scientific reports

Check for updates

OPEN

Leveraging assistive technology for visually impaired people through optimal deep transfer learning based object detection model

Mahir Mohammed Sharif Adam[1✉], Nojood O. Aljehane[2], Mohammed Yahya Alzahrani[3] & Samah Al Zanin[4,5]

Visual impairment, such as blindness, can have a profound impact on an individual's cognitive and psychological functioning. Therefore, the use of assistive techniques can help alleviate the adverse effects and enhance the quality of life for people who are blind. Most existing research primarily focuses on mobility, navigation, and object detection, with aesthetics receiving comparatively less attention, despite notable advancements in smart devices and innovative technologies for visually impaired individuals. Object detection is a crucial aspect of computer vision (CV), which involves classifying objects within images, enabling applications such as image retrieval, augmented reality, and many more. In recent times, deep learning (DL) techniques have become a powerful approach for extracting feature representations from data, leading to significant advancements in the field of object detection. In this paper, an enhanced assistive Technology for Blind People through Object Detection Using a Hiking optimization algorithm (EATBP-ODHOA) technique is proposed. The primary objective of the EATBP-ODHOA technique is to develop an effective object detection model for visually impaired individuals by utilizing advanced DL techniques. The image pre-processing stage initially employs an adaptive bilateral filtering (ABF) technique to improve image quality by removing unwanted noise. Furthermore, the Faster R-CNN model is used for the object detection process. Moreover, the EATBP-ODHOA method utilizes fusion models such as ResNet and DenseNet-201 for the feature extraction process. Additionally, the bidirectional gated recurrent unit (Bi-GRU) method is employed for the classification process. Finally, the parameter tuning of the fusion models is performed by using the Hiking Optimisation Algorithm (HOA) method. The experimentation of the EATBP-ODHOA model is performed under the indoor object detection dataset. The comparison analysis of the EATBP-ODHOA model revealed a superior accuracy value of 99.25% compared to existing approaches.

Vision is one of the most significant senses, and it is crucial for everyday life activities. According to data from the World Health Organisation (WHO), approximately 2.2 billion people experience blindness or vision impairment[1]. For people who are visually impaired or blind, assistive technology has emerged to support them. Because they need freedom and mobilization, however, most of this technology is not accessible and also costly[2]. Hence, they need lightweight, inexpensive, flexible, and practical solutions that operate in real-time with great accuracy. Technology improvements allow systems to analyze and look at environments as people do, with the use of CV[3]. For visually impaired people, CV can deliver visual data that enables them to recognize the world. At the same time, it may not entirely replace human vision, which relies on the eyes and brain to see. Object detection is a crucial task in CV, which allows for the gain of clever abilities and security for humans, robots/machines, and autonomous vehicles[4]. Object detection primarily focuses on classifying and identifying several objects in video

[1]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia. [2]Department of Computer Science, Faculty of Computers and Information Technology, University of Tabuk, Tabuk, Saudi Arabia. [3]Faculty of Computing and Information, Al-Baha University, Al-Aqiq, Saudi Arabia. [4]Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Kharj, Saudi Arabia. [5]King Salman Centre for Disability Research, Kingdom of Saudi Arabia, Riyadh 11614, Saudi Arabia. ✉email: m.adam@psau.edu.sa

frames and digital images. Additionally, Object recognition is one of the essential modules in computer vision, addressing the classification of objects within images, which enables applications such as augmented reality and content-based image retrieval[5]. Over the past few years, Object detection has been a vital and long-standing CV problem, which remains a main active research field[6]. Several applications utilize object detection, including face recognition, face detection, security systems, self-driving cars, pedestrian counting, and vehicle detection[7]. Some of the CV terms, such as identifying, object positioning, and recognizing, are interconnected to the object detection process. However, it is a critical requirement for many complex CV tasks, namely event detection, behaviour analysis, scene semantic understanding, and target tracking[8]. Additionally, it is widely utilized in video and image retrieval, intelligent video surveillance, industrial inspection, autonomous vehicle driving, and other fields[9]. Each object has specific characteristics that help identify its class. A conventional detection algorithm for physically removing characters for unique task recognition. The DL approach has become a significant method for obtaining characteristic demonstrations from information in a straightforward manner, resulting in essential developments in the field of object detection[10].

In this paper, an enhanced assistive Technology for Blind People through Object Detection Using a Hiking optimization algorithm (EATBP-ODHOA) technique is proposed. The primary objective of the EATBP-ODHOA technique is to develop an effective object detection model for visually impaired individuals by utilizing advanced DL techniques. The image pre-processing stage initially employs an adaptive bilateral filtering (ABF) technique to improve image quality by removing unwanted noise. Furthermore, the Faster R-CNN model is used for the object detection process. Moreover, the EATBP-ODHOA method utilizes fusion models such as ResNet and DenseNet-201 for the feature extraction process. Additionally, the bidirectional gated recurrent unit (Bi-GRU) method is employed for the classification process. Finally, the parameter tuning of the fusion models is performed by using the Hiking Optimisation Algorithm (HOA) method. The experimentation of the EATBP-ODHOA model is performed under the indoor object detection dataset. The significant contribution of the EATBP-ODHOA model is listed below.

- The EATBP-ODHOA approach utilizes the ABF model during image pre-processing to effectively mitigate noise while preserving crucial edge details, resulting in clearer and higher-quality images. This enhances the accuracy of subsequent object detection and classification steps, ultimately resulting in more reliable performance in challenging visual conditions.
- The EATBP-ODHOA model employs Faster R-CNN to perform precise and efficient object detection, effectively managing intrinsic and cluttered environments. Its region proposal network improves accuracy by generating high-quality candidate regions, which speeds up detection without sacrificing performance. This approach enhances the model's capability to detect multiple objects in real-time, making it suitable for dynamic scenarios.
- The EATBP-ODHOA methodology combines ResNet and DenseNet-201 models to extract rich and complementary feature representations, capturing both deep hierarchical and dense connectivity patterns. This integrated technique improves feature diversity and the overall discriminative power for complex object recognition tasks. By leveraging the merits of both architectures, the model achieves more robust and accurate feature extraction, leading to improved detection and classification performance.
- The EATBP-ODHOA method utilizes Bi-GRU to capture bidirectional dependencies within sequential data, thereby enhancing context understanding from both past and future inputs. The model also effectually learns intrinsic temporal patterns and improves the robustness of sequence classification. The model enhances the capability of handling varied and dynamic data sequences.
- The EATBP-ODHOA model utilizes the HOA technique to optimize hyperparameters, thereby enhancing the overall training process efficiently. This method enables faster convergence by effectively exploring the search space and avoiding local minima. As a result, it improves performance and accuracy, and also ensures a more reliable and fine-tuned model suitable for complex tasks.
- The EATBP-ODHOA approach uniquely incorporates ABF-based pre-processing, dual fusion-based feature extraction using ResNet and DenseNet-201, Bi-GRU-based classification, and HOA-based tuning. This integration is specifically optimized to present accurate and real-time object detection and classification for visually impaired users. The novelty is in the seamless fusion of these advanced techniques to address both performance and efficiency challenges in assistive technology applications.

