



OPEN Leveraging retinanet based object detection model for assisting visually impaired individuals with metaheuristic optimization algorithm

Alaa O. Khadidos¹ & Ayman Yafoz^{1,2}✉

Visually impaired individuals suffer many problems in handling their everyday activities like road crossing, writing, finding an object, reading, and so on. However, many navigation methods are available, and efficient object detection (OD) methods for visually impaired people (VIP) need to be improved. OD is the most crucial contribution of computer vision (CV) and plays an important part in recognizing and finding objects in the image. The aged and visually challenged people can identify different objects accurately under some constraint features like scaled, occlusion, illuminated, and blurred nature in different real world. A considerable number of studies are performed in the domain of real-time object recognition using deep learning (DL). The DL-based methods remove the features autonomously and classify and detect the objects. This paper proposes a novel Object Detection Model for Visually Impaired Individuals with a Metaheuristic Optimization Algorithm (ODMVII-MOA) technique. The proposed ODMVII-MOA technique aims to improve OD methods in real-time with advanced techniques to detect and recognize objects for disabled people. At first, the image pre-processing stage applies the Weiner filter (WF) method to enhance image quality by eliminating the unwanted noise from the data. Furthermore, the RetinaNet technique is utilized for the OD process to recognize and locate objects within an image. Besides, the proposed ODMVII-MOA method employs the EfficientNetB0 method for the feature extraction process. For the classification process, the LSTM-AE method is employed. Finally, the Dandelion Optimizer (DO) method adjusts the hyperparameter range of the LSTM-AE method optimally and results in better performance of classification. The experimental validation of the ODMVII-MOA model is verified under the indoor OD dataset and the outcomes are determined regarding different measures. The comparison study of the ODMVII-MOA model portrayed a superior accuracy value of 99.69% over existing techniques.

Keywords Object detection, Visually impaired individuals, Dandelion optimizer, Feature extraction, Deep learning, Computer vision

A human being's essential feature is often his vision ability¹. The capacity of being able to see objects with the eyes is considered a gift and a major factor in everyday activities. The significant challenge for numerous VIPs is not able to be completely independent and being restricted by their vision². Visually impaired individuals mostly encounter significant threats with everyday activities, and object recognition is a significant feature they depend on regularly. They specifically face difficulty with recognizing movement and objects in their environment, particularly while walking on the streets³. Vision loss usually occurs around the age of 50, a concern that is becoming more prevalent due to the ageing population. Vision impairment can cause diverse reasons like diabetic retinopathy, uncorrected refractive errors, unaddressed presbyopia, glaucoma, corneal opacity, and cataracts⁴. Numerous methods are intended to assist VIPs and enhance their living standards. Unfortunately, most of these methods are restricted in their capability. Artificial intelligence (AI) is initiating novel methods for individuals with disabilities to access the globe. OD is a significant task of CV, which handles identifying samples

¹Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. ²King Salman Center for Disability Research, Riyadh 11614, Saudi Arabia. ✉email: ayafoz@kau.edu.sa

of visual objects of certain classes in digital images⁵. The purpose of OD is to progress computational techniques and models that offer the most fundamental pieces of knowledge required by the application of CV: What objects are where? The dual important metrics for OD are speed and precision⁶.

OD acts as an origin for multiple other tasks of CV, like object tracking, instance segmentation, image captioning, and more. Over recent years, the faster growth of DL models has greatly endorsed the enhancement of OD, resulting in remarkable innovations and propelling it to an investigation hotspot with unprecedented attention⁷. OD is broadly employed in various real-time applications namely robot vision, video surveillance, autonomous driving, and so on. Different methodologies have been utilized to identify objects effectively and accurately in various applications. The DL models remove the features individually to classify and detect the objects. Since a model is proficient with multiple objects, a DL-based OD method offers a precise outcome⁸. Internet of Things (IoT) gadgets, for example, microcontrollers can run neural networks effectively. Therefore, a DL technique is combined with IoT methods to progress it as an assistive gadget for visually challenged individuals. Numerous effective DL-based OD structures are accessible. The growing population of VIPs faces growing challenges in daily life, specifically in navigating their environment and recognizing objects⁹. With improvements in AI and CV, there is a significant opportunity to develop systems that can assist these individuals in real-time, improving their independence. Object detection, significant for situational awareness, is optimized with advanced techniques to give reliable feedback. Metaheuristic optimization algorithms and advanced object detection models can improve assistive technologies for VIPs, enhancing accuracy, safety, and autonomy. These innovations enable real-time object recognition, boosting confidence and independence in navigating their environment¹⁰.

This paper proposes a novel Object Detection Model for Visually Impaired Individuals with a Metaheuristic Optimization Algorithm (ODMVII-MOA) technique. The proposed ODMVII-MOA technique aims to improve OD methods in real-time with advanced techniques to detect and recognize objects for disabled people. At first, the image pre-processing stage applies the Weiner filter (WF) method to enhance image quality by eliminating the unwanted noise from the data. Furthermore, the RetinaNet technique is utilized for the OD process to recognize and locate objects within an image. Besides, the proposed ODMVII-MOA method employs the EfficientNetB0 method for the feature extraction process. For the classification process, the LSTM-AE method is employed. Finally, the Dandelion Optimizer (DO) method adjusts the hyperparameter range of the LSTM-AE method optimally and results in better performance of classification. The experimental validation of the ODMVII-MOA model is verified under the indoor OD dataset and the outcomes are determined regarding different measures. The key contribution of the ODMVII-MOA model is listed below.

- The ODMVII-MOA model employs WF-based image pre-processing to enhance image quality by mitigating noise and enhancing clarity. This step confirms more accurate OD by giving cleaner input data. By minimizing noise interference, the approach strengthens the overall performance of the object recognition process.
- The ODMVII-MOA technique utilizes RetinaNet-based OD to effectually detect and localize objects within images. This approach improves the accuracy of the detection process by concentrating on the focal loss function, which addresses class imbalance issues. It allows for precise object recognition, even in complex environments with varying object sizes.
- The ODMVII-MOA method implements EfficientNetB0-based feature extraction to generate effectual and high-quality feature representations. This method significantly mitigates computational costs while maintaining performance. By optimizing the feature extraction process, it improves the capability of the model to detect and classify objects with greater accuracy.
- The ODMVII-MOA methodology employs DO-based tuning to fine-tune the parameters and improve detection accuracy. This optimization technique effectively enhances convergence speed and model performance. By adjusting parameters in a dynamic and adaptive manner, it results in more precise object detection in diverse scenarios.
- Integrating the DO with RetinaNet and EfficientNetB0 for OD, integrated with WF-based pre-processing, creates a novel approach to improving accuracy and efficiency. This unique integration addresses challenges such as noise reduction, class imbalance, and computational efficiency. The use of DO additionally improves optimization, allowing for enhanced performance in dynamic and complex environments.

Literature survey

In¹¹, a real-world edge-based hazardous OD method is projected that employs a lightweight DL method for classifying objects taken by a camera positioned on Raspberry Pi edge gadgets. By alerting and identifying VIPs to the existence of hazardous objects in their settings, this method can considerably enhance their independence and safety. Chinni et al.¹² presented a transformative OD method intended to upgrade the navigational capability of VIPs over the application of sophisticated CV technology. Employing the YOLO technique, connected to the comprehensive object collection (COCO) dataset, this method offers real-world, precise classification and OD. The core capability of the application permits the processing of either static images or live video feeds, allowing blind users to receive auditory announcements of neighbouring objects. In¹³, a smart assistive navigation method is projected that integrates voice-over with OD. The mobile application employs audio and voice only to offer navigational guidance to consumers. Likewise identifying objects, the built prototype permits the user and access the environmental data namely surface wetness, item distance, and orientation. Talaat et al.¹⁴ projected SightAid. This novel wearable vision method employs a DL-based structure. SightAid includes a seven-stage architecture with sophisticated DNN with many convolutional and FC layers. Interaction with users assisted through haptic feedback or audio. A continuous learning mechanism, combining new data and user feedback, guaranteeing the system's adaptability and ongoing refinement. More et al.¹⁵ developed an OD model particularly described for VIP utilizing the YOLO-v8. The application of this OD enlarges multiple features of day-to-day

life, accessibility, and promoting inclusivity for VIPs and bridging the gap between progressive investigation and practical assistive technology with the assistance of auditory feedback and edge devices. Sugashini and Balakrishnan¹⁶ projected an innovative economically effective, robust, handy, and simple solution with the help of smartphones. After identifying another stage is to evaluate the distance among the objects and VIPs to interact with them utilizing the headphones or audio devices. Arifando et al.¹⁷ presented the Improved-YOLO-v10, an innovative method intended to address challenges in pov classification and bus identification by combining Adaptive Kernel Convolution (AKConv) and co-ordinate attention (CA) into the YOLO-v10 structure. The Improved YOLO-v10 progresses the YOLO-v10 framework over the integration of CA that improves spatial awareness, AKConv, and longer-range dependency modelling that are dynamically fine-tuned convolutional kernels for extracting features. Gomaa and Saad¹⁸ propose a novel residual channel-attention (RCA) network for scene classification, integrating a lightweight residual structure for multi-scale feature capture with a channel attention mechanism to highlight relevant features. Furthermore, a squeeze-and-excitation (SE) mechanism is integrated to prioritize key features and minimize background noise.

