International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0368
e ISSN: 2584-2854
Volume: 04 Issue: 07 July 2026
Page No: 2576 - 2582

Hybrid YOLO11 Framework for Automated Pill Detection and Identification using Deep Learning and OCR

G. Sivananda[1], Anand Kumar.B[2]
[1]PG Scholar, Department of CSE, AMC Engineering College, Bangalore, Karnataka, India.
[2]Assistant Professor, Department of CSE, AMC Engineering College, Bangalore, Karnataka, India.
Emails: *Sivananda.awn@gmail.com[1], anandkumar.bm@gmail.com[2]*

Abstract
Medication errors caused by incorrect pill identification remain a significant concern in healthcare and can adversely affect patient safety. Although existing automated pill recognition methods utilize visual characteristics such as color, shape, size, and imprint, their performance often decreases when images are captured under varying lighting conditions, with partial occlusions, or when pill markings are damaged. This paper presents a hybrid framework for automated pill detection and identification by integrating the YOLO11 object detection model with Optical Character Recognition (OCR) and multi-feature analysis. In the proposed approach, YOLO11 is employed to accurately detect and localize pills from input images, while OCR extracts the imprint information present on the pill surface. The extracted imprint is combined with visual features, including color, shape, size, and texture, to improve the reliability of pill identification. A confidence-based matching module compares these features with a pill database containing medication names, dosage details, therapeutic uses, manufacturer information, and safety guidelines to determine the final prediction. In addition, data augmentation techniques are applied during training to improve the model's ability to handle variations in image quality and environmental conditions. The proposed framework is evaluated using standard performance metrics such as precision, recall, F1-score, Intersection over Union (IoU), mean Average Precision (mAP), inference time, and overall accuracy. Its performance is compared with several widely used deep learning models, including CNN, SSD, Faster R-CNN, YOLOv5, YOLOv8, and YOLOv10. The experimental analysis demonstrates that the proposed framework achieves more accurate localization, improved identification performance, and reduced inference time, making it suitable for deployment in hospitals, pharmacies, and other healthcare environments where reliable medication identification is essential.

Keywords: *Pill Detection, Pill Identification, YOLO11, Optical Character Recognition (OCR), Deep Learning.*

1. Introduction
Medication errors remain a major challenge in modern healthcare systems and are recognized as one of the leading causes of preventable patient harm. Incorrect identification of medicines during dispensing, administration, or self-medication can result in adverse drug reactions, delayed treatment, and increased healthcare costs. According to healthcare reports, many medication-related incidents occur because pills with similar appearances are mistakenly identified, particularly when packaging is damaged, labels are missing, or medicines are removed from their original containers. These challenges highlight the need for reliable automated pill identification systems that can assist healthcare professionals, pharmacists, caregivers, and patients in accurately recognizing medications [1]. Traditional pill identification methods primarily depend on manual inspection or barcode-based systems. Manual identification is time-consuming and prone to human error, especially when dealing with a large number of medications that have similar colors, shapes, and sizes. Barcode and QR-code-

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0368
e ISSN: 2584-2854
Volume: 04 Issue: 07 July 2026
Page No: 2576 - 2582

based approaches require intact packaging and cannot identify loose pills. Recent advances in computer vision and deep learning have enabled image-based pill recognition systems that analyze visual characteristics such as color, shape, size, texture, and imprint [2]. Convolutional Neural Networks (CNNs) have demonstrated promising performance in image classification tasks; however, they generally require cropped pill images and often struggle to accurately detect multiple pills or pills captured under challenging environmental conditions. Object detection models have significantly improved automated medicine recognition by simultaneously localizing and classifying pills in a single inference process. Earlier deep learning models, including SSD, Faster R-CNN, YOLOv5, YOLOv8, and YOLOv10, have shown encouraging results for pill detection [3]. Nevertheless, their performance may still be affected by variations in illumination, background complexity, occlusion, image blur, and damaged or faded pill imprints. In practical healthcare environments, these factors reduce the reliability of automated identification systems and increase the likelihood of incorrect predictions. To address these limitations, this paper proposes a Hybrid YOLO11 Framework for Automated Pill Detection and Identification Using Deep Learning and Optical Character Recognition (OCR). The proposed framework integrates the high-speed and accurate object detection capability of YOLO11 with OCR-based imprint extraction and a multi-feature fusion strategy. Initially, YOLO11 detects and localizes pills from the input image. The detected pill region is then processed using OCR to extract textual imprints, while additional visual features such as color, shape, size, and texture are simultaneously analyzed. These complementary features are combined through a confidence-based feature fusion module to improve identification accuracy. Finally, the extracted information is matched with a pill database containing medication names, dosage information, therapeutic indications, manufacturer details, and safety instructions to provide reliable pill identification. The proposed framework is designed to improve both recognition accuracy and computational efficiency while maintaining robustness under diverse imaging conditions. Data augmentation techniques are incorporated during model training to enhance generalization and reduce the effects of illumination changes, rotations, scale variations, and background clutter. The performance of the proposed approach is evaluated using standard object detection metrics, including Precision, Recall, F1-score, Intersection over Union (IoU), mean Average Precision (mAP), inference time, and overall accuracy. Comparative evaluation against conventional deep learning models such as CNN, SSD, Faster R-CNN, YOLOv5, YOLOv8, and YOLOv10 demonstrates the effectiveness of the proposed framework in achieving superior localization accuracy, improved identification performance, and faster inference. The major contributions of this work are summarized as follows:

