

Deep learning for daily care: medicine recognition and reminder systems for the visually impaired

Uttam Waghmode¹, Pooja Bagane^{2,*},
Ashwini Naik³, Rajendra Balaso
Mohite⁴ and Aasha Mahesh Chavan¹

¹Department of Computer Engineering, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, India

²Department of Computer Science and Engineering, Symbiosis Institute of Technology-Pune Campus, Symbiosis International (Deemed University), Pune, India

³Department of Electronic & Communication Engineering, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, India

⁴Department of Electronics and Telecommunication Engineering, Bharati Vidyapeeth College of Engineering, Navi Mumbai, India

*E-mail: pooja.bagane@gmail.com

Received for publication
March 18, 2025.

Abstract

This research uses a deep learning-based approach that is helpful for the visually impaired humans in their medications, regular life activities without any support from other normal humans/nurses/caretakers. In this study, the inception model is introduced and is trained by 4000 images of the targeted medicines as follows: Dolo, Strepsils, Jovees shampoo, and Volini. The purpose of using many images in this study is to train the model with the training images of the different insights, variations in the distance, and inclination of medicines. After many experimentations of model training with tuning their hyperparameters, the model is trained for 98% training accuracy and 96.98% validation accuracy, which highlights the effectiveness of the system. To create an interactive UI for demonstration, the purpose of the local host is created, where pressing any key on the keyboard activates the camera input. The user is supposed to press any key on keyboard and keep the medicines in front of camera in a random way. The camera will capture the image and take it as input frame for processing purpose through the inception model. After processing the audio, output information will be announced through speaker about the medicine uses and doctor prescription for that particular medicines. The extra facility is also merged in this study to add on the timing and the name of the medicine by nurses that is to be helpful for the blind persons. The uses of the inception model for the medicine recognition and management create the awareness about the advanced program utilization to make easy for the visually impaired persons. All the medicines used in this study are trademark company products of respective owners. The products used in this study are only for research purposes and does not imply any endorsement or affiliation.

Keywords

CNN, inception, visually impaired, video processing

1. Introduction

The visually impaired community faces numerous challenges in their daily lives, and one of which is the identification and management of medications. The solutions to the problems related to the daily life of the visually impaired persons are hidden in the embedding

of advanced technology, artificial intelligence (AI), and deep learning, which aims to make ease of their lives. The demonstration of proof of concepts in this study shows that the AI, more specifically, we can say the deep learning models, to build an intelligent system can assist blind persons in identifying and managing their medications more effectively.

The innovative applications are developed in the field of AI, machine learning, and deep learning, which shows the capabilities of the computer vision, natural language processing, robotics, etc. The intelligent system in future with the features, such as voice recognition, sentiment recognition-based, and scene description in text using the NLP, can handle the complex environments, which are suited for the visually impaired humans. The system whose proof of concept shown in this research work utilizes the deep neural networks inception model to identify the medications based on their visual appearance.

By preprocessing the images of the medicines, it will get converted into the desired format which would be understood by the model. This improves the reliability of the system in accurately predicting the medicines given as input images from the camera while testing.

This system will also incorporate a user-friendly interface, allowing visually impaired individuals to interact with it easily. To understand the concept, it is shown with the user interface, making it easy to visualize for a normal person and get to know the exact area where the proposed system is creating impact. As an add-on to the system, it can provide additional information about the medicines, such as their usage and purpose for which it is generally recommended to patients. The laptop speakers are responsible for the audio outputs, which enable the visually impaired individuals to hear and manage their medications independently.

The features, such as medicine recommendations, medicine timing reminders, medicine information system, and easy UI for the caretakers, will be impactful in the use of the proposed system for the visually impaired people.

Overall, we can summarize that the system is needed to tackle the issues of the visually impaired people, such as reminding to take medicines, identifying the exact medicine from the list of the different medicines in the box, and understanding the uses of the medicines. The use of advanced algorithms embedded with the new applications is helpful and effective in making the life ease of the visually impaired persons. The detailed understanding of the images is nourished by the neural network layers through the use of filters. Various filters are responsible for extracting features and providing the input to the fully connected neural network. Some advanced functions, such as softmax, etc., are also helpful in creating the vectors for the processing purpose. The dense layer is where the features that were extracted from the convolution layers are combined into a

single vector. This one-dimensional vector has a very large number of parameters that are discussed in this study.

II. Literature Review

The literature survey demonstrates a wide range of methodologies and techniques intended to support blind individuals, emphasizing on strengthening their independence and quality of life.

One tactic involves integrating deep learning and imagination to detect items and offer speech responses for harmless navigation [1]. Another article study emphasizes the identification of money and clothing for visually impaired individuals in daily activities [2]. These novelties highlight the important progressions in assistive tools for the visually impaired individuals. Computer vision in AI plays a vital role in understanding photographic data, with applications in aiding visually impaired individuals. For example, a learning concentration on using a convolutional neural network (CNN) designed for a mobile banknote recognition app, aiding visually impaired individuals [3]. This article reading underlines the significance of dataset composition and augmentation in accomplishing high accuracy rates for real-time money identification. One more central characteristic is contact-free heart rate measurement, which is vital in several fields. A study suggests a multi-person approach using CNNs for heart rate approximation, refining runtime and accurateness for huge groups [4]. This method demonstrates how deep learning can provide contactless health monitoring.

