary from a large number of dictionaries representing various environmental conditions for recognition. Their findings revealed that their proposed strategy outperformed the current single dictionary-based approach (76.4%) in terms of recognition rate (89.8%).

Thi et al. [20] developed a system that extracts Vietnamese book titles from book cover images. To detect text, they used a single-shot arbitrarily shaped text scheme (SAST) that was built on a fully convolutional network. They also used Transformer OCR to recognize text. They compiled a dataset by combining nearly 7800 book cover images. They were 80.7 percent accurate in text detection and 82.4 percent accurate in text recognition. In Bangdiwala et al. [21], the authors created a book recognition system using Tesseract OCR modules. They achieved an 82.5 percent accuracy rate in book recognition. The sample size is also small (80 book images).

In Huang et al. [22], the authors used a Baidu OCR-based character recognition system to identify medicine boxes with an accuracy of 88.3%. Their dataset has a small number of images (less than 100). Magalhães et al. [23] used OCR technology to recognize medicine names from drug images. They had an 80% accuracy rate in detecting drug names. Their dataset includes two hundred eleven images of fifty-two different drugs. Fu et al. [24] developed a food ingredient recognition system from food images using CNN technology. Their single ingredient dataset contains 10,750 food images and achieves 86% accuracy with a ResNet-18-based CNN model. Datta et al. [25] used regions with convolutional neural networks to detect road objects from real-time video data and nineteen object categories. Their proposed method had a detection accuracy of 86%. They tested a total of 199 instances. Sarika et al. [26] presented a CNN-based OCR model for handwriting recognition. Their dataset includes 1600 Telugu characters, and handwritten character recognition has an accuracy of nearly 90%. Fatema et al. [27] developed a Tesseract OCR-based book category detection system that achieved an overall accuracy of 82%. However, the dataset is extremely small (e.g., 50 book images). They did not work with both the medicine box detection and recognition systems. Yu et al. [28] developed a Chinese scene text detection method based on the universal attention mechanism and YOLOv5. The proposed method has a precision of 67.7% and a recall of 51.4%. They used an RCTW-17 dataset of twelve thousand two hundred and sixty-three images to detect scene text. They did not perform any text recognition work. Wang et al. [29] used YOLO v3 with GCN (graph convolution neural network) to detect straw-burning flames in images, achieving an accuracy of nearly 76%. Their dataset includes 5,000 images. Punia et al. [30] studied support vector machine (SVM)-based OCR model. They did not specify how many images were in the dataset or how accurate their system was. They only talked about the theoretical architecture of handwritten character recognition with the SVM classifier. Suwattanapunkul et al. [31] used the YOLO model and a hybrid dataset to identify traffic signs in Taiwan. They worked with an 837-image dataset of Taiwan traffic signs. They showed that their proposed hybrid model achieves 86 percent accuracy and 65 percent recall for traffic sign recognition.

Quoc et al. [32] proposed a framework for identifying books on shelves based on the titles on the spines of volumes. The framework has two modules: control and recognition. The control module positions the camera for image capture, while the recognition module examines the captured images to determine which books are on the shelves. The initial photographs in this series were taken at random locations. Then they separate it into two sections: book section and non-book section. They also applied Otsu's thresholding method to the book title removal and recognition processes.

Tabassum et al. [33] proposed a technique for detecting book spines. Their method can also detect book titles in bookshelf images (multiple rows). The dataset is also quite small (less than fifty images). They did not provide accurate results for the proposed method. Yang et al. [34] used supervised deep learning techniques for text recognition and book spine retrieval activities. They didn't work on medicine box text detection. In Xia et al. [35], the authors developed a deep learning framework for detecting and segmenting objects in chromium-based super alloys. They used the YOLO v5 and SegFormer structures to identify and segment objects in EM images, respectively. To aid in wildfire event prediction, Cheng et al. [36] used three-dimensional auto encoders to generate spatial-temporal features of wildfire regions. Cheng et al. [37] used a generalized latent assimilation approach to find appropriate parameters in a wildfire prediction model.

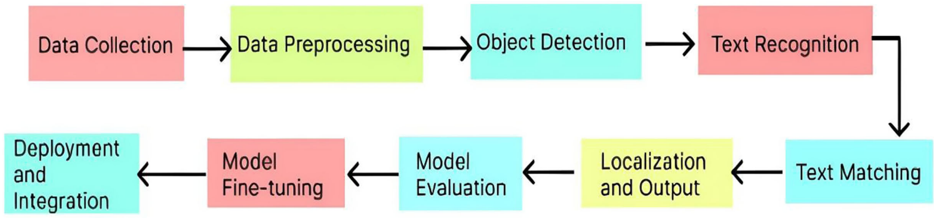
The preceding discussion shows that previous research did not employ ML techniques to investigate both drug name detection and drug self-recognition. Furthermore, several related works use very small datasets. Furthermore, previous research did not yield a real-time dataset for drug recognition using drug self-reports. Furthermore, medical detection and recognition work has a very low accuracy rate. Unlike previous research, this study presented a high-accuracy ML-based drug detection approach from the drug itself, employing a faster R-CNN for medicine box detection and a Tesseract OCR scheme for medicine name recognition from the medicine box.

## 3. Proposed Framework

Figure 1 devises the proposed ML-based framework for the detection and recognition of medicine boxes from a medicine shelf. The framework utilizes some highly efficient object detection and text recognition models to automate the process of identifying specific medicines from a collection of medicine boxes.