Literature review

Aniedu et al.[11] presented an AI-based navigation method employing the Raspberry Pi single-board computer. The progression emulates an agile model, highlighting flexibility. This paper examines a significant technology, which is divided into two major subsystems: software and hardware. The processed images were identified and classified employing the You Only Look Once version 3 (YOLO-v3) model. Arsalwad et al.[12] projected a novel AI-powered real-world assistive method. By utilizing advanced AI models, it provides navigation support, personalized assistance, and flawless communication for users with visual challenges. With an intuitive, adaptive technology and user-friendly interface, the solution enables individuals with visual impairments to navigate spaces independently and with greater confidence. Yu and Saniie[13] developed the Visual Impairment Spatial Awareness (VISA) method, intended to aid visually impaired consumers in navigating indoor activities across multi-level and structured environments. The progressive level synthesizes this capability to improve spatial awareness, allowing users to navigate intricate settings. Baig et al.[14] advanced a new wearable vision assistive method that has a hat-mounted camera with AI technology. The key aspect of this method is a user-friendly process for identifying novel objects or people, which involves a one-click procedure that enables users to add data on new objects or individuals for later identification. Sabit[15] progressed the development and design of an AI-based Smart Security System utilizing IoT technologies for smart home applications. Joshi et al.[16] proposed

an AI-based novel wearable assistive gadget, AI-SenseVision, to examine sensory and visual data about objects and complications in the scene, thereby recognizing the surrounding environment. The gadget is a wide-ranging amalgamation of sensors and CV-based technology, which creates auditory data with the name of recognized audio or object warnings for identifying difficulties. Sophia et al.[17] developed an automated indoor navigation method designed for people who are blind or visually impaired. This technique also projects the execution of a navigation model employing visible light communication technology. Finally, this method is enhanced with an APR voice to utilize a speaker and provide information to blind users.

Naz and Jabeen[18] presented a comprehensive survey of object detection techniques for visually impaired individuals, integrating both traditional and DL classification methods. Malkocoglu and Samli[19] proposed the deep channel attention super resolution (DCASR) model to improve dangerous object detection by enhancing image quality before classification. The method also integrates the YOLOv5 detection framework. Zhang et al.[20] proposed a CNN-Transformer Fusion Network (CTF-Net) methodology that incorporates convolutional neural networks (CNNs) for local features and transformers for global context. The cross-domain fusion module (CDFM), boundary-aware module (BAM), and feature enhancement module (FEM) are also utilized to enhance accuracy. Zin et al.[21] proposed a computer vision-based action recognition system that supports ambient assisted living by monitoring elderly individuals using stereo depth cameras. It incorporates depth motion appearance (DMA), depth motion history (DMH), and histogram of oriented gradients (HOG) with distance-based features to recognize actions from long depth frame sequences. Khadidos and Yafoz[22] proposed an object detection model for visually impaired individuals using a metaheuristic optimization algorithm (ODMVII-MOA) technique to enhance real-time object detection for visually impaired users. The model integrates the Wiener filter (WF) for image pre-processing, RetinaNet for object detection, EfficientNetB0 for feature extraction, and LSTM autoencoder (LSTM-AE) for classification. Aung et al.[23] presented an object recognition system integrating the YOLOv4 method, specifically designed to assist visually impaired individuals. Alsolai et al.[24] presented the sine cosine optimization with DL–driven object detector (SCODL–ODC) methodology by incorporating YOLOv5 for object detection, sine cosine optimization (SCO) for tuning, and a deep stacked autoencoder (DSAE) for classification and Harris Hawks Optimisation (HHO) model for optimizing the DSAE model. Bhandari et al.[25] proposed a model by using YOLOv8 and TL to detect new objects crucial for visually impaired users. Venkatkrishnan et al.[26] proposed an IoT-based voice-driven smart finder that integrates natural language processing (NLP), YOLOv8 for real-time object detection, to assist visually impaired individuals in locating household objects via voice commands. Srivastava et al.[27] presented a real-time object detection system for visually impaired individuals using the YOLO dataset and ML in Python.

The existing models are effective in AI-based assistive systems, but still encounter various difficulties, such as limited adaptability in dynamic real-world settings. Various techniques primarily focus on indoor or structured environments. In contrast, a few methods lack integration with efficient real-time feedback mechanisms, such as voice assistance, which is crucial for practical usability. Furthermore, most techniques heavily depend on predefined datasets, which limits the detection of novel or uncommon objects. The computational complexity of advanced models also poses challenges for deployment on low-power devices. The research gap lies in developing robust, real-time, and adaptable OD systems that strike a balance between accuracy, speed, and usability, while effectively addressing these constraints.

Proposed model
In this paper, the EATBP-ODHOA methodology is proposed. The main aim of the EATBP-ODHOA methodology is to develop an effective object detection model for visually impaired people using an advanced DL technique. Figure 1 portrays the workflow of the EATBP-ODHOA model.

Image pre-processing
Initially, the image pre-processing stage employs an ABF technique to enhance the image quality by removing unwanted noise[28]. This model is chosen for its superior capability in reducing noise while preserving crucial edges and fine details, which are significant for accurate object detection. This technique adaptively adjusts filtering based on local image characteristics, maintaining structural integrity, unlike conventional smoothing filters that often blur edges. This results in improved image quality without loss of crucial data, thereby enhancing downstream feature extraction and detection accuracy. Compared to other methods, such as Gaussian or median filtering, ABF provides a better balance between noise reduction and edge preservation, making it highly effective for pre-processing images in complex and noisy environments.

ABF is a progressive image processing model that enhances object detection by preserving edges while reducing noise. Unlike conventional bilateral filtering, ABF dynamically adjusts filter parameters based on local image features, thereby enhancing feature preservation in fluctuating illumination and texture conditions. This makes it mainly beneficial in tasks of object detection where upholding edge facts is vital. By adaptively smoothing areas while preserving sharp changes, ABF enhances the clarity of perceived objects, thereby improving the accuracy of ML and CV methods. It is widely used in applications such as medical imaging, image segmentation, and assistive technologies for individuals with visual impairments.