In¹⁹, a resilient semi-automatic method for unsupervised object detection, incorporating YOLOv8 with background subtraction and clustering is introduced. The approach eliminates manual labelling by refining background subtraction results and utilizing them to fine-tune YOLOv8 for OD and classification. Navya et al.²⁰ utilize ML and CV to automatically detect hypochromic microcytes in PBS images for improved IDA diagnosis, utilizing ResNet50 for classification and YOLOv7 for localization. Gomaa, Abdelwahab, and Abo-Zahhad²¹ propose a real-time approach for vehicle detection and tracking in aerial videos, incorporating Top-hat and Bottom-hat transformations for detection and motion analysis utilizing KLT tracker and K-means clustering. An efficient algorithm assigns vehicle labels to corresponding trajectories. Wei et al.²² propose a Refined-Dilated Hybrid Bridge Feature Pyramid Network-Small Object Detection YOLOv8 (RHS-YOLOv8) method, which improves feature extraction with Refined-Dilated convolution, Hybrid Bridge Feature Pyramid Network (HBFPN), and Efficient Localization Attention (ELA), improves spatial relationships with an Involution module, and adds a small OD branch for improved performance. Gomaa and Abdalrazik²³ present a semi-automatic method integrating modified YOLOv4 and background subtraction for unsupervised object detection, eliminating manual labelling by using motion information to generate automatically labelled data. Zhao, Feng, and Wang²⁴ present a remote sensing object detection network (SCENE-YOLO) technique, a remote sensing OD network built on YOLOv8 with scene supervision. It integrates a scene information gathering and distribution network (SGD) model for high-level semantic injection, an omni-dimensional dynamic convolution (ODConv) method for feature weight redistribution, a scene label generation algorithm (SLGA) technique for model supervision, and a scene-assisted detection head (SADHead) method for improved performance in complex backgrounds. Abdalrazik, Gomaa, and Afifi²⁵ propose a tri-band elliptical patch antenna for L2, L5, and L1 GNSS bands, utilizing a probe-fed patch for circular polarization and an eye-shaped aperture to improve beamwidth, with a stacked parasitic element improving performance at the L1 band. Wang et al.²⁶ present a dynamic adversarial attack with transferable patches for non-rigid objects, optimized utilizing a cascade module and momentum technique, and improved with distance-adaptive generation and perspective transformation for improved robustness. Patil et al.²⁷ improve OD by mitigating image size with Gaussian Blur and utilizing colour descriptors and Fourier transforms for feature extraction, surpassing conventional restricted Boltzmann machine (RBM) and deep neural network (DNN) methods.

Despite crucial improvements in OD systems for VIPs, various limitations still exist. Many methods still face difficulty with real-time processing and accuracy in dynamic environments, specifically in complex backgrounds and diverse object types. Existing models often need precise labelling and manual interventions, restricting their scalability and adaptability. Moreover, the integration of advanced techniques such as RBM and DNN in OD models is still constrained by computational complexity and feature extraction inefficiency. Moreover, many systems encounter threats in object classification under varying conditions such as lighting or occlusions. The requirement for more robust, real-time, and adaptable systems remains evident across several domains, from surveillance to assistive technologies for VIPs.

The proposed methodology

In this study, a novel ODMVII-MOA technique is introduced. The proposed ODMVII-MOA technique aims at improving the OD method in real-time with advanced techniques to detect and recognize objects for disabled people. It holds image preprocessing, OD, feature extractor, hybrid classification, and parameter tuning. Figure 1 depicts the entire flow of the ODMVII-MOA model.

Image preprocessing

At first, the image pre-processing stage applies the WF method to enhance image quality by eliminating the unwanted noise from the data²⁸. This method is chosen for its capability to effectively mitigate noise and enhance image quality, which is significant for accurate OD. Unlike other methods, the WF adapts to local variations in the image, making it highly effective for removing noise without blurring crucial features. This is particularly beneficial in real-world scenarios where images may have varying levels of distortion or noise. The method is computationally efficient, enabling faster preprocessing while maintaining high-quality input for the detection model. Its capacity to improve image clarity enhances the overall performance of downstream models, making it more reliable in complex and noisy environments. Additionally, the WF model is a well-established technique, ensuring robustness and consistency across diverse datasets. Figure 2 illustrates the overall structure of the WF technique.

WF is a signal processing method applied for image restoration and denoising, making it beneficial in OD applications. In OD, WF aids in increased feature extraction by improving edges and significant details, resulting in improved detection precision. It is mainly beneficial in pre-processing stages for DL methods such as

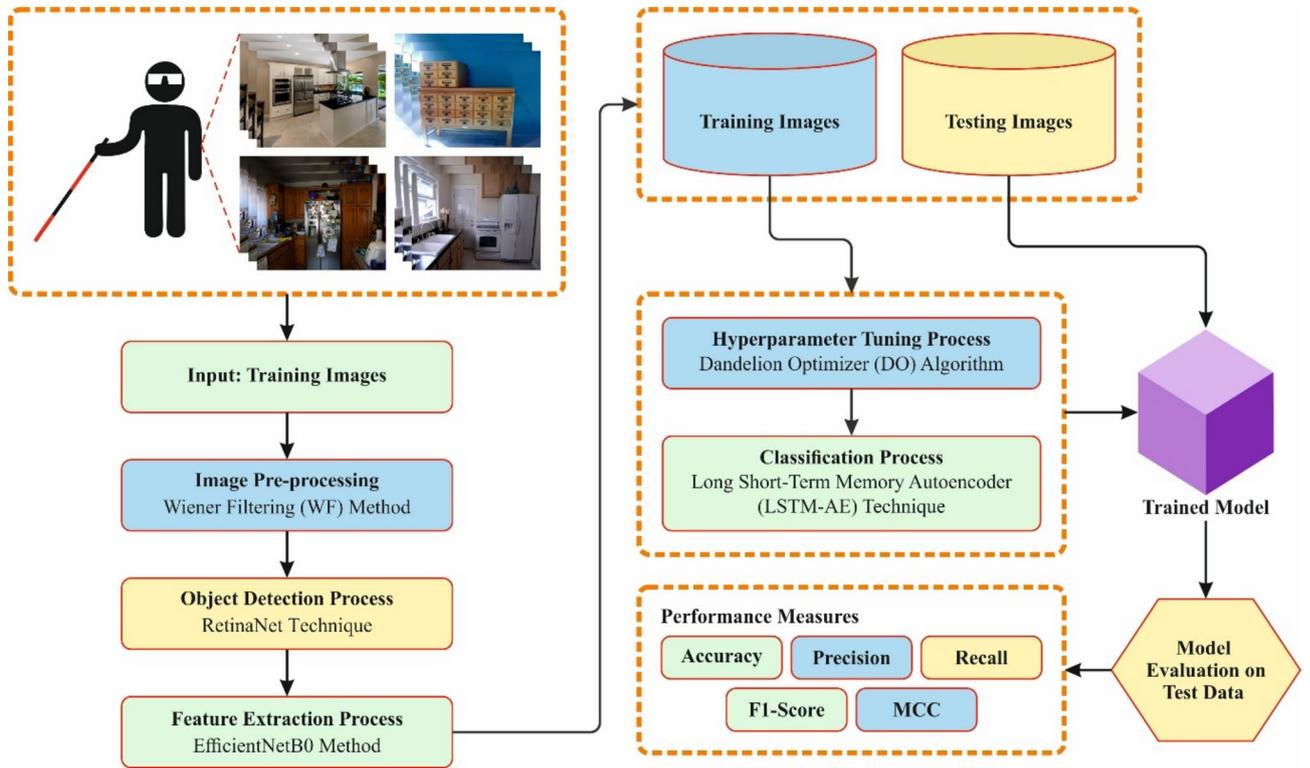


Fig. 1. Overall flow of ODMVII-MOA model.

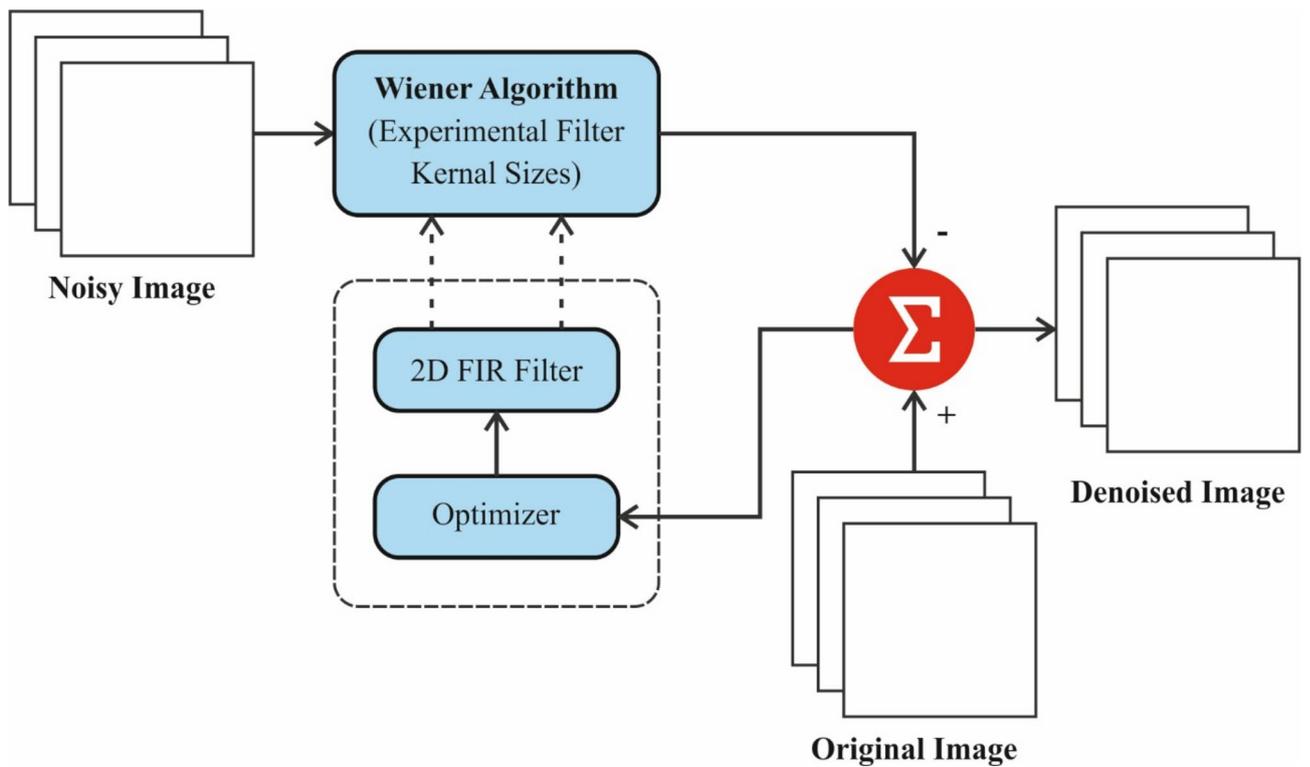


Fig. 2. Structure of the WF method.