By leveraging advanced computer vision techniques, the framework enables efficient and accurate detection of medicine

**Figure 1**
**Design overview of the proposed model**




boxes and extraction of text information from these boxes. This automated approach offers significant advantages over manual inspection, saving time, and reducing the risk of human error. The proposed system is developed to detect and recognize medicine boxes from a drug shelf, extract the text information, and match it with the given medicine name. A flowchart diagram of the whole process is given in Figure 2.

### 3.1. System design

This work begins with the collection of a comprehensive dataset consisting of images containing medicine shelves. These images were carefully annotated, marking the bounding boxes around the medicine boxes and their corresponding text. The dataset served as the foundation for training and evaluating the object detection and text recognition models. The collected dataset was meticulously prepared, undergoing a series of preprocessing steps. We have split the dataset into a training dataset and a testing dataset. We have ensured the consistency and compatibility of the images and annotations. To enhance the generalization and handling work regarding variations in real-world scenarios, we have used data augmentation techniques. Data augmentation helps to reduce overfitting by increasing the effective size and diversity of the training dataset. The model is exposed to a wider range of examples by introducing various transformations such as flipping, rotating, and adding noise. This forces the model to learn broader features rather than memorizing specific details from the training samples. Data augmentation improves robustness by simulating variations during the training process. For example, in medical box recognition tasks, adding random noise, changing brightness, or applying slight blurs to images during training can prepare the model to deal with similar disturbances in real-world images.

Multiple object detection models (i.e., including both Faster R-CNN and YOLOv5) are implemented and trained on the annotated dataset. The models underwent transfer learning, leveraging pre-trained weights from large-scale datasets and fine-tuning them specifically for the medicine detection task. This approach resulted in highly accurate and efficient models capable of detecting medicine boxes with remarkable precision. Tesseract is a widely used open-source OCR engine capable of extracting text from images. It supports multiple languages and provides accurate text recognition capabilities. By leveraging Tesseract, we can extract the textual information present on the medicine boxes, including medicine names, dosages, and other relevant details. The integration of Tesseract into our methodology enables us to accurately recognize and extract the text from the detected medicine boxes. This text will serve as valuable information for further processing and analysis. The framework's performance was

**Figure 2**
**Methodology of the proposed model**

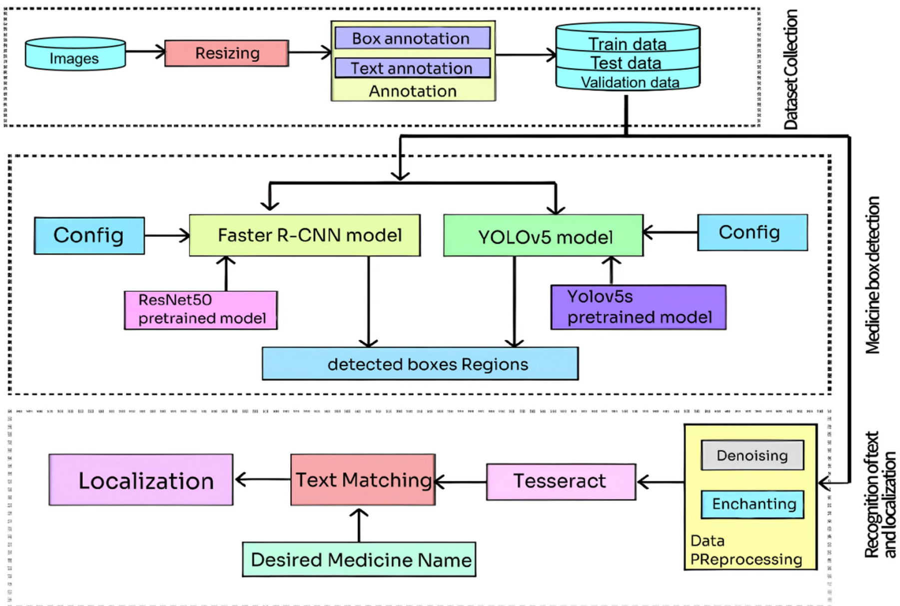


**Figure 3**
**Image of medicine shelf**


extensively evaluated, employing rigorous assessment measures to determine its accuracy, speed, and robustness. Performance evaluation metrics like precision, accuracy, recall, and mean average precision were utilized to check whether the models perform accurately. Real-world scenarios were also simulated to validate the framework's effectiveness and usability in practical applications.

## 3.2. Detail explanation of the proposed system

The methodology for this work involves several key steps to achieve drug detection on a drug shelf using ML. The overall process can be divided into the following stages.

### *3.2.1. Dataset development*

In this work, a custom dataset was collected by visiting medicine dispensary shops and capturing images using a personal camera. The dataset acquisition process involved the following steps. A diverse set of images (a total of 1749 images) was captured by taking photographs of medicine shelves from different angles and distances. These images were captured using a digital camera or a smartphone, ensuring a diverse range of perspectives and conditions to mimic real-world scenarios. Figure 3 shows the sample medicine box images of the medicine shelf. Each image in the dataset features a medicine shelf with multiple medicine boxes arranged in different configurations. The medicine boxes vary in size, shape, and label design, representing a wide range of medication types, such as tablets, capsules, syrups, and ointments. The collected images were annotated using the VGG Image Annotator (VIA) tool. Bounding boxes were drawn around each medicine box to mark the regions of interest (ROI) on the shelves. These ROIs were labeled as "medicine boxes" to indicate their significance.