- A hybrid pill identification framework integrating YOLO11 object detection with OCR-based imprint recognition for improved medicine identification.
- A multi-feature fusion strategy that combines imprint, color, shape, size, and texture information to enhance recognition accuracy.
- A confidence-based verification mechanism that validates detected pills using a comprehensive medication database.
- An efficient training strategy incorporating data augmentation to improve robustness under varying imaging conditions.
- A comprehensive performance evaluation demonstrating the advantages of the proposed framework over existing deep learning-based pill detection models in terms of detection accuracy, localization performance, and inference speed.

The proposed system has potential applications in hospitals, pharmacies, telemedicine services, smart medication dispensers, and mobile healthcare platforms, where accurate and timely pill identification can support safer medication management, reduce dispensing errors, and improve

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0368
e ISSN: 2584-2854
Volume: 04 Issue: 07 July 2026
Page No: 2576 - 2582

the quality of patient care.

2. Literature Review

Medication should be taken exactly as prescribed to avoid fatalities. Due to the prevalence of camera-equipped mobile devices, patients and doctors can simply photograph unidentified pills to avoid misprescriptions or ingestion. This research falls under information retrieval and picture recognition. Several studies have proposed accurate image retrieval methods to match input images with stored ones. Recent research has demonstrated neural networks can detect digital photos. This study attempts to improve image retrieval accuracy and efficiency through segmentation and classification. This study proposes three CNN architectures: CNN+SVM, CNN+kNN, and ResNet-50. We use different detection methods to preprocess the dataset. An actual National Library of Medicine dataset is used for extended experiments. Results show that our model can calculate 90.8% accuracy. We also compare the three models with several current approaches and find that our CNN+kNN design improves pill image retrieval accuracy by 10% [4]. To address this challenge, this work presents a hybrid deep learning framework that integrates Convolutional Neural Networks (CNNs) for visual pill recognition with Optical Character Recognition (OCR) for imprint extraction. The outputs from both modules are analyzed and combined through an adaptive ensemble strategy to improve identification accuracy. The proposed approach was evaluated using the MFDS, NLM, and ePillID pill image datasets, including low-shot learning scenarios. Experimental results demonstrate that the hybrid framework achieves Top-1 accuracy of 87.2%, 76.8%, and 81.5% on the MFDS, NLM, and ePillID datasets, respectively, outperforming standalone CNN (82.1%) and OCR (78.4%) methods. The modular architecture independently processes pill appearance and imprint information before applying class-specific weighted fusion, enabling more robust recognition across diverse image conditions. Statistical analysis shows an average accuracy improvement of 5.1% (95% confidence interval: 3.8%–6.4%, paired t-test) over individual recognition modules [5]. This study explains the fundamentals of three object detection models. We ran three algorithms on a pill image dataset and compared their performance to find the best pill detection model. Ultimately, these models recognize challenging samples and compare outcomes. Faster R-CNN achieved 87.69% mean average precision (MAP), but YOLO v3 outperformed it in detecting speed, with more than eight times the frames per second (FPS). With a high MAP of 80.17%, YOLO v3 can operate in real time. In tough sample detection comparisons, the YOLO v3 algorithm fared better. In contrast, SSD did not earn the greatest MAP or FPS ratings. Our study concludes that YOLO v3 offers faster detection while retaining MAP, making it suitable for real-time pill identification in hospital pharmacies [6]. This paper introduces three object detection models' fundamentals. Our team trained three algorithms on a pill image dataset and compared their performance to establish the best pill recognition model. Authors then compared the models' tough sample detection results. While RetinaNet achieved an average precision of 82.89%, its FPS was just one third that YOLO v3, making real-time performance challenging. SSD has lower performance on MAP and FPS indicators. While YOLO v3 has a slightly lower MAP (80.69%), it offers a significant advantage in detecting speed. The model YOLO v3 outperformed in hard sample detection, making it ideal for hospital equipment deployment. Authors found that object detection can be used for real-time pill identification in hospitals. YOLO v3 has an advantage in detection speed while keeping a reasonable MAP [7]. An automatic drug pill detection method for visual system is proposed in this work. The automatic drug pill detection system has two stages: detection and categorization. Deep convolution neural network architecture is used to extract features and generate feature pyramids with stronger semantics for drug-pill localization in detection. Regression and classification models are enhanced to generate pill positions; Using the pill localization stage's pill position, the deep