As a result of large developments in assistive technology, the compact computationally constrained portable aid was developed using a Raspberry Pi to guide blind individuals in detecting objects and recognizing information about them [5].

This development showcases the real-world applications of deep learning in aiding visually impaired persons in their everyday lives. There is a high need for the development of solutions to the communication challenges faced by individuals with auditory as well as visual impairments. One study suggested an original method using a combination of CNN and long short-term memory (LSTM) network on Raspberry Pi to facilitate communication for the visually and hearing impaired [6].

Diagnostic examinations are essential for evaluating development in building deep learning systems that can reason and respond to visual input enquiries. Visual question answering (VQA) tackles the challenge of providing feature descriptions of web images

to the visually impaired individuals by responding to their questions [7]. This study explores VQA using CNN and LSTM with the VQA-2.0 dataset, highlighting its strong points and flaws associated with other models.

For the country's security, the boundary and surveillance systems are critical for detecting hazardous events. Deep learning provides an intelligent and autonomous approach for object detection using visual data [8]. This study explores a dataset for training and testing YOLO, attaining great accuracy in sensing ground pits, which can serve as hidden locations for illegal actions.

In the field of movable machine navigation, precise pedestrian trajectory forecast is vital for enhancing problem-escaping performance [9]. This study presents a new algorithm for pedestrian trajectory prediction built on a panoramic camera, signifying its potential for mobile robot navigation and obstacle avoidance in human-robot coexistence surroundings.

We can conclude from the above studies that the variety of solutions using deep learning and machine learning is available, which are part of assistive technology for visually impaired persons.

Table 1 shows the approaches and systems used in modern investigations to support visually impaired persons. These methods include deep learning, computer vision, and innovative technologies to improve navigation, object acknowledgment, communication, and healthcare for the visually impaired persons.

III. Main Challenges

The visually impaired persons face many issues in their daily life to perform daily activities, so there is a scope for building innovative solutions that will be useful for making their life independent and improving the quality of life.

Tools, such as deep learning, computer vision, and smart devices, can help resolve many complications. These technologies make it easy for people to identify objects, navigate their environment, communicate well, and access healthcare. They also help create inexpensive assistive devices for persons in need. The main encounters associated with these solutions are discussed below:

1. Object detection and recognition: Development of algorithms that are efficient and helpful for object detection in real-time scenarios used by visually impaired persons, ensuring that maximum daily life objects are recognized in different environments/scenarios.

2. Navigation in unfamiliar environments: A system that can guide and help blind persons in new places/environments needs to be developed through audio cues and object identification mechanisms.
3. Communication solutions: To assist with communication difficulties, different solutions can be created, similar to converting SMS emails into Braille for blind people and using special audible range devices for phone calls. These solutions make it easier for individuals to cooperate with their environment.
4. Healthcare and diagnostic tools: The use of deep learning models for eye disorders, such as glaucoma, can be accurately diagnosed with retinal images, which is helpful in early detection and treatment outcomes.
5. Cost-effective assistive devices: Emerging profitable solutions, such as smart glasses, for recognizing drugs, ensuring convenience for visually impaired persons without depending on costly technologies.
6. Enhancing navigation aids: Refining existing navigation supports for the blind, such as guide dogs, by incorporating deep learning algorithms for object recognition and grouping, enhancing security and independence.

IV. Proposed System Model

In this investigation work, a dataset is curated specifically for medications, such as Dolo, Strepsils, Volini gel, and Jovees shampoo. The dataset is vital for training the suggested model, which uses cutting-edge CNN models for image cataloguing. To generate the training dataset, videos capturing the medicines from many viewpoints (front, side, and back) are recorded. These videos are then treated to extract pictures at intervals of milliseconds, resulting in about 4000+ images for each drug. This dataset augmentation strategy improves the model's capability to identify and categorize these drugs precisely.

a. Methodology

This study presents a system for medicine recognition employing deep learning. Initially, the system augments the medicine image dataset using a probabilistic method for multilevel classification. It then extracts features from the augmented images through convolutional layers and a rectified linear unit (ReLU) activation function. Finally, the system