Faster R-CNN object detection
Besides, the Faster R-CNN method is employed for the object detection process[29]. This method is chosen for its efficiency in detecting objects. The model integrates Region Proposal Network (RPN) with a fast CNN-based detector, enabling end-to-end training and reducing the computational cost compared to earlier R-CNN variants. The technique also effectively balances speed and precision, making it appropriate for real-time applications. Compared to single-stage detectors like YOLO or SSD, it often provides better detection
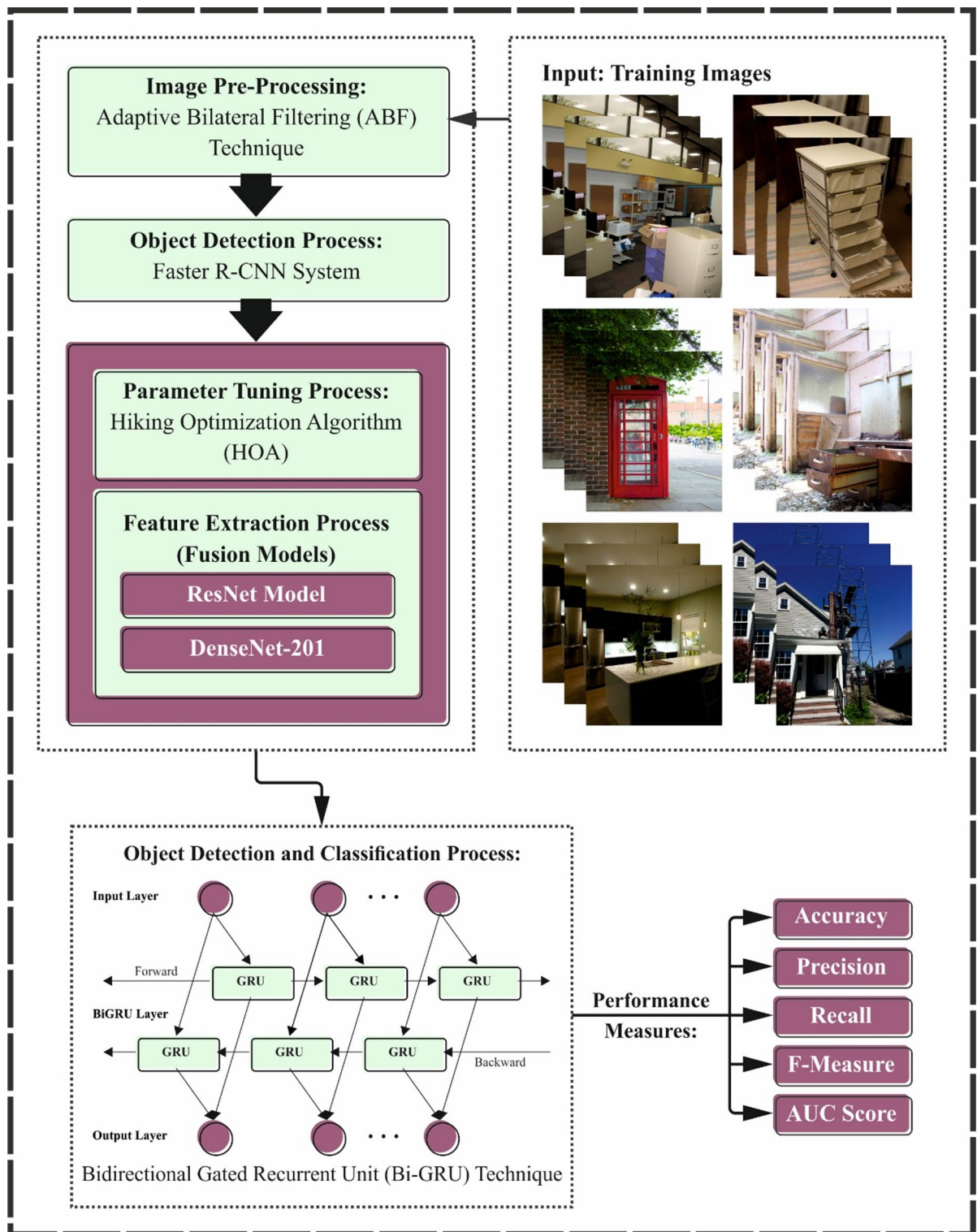
Fig. 1. Workflow of the EATBP-ODHOA model.

performance, particularly for small or overlapping objects. Its robustness in diverse environments and ability to localize objects precisely justify its selection over other techniques.

Faster R-CNN is an advanced object detection structure that enhances its predecessors by combining a region proposal network (RPN) to generate higher-quality proposals of areas. The structure includes dual major modules: the RPN, which is responsible for presenting candidate object bounding boxes, and the Faster R-CNN detector, which classifies these objects and improves their coordinates. By distributing convolutional

characteristics between the detection network and the RPN, Fast R-CNN achieves significant improvements in either precision or speed.

The function of loss applied in Fast R-CNN comprises two major modules: the bounding box regression loss and the classification loss, both of which are measured in the detection network and RPN modules. The complete loss L was described as demonstrated:

$$L_{FasterR-CNN} = L_{c1s}^{RPN} + L_{reg}^{RPN} + L_{c1s}^{Det} + L_{reg}^{Det} \tag{1}$$

Whereas L_{cls} represents loss of classification, usually calculated using $softmax$ cross-entropy for the forecast classes, and L_{reg} denotes loss of bounding box regression, which calculates the precision of the forecast bounding box coordinates. The loss of bounding box regression is frequently expressed as a smooth L1 loss.

Faster R-CNN serves as the backbone structure in most works that combine field adaptation components for object detection, owing to its modular and flexible design, which enables the incorporation of adaptation methods targeted at enhancing performance across various fields.

Fusion of feature extraction

For the feature extraction process, the EATBP-ODHOA method utilizes fusion models such as ResNet and DenseNet-201. The ResNet method effactually addresses the vanishing gradient problem, enabling deep networks to be trained with improved accuracy. Furthermore, DenseNet-201 promotes feature reuse and strengthens gradient flow, resulting in more effective learning and generalization due to its dense connectivity. Integrating these models utilizes both ResNet's robust feature extraction and DenseNet's efficient data propagation, resulting in a robust and comprehensive feature set. This fusion outperforms many single architectures by capturing diverse and discriminative features significant for accurate object detection.

ResNet model
ResNet is an extension of deep nets that transforms the CNN structural race by incorporating residual learning and an effective process for deep Net training[30]. ResNet is twenty times deeper than AlexNet and eight times more profound than VGG, expressing lower intricacy of computations than preceding Nets. Experimental analysis indicated that ResNet with 50/101/1521 layers performed better on image classification challenges than a plain Net with 34 layers. Moreover, ResNet enhanced the classification of the image benchmark dataset by 28 per cent, emphasizing the importance of depth in visual identification challenges. To overcome a problem through deep net training, ResNet utilizes bypass channels. The expression of ResNet is given:

$$g(x_i) = f(x_i) + x_i \tag{2}$$

$$f(x_i) = g(x_i) - x_i \tag{3}$$

Here, $f(x_i)$ specifies the transformed signal, χ_i represents the original input and is comprised of $f(x_i)$ over the bypass pathway. In basic terms, $g(x_i) - x_i$ accomplishes the residual learning. ResNet presents a shortcut connection in layers to allow interconnections through layers, while these connections are data-independent and do not necessitate parameters, unlike the Highway Network. ResNet constantly transfers residual data and doesn't extract shortcuts of identity. Residual links, also known as shortcut connections, accelerate the convergence of deep networks, enabling ResNet to circumvent the difficulties associated with gradient vanishing. VGG-19 is based on a ResNet 34-layer design, which is a simple network, and a shortcut connection is included afterwards. These shortcut connections change the design of the residual network. Figure 2 illustrates the structure of ResNet.