The text areas within the bounding boxes were labeled as "name region" to identify the portion where the medicine names were located. This annotation allowed for precise text recognition during the model training phase. This annotation process captures the precise location and extent of each medicine box within the image. There are some examples of annotated images in Figure 4.

**Figure 4**
**Annotated image of medicine shelf**

To ensure consistency and compatibility, the dataset was standardized. All images were resized to a resolution of 600 × 600 pixels, maintaining a uniform format across the dataset. Additionally, the images were renamed systematically to maintain order and facilitate data management. The annotations were saved in the comma-separated values (CSV) format, as supported by VIA. This format allowed for easy organization and retrieval of the annotation information. Each row in the CSV file corresponded to an image, listing the image filename, image width, and height, and the details of the annotated regions, including the bounding box coordinates and corresponding labels. The dataset includes not only training data but also both validation and testing subsets. The training set can be utilized regarding training the model. The validation set assisted in hyperparameter tuning and performance evaluation. The testing set provided an unbiased assessment of the final model. By completing these steps, a comprehensive dataset was developed, serving as a solid foundation for training the model. The dataset included a diverse range of medicine box images, accurately annotated bounding boxes, and labeled text. This dataset ensured the model's ability to detect and recognize medicine boxes from drug shelves accurately.

*3.2.2. Medicine box detection*

To detect the medicine box region on every shelf, we have examined the most popular object detection model such as the Faster R-CNN model and YOLOv5 model. Faster R-CNN (region-based convolutional neural network) is known as an object detection framework. The faster R-CNN includes main components like a region proposal network or RPN and a region-based convolutional network. The RPN can generate results like region proposals that may contain objects. The RPN operates on a sliding window of different sizes and its respective aspect ratios across the full image. It predicts the likelihood of an object's presence within each region. These proposed regions serve as potential candidate regions for object detection. Once the region proposals are obtained, the region-based convolutional network is applied to each proposal to perform object detection. This network is typically a CNN (convolutional neural network) architecture, such as ResNet, VGG, or similar networks. The region-based convolutional network extracts features from each proposed region and performs classification to determine the presence or absence of objects, as well as regression to refine the bounding box coordinates of the detected objects. The faster R-CNN training process is done by using two stages. In the 1st stage, the RPN is trained to generate accurate region proposals using labeled bounding boxes from the training data. In the 2nd stage, the region-based convolution network is fine-tuned using the proposals generated by the RPN, along with their corresponding ground truth labels. The entire framework is trained by using a combination of classification functions along with the bounding box regression loss functions. We installed the required module like tensorflow, tensorflowaddons, and tf-models-official for faster R-CNN implementation. We used pre-trained weights for the Faster R-CNN model, which has been trained on large-scale datasets like COCO. These weights will provide a good starting point. We used ResNet-50. We configured hyperparameters as follows. The learning rate is 0.001, batch size is 8. The num train step is 1000. We converted the annotated CSV file into the required TFRecord format for training. We implement the loss function for the Faster R-CNN model. We used keras loss reduction sum for loss reduction. Then, we trained the model by using the following parameters. The parameter learning rate is 0.001. The batch size is 8. The num train step is 1000. We evaluated the model using the training dataset and evaluation metrics. We prepared a separate testing dataset that contain unseen images (i.e., not used during training or validation step). We prepared the text record format for testing. Then we evaluate the performance of the proposed model by using the accuracy metric, precision metric, recall metric, and $F1$-score metric. Now, we will discuss YOLOv5 implementation. YOLOv5 is an object detection algorithm introduced by Ultralytics. It is the latest iteration in the YOLO (You Only Look Once) series of models. We installed the YOLOv5 framework in our Colab environment. We cloned the YOLOv5 repository from GitHub. The dataset is in the required format, such as YOLO annotation format (.txt files with bounding box coordinates). We modified the data.yaml file to specify the paths to our dataset, including the training and validation sets in the yolov5/directory. We trained the YOLOv5 model on our dataset for 50 epochs, where the batch size is 16 and the input image size is 600 × 600. We evaluated the model using the training dataset and evaluation metrics. We prepared a separate testing dataset that contains unseen images (not used during training or validation step). We prepared a text format for testing. Then, we evaluate the performance by using precision metric, recall metric, and $F1$-score metric. After the model was trained and evaluated, you used it to detect medicine boxes in new images.

*3.2.3. Text recognition using Tesseract*

Tesseract is known as an OCR engine which is open source and developed by Google. It is widely used for text recognition and extraction from images or scanned documents. Tesseract utilizes the deep learning-based approach for OCR. It is generally based on RNN or recurrent neural networks and LSTM or long short-term memory units. Tesseract enables the recognition of sequences of characters. We applied necessary image preprocessing techniques, such as denoising, enhancing the image to improve text recognition accuracy. Then we converted the image to grayscale for better text extraction. We used OpenCV for preprocessing the image. We perform text detection using an object detection model yolov5 to identify the ROI containing text on the medicine boxes applying the previous method for box detection. Then we extracted the bounding box coordinates for each ROI detected. We cropped the ROI regions from the original image based on the extracted bounding box coordinates, then applied Tesseract OCR engine to recognize the text within each cropped ROI, and then retrieved the recognized text output from Tesseract OCR.