Table 1: Literature review

Reference	Method/algorithm used	Merits	Demerits
[1]	YOLO and OpenCV	Provides real-time object detection and visual replacement for the blind.	YOLO may struggle with small or distant objects, and OpenCV's accuracy can vary based on lighting conditions and object complexity.
[2]	TensorFlow API, CNN, SSD, and MobileNet V2	Achieves high accuracy without needing an Internet connection.	May require significant computational resources, especially for training the model.
[3]	CNN	Utilizes data augmentation to achieve a 94% accuracy rate for banknote recognition.	Edge-detected images negatively affect accuracy, indicating a need for larger datasets and varied lighting conditions.
[4]	CNN	Improves runtime and accuracy for heart rate estimation in large groups.	Specific details about the adapted algorithm and its implementation are needed for a thorough evaluation.
[5]	YOLO and SSD	Uses Raspberry Pi devices for a compact travel aid, demonstrating real-world implementation.	The system may face limitations in detecting objects in complex environments or under varying lighting conditions.
[6]	CNN and LSTM	Uses Braille and sound bite hearing devices for communication, achieving high accuracy.	The system's effectiveness may depend on the user's familiarity and comfort with Braille.
[9]	FSP algorithm	Utilizes a panoramic camera for pedestrian trajectory prediction, improving real-time performance.	The algorithm's accuracy and performance in dynamic or crowded environments need further evaluation.
[10]	CNN	Achieves 98% accuracy in diagnosing glaucoma using retinal images.	The effectiveness of the technique in clinical settings and its generalizability to diverse populations need further validation.
[11]	Four-layered CNN	Detects and classifies objects with high accuracy and low response time.	The device's performance may vary based on the complexity of the environment and the types of objects present.
[12]	CNN and fuzzy logic	Provides auditory feedback for obstacle detection, enhancing interaction with surroundings.	The computational complexity of the algorithms may affect real-time performance.
[13]	CNN	Uses smart glasses to identify medicine, showing promise for real-world implementation.	The system's accuracy and reliability in identifying specific medications need further validation.

CNN, convolutional neural network; LSTM, long short-term memory.

classifies the medicine using fully connected layers and a softmax function. Similar to the CNN layers, the inception model is used with a greater number of layers.

The proposed system employs the inception CNN architecture for medicine recognition. A custom dataset of approximately 4000 images per medicine was created by recording videos from multiple

angles—front, side, and top views—of Dolo, Strepsils, Volini gel, and Jovees shampoo. The videos were processed using OpenCV's `cv2.VideoCapture()` method to extract frames at intervals of 200 milliseconds, generating consistent and diverse training images for each medicine. These extracted frames were then resized and normalized to 150×150 pixels to ensure uniformity before being fed into the model. The system was trained using TensorFlow on an NVIDIA GTX 1080 Ti GPU for 50 epochs with a batch size of 32, utilizing the RMSprop optimizer and categorical cross-entropy as the loss function.

b. Block diagram and working

A schematic representation of the system workflow is depicted in Figure 1. A CNN comprises input,

hidden, and output layers, with millions of parameters for learning complex patterns. The network subsamples input through convolution and pooling, followed by activation functions in hidden layers. These layers are partially connected, leading to a fully connected layer at the end and producing the output. The output maintains the original input image dimensions.

Figure 1 shows the blocks used to build the system. The data augmentation process is performed on the video shots taken for each of the medicines. This will create a number of dataset images that will be helpful for training the model. There is a need for hyperparameter tuning, which includes the number of layers, kernel size, activation function, loss function, optimizers, learning rate, etc. Preprocessing of the images is performed after creating the dataset of the medicine images.

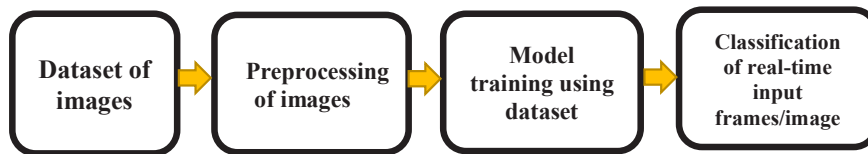


Figure 1: System block diagram.

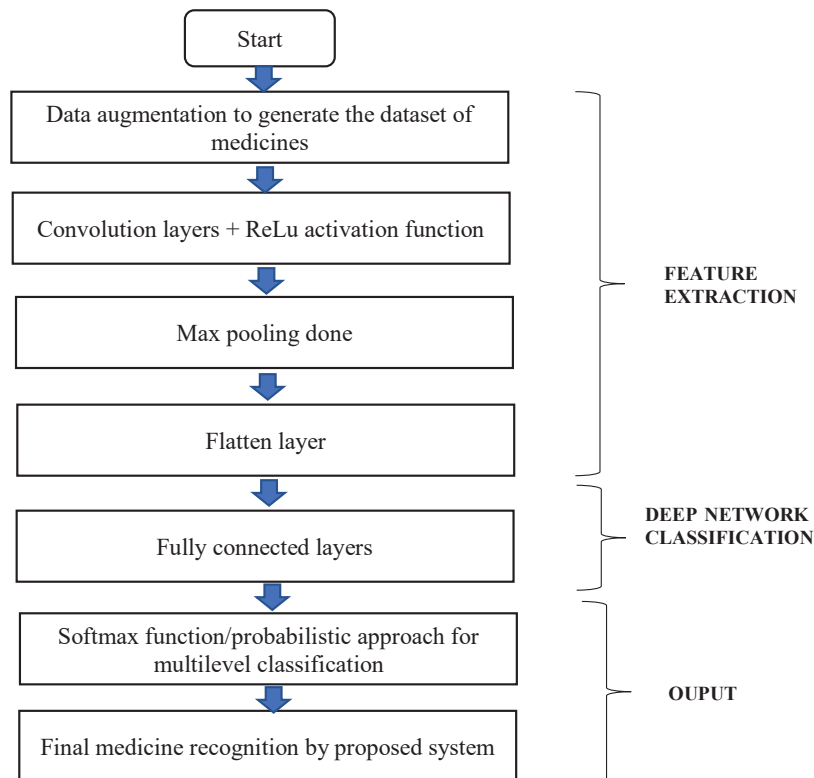


Figure 2: Flowchart of system/CNN layers. CNN, convolutional neural network.