Investigators employed dual attention loss to tackle the problem of data sample imbalance through experimentation. ResNet-50 has certain restrictions. One concern is that the convoluted design needs a substantial variety of memory and computer resources. Furthermore, training ResNet-50 is a challenging and time-consuming process.

DenseNet201 method
DenseNet-201is a CNN structure that has become a powerful device in the field of DL, mainly in CV tasks[31]. DenseNet-201 is proposed to rely on the theory of densely linked convolutional layers to improve feature propagation and boost feature reuse within the system. Unlike conventional CNNs, which accept inputs only from the previous layer, DenseNet-201 establishes direct connections among all layers within a dense block, thereby promoting rich feature representations. This feature not only enables the flow of gradients in training but also alleviates the problem of vanishing gradients, allowing more efficient training of deep systems. By its exceptional performance on benchmark datasets, such as ImageNet, it has become a foundation for various applications, including segmentation, image classification, and object detection. Besides, its dense, yet suggestive framework makes it suitable for use in resource-constrained environments.

Object detection and classification using BiGRU

Moreover, the Bi-GRU model is employed for the classification process[32]. This model is chosen for its effectual capability in capturing contextual data from both past and future data sequences, improving the comprehension of spatial and temporal dependencies in image features. This methodology is computationally efficient, enabling faster training and improved performance. The model also mitigates vanishing gradient issues, and their bidirectional structure provides a more comprehensive representation of features, which is particularly beneficial for complex object classification tasks. This results in enhanced accuracy and robustness compared
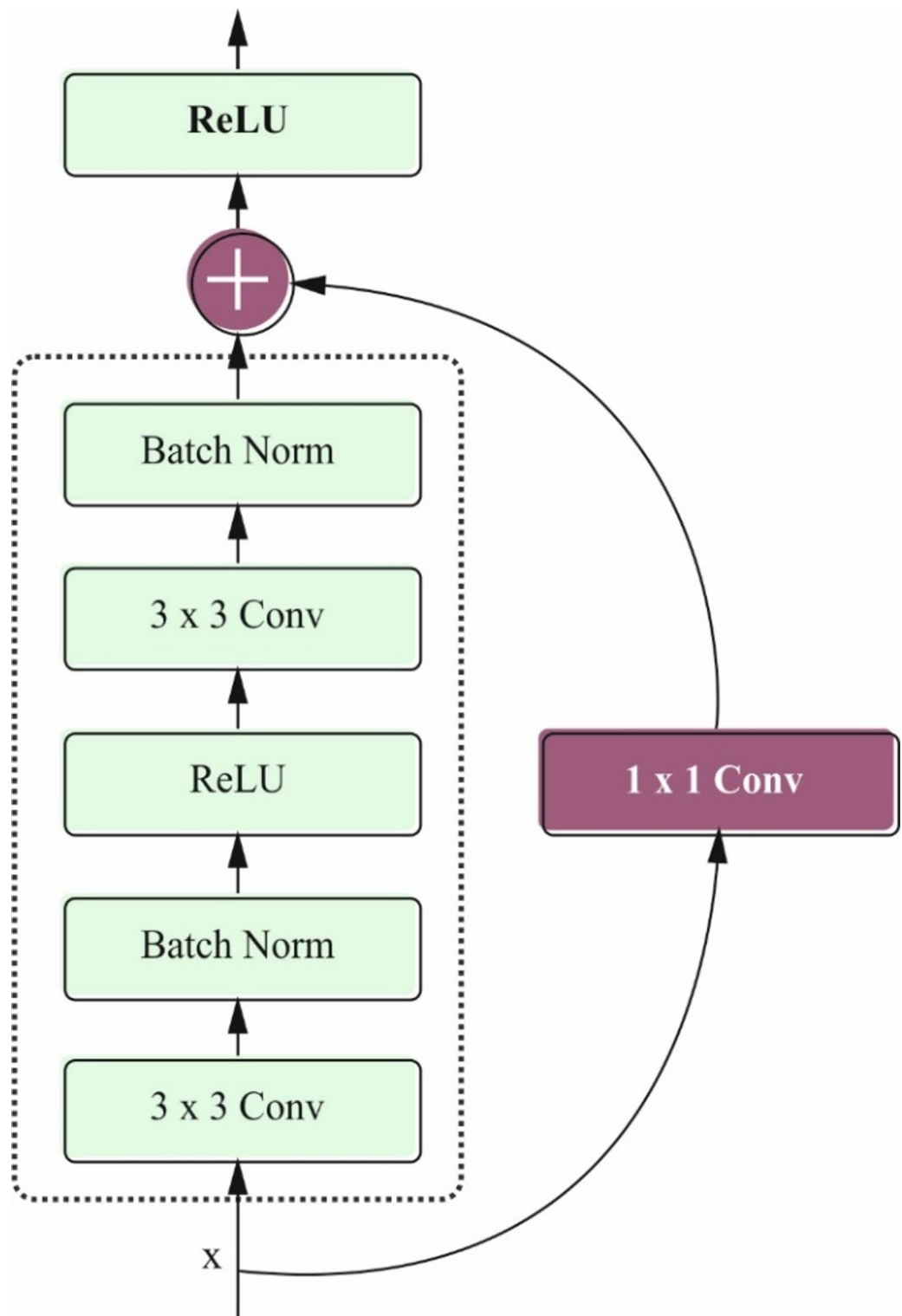
Fig. 2. Architecture of ResNet.

to unidirectional or simpler models, making BiGRU suitable for dynamic and sequential data in the object detection process.

For mapping the multivariable time series input data to the 1D output of the time series, a DL method is trained and developed, including BiGRU layers, that is suitable for the modelling of sequential data. This networking structure comprises dual layers of BiGRU, accompanied by a batch normalization (BN) layer for stabilizing learning, a dropout layer for preventing overfitting, and a dense layer for output combination. Incorporating

dual Bi-GRU layers improved the network's ability to discern composite relations and deep characteristics in the data; however, including more than two layers didn't increase performance due to a lack of data.