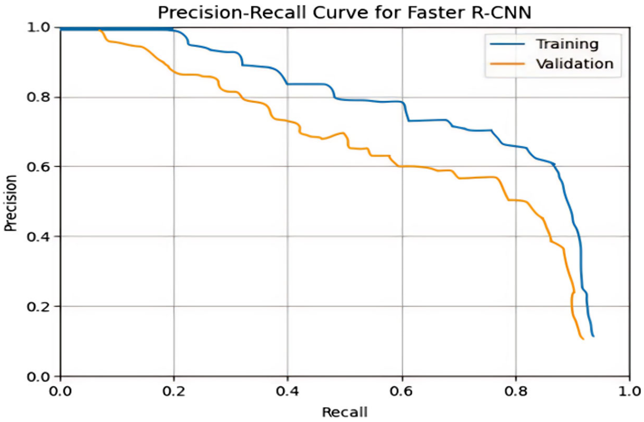
*3.2.4. Text matching and localization*

We used a string-matching algorithm to match text with the desired medicine name. Then, we selected the text region of the matched text. Next, our work retrieved the coordinates (bounding box) of the matched text region. We use the bounding box coordinates to extract the corresponding region (i.e., from the original image). We got the bounding box regions of the boxes from the medicine box detection step.

## 4. Evaluation Results

To evaluate our proposed system, we used performance evaluation matrices like accuracy metric, precision metric, recall metric, and $F1$ score metric. Precision can evaluate a model's ability to accurately identify a sample as positive. Recall can measure a model's ability to correctly identify and capture the positive samples in the dataset. The $F1$ score combines both precisions and recalls into a single value (harmonic mean). Accuracy measures the overall correctness of the model's

**Figure 5**
**Precision recall results regarding Faster R-CNN**




predictions. In this work, true positive (TP) means positive boxes that are correctly detected and positive text that are correctly recognized.

True negative (TN) means nonboxes region that is correctly detected as empty areas and Non-matched text that is correctly recognized as not matched. False positive (FP) means nonboxes region that is detected as boxes and non-matched text that is recognized as matched text. False negative (FN) means true boxes region that is detected as an empty area and true matched text that is recognized as not matched. The accuracy, precision, recall, and $F1$ score performance metrics are calculated by using following equations as follows:

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

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

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

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

Figure 5 represents the performance based on the precision-recall curve of the model when using a faster R-CNN model for box detection. The precision-recall curve of the model reveals a strong performance in terms of accurately detecting medicine boxes. The precision value indicates a high level of confidence in the model's predictions, as it has a low false positive rate. This is crucial for ensuring that the identified medicine boxes are indeed accurate and reliable. The recall value, although slightly lower, suggests that the model can capture a significant portion of the true positive instances. This combination of high precision and reasonable recall underscores the effectiveness of the model in accurately identifying medicine boxes. Overall, the precision-recall analysis demonstrates the model's capability to provide reliable and accurate results, making it a valuable tool for medicine detection tasks.

The accuracy curve regarding the Faster R-CNN model is visualized in Figure 6. The curve shows an upward trend, indicating that the model is effectively learning to identify and classify medicine boxes. The curve demonstrates how precision varies at different recall levels, indicating the model's ability to accurately detect medicine boxes while minimizing false positives. This high-accuracy value reflects the model's ability to accurately detect the presence of medicine boxes with a strong level of certainty. Overall, the accuracy curve demonstrates the effectiveness of the proposed model or Faster R-CNN model in achieving reliable and accurate results in medicine box detection. The precision curve for YOLOv5 is shown in Figure 7. As the number of epochs increases, both the training and validation precision exhibit a positive trend, indicating the model's ability to accurately detect medicine boxes. This suggests that the model is continuously learning and improving its precision over time. The consistent growth in precision demonstrates the effectiveness of YOLOv5 in accurately identifying medicine boxes, making it a reliable choice for our work.

The analysis of the recall curve (in Figure 8) for YOLOv5 in our work provides valuable insights into the model's performance in terms of recall. As the epoch number increases, both the training and validation recall values show an upward trend, indicating the model's ability to capture a higher proportion of positive instances (i.e., medicine boxes) in the dataset. This suggests that the model becomes more effective over time in identifying and recalling