RetinaNet, whereas higher-quality input images increase performance. The filter adjusts to local image variants, making it efficient in processing real-time noisy environments. Its combination with feature extraction methods can result in enhanced OD in complex environments.

OD using retinanet

Next, the RetinaNet technique is utilized for the OD process to recognize and locate objects within an image²⁹. This technique is chosen due to its efficiency in handling the class imbalance issue, which is a general threat in many real-world datasets. Its use of the focal loss function allows it to concentrate more on hard-to-detect objects, enhancing detection accuracy, particularly for small or rare classes. Unlike other models, RetinaNet strikes a balance between speed and accuracy, making it appropriate for real-time applications. It also performs well in both dense and sparse object scenarios, making it versatile for a wide range of detection tasks. The capability of the architecture to efficiently detect objects at diverse scales additionally improves its performance. Its proven track record in multiple domains, comprising autonomous driving and surveillance, underscores its reliability and robustness. Figure 3 represents the architecture of RetinaNet.

DL-based OD methods are separated into dual families. The 1st family is made of dual-phase recognition methods, like R-CNN and its elements. This type of model splits the OD task into a dual phase that includes making candidate areas before implementing regression and classification. Therefore, they naturally can attain higher detection precision, however, they frequently have slower detection speed. The 2nd family comprises -phase DL-based OD methods, like YOLO and its elements. These models explain OD to the end-to-end regression and classification task without making a candidate area. Therefore, they are normally faster but are frequently more uncertain in comparison with two-phase systems. How to attain either higher accuracy or speed became a difficult task in the development of DL-based OD analyses. Observed that these challenges mostly originate from the great imbalance normally discovered in sample class allocations. They presented the new function of loss called focal loss (FL) as a replacement for the conventional cross-entropy loss, and they established the RetinaNet networks for testing this FL task. FL is described in Eqs.(1-3)

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & otherwise \end{cases} \quad (1)$$

$$\alpha_t = \begin{cases} \alpha, & y = 1 \\ 1 - \alpha, & otherwise \end{cases} \quad (2)$$

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (3)$$

Whereas, y means the label, $y = 1$ signifies that the instance constitutes the positive class. $p \in [0,1]$ for binary classification epitomizes the possibility that the samples should be forecast as the positive instance. p_t denotes the method's projected possibility p for positive class, or else is the $1 - p$ for negative class. $\alpha \in [0,1]$ refers to the weighting factor applied to fine-tune the presentations to the FL from negative and positive samples. Akin to the description of p_t , α_t refers to α for the positive class, else is the $1 - \alpha$ for the negative class. $\gamma > 0$ represents the modulation feature applied to overwhelm the sample weight which is classified and to manage the method to concentrate on instances that are in addition problematic to categorize. Fixed $\alpha = 0.25$ and $\gamma = 2$ in this research. Due to advancements in FL, it was ultimately shown that RetinaNet can perform real-world object detection with either higher speed or accuracy.

EfficientNetB0 feature extractor

Besides, the proposed ODMVII-MOA method implements the EfficientNetB0 method for the feature extraction process³⁰. This technique is chosen due to its superior efficiency in balancing the performance and computational cost of the model. By utilizing a compound scaling method, it optimizes depth, width, and resolution simultaneously, attaining advanced accuracy with lesser parameters compared to other architectures. This makes EfficientNetB0 ideal for scenarios with restricted computational resources while maintaining high-quality feature extraction. Additionally, its lightweight nature allows for faster inference, which is significant

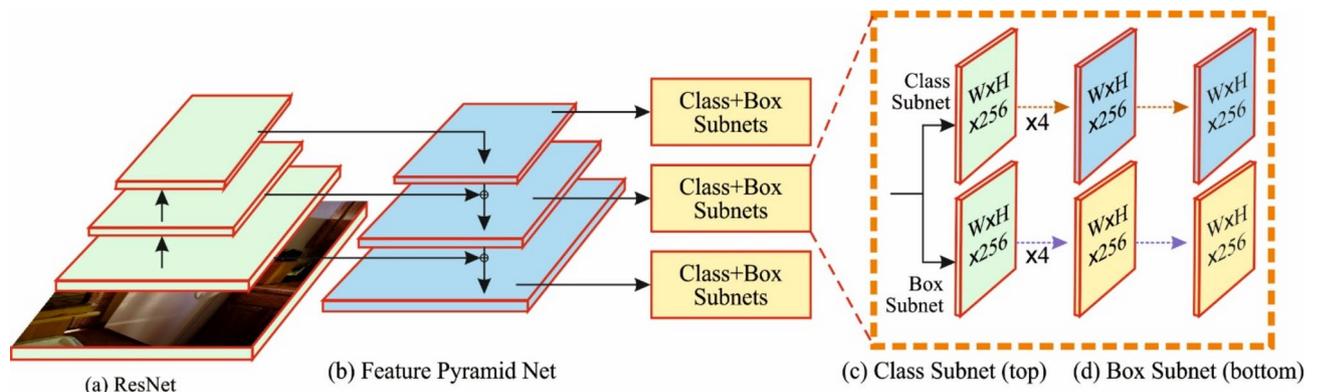


Fig. 3. Structure of RetinaNet.

in real-time applications. Unlike other deep models, EfficientNetB0 is designed to scale well, making it highly adaptable to diverse dataset sizes and complexities. Its efficiency confirms robust performance without the trade-offs in accuracy or speed commonly found in other models. Figure 4 demonstrates the EfficientNetB0 framework.

EfficientNet-B0 is the primary method of the EfficientNet family that begins with the layer of stem convolution. It is followed by numerous MBConv blocks that apply depth-wise separable convolutional. This convolution divided the normal convolution 1 into depth-wise convolutional individually filters the input channel for reading, and point-wise convolutional which combines the output of the depth-wise convolution. These separation type considerably reduces the parameter counts and computing cost by the features mapping, however, the feature extraction capacity stays very effective. The MBConv blocks in EfficientNet-B0 are more accurately allowed to identify the best features from the images like the minute parts in the images which are extremely valuable. Eventually, to combine each of the features gained from the extraction procedure, it combines Global Average Pooling. This layer down-samples the provided feature mapping from the preceding convolution layer in the average method and preserves simply the significant features. The result of these models is the lower comprehensive vector still concentrates on significant features. The last output is then transmitted to the dense classification layer but the presence should be recognized.

Hybrid LSTM-AE classification

For the classification process, the LSTM-AE method is employed³¹. This model is chosen for its capacity to effectually capture both temporal dependencies and spatial features in sequential data. LSTM outperforms at learning long-range dependencies, making it ideal for tasks comprising time series or sequential data. The AE component gives unsupervised feature learning, enabling the model to extract meaningful patterns from raw input data, which enhances the classification process. Integrating these two techniques allows the model to improve both feature extraction and sequence modelling capabilities, outperforming conventional methods that may face difficulty with complex, high-dimensional data. Furthermore, the hybrid approach mitigates overfitting by utilizing both supervised and unsupervised learning, making it more robust and efficient. This results in superior classification accuracy, particularly in dynamic environments with growing data. Figure 5 specifies the LSTM-AE approach.

AE is the form of ANN that learns how to effectively encode unlabeled data. It incorporates the succeeding dual portions: decoder and encoder. The initial portion addresses the reduction and generation of a few models of the characteristics from the first input x over the hidden layer (HL) h . The overall objective of the AE represents the transformation of higher- to lower-dimensional data. Determine the input layer $X \in R^l$, the HL $h \in R^p$, and the layer of output as $x \in R_l$. The input data maps are implemented based on Eq. (4), while to condense the concealed representations $p < l$, and ϕ means the function of activation, W , and b refers to the weighted matrix and biased vector. These representations of the concealed data are applied to rebuild the input data afterwards seeing Eq. (5).

$$h = \varphi \cdot (W \cdot x + b) \tag{4}$$



Fig. 4. Architecture of EfficientNetB0 methodology.

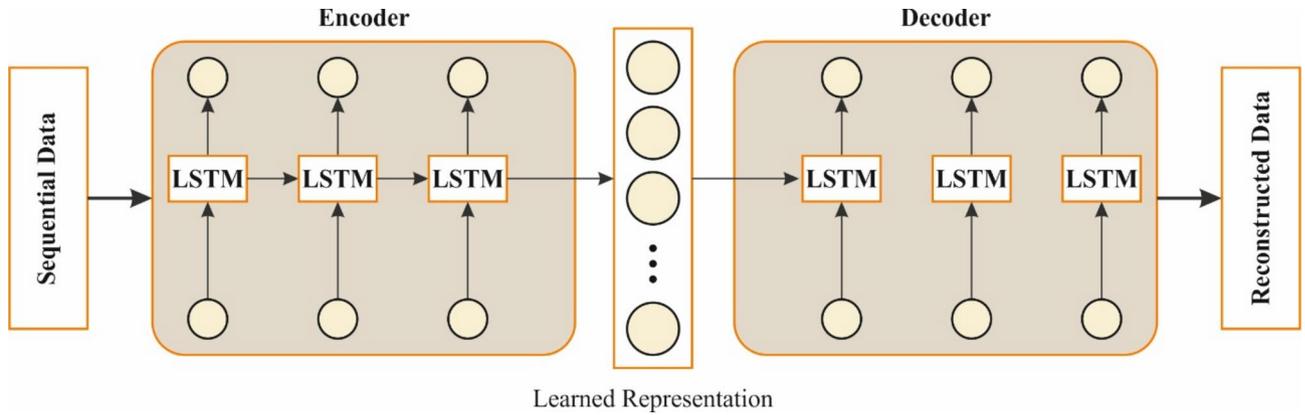


Fig. 5. Structure of LSTM-AE method.

$$\tilde{x} = \tilde{\phi} \cdot (\tilde{W} \cdot h + \tilde{b}) \tag{5}$$

The LSTM-AE is an incorporation of the AE and the encoding-decoding LSTM structure. The AE tries to identify the most related special designs within the training data by minimizing the latent area at the input $n < m$. It includes the succeeding 3 layers: the output layer, the input layer, and the hidden area. The training process contains various phases. During this next phase, the system learning for decoding moves back to the new input. The error of mean reconstruction is measured. The weighting is altered by the BP model.