The Bi-GRU structure can capture either forward or reverse temporal dependencies in the input data, making it suitable for a time-series model of gait measures. In comparison with LSTMs, BiGRUs offer comparable performance with reduced computational costs. However, Transformer-based methods have proven their effectiveness in numerous fields; they usually require large datasets for efficient training, which may not align with the size of the dataset. During the development of recurrent neural networks (RNNs), the GRU was presented. During this first state, the output vector begins at 0:

$$h_t = 0, \ t = 0 \ . \ \text{Formerly,}$$

$$z_t = sigmoid\left(W_z x_t + U_z h_{t-1} + b_z\right) \tag{4}$$

$$r_t = sigmoid\left(W_r x_t + U_r x_t + b_r\right) \tag{5}$$

$$\overline{h}_t = tanh(W_h x_t + U_h(r_t \odot \ h_{t-1}) + b_h \tag{6}$$

$$h_t = (1 - z_t) \odot \ h_{t-1} + z_t \odot \ \overline{h}_t \tag{7}$$

Whereas, $Error :: 0x0000$ explains dot-product, x_t refers to the vector of input at time-step t, h_t signifies the vector of output, $\overline{h}$ stands for the vector of the candidate activation, r_t and z_t symbolize reset and update gate vectors, W and U represent matrices of the parameters, and b denotes the vector of parameters. The update gate (z) utilizes the sigmoid function to determine how much previous information should be maintained, thereby affecting the sum of information being processed. The reset gate (r_t) selects the degree to which previous data is forgotten, allowing the method to remove unrelated information. The candidate's activation $\left(\overline{h}\right)$ recommends novel memory content by integrating novel input to prior data, which is fine-tuned by the reset gate. At last, the last memory update (h_t) combines the older status and the novel candidate activation, as regulated by the update gate, to update the present state successfully. In contrast to GRU, BiGRU is made by dual unidirectional GRUs that face opposite directions. The computation of BiGRU is expressed as shown.

$$\overrightarrow{h}_t = GRU_{fwd}\left(x_t, \ \overrightarrow{h}_{t-1}\right) \tag{8}$$

$$\overleftarrow{h}_t = GRU_{bwd}\left(x_t, \ \overleftarrow{h}_{t-1}\right) \tag{9}$$

$$h_t = \overrightarrow{h}_t \oplus \ \overleftarrow{h}_t \tag{10}$$

Now, $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ represent the state data of the forward and reverse GRU, respectively. GRU_{fwd} and GRU_{bwd} signify forward and reverse GRU. Either GRU follows the expression of Eq. (4) to (7). $\oplus$ signifies connecting the $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$.

Parameter fine-tuning using HOA

Finally, the parameter tuning of fusion models is employed by using the HOA method[33]. This method is chosen for its faster convergence and enhanced capabilities for exploration and exploitation. This technique helps find optimal hyperparameters with faster convergence and mitigates the risk of getting stuck in local minima. The HOA model replicates the adaptive and iterative behaviour of hikers, allowing for the dynamic adjustment of search steps to achieve better accuracy compared to conventional optimization techniques. The simplicity and low computational overhead of the method also make the model appropriate for fine-tuning intrinsic models without excessive resource consumption. By effectively balancing global and local search, HOA enhances model performance and robustness, outperforming many conventional algorithms in parameter optimization tasks.

HOA is a meta-heuristic model that imitates the collective knowledge of a group of hikers, while considering the ruggedness of the terrain and avoiding steep regions. The region concealed by the trip is equivalent to the searching region. Additionally, due to the complex nature of the land, the hiker might struggle to find the shortest path to the target, similar to the model getting stuck in a local optimum. The mathematical expression of HOA is as demonstrated:

Randomly initializing the primary place of the hiker.

$$\beta_{i_i \tau} = \varphi_j^1 + \delta_j\left(\varphi_j^2 - \varphi_j^1\right)\# \tag{11}$$

Whereas $\beta_{i,t}$ signifies initializing the primary place of the hiker. δ_j refers to uniformly distributed numbers between $(0,1)$. Additionally, φ_j^1 and φ_j^2 signify the minimal and maximal values of the jth size of the variables in the optimizer difficulty.

Define the primary speed of the hiker based on the steepness of the land.

$$\mathcal{W}_{i,t} = 6e^{-3.5|s_{i,t}+0.05|} \tag{12}$$

Here, $\mathcal{W}_{i,t}$ characterizes the hiker's speed i at the tth iteration; $s_{i,t}$ signifies the steepness of the land, and its computation equation is presented in Eq. (13).

Establish the steepness of the territory based on the inclination angle of the land.

$$S_{i,t} = \frac{dh}{dx} = tan\theta_{i,t} \tag{13}$$

Whereas dh signifies the vertical change in height and dx characterizes the horizontal distance travelled by the hiker. $\theta_{i,t}$ refers to the inclination angle of the land that ranges from 0 to 50 degrees.

$$\mathcal{W}_{i,t} = \mathcal{W}_{i,t-1} + \gamma_{i,t} \left(\beta_{best} - \alpha_{i,t}\beta_{i,t} \right) \tag{14}$$

Now, $\gamma_{i,t}$ means uniformly distributed number among (0,1). $\mathcal{W}_{i,f}$ and $\mathcal{W}_{i,t-1}$ signify the present and primary velocity of hiker i at the tth iteration, correspondingly. $\beta_{i,t}$ represents the location of the leading hiker at tth iteration, and $\alpha_{i,t}$ denotes the sweep feature of hiker i at tth iteration that lies between (1,3).

Upgrade the location of the hiker based on the present speed of the hiker.

$$\beta_{i,t+1} = \beta_{i,t} + W_{i,t} \tag{15}$$

In the HOA, the fitness selection is a significant factor that influences performance. The procedure's hyperparameter range contains the encoded technique to evaluate the effectiveness of the candidate outcome. The HOA reflects accuracy as the primary measure to project the FF.

$$Fitness = \max (P) \tag{16}$$

$$P = \frac{TP}{TP + FP} \tag{17}$$

Here, TP and FP represent the positive values of true positives and false positives, respectively.

Result analysis

The experimental analysis of the EATBP-ODHOA technique is examined using the indoor object detection dataset[34]. This dataset comprises 6,642 samples from 10 classes, as listed in Table 1.

Figure 3 presents the confusion matrices generated by the EATBP-ODHOA approach for TRAPA/TESPA ratios of 80:20 and 70:30. The results indicate that the EATBP-ODHOA methodology is efficient in detecting and recognizing all classes.