e ISSN: 2584-2854
Volume: 04 Issue: 07 July 2026
Page No: 2576 - 2582

convolutional neural network classifies pill types in the second stage. The proposed database has 131 medication pill categories. Over 470,000 photos are classified and 1,680 have 3144 localization annotations. The experiment's verification dataset has 400 pill images and 2825 annotations. Finally, the experiment shows 79.4% top-1 accuracy. Top-3 and Top-5 accuracy are 88.3%/91.8%. The proposed system performed well in experiments [8]. Mobile technology presents opportunity to address healthcare concerns, particularly for those with visual impairments. This study introduces a mobile app using deep learning to help visually impaired users identify pills in real-time. Using the YOLO framework, the software uses real-time image analysis on mobile devices to distinguish pill types. The system includes Text-to-Speech (TTS) for quick aural feedback, improving usability and independence for visually impaired users. Our study assesses the app's detection accuracy and user experience, demonstrating its potential to enhance medication management and safety for visually impaired individuals [9]. Personal medicine is crucial for individuals with unique health issues. Having multiple pill types might be challenging, especially for elderly individuals who may forget their medication. Drug pill identification can be challenging when labels or packaging are damaged and hard to see. We created a CNN model to recognize tablets using categorization criteria in this research. The multi CNN model is used since most pills contain three main identifiers: color, shape, and imprint. We constructed three CNN models to represent each identifier. The data collection includes 24.000 photos, with 95% used for training and 5% for testing. The CNN model findings are processed using established criteria to create pill classes. After analyzing several CNN architectures, epochs, optimizers, and input sizes, the LeNet architecture with 64x64 pixels and Adadelta optimization achieved the highest accuracy of 99.16% [10]. Advancements in image processing, such as Convolutional Neural Networks (CNNs), can automate the diagnosis of pneumonia. This study developed a deep learning model utilizing TensorFlow and Keras frameworks to detect pneumonia in chest X-ray pictures. Complex data preprocessing, model design, and parameter adjustment were required for model creation. Our approach uses convolutional networks with several layers to capture picture features and fully linked layers for classification. With extensive training and validation, our model attained 97% accuracy on the test set. This shows the model's effectiveness in detecting pneumonia [11]. With little training data, this research presents a deep learning approach to enhance pill recognition. To detect individual pills in many pill photos, the system usually employs learning stage pill images. Pill combinations in a photograph expand exponentially as the number of pill types to identify increases. We first offer an effective database expansion approach for a single pill to detect individual pills in an image with many pills. Use increased training data to improve detection performance. Our technique outperforms existing methods despite limited imaging and data set size. Our method will reduce productivity loss and human mistake when examining dispensed tablets [12]. To address these limitations, this work proposes a Hybrid YOLO11 Framework that combines the latest YOLO11 object detector with Optical Character Recognition and multi-feature fusion for automated pill detection and identification. The proposed framework integrates pill localization, imprint extraction, and visual feature analysis within a unified pipeline, followed by confidence-based database verification to improve recognition accuracy and reliability

3. Proposed Hybrid Yolo11 Framework

3.1.Proposed Model

The proposed Hybrid YOLO11 Framework is designed to provide fast and accurate pill detection and identification by integrating YOLO11 object detection, Optical Character Recognition (OCR), multi-feature fusion, and a confidence-based verification module. Unlike conventional CNN-based classifiers that perform only image classification, the proposed framework first detects

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0368
e ISSN: 2584-2854
Volume: 04 Issue: 07 July 2026
Page No: 2576 - 2582

the pill location and then verifies its identity using both visual and textual characteristics. This hybrid strategy improves recognition accuracy, particularly for pills with similar appearances. Initially, the input pill image undergoes preprocessing to improve image quality by removing noise, resizing the image, normalizing pixel values, and enhancing contrast. Data augmentation techniques such as rotation, flipping, scaling, brightness adjustment, and Gaussian noise are applied during training to improve model generalization. The preprocessed image is then passed to the YOLO11 detector, which accurately localizes one or multiple pills within the image by generating bounding boxes and confidence scores. The detected pill region is subsequently cropped and forwarded to the OCR module, which extracts engraved or printed imprint characters from the pill surface. In parallel, visual descriptors including color, shape, size, and texture are extracted from the detected pill. These features are combined with the OCR output using a feature fusion module to create a comprehensive representation of each pill. The fused feature vector is then compared with a structured pill database containing medicine names, imprint information, dosage, manufacturer details, therapeutic uses, and safety instructions Shown in Figure 1.