The input is $X \in R^m$, whereas the encoding compress x to get the encoding representation $z = e(x) \in R^n$. Formerly, the decoder rebuilds this model to get $\tilde{X} = d(z) \in R^m$ as the output. The training procedure aims to reduce the reconstruction error, as presented in Eq. (6).

$$L = \frac{1}{2} \sum_x \|x - \hat{x}\|^2 \tag{6}$$

Parameter tuning using DO

Finally, the DO approach fine-tunes the hyperparameter range of the LSTM-AE model optimally and results in better performance of classification³². This approach is chosen due to its capability to effectually optimize complex, non-convex objective functions. Unlike conventional optimization algorithms like gradient descent, DO does not depend on derivative information, making it appropriate for problems where gradient-based methods struggle, such as in noisy or multimodal landscapes. It gives a robust global search capability that averts local minima, resulting in improved solutions in challenging optimization scenarios. Additionally, DO adapts dynamically, adjusting its search strategy during optimization to improve convergence speed and accuracy. The capability to fine-tune model parameters in real-time enhances overall model performance and generalization. By utilizing DO, the parameter tuning process becomes more effectual, ensuring optimal model configurations even in highly variable environments. Figure 6 indicates the steps involved in the DO model.

The DO is a bio-inspired optimizer model derived from the dispersion procedure of dandelion seeds. The DO model transforms these biological phenomena into the optimizer structure which balances exploitation and exploration, making it suitable for resolving composite and higher-dimensional optimizer problems. The DO works in the following 3 major steps: rising, decline, and landing phases. All phases symbolize an alternative phase of the optimizer procedure, facilitating either exploration or exploitation of the searching region.

Rising phase

Dandelion seeds are spread in the air, imitating the exploration stage. The model presents uncertainty over parameters like the scaling feature λ , angle θ , and random vectors NEW. The location of every seed is upgraded as presented in Eq. (7):

$$P(i, j) = P(i, j) + \alpha \cdot v_x \cdot v_y \cdot (\lambda, 0, 1) \cdot (NEW(1, j) - P(i, j)) \tag{7}$$

Whereas α refers to the scaling factor; v_x and v_y mean velocity elements; λ stands for random scaling feature; and NEW denotes randomly produced vector.

Decline phase

Once the seeds have grown, they start to run down, representing the migration from exploration to exploitation. Here, the population is fine-tuned according to the mean location of the seeds and a random feature β , guaranteeing that seeds determine nearer feasible locations, as demonstrated in Eq. (8):

$$P(i, j) = P(i, j) - \beta(i, j) \cdot \alpha \cdot (P_{mean}(1, j) - \beta(i, j) \cdot \alpha \cdot P(i, j)) \tag{8}$$

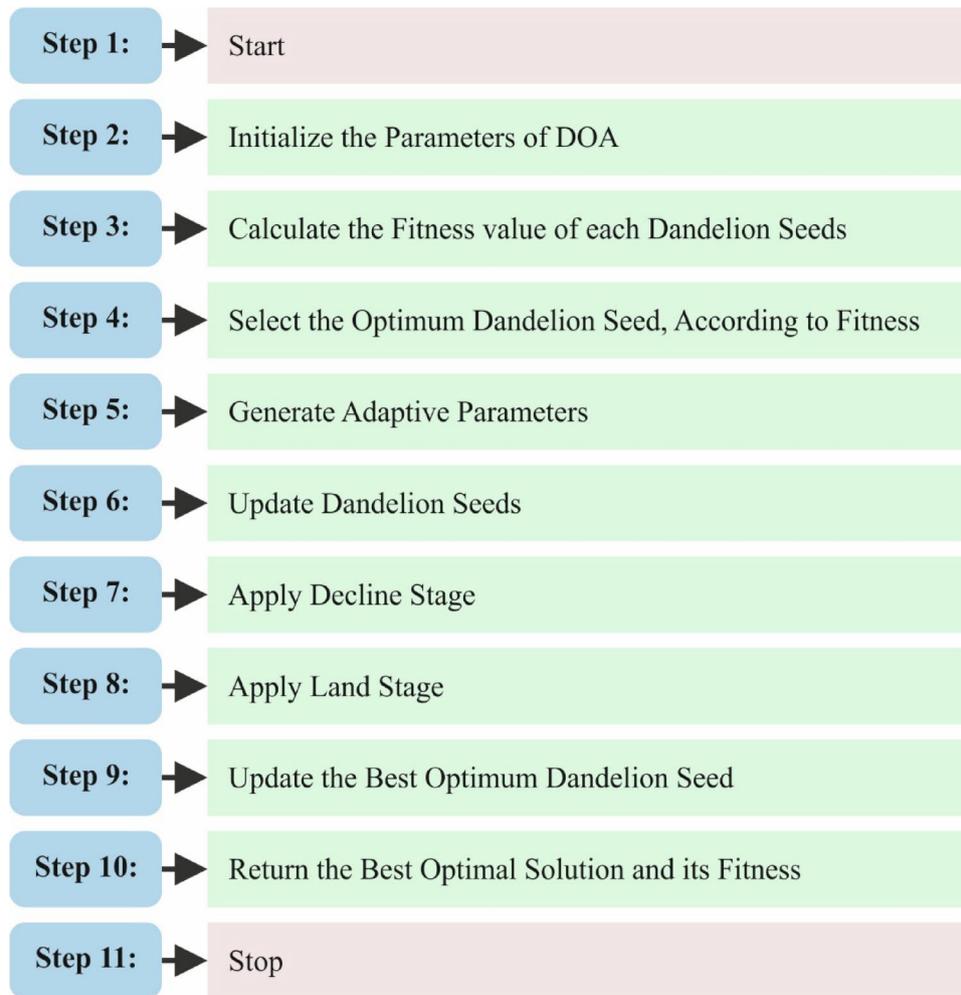


Fig. 6. Steps involved in the DO methodology.

Here, p_{mean} characterizes the mean location of the population, and β denotes randomly formed features which control the fall. This phase improves the search and utilizes the best-discovered solutions.

Landing phase

During this last phase, the seeds land and settle into the best locations. This phase is focused on a Lévy flight (LF) that improves global exploration by presenting large, arbitrary stages. The location upgrade within the landing phase is provided, as exposed in Eq. (9):

$$P(i, j) = Elite(i, j) + Step\ length(i, j) \cdot \alpha \cdot \left(Elite(i, j) - P(i, j) \cdot \frac{21}{Max\ iter} \right) \quad (9)$$

Whereas $Elite(i, j)$ characterizes the top individuals inside the population; $Step\ length(i, j)$ refers to step size established by the distribution of LF; l denotes present iteration, and $Max\ iter$ signifies maximal iteration counts.

Exploration and exploitation balance

The DO successfully balances exploitation and exploration over its 3-phase procedure. The increasing phase underlines exploration by permitting the model to search new areas of the searching region. The landing and decline phase move the concentration toward exploitation, refines the solutions, and converges near the best outcomes.

LF mechanism

LFs present randomness with a likelihood distribution which allows for larger, arbitrary stages, assisting the model escaping local ideals and exploring novel regions of the search area. This method is mainly useful in higher-dimensional and multi-modal optimizer issues, while the searching region is vast and complex. The DO is summarized accurately over the succeeding stages: Rising phase

$$P(i, j) = P(i, j) + \alpha \cdot v_x \cdot v_y \cdot (\lambda, 0, 1) \cdot (NEW(1, j) - P(i, j)) \quad (10)$$

Decline phase

$$P(i, j) = P(i, j) - \beta(i, j) \cdot \alpha \cdot (P_{mean}(1, j) - \beta(i, j) \cdot \alpha \cdot P(i, j)) \quad (11)$$

Landing phase

$$P(i, j) = Elite(i, j) + Steplength(i, j) \cdot \alpha \cdot \left(Elite(i, j) - P(i, j) \cdot \frac{2L}{Max\ iter} \right) \quad (12)$$

1. Initialization:

- Define the optimization problem (objective function to minimize/maximize).
- Initialize a population of dandelions (candidate solutions) randomly in the solution space. Every dandelion corresponds to a potential solution.
- Set the maximum number of iterations (or stopping criteria).
- Define parameters such as the population size, search space boundaries, and control parameters (e.g., seed spread range, mutation rate).

2. Fitness Evaluation:

- Compute the fitness of every dandelion in the population by assessing the objective function's value for each candidate solution.
- Choose the best dandelion (with the highest/lowest fitness value relying on the objective function).

3. Foraging Phase:

- Each dandelion performs a foraging process to find optimal solutions (similar to seeds spreading in nature).
- Update the position of each dandelion by considering both local and global search strategies:
 - **Local Search:** Update based on the best solution within a local neighbourhood.
 - **Global Search:** Update based on the best solution found by the entire population.

The update can be represented as:

$$X_i(t+1) = X_i(t) + \alpha \times (X_{best}(t) - X_i(t)) + \beta \times Rand(Spread)$$

Where:

- $X_i(t)$ is the position of dandelion i at time step t ,
- $X_{best}(t)$ is the position of the best solution found so far,
- α is a parameter controlling the influence of the best solution,
- β is a parameter controlling the spread,
- $Rand(Spread)$ is an arbitrary value that simulates the spread of dandelion seeds.