In this preprocessing step, we have to decide the dimension for the input image that is used. Also, if there is any noise present, then we have to remove it. The third and fourth steps involve model training and classification, where feature extraction, parameter processing, batch processing, pooling, and multilevel classification are performed.

c. Processing of the system

Figure 2 provides a detailed flowchart depicting the sequential processes involved in the CNN-based recognition system.

- **Data augmentation:** This step is crucial for expanding the dataset of medicine images. It involves creating new variations of existing images by applying random transformations, such as cropping, rotating, and flipping. These transformations help make the dataset more diverse, which can improve the performance and reliability of the deep learning model trained on it. Figures 3–5 display the image datasets of Dolo, Volini, and Strepsils medicines, respectively, captured from multiple angles for robust model training.
- **Feature extraction:** This step is essential for medicine recognition. It involves extracting relevant

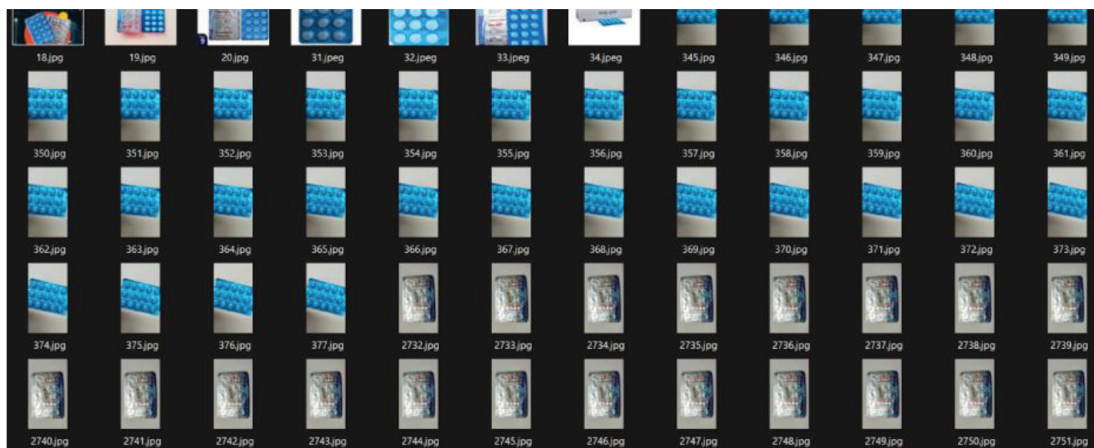


Figure 3: Dataset of Dolo medicine.

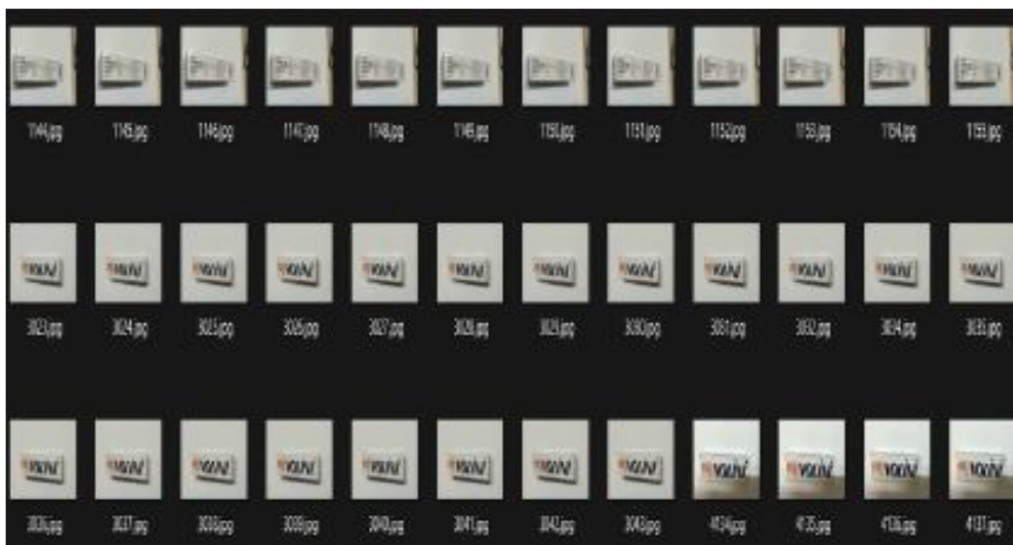


Figure 4: Dataset of Volini medicine.



Figure 5: Dataset of Strepsils medicine.

features from augmented images. Convolutional layers are used to learn these features from the images. The ReLU activation function is then applied to introduce nonlinearity into the network, helping it capture complex patterns. Max pooling is used to reduce the output's dimensionality and make the network more robust to image variations. Finally, the flatten layer converts the output of the convolutional layers into a 1D array for further processing. Other pooling methods, such as average pooling, global average pooling, min pooling, stochastic pooling, etc., are also available; however, max pooling is the most popular due to its invariance to small translations, dimensionality reduction, feature selection, simplicity, and efficiency.