Table 2 represents the object detection results of the EATBP-ODHOA approach under an 80:20 TRAPA/TESPA split. The outcome stated that the EATBP-ODHOA approach has appropriately classified each of the dissimilar classes. Concerning 80% TRAPA, the EATBP-ODHOA approach achieves an average $accu_y$ of 99.04%, $prec_n$ of 80.56%, $reca_l$ of 73.30%, $F_{Measure}$ of 75.34%, and AUC_{score} of 86.31%. Following this, on 20% TESPA, the EATBP-ODHOA technique achieves an average $accu_y$ of 99.25%, $prec_n$ of 84.88%, $reca_l$ of 72.60%, $F_{Measure}$ of 75.08%, and AUC_{score} of 86.02%.

In Fig. 4, the training (TRAN) $accu_y$ and validation (VALN) $accu_y$ results of the EATBP-ODHOA model under 80:20 are demonstrated. The outcome highlighted that both $accu_y$ analysis exhibits a rising trend, which indicates the competency of the EATBP-ODHOA approach with maximal performance across numerous iterations. Additionally, both $accu_y$ remain relatively close across the epochs, indicating lesser overfitting and demonstrating the optimal performance of the EATBP-ODHOA approach.

In Fig. 5, the TRAN loss and VALN loss graphs of the EATBP-ODHOA approach with an 80:20 ratio are exhibited. It is noted that both values exhibit a reducing trend, indicating the capacity of the EATBP-ODHOA model in balancing a trade-off between data fitting and generalization. The constant reduction in loss values also ensures a more favourable outcome for the EATBP-ODHOA model, improving prediction accuracy.

Table 3 presents the object detection results of the EATBP-ODHOA methodology under a 70:30 TRAPA/TESPA ratio. The outcome stated that the EATBP-ODHOA technique has correctly classified all the dissimilar classes. Depending on 70% TRAPA, the EATBP-ODHOA technique achieves an average $accu_y$ of 99.61%,

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**

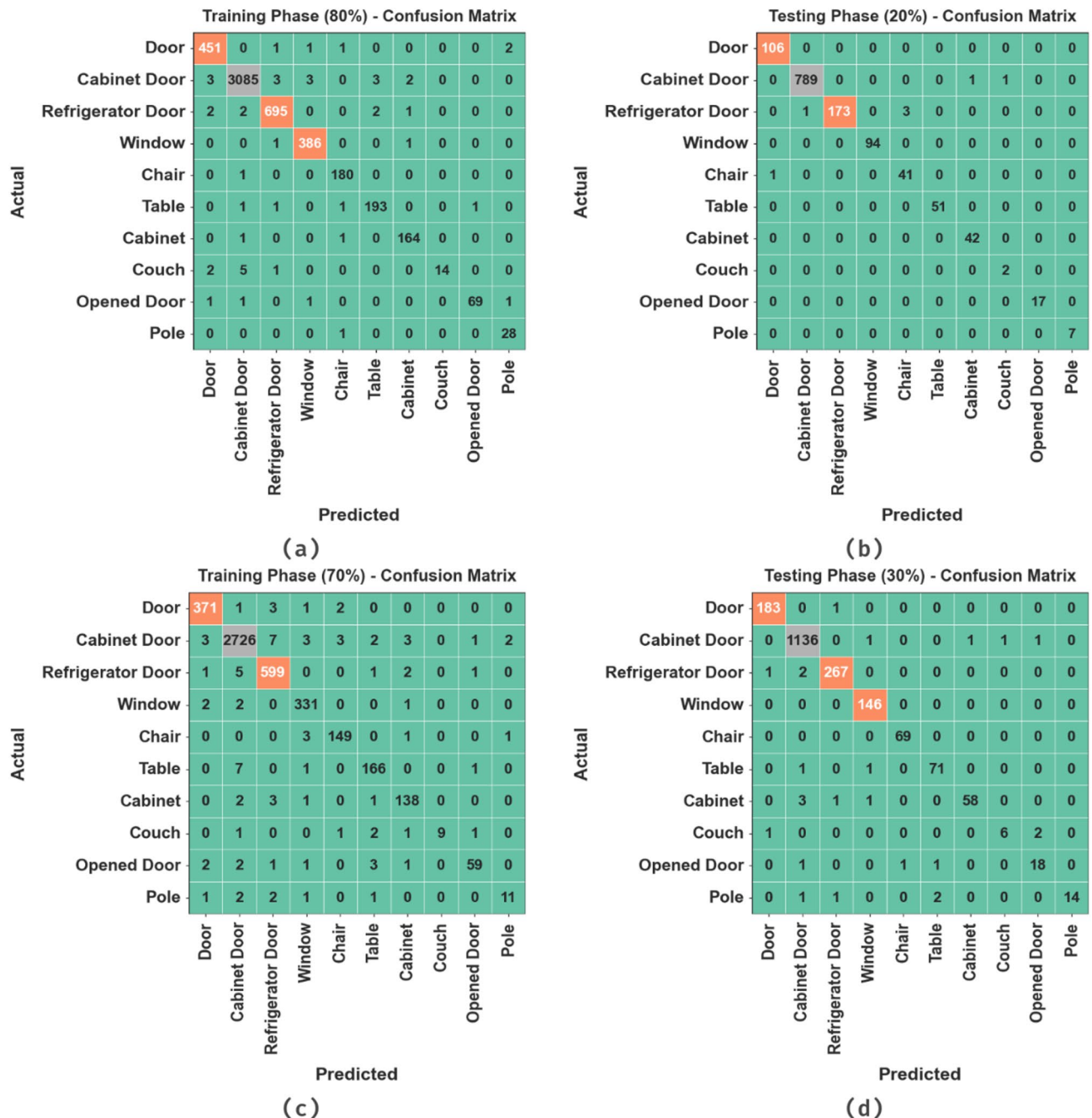**Table 1**. Dataset details.

Fig. 3. Confusion matrix of EATBP-ODHOA model (**a,b**) 80:20 and (**c,d**) 70:30 of TRAPA/TESPA.

$prec_n$ of 94.76%, $reca_l$ of 88.75%, $F_{Measure}$ of 91.13%, and AUC_{score} of 94.24%. Simultaneously, on 30% TESPA, the EATBP-ODHOA approach achieves an average $accu_y$ of 99.75%, $prec_n$ of 95.93%, $reca_l$ of 91.75%, $F_{Measure}$ of 93.57%, and AUC_{score} of 95.78%.

In Fig. 6, the TRAN $accu_y$ and VALN $accu_y$ results of the EATBP-ODHOA techniques under a 70:30 ratio are exemplified. The outcome highlighted that both $accu_y$ analyses exhibit a growing tendency, which informed the capacity of the EATBP-ODHOA model, achieving maximum performance across multiple iterations. At the same time, both $accu_y$ remain closer across the epochs, indicating lesser overfitting and a more favourable outcome for the EATBP-ODHOA model, which guarantees consistent predictions on unseen samples.