4. Mutation Phase:

- To maintain diversity in the population, a mutation step is applied to a subset of the dandelions.
- The mutation can involve random changes in the solution, such as altering the position of the dandelions by a small random amount:

$$X_i(t+1) = X_i(t) + \gamma \times Rand(Spread)$$

where γ is a mutation rate and $Rand(Spread)$ is a random value.

5. Reproduction Phase:

- After every iteration, a new generation of dandelions is created by choosing the best solutions based on fitness evaluation.
- The best solutions are passed on to the next generation, and weaker solutions are discarded.

6. Termination:

- The algorithm terminates when the stopping criteria are met (e.g., maximum number of iterations or convergence threshold).
- The solution corresponding to the best dandelion found is returned as the optimal solution.

Algorithm 1. DO method

The DO has proven to function well on various benchmark optimizer difficulties, mainly those that are higher-dimensional and multi-modal. Its capability to balance exploitation and exploration, incorporated with the LF method, makes it a stronger and more useful model to resolve composite real-time difficulties. The DO model initiates a fitness function (FF) to reach the boosted outcome of classifications. It establishes a progressive number to characterize the greater performance of the candidate solutions.

$$\begin{aligned} \text{fitness}(x_i) &= \text{ClassifierErrorRate}(x_i) \\ &= \frac{\text{no of misclassified instances}}{\text{Total no of instances}} \times 100 \end{aligned} \quad (13)$$

Result analysis and discussion

The experimental analysis of the ODMVII-MOA model is examined under the indoor OD dataset³³. This dataset contains 6642 samples under 10 objects. The complete details of the dataset are given below in Table 1.

Figure 7 presents the classifier outcomes of the ODMVII-MOA approach. Figure 7a and b shows the confusion matrices with truthful classification of each class label under 70%TRAPHA and 30%TESPHA. Figure 7c shows the PR values, showing optimal performance over every class. Besides, Fig. 7d demonstrates the ROC values, signifying accomplished outcomes with maximum ROC values for dissimilar classes.

Table 2; Fig. 8 established the OD of the ODMVII-MOA technique under 70:30 of TRAPHA/TESPHA. Based on 70% TRAPHA, the proposed ODMVII-MOA technique gains an average $accu_y$ of 99.69%, $prec_n$ of 94.40%, $reca_l$ of 89.63%, $F1_{score}$ of 91.22%, and MCC of 91.39%. Besides, depending upon 30%TESPHA, the proposed ODMVII-MOA approach accomplishes an average $accu_y$ of 99.67%, $prec_n$ of 96.43%, $reca_l$ of 87.72%, $F1_{score}$ of 90.53%, and MCC of 91.07%.

In Fig. 9, the training (TRAN) $accu_y$ and validation (VALN) $accu_y$ outcomes of the ODMVII-MOA model are demonstrated. The figure highlighted that both $accu_y$ analyses exhibit a growing trend which informed the capability of the ODMVII-MOA with better performance across distinct iterations. Moreover, both $accu_y$ remain closer across the epochs, which shows lesser overfitting and displays the greater solution of the ODMVII-MOA.

In Fig. 10, the TRAN loss (TRANLOSS) and VALN loss (VALNLOSS) analysis of the ODMVII-MOA model is exhibited. It is denoted that both values illustrate a lessening tendency, updating the capacity of the ODMVII-MOA in balancing a trade-off between data fitting and generalization. The continual reduction in loss values additionally promises the enhanced performance of the ODMVII-MOA model and tunes the prediction outcomes.

Table 3; Fig. 11 illustrate the comparative results of the ODMVII-MOA technique with existing models^{34–36}. The results emphasized that the Yolo-V8, Faster R-CNN, Yolo-v5n, GSDN, FCAF3D, Instant-NGP, and 3DGS-DR models have reported lesser performance. Whereas, the ODMVII-MOA approach stated the highest performance with high $accu_y$, $prec_n$, $reca_l$, and $F1_{score}$ of 99.69%, 94.40%, 89.63%, and 91.22%, respectively.

The time complexity (TC) result of the ODMVII-MOA technique is illustrated in Table 4; Fig. 12. Based on TC, the ODMVII-MOA method presents a minimal TC of 10.86 s while the Yolo-V8, Faster R-CNN, Yolo-v5n, GSDN, FCAF3D, Instant-NGP, and 3DGS-DR methodologies achieve higher TC values of 21.28 s, 18.37 s, 13.98 s, 17.35 s, 25.43 s, 18.13 s, and 18.40 s, correspondingly.

Conclusion

In this paper, a novel ODMVII-MOA methodology is introduced. The proposed ODMVII-MOA methodology aims at improving the OD method in real-time with advanced techniques to detect and recognize objects for disabled people. At first, the image pre-processing stage applies the WF method to enhance image quality by eliminating the unwanted noise from the data. Furthermore, the RetinaNet technique is utilized for the

Objects	Count
Door	562
Cabinet Door	3890
Refrigerator Door	879
Window	482
Chair	223
Table	248
Cabinet	208
Couch	24
Opened Door	90
Pole	36
Total	6642

Table 1. Dataset details.

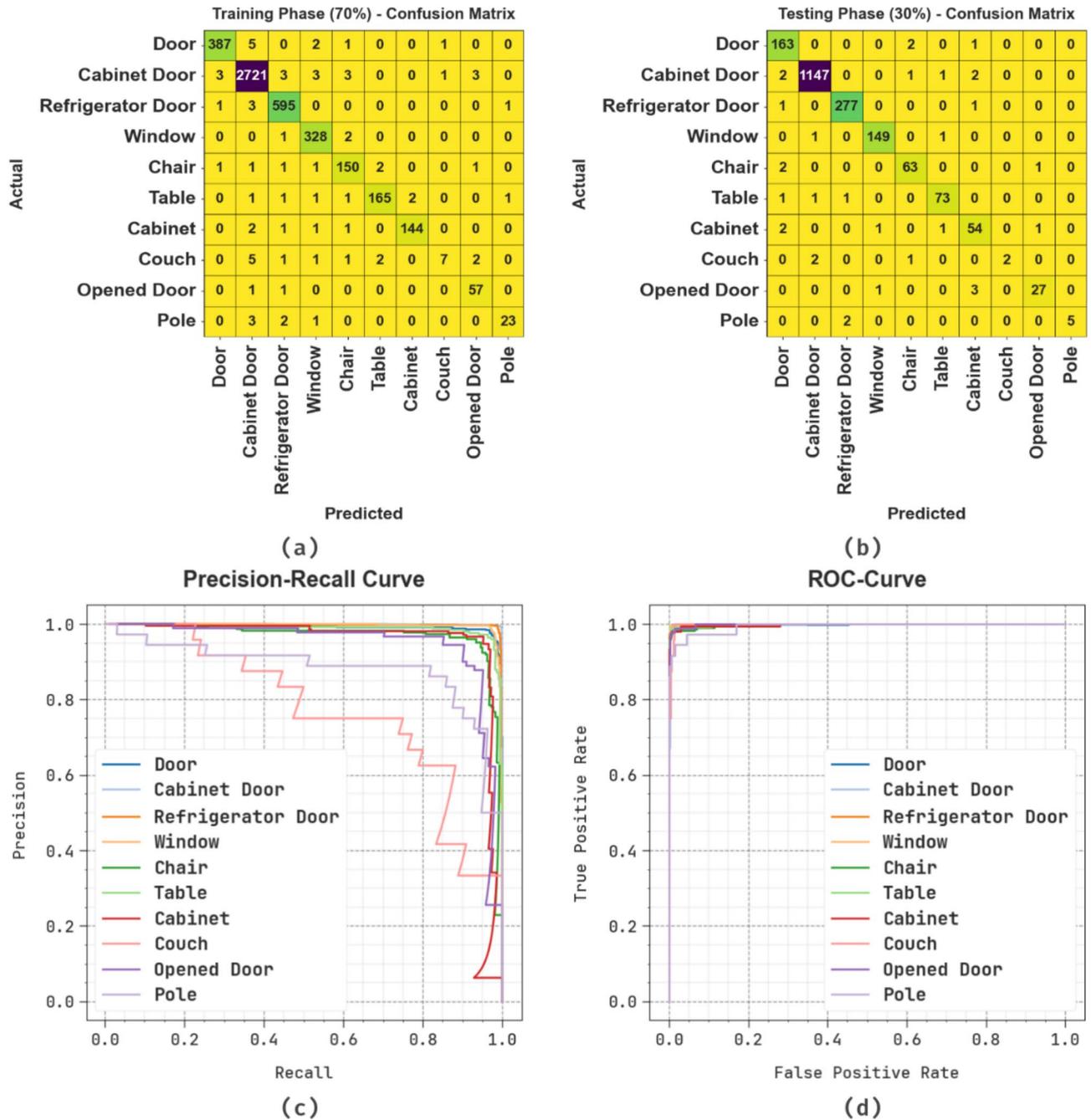


Fig. 7. Classifier outcome of ODMVII-MOA method (a-b) 70%TRAPHA and 30%TESPHA and (c-d) Curves of PR and ROC.