- **Classification:** In this step, a deep neural network is used to classify the medicine in the image. Fully connected layers are employed to establish relationships between the extracted features and the different medicine classes. These layers enable the network to understand the complex patterns and characteristics specific to each medicine, ensuring accurate classification. This ability is crucial for identifying medicines based on the features extracted from the image, which is useful for tasks, such as medication reminders and inventory management. Softmax is used in this study because the softmax function normalizes the output probabilities, ensuring that they sum up to 1. This property is crucial for interpreting the output as probabilities, making it easier to determine the most likely class for the medicine being recognized, and multiclass classification feature. As seen in Figure 6, the inception model achieves stable accuracy after several training epochs.

V. Mathematical Expression for CNN Algorithm

Figure 7 illustrates the architecture of the CNN model implemented for medicine classification. For

convolution layers, the formula used in this study is as follows:

$$\{(n - k + 2p)/S\} + 1 \quad (1)$$

where n is the dimension of the input image, k is the filter size, p is padding, and S is stride.

Output for first layer convolution layer is as follows:

$$\{[(150 - 3 + 2 \times 0)/1] + 1\} = 148$$

$$\text{Output of the max pooling layer: } \{[(148 - 2 + 2 \times 0)/2] + 1\} = 74$$

In this study, the dimension is decreasing after each of the convolution layer and max pooling layer.

The output of the flatten layer is a one-dimensional vector of the features extracted through the various convolution and max pooling layers.

Flatten layer output will be: $17 \times 17 \times 64 = 18496$.

Dense layer output will be: 512.

Dense layer 2 output will be: 4.

At the output, we have the four-level classification of the results with a probabilistic approach using the softmax function. In the inception model, in total, 50 layers are used for building a model.

It takes a lot of time to train the model, but the validation accuracy is very good with this model.

VI. Results and Discussion

In this study, the input image size is 150×150 . The kernel size is 3×3 , and the activation function used is ReLU. In this study, the loss function used is cross entropy, the optimizer used is RMSprop, and the learning rate is set to 0.01. This hyperparameter training will give us a resultant summary as shown in Figure 8. Sample recognition results for different medicines processed by the system are illustrated in Figures 9–12.

• Success rate of the proposed system:

Table 2 summarizes the success rate of the proposed system across different medicines, with all achieving over 94% accuracy.

```
Epoch 1/5
172/172 [=====] - 592s 3s/step - loss: 1.4123 - acc: 0.9751 - val_loss: 0.9324 - val_acc: 0.9623
Epoch 2/5
172/172 [=====] - 442s 3s/step - loss: 0.3626 - acc: 0.9886 - val_loss: 3.0322 - val_acc: 0.9396
Epoch 3/5
172/172 [=====] - 441s 3s/step - loss: 0.3249 - acc: 0.9920 - val_loss: 1.9831 - val_acc: 0.9736
Epoch 4/5
172/172 [=====] - 522s 3s/step - loss: 0.2428 - acc: 0.9939 - val_loss: 1.5506 - val_acc: 0.9472
Epoch 5/5
172/172 [=====] - 558s 3s/step - loss: 0.2048 - acc: 0.9953 - val_loss: 2.2591 - val_acc: 0.9698
```

Figure 6: Epochs result for the inception model.

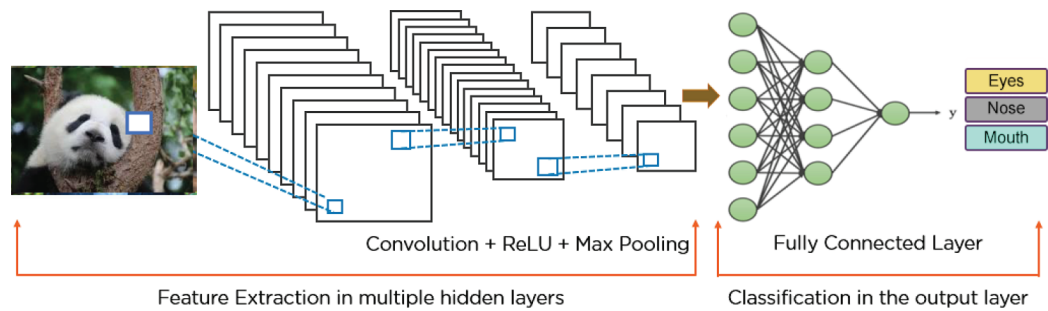


Figure 7: CNN architecture. CNN, convolutional neural network.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 148, 148, 16)	448
max_pooling2d (MaxPooling2D)	(None, 74, 74, 16)	0
conv2d_1 (Conv2D)	(None, 72, 72, 32)	4640
max_pooling2d_1 (MaxPooling2D)	(None, 36, 36, 32)	0
conv2d_2 (Conv2D)	(None, 34, 34, 64)	18496
max_pooling2d_2 (MaxPooling2D)	(None, 17, 17, 64)	0
flatten (Flatten)	(None, 18496)	0
dense (Dense)	(None, 512)	9470464
dense_1 (Dense)	(None, 4)	2052

=====
Total params: 9,496,100
Trainable params: 9,496,100
Non-trainable params: 0

Figure 8: Summary of parameters and trained CNN model. CNN, convolutional neural network.