**Figure 6**
**Accuracy results regarding Faster R-CNN**

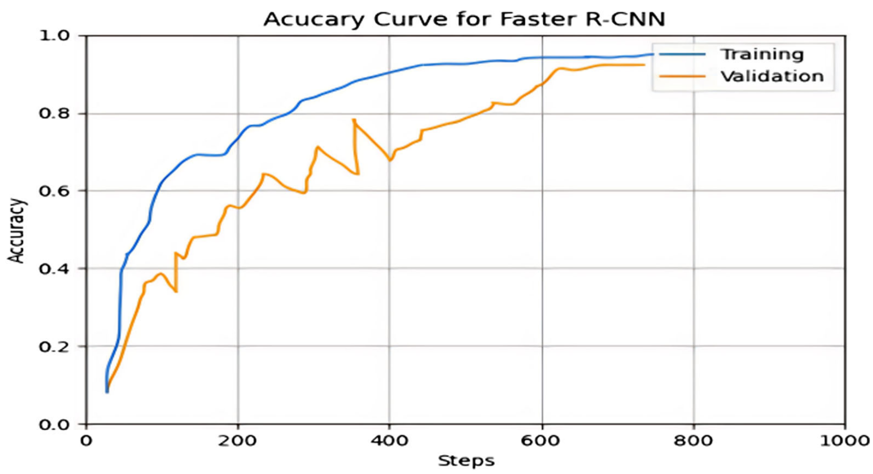


**Figure 7**
**Precision curve of yolov5**

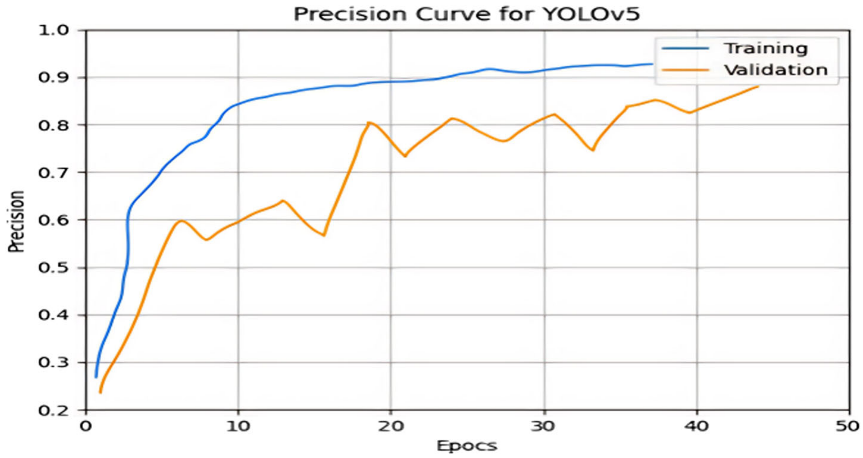
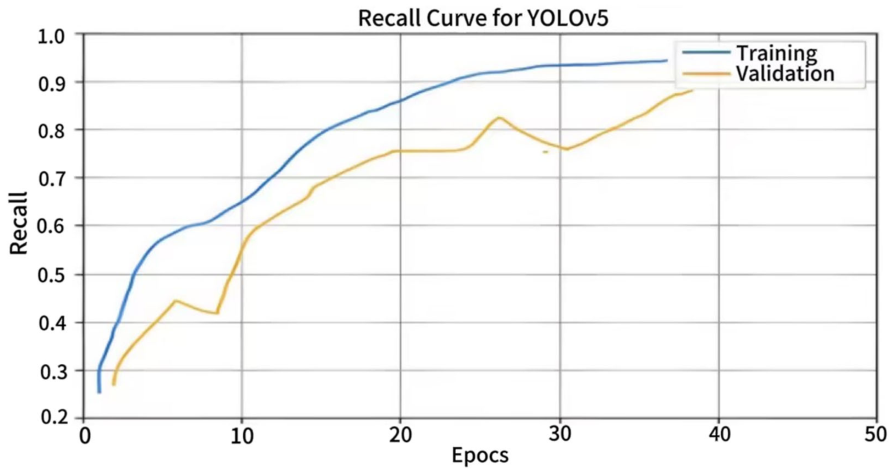



**Figure 8**
**Recall curve of yolov5**




medicine boxes. The consistent improvement in recall values indicates the robustness of YOLOv5 in capturing a larger number of relevant instances, making it a suitable choice for our work's requirements.

The accuracy curve for YOLOv5 (see Figure 9) provides insights into the model's performance in terms of accuracy. As the number of epochs increases, both the training and validation accuracy values show a gradual improvement, indicating the model's ability to correctly classify and localize medicine boxes. The increasing accuracy values demonstrate the effectiveness of YOLOv5 in accurately detecting and identifying medicine boxes in the images. The high-accuracy values achieved by the model validate its capability to accurately recognize medicine boxes in our work.

Table 1 shows that the faster R-CNN works better in comparison with yolov5. Faster R-CNN achieves higher accuracy by utilizing a two-stage detection approach. It incorporates an RPN to generate region proposals and then performs object detection within these proposed regions. This two-stage architecture allows for better accuracy as it refines the region proposals before performing object classification and localization. However, the YOLOv5 model relies on a one-stage object detection algorithm that could strike a balance between speed and accuracy. From a grid-based approach, it can predict the bounding boxes and class probabilities directly. Thus, YOLOv5 results in faster inference times but with potentially slightly lower accuracy compared to two-stage methods like Faster R-CNN. The output bounding box of a desired image is shown in Figure 10. This model shows a bounding box in the image of a desired medicine and gives the coordinate of the medicine box. Table 2 gives the system accuracy comparison results. Table 2 demonstrates that Faster R-CNN provides higher accuracy in medicine box detection, resulting in more precise identification of the target objects. Please also note that the Faster R-CNN model is relatively slower in terms of processing time due to its two-stage detection model. Whereas, the YOLOv5 the model exhibits faster inference speed due to its single-stage architecture, making it more suitable for real-time applications. However, it is associated with a slightly lower accuracy compared to Faster R-CNN. Furthermore, Table 2 also shows that the integration of text recognition algorithms, such as Tesseract OCR enhances the system's capabilities in extracting medicine names from the detected boxes. R-CNN performs better in object detection tasks due to its refined region proposal mechanism. YOLOv5 is