In Fig. 7, the TRAN loss and VALN loss analysis of the EATBP-ODHOA approach with a ratio of less than 70:30 is shown. Both values exhibit a diminishing trend, indicating the EATBP-ODHOA approach's ability to balance the trade-off between generalization and data fitting. The constant reduction in loss values ensures a higher solution for the EATBP-ODHOA model and improves the prediction outcomes.

Table 4; Fig. 8 present a comparative analysis of the EATBP-ODHOA approach with existing techniques[19,20,35–37]. The EATBP-ODHOA approach outperforms others with an $accu_y$ of 99.25%, $prec_n$

Class labels	$Accu_y$	$Prec_n$	$Reca_l$	$F_{Measure}$	AUC_{score}
TRAPA (80%)					
Door	99.10	93.97	95.61	94.78	97.52
Cabinet Door	98.04	97.83	98.84	98.33	97.88
Refrigerator Door	98.49	93.19	95.58	94.37	97.26
Window	98.93	91.69	93.81	92.74	96.57
Chair	99.08	88.82	83.43	86.04	91.53
Table	99.17	86.96	91.37	89.11	95.42
Cabinet	99.06	86.25	83.13	84.66	91.35
Couch	99.59	00.00	00.00	00.00	50.00
Opened Door	99.44	89.09	67.12	76.56	83.50
Pole	99.55	77.78	24.14	36.84	62.05
Average	**99.04**	**80.56**	**73.30**	**75.34**	**86.31**
TESPA (20%)					
Door	99.32	94.50	97.17	95.81	98.34
Cabinet Door	98.50	98.13	99.37	98.74	98.29
Refrigerator Door	98.80	95.48	95.48	95.48	97.39
Window	99.40	93.00	98.94	95.88	99.18
Chair	98.87	86.49	76.19	81.01	87.90
Table	99.32	88.89	94.12	91.43	96.82
Cabinet	99.32	92.31	85.71	88.89	92.74
Couch	99.85	00.00	00.00	00.00	50.00
Opened Door	99.55	100.00	64.71	78.57	82.35
Pole	99.55	100.00	14.29	25.00	57.14
Average	**99.25**	**84.88**	**72.60**	**75.08**	**86.02**

Table 2. Object detection of EATBP-ODHOA model under 80:20 of TRAPA/TESPA.
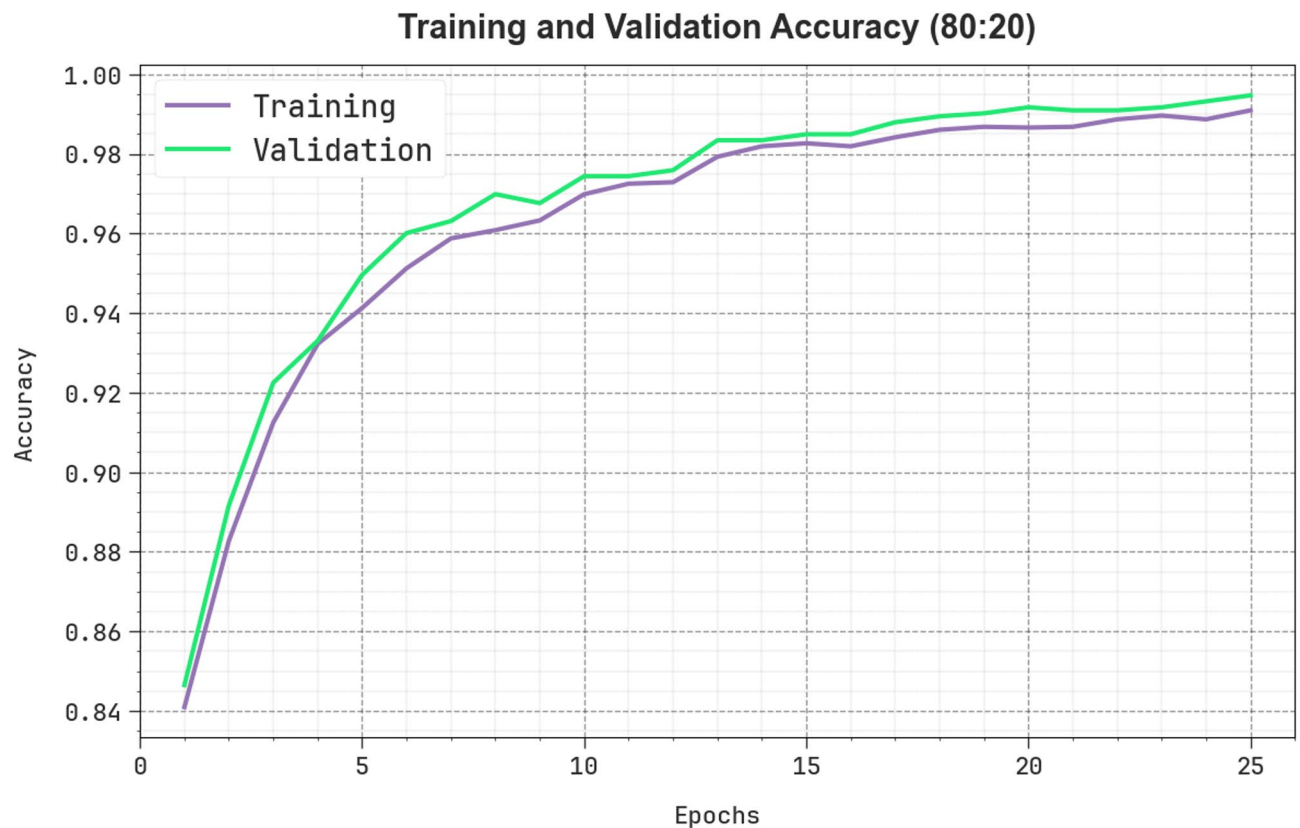
Fig. 4. $Accu_y$ curve of the EATBP-ODHOA method below 80:20.
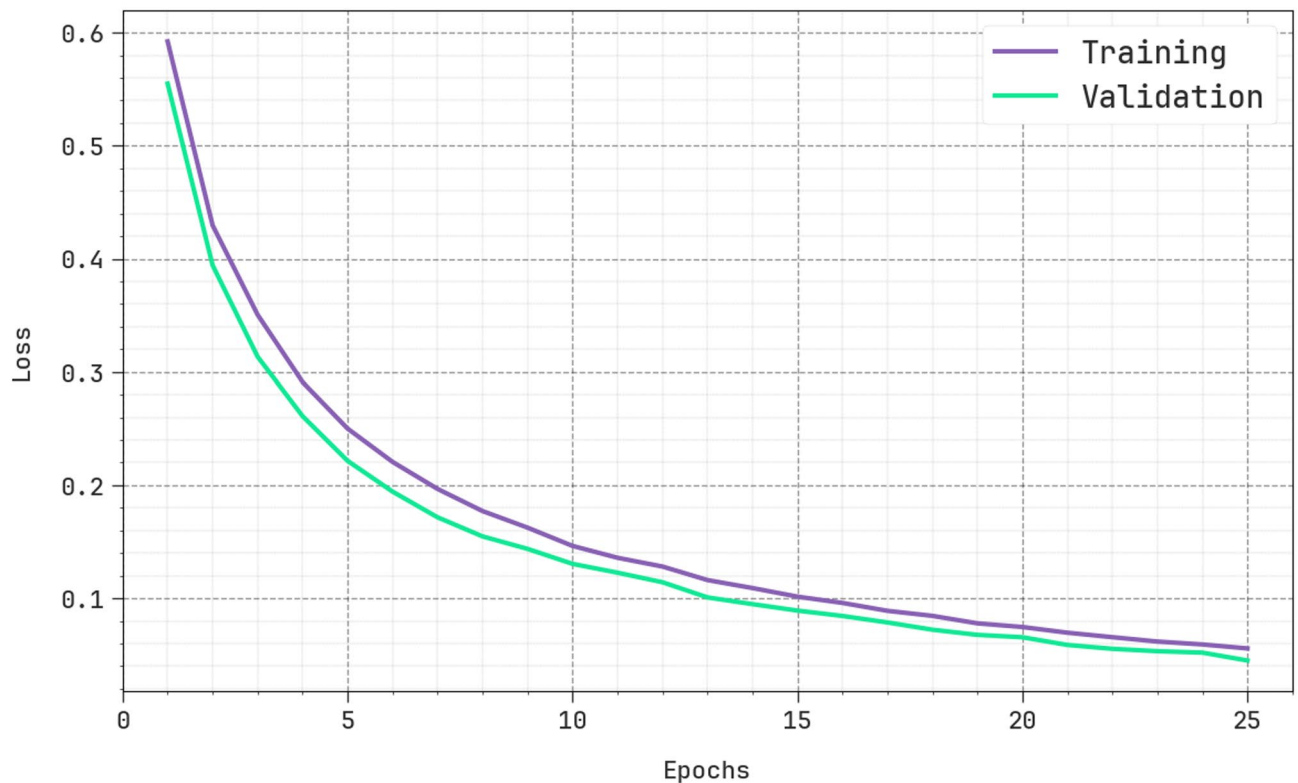
Fig. 5. Loss graph of the EATBP-ODHOA method below 80:20.