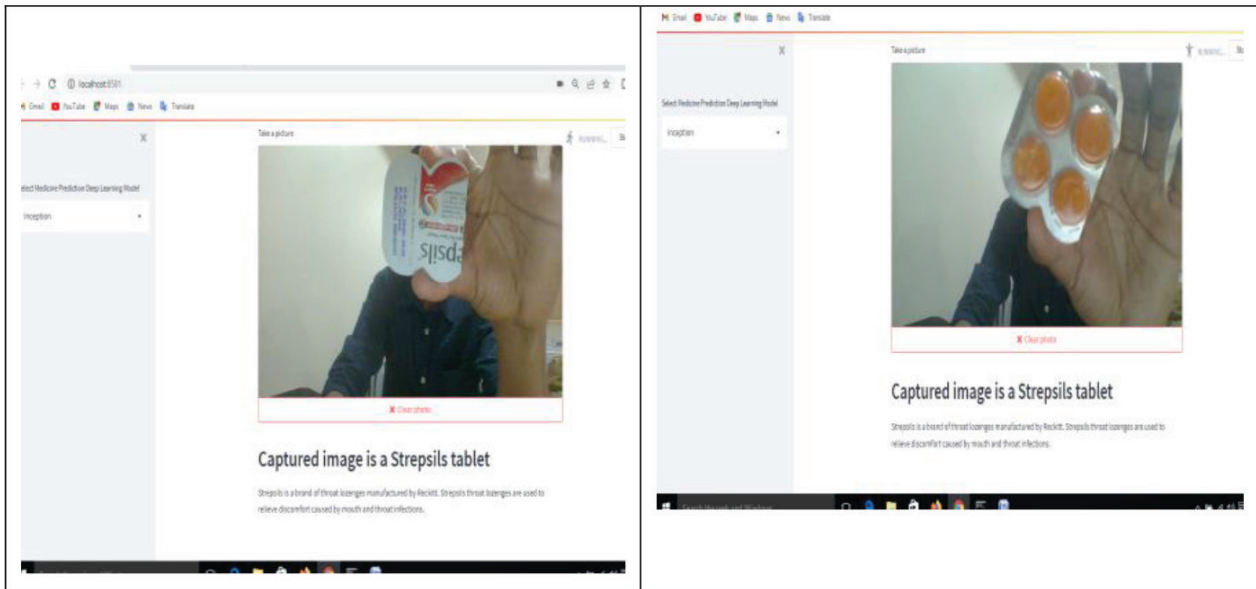


Figure 9: Output results for Strepsils.

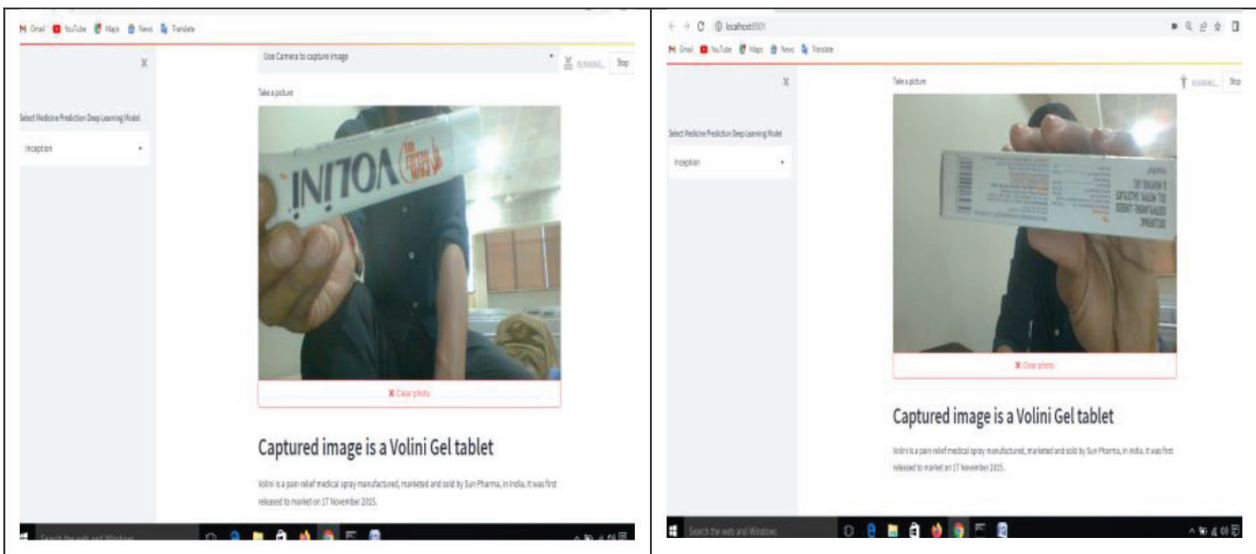


Figure 10: Output results for Volini gel.

Table 3 presents a comparative analysis between the proposed system and three recent similar works in the domain of assistive technologies using deep learning. The comparison focuses on key parameters, such as the method used, dataset size, achieved accuracy, number of medicines or categories handled, and distinct advantages offered by each system. The proposed system demonstrates superior accuracy and practical utility through audio-based feedback and an integrated medication

reminder feature. Unlike prior works which either focus on a limited number of items or different domains (e.g., currency recognition), this system directly addresses the medication needs of visually impaired individuals.