OD process to recognize and locate objects within an image. Besides, the proposed ODMVII-MOA method employs the EfficientNetB0 method for the feature extraction process. For the classification process, the LSTM-AE method is employed. Finally, the DO method adjusts the hyperparameter range of the LSTM-AE method optimally and results in better performance of classification. The experimental validation of the ODMVII-MOA model is verified under the indoor OD dataset and the outcomes are determined regarding different measures. The comparison study of the ODMVII-MOA model portrayed a superior accuracy value of 99.69% over existing techniques. The limitations of the ODMVII-MOA model comprise various aspects that could affect its generalizability and performance in real-world applications. First, the approach may be sensitive to discrepancies in environmental conditions, such as lighting changes or sensor noise, which could mitigate its robustness. Second, the reliance on a fixed dataset limits the adaptability of the technique to unseen scenarios or diverse environments. Furthermore, the computational complexity of the algorithm may affect real-time processing in resource-constrained systems. The requirement for large training datasets and the time-consuming nature of model tuning additionally restrict its scalability. Future works may concentrate on improving the efficiency of

Class	$Accu_y$	$< Prec_n$	$Recal_l$	$F1_{score}$	MCC
TRAPHA (70%)					
Door	99.70	98.72	97.73	98.22	98.06
Cabinet Door	99.20	99.23	99.42	99.32	98.36
Refrigerator Door	99.66	98.18	99.17	98.67	98.48
Window	99.72	97.04	99.09	98.06	97.91
Chair	99.66	94.34	95.54	94.94	94.76
Table	99.76	97.63	95.93	96.77	96.66
Cabinet	99.85	98.63	96.64	97.63	97.55
Couch	99.70	77.78	36.84	50.00	53.41
Opened Door	99.83	90.48	96.61	93.44	93.41
Pole	99.83	92.00	79.31	85.19	85.34
Average	99.69	94.40	89.63	91.22	91.39
TESPHA (30%)					
Door	99.45	95.32	98.19	96.74	96.45
Cabinet Door	99.50	99.65	99.48	99.57	98.97
Refrigerator Door	99.75	98.93	99.28	99.11	98.96
Window	99.80	98.68	98.68	98.68	98.57
Chair	99.65	94.03	95.45	94.74	94.56
Table	99.70	96.05	96.05	96.05	95.90
Cabinet	99.40	88.52	91.53	90.00	89.70
Couch	99.85	100.00	40.00	57.14	63.20
Opened Door	99.70	93.10	87.10	90.00	89.90
Pole	99.90	100.00	71.43	83.33	84.47
Average	99.67	96.43	87.72	90.53	91.07

Table 2. OD of ODMVII-MOA approach at 70:30 of TRAPHA/TESPHA.

the model through transfer learning and integrating adaptive mechanisms for enhanced robustness in dynamic environments. Additionally, exploring real-time deployment on edge devices and combining multi-modal data sources could result in a more versatile system.

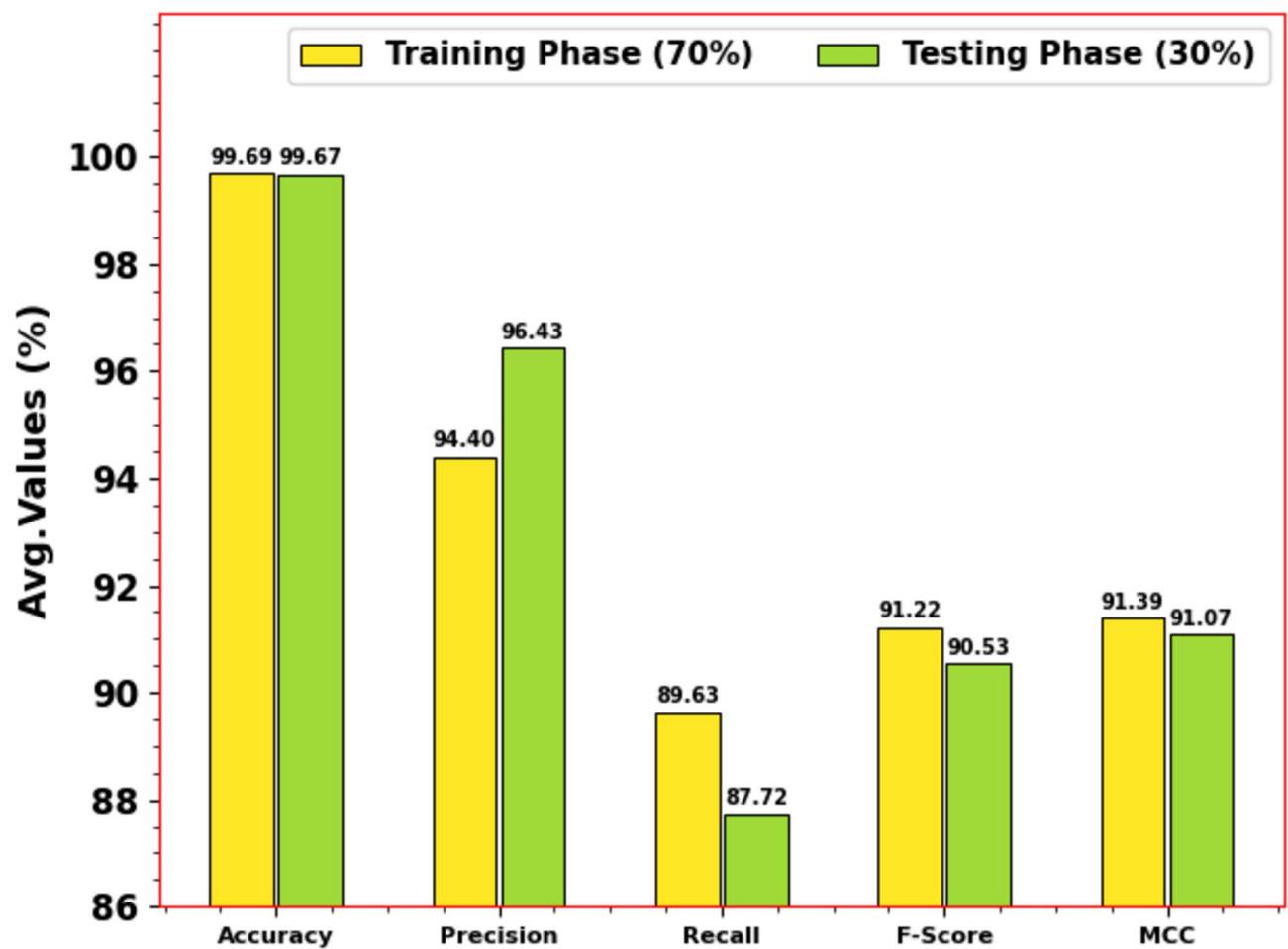


Fig. 8. Average of ODMVII-MOA model under 70:30 of TRAPHA/TEPHA.

Training and Validation Accuracy

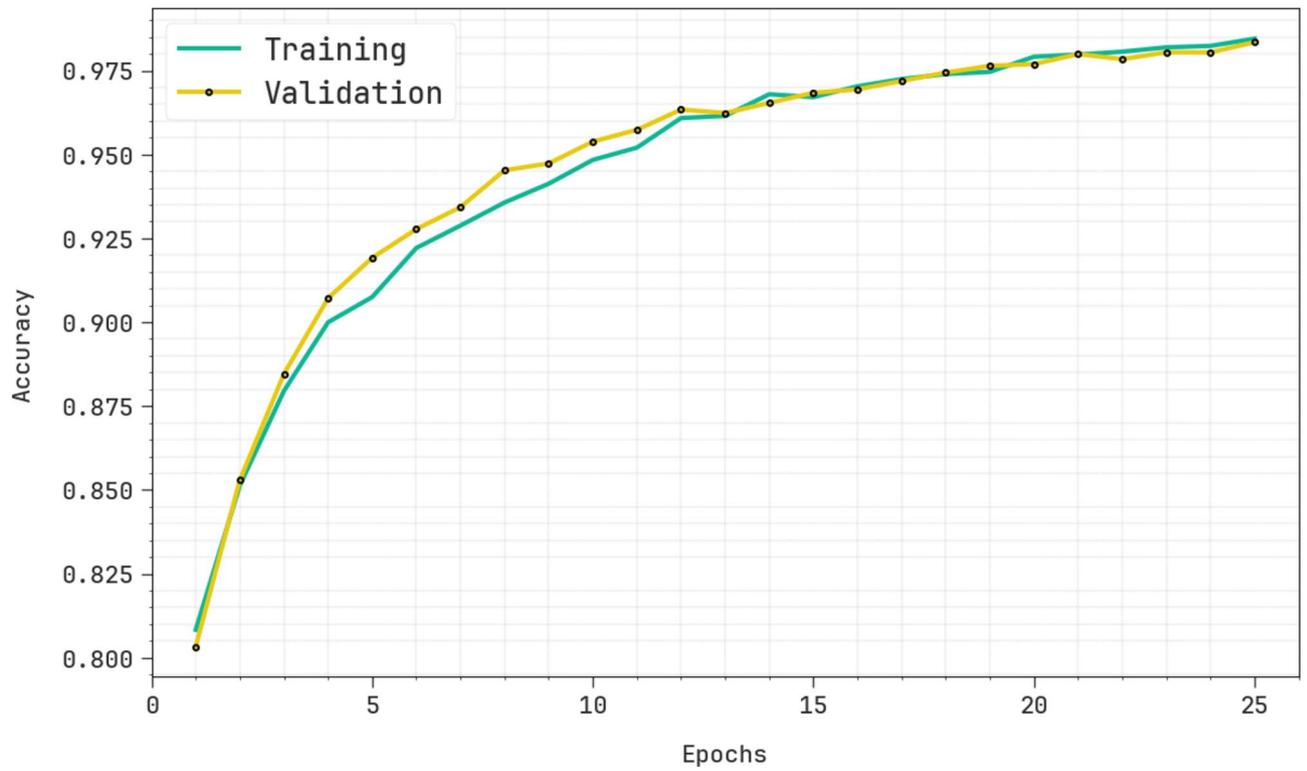


Fig. 9. $Accu_y$ curve of the ODMVII-MOA method.