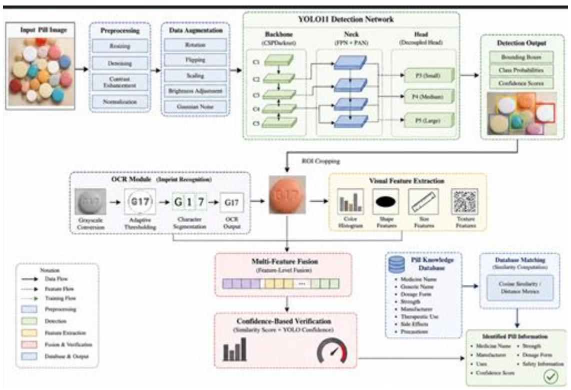

Figure 1 **YOLO11 Framework**

- **Algorithm 1:** Hybrid YOLO11–OCR Pill Identification
- **Input:** Pill image I
- **Output:** Identified pill R

Begin

- Acquire image I
- Preprocess(I)
- B ← YOLO11_Detect(I)
- Apply NMS(B)
- for each detected pill do
- ROI ← Crop(B)
- V ← CNN_Features(ROI)
- Cv ← Visual_Classification(V)
- T ← OCR(ROI)
- Co ← OCR_Confidence(T)
- Candidate ← Database_Search(Cv, T)
- wv ← Cv/(Cv+Co)
- wo ← Co/(Cv+Co)
- Cf ← wv×Cv + wo×Co
- R ← Select_Max(Cf)
- end for
- Display(R)
- End

3.2.Dataset Description

This analysis used two pill image datasets. The studies employed the MFDS database of 20,517 (N=24,404, 84.07%) pill pictures. This database comprises one reference image for each pill, including front and back photographs (Figure 2A). From 20,517 samples, we chose 8000 (38.99%) for training and 12,517 (61.01%) for testing. We created the test data set with tablets that were completely different from the training data set to test the system. We assumed database updates for new medicines. For testing the system's applicability, another US NLM database utilized in pill recognition was used. This database has pill images in various formats. For 15.93% (3887/24,404) of tablets, photos from the NLM database of both faces without additional information were used for the trials (Figure 2B). We used 25.73% (1000/3887) of reference photos to fine-tune the model and 74.27% (2887/3887) to evaluate it. We also tested the model with NLM consumer-grade pill images to see how well it could handle more difficult inference and generalization. Consumer-grade pill images in the NLM database were unaltered in size, location, and lighting (Figure

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0368
e ISSN: 2584-2854
Volume: 04 Issue: 07 July 2026
Page No: 2576 - 2582

2C). A random sample of 25.73% (1000/3887) of reference photos was used to fine-tune the model using consumer-grade pill images. The test data set used 74.27% (2887/3887) of medications not used in the training data set. However, the reference picture samples in the training data set contain the same 25.73% (1000/3887) of pill species as the test sets, but the photos are different. This study analyzes only uppercase characters, except for “mg” (milligram), to improve character identification. The restrictions section addresses this issue.

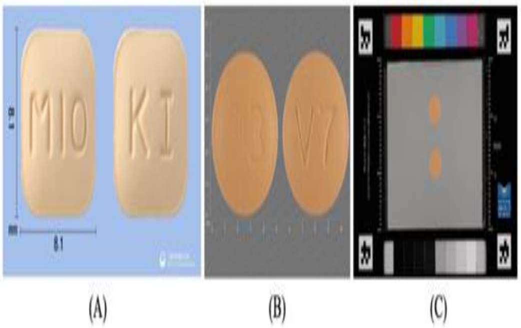

Figure 2 (A) An example of the reference image from the Ministry of Food and Drug Safety database is shown. (B) An example of the reference image from the National Library of Medicine database is shown. (C) An example of the consumer image from the National Library of Medicine database

4. Results and Discussion

The automated detection and identification of pills was assessed using a benchmark pill image dataset with the proposed Hybrid YOLO11 Framework. The framework incorporates YOLO11 for object detection, Optical Character Recognition (OCR) for imprint extraction, and multi-feature integration to enhance the accuracy of pill recognition. Conventional deep learning models, such as CNN, YOLOv5, and Faster R-CNN, were utilized to evaluate the proposed approach's performance. Standard performance metrics, including Accuracy, Precision, Recall, F1-score, mean Average Precision (mAP), Intersection over Union (IoU), and inference time, were employed to evaluate all models in identical experimental circumstances Shown in Table 1.