VII. Conclusion

In this study, we proposed a computer vision and audio feedback-based, deep learning-assisted guide

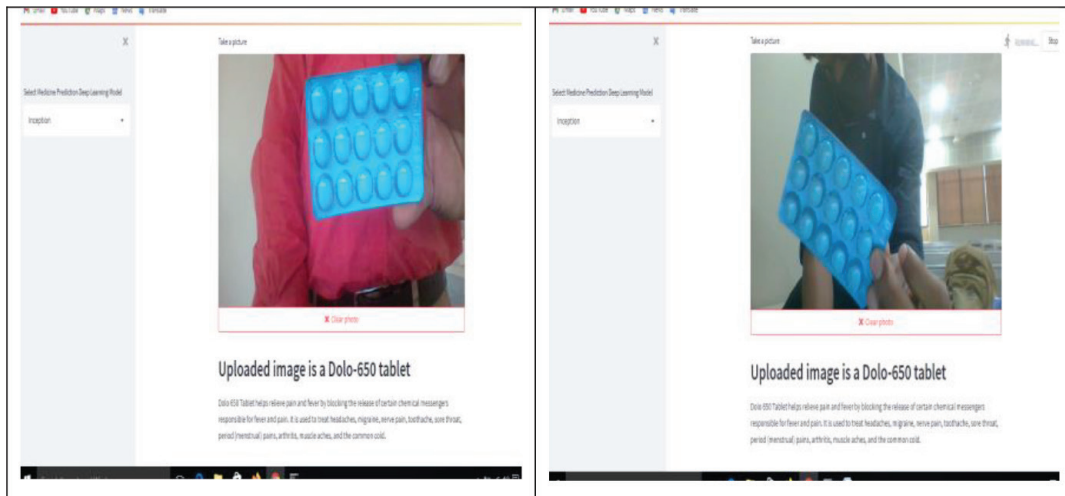


Figure 11: Output results for Dolo.

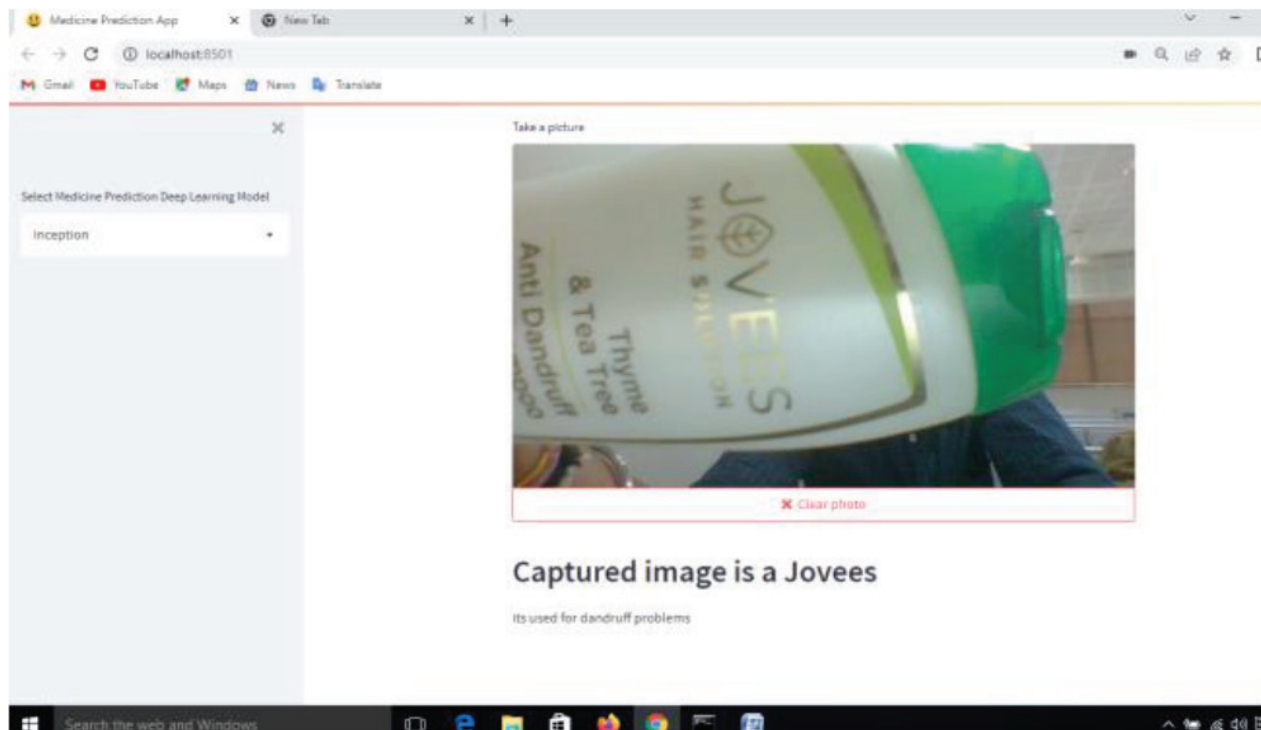


Figure 12: Output results for Jovees shampoo.