**Figure 9**
**Accuracy curve of yolov5**

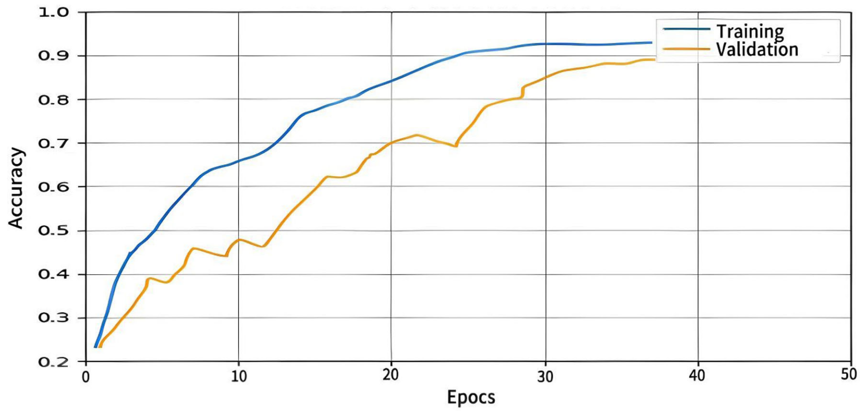



**Table 1**
**Evaluation results**

| Framework | Accuracy | Precision | Recall | *F*1 score |
|---|---|---|---|---|
| Faster R-CNN | 0.92 | *0.85* | *0.92* | *0.88* |
| YOLO v5 | 0.89 | 0.87 | 0.88 | 0.88 |
| Tesseract | 0.88 | 0.83 | 0.88 | 0.83 |

**Table 2**
**Overall system accuracy comparison**

| Box detection framework | Text detection framework | Proposed system accuracy |
|---|---|---|
| Faster R-CNN | Tesseract | *86%* |
| YOLO v5 | Tesseract | 82% |

generally faster but less accurate than Faster R-CNN, which performs better with smaller datasets. To determine the computational time, we ran 100 simulations and averaged the computational time for the box and text recognition trials. The average computational time for R-CNN-based activity is 64.5 ms, while Tesseract-based activity takes 6.2 ms (on average). Our system has a computational complexity of O (m*n), where m is the total number of images and *n* is the total number of steps in our proposed system for detecting medicine boxes and text recognition.

Table 3 compares our proposed scheme with existing works. For this simulation, we used our own dataset. The table shows that the proposed scheme (with faster R-CNN and Tesseract)

**Figure 10**
**Output bounding box of a desired image**

**Table 3**
**Comparison with existing works**

| Scheme | Method used | Accuracy |
|---|---|---|
| Thi et al. [20] | Single-shot arbitrarily shaped text scheme (SAST) based on a fully convolutional network | 80.7 percent (text detection accuracy) |
| Yu et al. [28] | Universal attention with YOLOv5 | 82 percent (text detection accuracy) |
| Wang et al. [29] | YOLO V3 with GCN | 76 percent (text detection accuracy) |
| Proposed scheme (faster R-CNN with Tesseract) | Faster R-CNN with Tesseract | 92 percent medicine box detection accuracy and 86% overall accuracy (box and text recognition) |

outperforms the compared works in terms of drug box detection and recognition accuracy. It should be noted that these works only detect text. However, unlike previous works, our proposed scheme detects and recognizes drug names from medicine boxes in the drug itself. Table 3 shows that Thi et al. [20] achieved an 80.7 percent text detection accuracy using the SAST (single-shot arbitrarily text-shaped) scheme and a convolution network. Yu et al. [28] achieve an 82% text detection accuracy using yolov5.

Wang et al. [29] used YOLOV3 with a GCN scheme to achieve a text detection accuracy of 76%. In contrast, our proposed scheme combines faster R-CNN with the Tesseract method. In our proposed scheme, the accuracy of detecting medicine boxes is 92%, and the overall system accuracy is 86%. The results clearly indicate the proposed system offers at least 4 percent better accuracy than the existing works.

## 5. Conclusion

This paper describes a ML-based medicine or drug detection system for medical shelves. The implemented framework, which included the Faster R-CNN, YOLOv5, and Tesseract models, produced outstanding accuracy and performance metrics. The Faster R-CNN model enabled precise detection of medicine boxes, allowing for efficient inventory management and quality control in pharmaceutical settings. The YOLOv5 model offered a quick and accurate alternative to object detection, broadening the possibilities for real-time applications. The use of Tesseract for text recognition aided in the extraction of medicine names from detected boxes, thereby contributing to prescription verification and counterfeit detection. The system showed consistent performance in recognizing small and varying text sizes, increasing its usefulness in real-world scenarios. The evaluation results showed that the Faster R-CNN model can achieve high accuracy of 92% as well as satisfactory precision, recall, and $F1$ scores for medicine box detection. The results show that the proposed system is at least 4% more accurate than previous works. Future work may include further model optimization, integration with robotic systems, and scalability for large-scale deployment. Another future research focus would be to collect a larger and more diverse dataset that includes various medicine boxes, labels, and text variations that can aid in addressing issues such as different font styles, languages, and packaging designs. This work can be used to detect real-time medicine box names from drug labels. In the future, we hope to learn more about system privacy analysis, more noisy scenarios, advanced data augmentation techniques, spatial-temporal feature analysis with deep learning, advanced parameter selection techniques, and more difficult scenarios. In the future, a user-friendly interface (e.g., a mobile application) that allows for simple interaction and seamless integration with existing pharmacy management systems can improve usability and adoption. Future work should also focus on meeting regulatory and ethical requirements, as well as ensuring data privacy and security through the use of advanced blockchain and deep learning techniques.

## Recommendations

Our finding revealed that faster R-CNN model can be suitable for detection of drugs in drug self.


## Acknowledgement

The authors are grateful to CUET, CSE department faculties for their necessary suggestions.


## Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

## Data Availability Statement

The GitHub data that support the findings of this study are openly available in https://github.com/argman/EAST. The GitHub data that support the findings of this study are openly available in https://github.com/clovaai/CRAFTpytorch. The GitHub data that support the findings of this study are openly available in https://github.com/bgshih/crnn. The GitHub data that support the findings of this study are openly available in https://github.com/MhLiao/TextBoxes. The GitHub data that support the findings of this study are openly available in https://tesseract-ocr.github.io/tessdoc/. The GitHub data that support the findings of this study are openly available in https://github.com/tesseract-ocr/tesseract.

## Author Contribution Statement

**Md. Nazmul Sakib:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Visualization. **Mahfuzulhoq Chowdhury:** Conceptualization, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration.

## References


[1] Chowdhury, F. R., Rahman, M. M., Huq, M. F., & Begum, S. (2006). Rationality of drug uses: Its Bangladeshi perspectives. *Mymensingh Medical Journal*, *15*(2), 215–219.

[2] Koumpagioti, D., Varounis, C., Kletsiou, E., Nteli, C., & Matziou, V. (2014). Evaluation of the medication process in pediatric patients: A meta-analysis. *Jornal de pediatria*, *90*(4), 344–355. https://doi.org/10.1016/j.jped.2014.01.008

[3] Gates, P. J., Baysari, M. T., Mumford, V., Raban, M. Z., & Westbrook, J. I. (2019). Standardising the classification of harm associated with medication errors: The harm associated with medication error classification (HAMEC). *Drug Safety*, *42*(8), 931–939. https://doi.org/10.1007/s40264-019-00823-4

[4] Mekonnen, A. B., Alhawassi, T. M., McLachlan, A. J., & Brien, J. A. E. (2018). Adverse drug events and medication errors in African hospitals: A systematic review. *Drugs-Real World Outcomes*, *5*, 1–24. https://doi.org/10.1007/s40801-017-0125-6

[5] Tariq, R. A., Vashisht, R., Sinha, A., & Scherbak, Y. (2024). *Medication dispensing errors and prevention*. USA: StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK519065/

[6] Kumar, S., & Chuli, A. (2023). Optimizing pharmaceutical inventory management with YoloV7 and easy OCR on medicine strips. *International Journal of Science and Research*, *12*(8), 1662–1669. http://dx.doi.org/10.21275/SR23804233813

[7] Nayak, N., Prarthana, T., Joshi, R., Vaibhavi, S., & Swathi, S. (2023). Medical prescription recognition using machine learning: A survey. *International Research Journal of Modernization in Engineering Technology and Science*, *5*(4), 7279–7284.

[8] Martinez-Martin, E., Ferrer, E., Vasilev, I., & Del Pobil, A. P. (2021). The UJI aerial librarian robot: A quadcopter for visual library inventory and book localisation. *Sensors*, *21*(4), 1079. https://doi.org/10.3390/s21041079

[9] Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., & Ahsan, M. J. (2022). Machine learning in drug discovery: A review. *Artificial Intelligence Review*, *55*(3), 1947–1999. https://doi.org/10.1007/s10462-021-10058-4

[10] Zhou, X., Yao, C., Wen, H., Wang, Y., Zhou, S., He, W., & Liang, J. (2017). East: An efficient and accurate scene text detector. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5551–5560.

[11] Baek, Y., Lee, B., Han, D., Yun, S., & Lee, H. (2019). Character region awareness for text detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9365–9374.

[12] Shi, B., Bai, X., & Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(11), 2298–2304. https://doi.org/10.1109/TPAMI.2016.2646371

[13] Liu, X., Liang, D., Yan, S., Chen, D., Qiao, Y., & Yan, J. (2018). FOTS: Fast oriented text spotting with a unified network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* 5676–5685.

[14] Liao, M., Shi, B., & Bai, X. (2018). Textboxes++: A single-shot oriented scene text detector. *IEEE Transactions on Image Processing*, *27*(8), 3676–3690. https://doi.org/10.1109/TIP.2018.2825107

[15] Bonini, A. (2018). Supersymmetric 4d gauge theories and Integrability. *arXiv Preprint:1807.09695*. https://doi.org/10.48550/arXiv.1807.09695

[16] Ting, H. W., Chung, S. L., Chen, C. F., Chiu, H. Y., & Hsieh, Y. W. (2020). A drug identification model developed using deep learning technologies: Experience of a medical center in Taiwan. *BMC Health Services Research*, *20*, 1–9. https://doi.org/10.1186/s12913-020-05166-w

[17] Chen, R. C., Pao, C. T., Chen, Y. H., & Jian, J. C. (2010). Automatic drug image identification system based on multiple image features. *Computational Collective Intelligence. Technologies and Applications*, *2*(2), 249–257. https://doi.org/10.1007/978-3-642-16732-4_27

[18] Rupa, C., Srivastava, G., Ganji, B., Tatiparthi, S. P., Maddala, K., Koppu, S., & Chun-Wei Lin, J. (2022). Medicine drug name detection based object recognition using augmented reality. *Frontiers in Public Health*, *10*, 881701. https://doi.org/10.3389/fpubh.2022.881701

[19] Sawaki, M., Murase, H., & Hagita, N. (1998). Character recognition in bookshelf images by automatic template selection. In *Proceedings Fourteenth International Conference on Pattern Recognition*, *2*, 1117–1120. https://doi.org/10.1109/ICPR.1998.711890