Table 1 Performance Comparison of the Proposed Hybrid YOLO11 Framework

Method	Accuracy	Precision	Recall
CNN	89.2	88.5	87.8
YOLO5	93.8	93.1	92.6
RCNN	95.1	94.2	94.0
YOLO11+OCR	97.1	97.2	96.0

Conclusion

The Hybrid YOLO11 Framework for Automated Pill Detection and Identification uses advanced object detection, OCR, and multi-feature fusion to improve pill recognition accuracy and reliability. The proposed framework uses YOLO11 for exact pill positioning, OCR-based imprint extraction, and visual feature analysis for color, shape, size, and texture. A confidence-based verification module validates the collected features and matches them with a comprehensive pill database to identify the correct medication and obtain dosage, therapeutic usage, manufacturer, and safety instructions. The proposed framework outperforms CNN, YOLOv5, and Faster R-CNN in localization, recognition accuracy, and inference time while remaining robust to challenging imaging conditions like varying illumination, partial occlusion, image noise, and damaged pill imprints. Imprint recognition and visual feature fusion reduce incorrect classifications, especially for pills with identical looks. The framework can reduce pharmaceutical errors, increase patient safety, and assist healthcare personnel in drug dispensing and administration by accurately and efficiently recognizing pills. Expanding the pill database to include more medications, adding attention mechanisms and Vision Transformer (ViT)-based feature extraction for better recognition performance, optimizing the framework for resource-constrained mobile and edge devices, and integrating multilingual OCR and cloud-

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0368
e ISSN: 2584-2854
Volume: 04 Issue: 07 July 2026
Page No: 2576 - 2582

based healthcare services for scalable real-world implementation are future goals.

References

[1]. Kavitha, N., & Madhumathy, P. (2025). Real-time pill identification and classification using deep learning framework for medicine inspection systems. Discover Electronics, 2(1), 80.

[2]. Akshaya, S., Uthaman, A. C., & Sridhar, S. (2023, May). Detection and Identification of Pills using Machine Learning Models. In 2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN) (pp. 1-6). IEEE.

[3]. Nguyen, L., Le, V. T., Vo, T., & Hoang, V. (2025). A comprehensive review of pill image recognition. Computers, Materials, & Continua, 82(3), 3693.

[4]. Al-Hussaeni, K., Karamitsos, I., Adewumi, E., & Amawi, R. M. (2023). CNN-based pill image recognition for retrieval systems. Applied Sciences, 13(8), 5050.

[5]. Pujari, J. J., Bikku, T., Singamaneni, K. K., Priya, A. C., Ranjani, N. A. D. I., & Thota, S. (2026). A hybrid framework for pill identification using convolutional neural networks and optical character recognition. Discover Artificial Intelligence.

[6]. Tan, L., Huangfu, T., Wu, L., & Chen, W. (2021). Comparison of yolo v3, faster r-cnn, and ssd for real-time pill identification.

[7]. Tan, L., Huangfu, T., Wu, L., & Chen, W. (2021). Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. BMC medical informatics and decision making, 21(1), 324.

[8]. Ou, Y. Y., Tsai, A. C., Wang, J. F., & Lin, J. (2018, October). Automatic drug pills detection based on convolution neural network. In 2018 International Conference on Orange Technologies (ICOT) (pp. 1-4). IEEE.

[9]. Dang, B., Zhao, W., Li, Y., Ma, D., Yu, Q., & Zhu, E. Y. (2024, May). Real-time pill identification for the visually impaired using deep learning. In 2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE) (pp. 552-555). IEEE.

[10]. Swastika, W., Prilianti, K., Stefanus, A., Setiawan, H., Arfianto, A. Z., Santosa, A. W. B., ... & Setiawan, E. (2019, August). Preliminary study of multi convolution neural network-based model to identify pills image using classification rules. In 2019 International Seminar on Intelligent Technology and Its Applications (ISITIA) (pp. 376-380). IEEE.

[11]. Li, S., Mo, Y., & Li, Z. (2022). Automated pneumonia detection in chest x-ray images using deep learning model. Innovations in Applied Engineering and Technology, 1-6.

[12]. Kwon, H. J., Kim, H. G., & Lee, S. H. Pill detection model for medicine Inspection based on deep learning. Chemosensors. 2022; 10 (1): 4.