system for visually impaired people to identify and manage drugs. Using the inception 919 model [12] trained using 16,141 image samples across four categories of medicine, Dolo, Strepsils, Volini, and Jovees Shampoo, 98% of training accuracy and 96.98% of

validation accuracy were achieved. The user interface hosted locally enables users to activate the system by a key press, taking pictures with the camera in real-time and recognizing images. The feedback of identifying medicine is via audio, where it mentions the

Table 2: Success rate of the proposed system

Sr. No.	Medicine name	No of trails		Success rate (%)
		Succeed	Failed	
1.	Strepsils	46	4	92
2.	Volini gel	48	2	96
3.	Dolo	47	3	94
Success rate $\geq$				94

Table 3: Comparative study for the success rate of the proposed system with existing work

Reference	Method used	Dataset size	Accuracy (%)	Medicines handled	Key advantage
[13]	CNN	~2000 images	92	3	Smart glasses integration
[6]	CNN + LSTM	~3000 images	94	5	Includes Braille output
[3]	CNN	Augmented banknote data	94	Not medicine-specific	Focused on money recognition
Proposed system	Inception CNN	16,000 images	96.98	4	Audio output + medication reminder

CNN, convolutional neural network; LSTM, long short-term memory.

name and use of the detected medicine; moreover, it also enables caregivers to help remind patients by setting the medicine's name and schedule.

The future work would further extend the system by adding larger varieties of medicines to the dataset to improve generalizability. There is also potential for greater integration with mobile apps, wearable devices, and real-time voice interaction for further autonomy. Usability studies with visually impaired users will provide further iterations and refinements as well.

References

- [1] Logesh, K., Karthikeyan, P., Abishek, T., & Jayamani, S. "Assistance for Visually Impaired People Using Deep Learning." 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN), IEEE, 2023, India.
- [2] Haslinah Mohd Nasir et al. "Android based application for visually impaired using deep learning approach." IAES International Journal of Artificial Intelligence (IJ-AI) (2021). DOI: 10.11591/ijai.v10.i4.pp879-888
- [3] Maria Thomas and Kevin Meehan, "Banknote Object Detection for the Visually Impaired using a CNN," 32nd Irish Signals and Systems Conference (ISSC), Letterkenny, Ireland, 2021. DOI: 10.1109/ISSC52156.2021.9467850.
- [4] Thomas Pursche, et.al., "Concurrent video-based heart-rate measurement from human faces for large groups of participants - an improved CNN approach," 32nd Irish Signals and Systems Conference (ISSC), University of Wuppertal, School of Electrical, Information and Media Engineering, Automatic Control, Wuppertal 42119, Germany, 2021, DOI: 10.1109/ISSC52156.2021.9467850.
- [5] Kanak Manjari, Madhushi Verma, and Gaurav Singal, "CREATION: Computational constrained travel Aid for object detection In Outdoor environment," 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Department of Computer Science Engineering,

Bennett University, Greater Noida, India, 2019, DOI: 10.1109/SITIS.2019.00049.

[6] Chellaswamy C, Dhelipan Raj A, Geetha T S, Archana K, Hariharan K, and Babitharani S, "Deep Learning-Based Braille Technology for Visual and Hearing Impaired People," 2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES), Department of ECE, SRM TRP Engineering College Tiruchirappalli, India, DOI: 10.1109/ICSSES58299.2023.10199935.

[7] Shruthi G, Pradyumna Patil, and Krishna Raj P M, "Design and Development of Visual Question Answering model for the Visually Impaired," International Conference on Recent Advances in Science & Engineering Technology (ICRASET), Ramaiah Institute of Technology, Bengaluru, 2023.

[8] Ankit Pandey, Naeem Ahmad, and Dibakar Saha, "Detection of ground holes using Deep Learning for Surveillance," 21st OITS International Conference on Information Technology (OCIT 2023), Department of Computer Applications, National Institute of Technology, Raipur, India.

[9] Juncen Long, Jie Mei, and Guangfu Ma, "Egocentric Two-Frame Pedestrian Trajectory Prediction

Algorithm Based on a Panoramic Camera," IEEE Transactions on Instrumentation and Measurement, vol. 72, 2023.

[10] Mohanram, S., Goventhan, M., Balamurugan, V., & Rajeshvaran, N., "Glaucoma detection using CNN ensemble classifier and OTSU threshold algorithm." In International Conference on Circuit Power and Computing Technologies (ICCPCT), 2023.

[11] Yadav, D. K., Mookherji, S., Gomes, J., & Patil, S., "Intelligent Navigation System for the Visually Impaired - A Deep Learning Approach", In Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020), IEEE Xplore, 2020.

[12] Punit Srivastava, Priyanshi Shrivastav, Nikhil Chaturvedi, and R.L. Yadava. "Performance Analysis of AI-based Optical guiding Device for visually compromised person." In 2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Greater Noida, India.

[13] Monica. R, Nirmala.K, Divya.B, and Suganthi.L, "Recognition of Medicine using CNN for Visually Impaired," in International Conference on Communication and Signal Processing, July 28 - 30, 2020, India.