[20] Thi, T. N., Do, T. H., & Yoo, M. (2023). Implementation of OCR system on extracting information from Vietnamese book cover images. In *2023 International Conference on Advanced Technologies for Communications*, 427–432. https://doi.org/10.1109/ATC58710.2023.10318889

[21] Bangdiwala, M., Mahadik, S., Mehta, Y., Salunke, A., & Das, R. (2023). Automated library management system using face recognition and OCR. In *2023 4th International Conference for Emerging Technology*, 1–4. https://doi.org/10.1109/INCET57972.2023.10170541

[22] Huang, J., Yan, M., Zhou, J., Zhang, X., Jiang, Y., & Tao, Z. (2022). Design and development of a medicine box recognition system based on machine vision. In *2022 8th International Conference on Systems and Informatics*, 1–6. https://doi.org/10.1109/ICSAI57119.2022.10005527

[23] Magalhães, L., Ribeiro, B., Alves, N., & Guevara, M. (2017). A three-staged approach to medicine box recognition. In *2017 24º Encontro Português de Computação Gráfica e Interação*, 1–7. https://doi.org/10.1109/EPCGI.2017.8124317

[24] Fu, K., Dai, Y., & Zhu, Z. (2023). CNN-based visible ingredients recognition in a food image using decision making schemes. In *2023 IEEE International Conference on Systems, Man, and Cybernetics*, 2427–2432. https://doi.org/10.1109/SMC53992.2023.10394513

[25] Datta, A., Meghla, T. I., Khatun, T., Bhuiya, M. H., Shuvo, S. R., & Rahman, M. M. (2020). Road object detection. In *2020 IEEE International Women in Engineering Conference on Electrical and Computer Engineering*, 348–351. https://doi.org/10.1109/WIECON-ECE52138.2020.9397954

[26] Sarika, N., Sirisala, N., & Velpuru, M. S. (2021). CNN based optical character recognition and applications. In *2021 6th International Conference on Inventive Computation Technologies*, 666–672. https://doi.org/10.1109/ICICT50816.2021.9358735

[27] Fatema, K., Ahmed, M. R., & Arefin, M. S. (2022). Developing a system for automatic detection of books. *Second International Conference on Image Processing and Capsule Networks*, *300*, 309–321. https://doi.org/10.1007/978-3-030-84760-9_27

[28] Yu, Z. (2023). Sim-YOLO: A real-time Chinese scene text detection method. In *IEEE 7th Information Technology and Mechatronics Engineering Conference*, *7*, 2305–2309. https://doi.org/10.1109/ITOEC57671.2023.10291444

[29] Wang, Y., Yang, G., & Jiang, K. (2023). Detection of straw burning based on graph-YOLOV3 algorithm. In *2nd*

*International Conference on Machine Learning, Cloud Computing and Intelligent Mining*, 227–232. https://doi.org/10.1109/MLCCIM60412.2023.00038
[30] Punia, A. (2023). Recognition of handwritten character using recognition model based on SVM. In *2023 7th International Conference on I-SMAC*, 444–449. https://doi.org/10.1109/I-SMAC58438.2023.10290409
[31] Suwattanapunkul, T., & Wang, L. J. (2023). The efficient traffic sign detection and recognition for Taiwan road using YOLO model with hybrid dataset. In *2023 9th International Conference on Applied System Innovation*, 160–162. https://doi.org/10.1109/ICASI57738.2023.10179493
[32] Quoc, N. H., & Choi, W. H. (2009). A framework for recognition books on bookshelves. *Emerging Intelligent Computing Technology and Applications*, *5*, 386–395. https://doi.org/10.1007/978-3-642-04070-2_44
[33] Tabassum, N., Chowdhury, S., Hossen, M. K., & Mondal, S. U. (2017). An approach to recognize book title from multi-cell bookshelf images. In *2017 IEEE International Conference on Imaging, Vision & Pattern Recognition*, 1–6. https://doi.org/10.1109/ICIVPR.2017.7890886
[34] Yang, X., He, D., Huang, W., Ororbia, A., Zhou, Z., Kifer, D., & Giles, C. L. (2017). Smart library: Identifying books on library shelves using supervised deep learning for scene text reading. In *ACM/IEEE Joint Conference on Digital Libraries*, 1–4. https://doi.org/10.1109/JCDL.2017.7991581
[35] Xia, Z., Ma, K., Cheng, S., Blackburn, T., Peng, Z., Zhu, K., . . . , & Arcucci, R. (2023). Accurate identification and measurement of the precipitate area by two-stage deep neural networks in novel chromium-based alloys. *Physical Chemistry Chemical Physics*, *25*(23), 15970–15987. https://doi.org/10.1039/D3CP00402C
[36] Cheng, S., Guo, Y., & Arcucci, R. (2023). A generative model for surrogates of spatial-temporal wildfire nowcasting. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *7*(5), 1420–1430. https://doi.org/10.1109/TETCI.2023.3298535
[37] Cheng, S., Jin, Y., Harrison, S. P., Quilodrán-Casas, C., Prentice, I. C., Guo, Y. K., & Arcucci, R. (2022). Parameter flexible wildfire prediction using machine learning techniques: Forward and inverse modelling. *Remote Sensing*, *14*(13), 3228. https://doi.org/10.3390/rs14133228

**How to Cite:** Sakib, M. N., & Chowdhury, M. (2024). A Machine Learning-Based Approach for the Detection of Drugs in Drug Self. *Artificial Intelligence and Applications*. https://doi.org/10.47852/bonviewAIA42022666