Training and Validation Loss

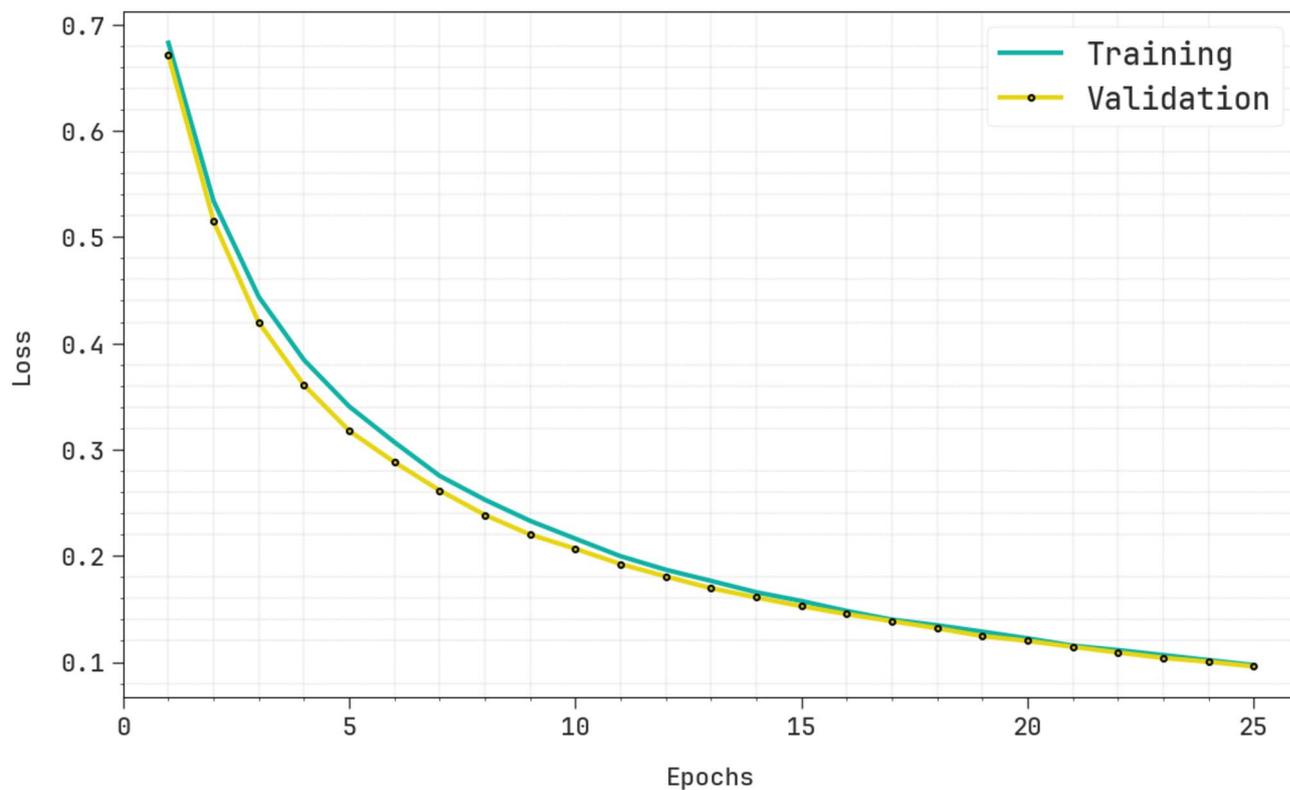


Fig. 10. Loss graph of ODMVII-MOA method.

Model	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal</i>	<i>F1_{score}</i>
Yolo-V8	98.38	84.88	78.57	82.83
Faster R-CNN	97.99	85.36	82.45	84.03
Yolo-v5n	92.01	85.77	82.52	82.64
GSDN Model	90.06	89.54	78.83	81.42
FCAF3D	88.50	89.22	78.77	89.80
Instant-NGP	92.90	89.13	77.30	83.09
3DGS-DR	97.34	89.54	88.99	83.86
ODMVII-MOA	99.69	94.40	89.63	91.22

Table 3. Comparative results of the ODMVII-MOA model with existing approaches^{34–36}.

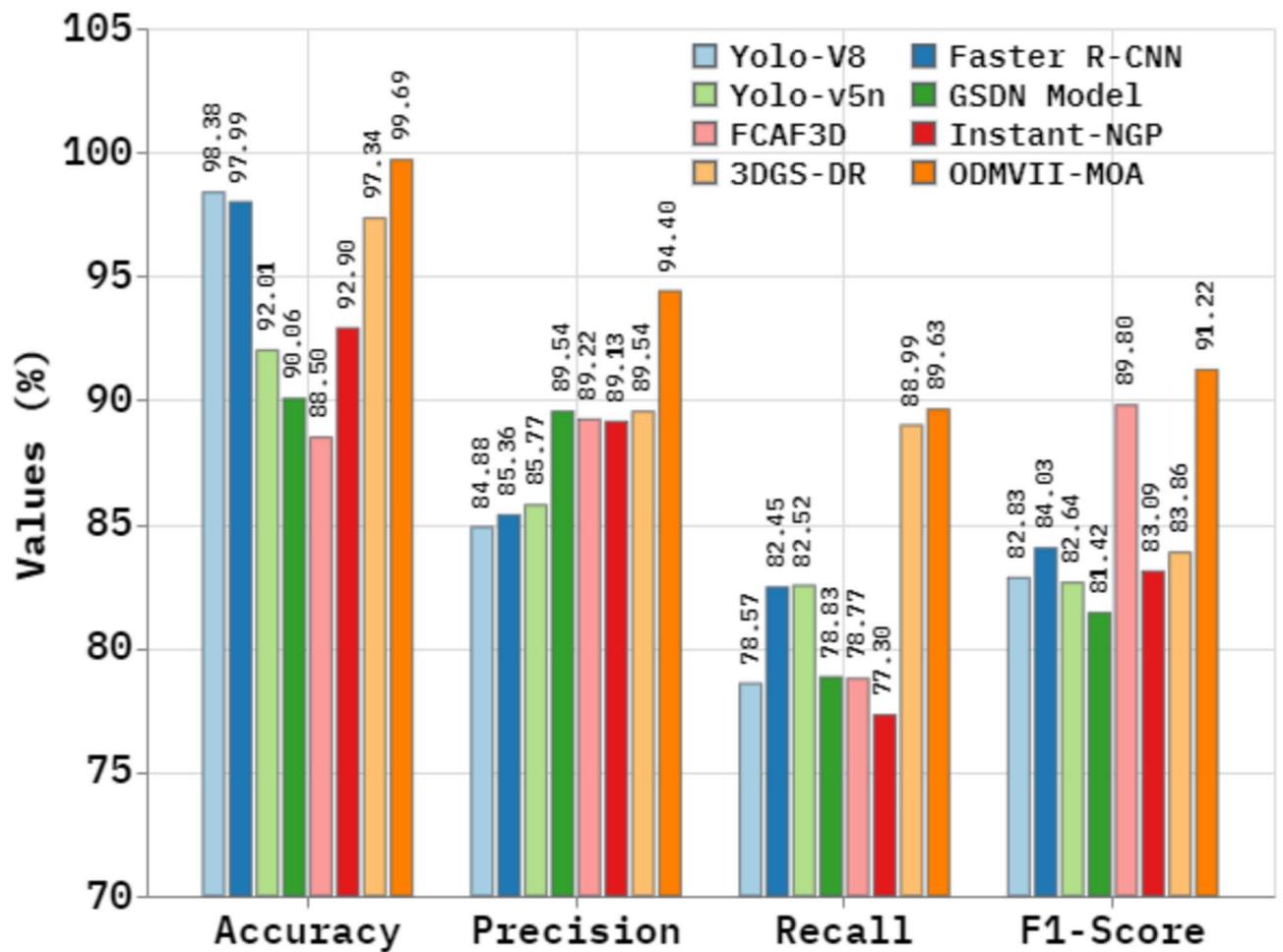


Fig. 11. Comparative analysis of the ODMVII-MOA model with existing approaches.

Technique	TC (sec)
Yolo-V8	21.28
Faster R-CNN	18.37
Yolo-v5n	13.98
GSDN Model	17.35
FCAF3D	25.43
Instant-NGP	18.13
3DGS-DR	18.40
ODMVII-MOA	10.86

Table 4. TC outcome of ODMVII-MOA technique with existing methods.

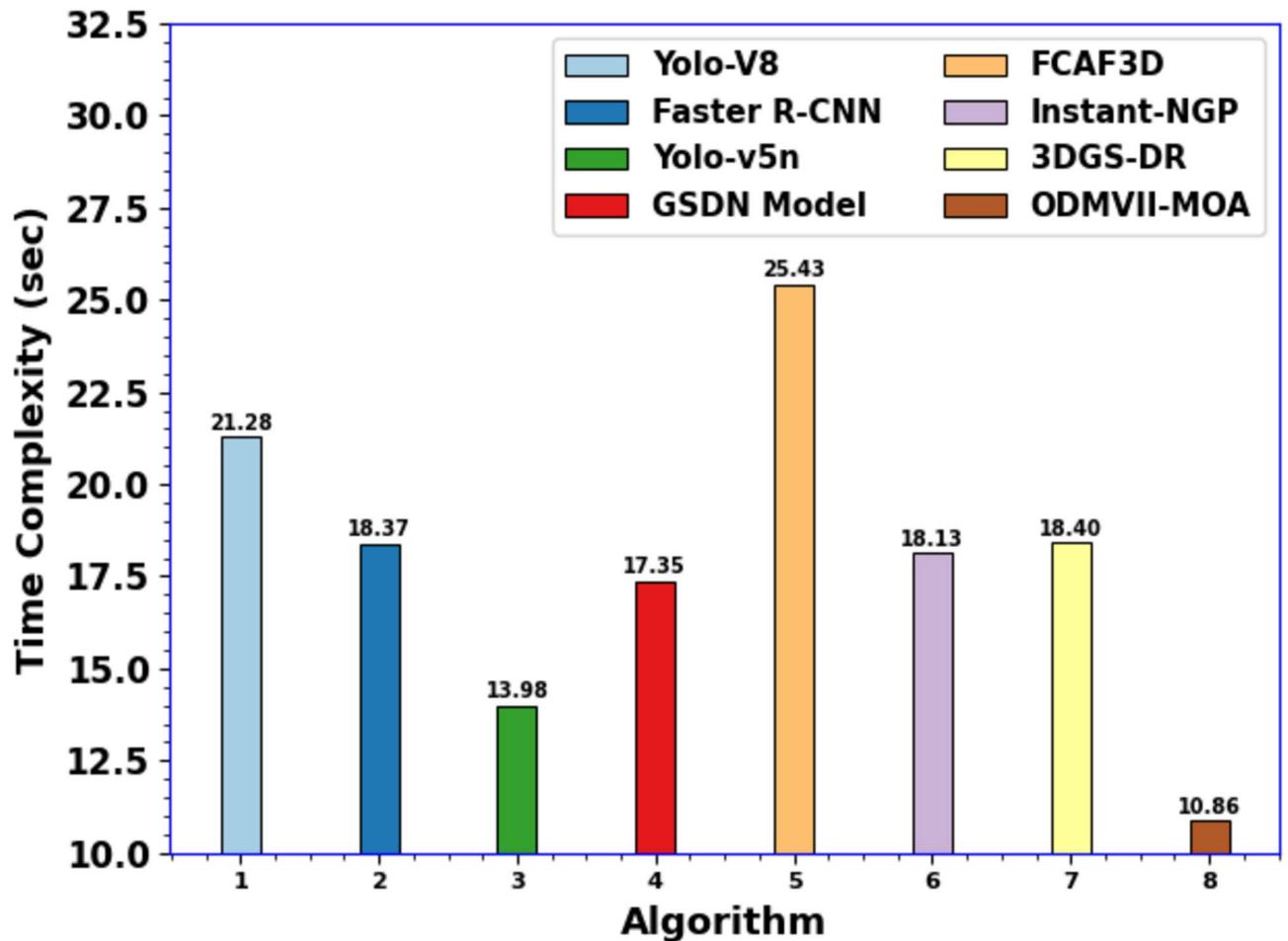


Fig. 12. TC outcome of ODMVII-MOA technique with existing methods.

Data availability

The data that support the findings of this study are openly available in the Kaggle repository at <https://www.kaggle.com/datasets/thebordindoor-object-detection>, reference number [33].

Received: 13 March 2025; Accepted: 23 April 2025

Published online: 08 May 2025

References

- Bhandari, A., Prasad, P. W. C., Alsadoon, A. & Maag, A. Object detection and recognition: using deep learning to assist the visually impaired. *Disabil. Rehabilitation: Assist. Technol.* **16** (3), 280–288 (2021).
- Naqvi, K., Hazela, B., Mishra, S. & Asthana, P. Employing real-time object detection for visually impaired people. In *Data Analytics and Management: Proceedings of ICDAM* (pp. 285–299). Springer Singapore. (2021).
- Theodorou, L. et al. October. Disability-first dataset creation: Lessons from constructing a dataset for teachable object recognition with blind and low vision data collectors. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1–12). (2021).
- Abid Siddique, R. V. & Naik, S. A Survey on Gesture Control Techniques for Smart Object Interaction in Disability Support. *Electronics*, **12**, p.512. (2023).
- Lecrosnier, L. et al. Deep learning-based object detection, localisation and tracking for smart wheelchair healthcare mobility. *International journal of environmental research and public health*, **18**(1), p.91. (2021).
- Joshi, R. C., Yadav, S., Dutta, M. K. & Travieso-Gonzalez, C. M. Efficient multi-object detection and smart navigation using artificial intelligence for visually impaired people. *Entropy*, **22**(9), p.941. (2020).
- Masud, U., Saeed, T., Malaikah, H. M., Islam, F. U. & Abbas, G. Smart assistive system for visually impaired people obstruction avoidance through object detection and classification. *IEEE Access*. **10**, 13428–13441 (2022).
- Jubair, M. A. et al. Multi-objective optimization in Satellite-Assisted UAVs. *J. Intell. Syst. Internet Things*, **16**(1). (2025).
- Nasreen, J., Arif, W., Shaikh, A. A., Muhammad, Y. & Abdullah, M. December. Object detection and narrator for visually impaired people. In *2019 IEEE 6th international conference on engineering technologies and applied sciences (ICETAS)* (pp. 1–4). IEEE. (2019).
- Madhuri, A. & Umadevi, T. Role of context in visual Language models for object recognition and detection in irregular scene images. *Fusion: Pract. Appl.*, **15**(1). (2024).
- Kadam, U. et al. Hazardous object detection for visually impaired people using edge device. *SN Comput. Sci.* **6** (1), 1–13 (2025).
- Chinni, N. P. K. et al. July. Vision Sense: Real-Time Object Detection And Audio Feedback System For Visually Impaired Individuals. In *2024 2nd World Conference on Communication & Computing (WCONF)* (pp. 1–6). IEEE. (2024).

13. Okolo, G. I., Althobaiti, T. & Ramzan, N. Smart Assistive Navigation System for Visually Impaired People. *Journal of Disability Research*, 4(1), p.20240086. (2025).
14. Talaat, F. M., Farsi, M., Badawy, M. & Elhosseini, M. SightAid: empowering the visually impaired in the Kingdom of Saudi Arabia (KSA) with deep learning-based intelligent wearable vision system. *Neural Comput. Appl.* **36** (19), 11075–11095 (2024).
15. More, S. S. et al. Empowering the visually impaired: YOLOv8-based object detection in android applications. *Procedia Comput. Sci.* **252**, 457–469 (2025).
16. Sugashini, T. & Balakrishnan, G. Visually impaired object segmentation and detection using hybrid Canny edge detector, Hough transform, and improved momentum search in YOLOv7. *Signal, Image and Video Processing*, 18(Suppl 1), pp.251–265. (2024).
17. Arifando, R., Eto, S., Tibyani, T. & Wada, C. January. Improved YOLOv10 for visually impaired: balancing model accuracy and efficiency in the case of public transportation. In *Informatics* (Vol. 12, No. 1, 7). MDPI. (2025).
18. Gomaa, A. & Saad, O. M. Residual Channel-attention (RCA) network for remote sensing image scene classification. *Multimedia Tools Appl.*, pp.1–25. (2025).
19. Gomaa, A. October. Advanced domain adaptation technique for object detection leveraging semi-automated dataset construction and enhanced yolov8. In *2024 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES)* (pp. 211–214). IEEE. (2024).
20. Verma, K. T. N., Prasad, S. & Kumar Singh, B. M. K. and Efficient diagnostic model for iron deficiency anaemia detection: a comparison of CNN and object detection algorithms in peripheral blood smear images. *Automatika*, 66(1), pp.1–15. (2025).
21. Gomaa, A., Abdelwahab, M. M. & Abo-Zahhad, M. Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis. *Multimedia Tools Appl.* **79** (35), 26023–26043 (2020).
22. Wei, Y., Tao, J., Wu, W., Yuan, D. & Hou, S. RHS-YOLOv8: A Lightweight Underwater Small Object Detection Algorithm Based on Improved YOLOv8. *Applied Sciences*, 15(7), p.3778. (2025).
23. Gomaa, A. & Abdalrazik, A. Novel deep learning domain adaptation approach for object detection using semi-self building dataset and modified yolov4. *World Electric Vehicle Journal*, 15(6), p.255. (2024).
24. Zhao, T., Feng, R. & Wang, L. SCENE-YOLO: A One-stage remote sensing object detection network with scene supervision. *IEEE Trans. Geosci. Remote Sens.* (2025).
25. Abdalrazik, A., Gomaa, A. & Afifi, A. Multiband circularly-polarized stacked elliptical patch antenna with eye-shaped slot for GNSS applications. *Int. J. Microw. Wirel. Technol.* **16** (7), 1229–1235 (2024).
26. Wang, X. et al. Transferable and robust dynamic adversarial attack against object detection models. *IEEE Internet Things J.* (2025).
27. Patil, S. et al. May. Object Recognition and Tracking System to Assist Visually Impaired: A Neural Network-Based Deep RBM Technique. In *2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)* (pp. 1–6). IEEE. (2024).
28. Yu, M., Su, J., Wang, Y. & Han, C. A noise reduction method for rolling bearing based on improved wiener filtering. *Rev. Sci. Instrum.*, **96**(2). (2025).
29. Wang, M., Liu, R., Luttrell, I. V., Zhang, J. & Xie, J. C. and Detection of masses in mammogram images based on the enhanced retinanet network with inbreast dataset. *J. Multidisciplinary Healthc.*, pp.675–695. (2025).
30. Gombasivam, G., Prabu kanna, G., Chauhan, M. S., Raja, P. & Kumar, Y. A hybrid deep learning model approach for automated detection and classification of cassava leaf diseases. *Scientific Reports*, 15(1), p.7009. (2025).
31. Duraj, A., Szczepaniak, P. S. & Sadok, A. Detection of Anomalies in Data Streams Using the LSTM-CNN Model. *Sensors*, 25(5), p.1610. (2025).
32. Al Hwaitat, A. K., Fakhouri, H. N., Zraqou, J. & Sirhan, N. Hybrid Optimization Algorithm for Solving Attack-Response Optimization and Engineering Design Problems. *Algorithms*, 18(3), p.160. (2025).
33. <https://www.kaggle.com/datasets/thebordin/indoor-object-detection>
34. Nimma, D. et al. Object detection in real-time video surveillance using attention based transformer-YOLOv8 model. *Alexandria Eng. J.* **118**, 482–495 (2025).
35. Yang, Y. Q. et al. Swin3d: A pretrained transformer backbone for 3d indoor scene Understanding. *Comput. Visual Media*. **11** (1), 83–101 (2025).
36. Cao, J., Cui, J. & Schwertfeger, S. GlassGaussian: extending 3D Gaussian splatting for realistic imperfections and glass materials. *IEEE Access*. (2025).

Acknowledgements

The authors extend their appreciation to the King Salman center for Disability Research for funding this work through Research Group no KSRG-2024-070.

Author contributions

Alaa O. Khadidos: Conceptualization, methodology, validation, investigation, writing—original draft preparation, funding Ayman Yafoz: Conceptualization, methodology, writing—original draft preparation, writing—review and editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025