of 84.88%, $reca_l$ of 72.60%, and $F_{Measure}$ of 75.08%. This surpasses the next best model, CTF-Net, which achieved $accu_y$ of 98.81%, $prec_n$ of 83.17%, $reca_l$ of 70.34%, and $F_{Measure}$ of 72.62%. Other notable models include YOLOv5 with $accu_y$ of 97.95%, $prec_n$ of 76.72%, $reca_l$ of 71.46%, and $F_{Measure}$ of 73.13%, as well as RSNet with accuracy of 97.26%, $prec_n$ of 76.15%, $reca_l$ of 71.88%, and $F_{Measure}$ of 74.37%. The results highlight the superiority of the EATBP-ODHOA model in balancing all key metrics, indicating improved overall effectiveness compared to existing models such as DCASR, PointNet++, and Faster R-CNN.

Furthermore, the EATBP-ODHOA model reported improved performance with maximum $accu_y$, $prec_n$, $reca_l$, and $F_{Measure}$ of 99.25%, 84.88%, 72.60%, and 75.08%, respectively. The execution time (ET) of the EATBP-ODHOA model with existing techniques is shown in Table 5; Fig. 9. Based on ET, the EATBP-ODHOA method provides a lesser ET of 10.87 s, whereas the DCASR, YOLOv5, CTF-Net, PointNet++, RSNet, MCNN, YOLOV3, SSD, Faster R-CNN, and LCCP techniques attain greater ET of 25.88 s, 26.00 s, 17.91 s, 29.17 s, 12.12 s, 22.50 s, 12.86 s, 17.77 s, 24.51 s, and 18.55 s, correspondingly.

Conclusion

In this manuscript, the EATBP-ODHOA approach is proposed. The main aim of the EATBP-ODHOA approach is to develop an effective object detection model for visually impaired people using an advanced DL technique. The image pre-processing stage initially employs an ABF model to enhance the image quality by removing unwanted noise. Furthermore, the Faster R-CNN model is used for the object detection process. The EATBP-ODHOA model utilizes fusion models, such as ResNet and DenseNet-201, for feature extraction. Moreover, the Bi-GRU method is employed for the classification process. Finally, the HOA model is utilized for hyperparameter tuning. The experimentation of the EATBP-ODHOA model is performed under the indoor object detection dataset. The comparison analysis of the EATBP-ODHOA model revealed a superior accuracy value of 99.25% compared to existing approaches. The limitations of the EATBP-ODHOA model include the dependence on a limited dataset, which may affect the generalizability across diverse real-world scenarios. The technique also encounters threats in detecting objects under extreme lighting conditions and heavy occlusions. The study also lacks appropriate analysis in outdoor environments. Real-time processing speed could be further optimized for deployment on low-power devices. Future work involves expanding the dataset with more varied and complex scenes, enhancing robustness to challenging conditions, and integrating multimodal sensory data, such as audio and tactile feedback, to improve user interaction and experience. Exploring lightweight architectures for faster inference on embedded systems is also planned to increase accessibility.

Class labels	$Accu_y$	$Prec_n$	$Reca_l$	$F_{Measure}$	AUC_{score}
TRAPA (70%)					
Door	99.66	97.63	98.15	97.89	98.97
Cabinet Door	99.01	99.20	99.13	99.16	98.98
Refrigerator Door	99.44	97.40	98.36	97.88	98.98
Window	99.66	96.78	98.51	97.64	99.13
Chair	99.76	96.13	96.75	96.44	98.31
Table	99.59	94.32	94.86	94.59	97.32
Cabinet	99.66	93.88	95.17	94.52	97.49
Couch	99.87	100.00	60.00	75.00	80.00
Opened Door	99.70	93.65	85.51	89.39	92.71
Pole	99.78	78.57	61.11	68.75	80.52
Average	**99.61**	**94.76**	**88.75**	**91.13**	**94.24**
TESPA (30%)					
Door	99.85	98.92	99.46	99.19	99.67
Cabinet Door	99.40	99.30	99.65	99.47	99.36
Refrigerator Door	99.70	98.89	98.89	98.89	99.36
Window	99.85	97.99	100.00	98.98	99.92
Chair	99.95	98.57	100.00	99.28	99.97
Table	99.75	95.95	97.26	96.60	98.55
Cabinet	99.70	98.31	92.06	95.08	96.01
Couch	99.80	85.71	66.67	75.00	83.31
Opened Door	99.70	85.71	85.71	85.71	92.78
Pole	99.80	100.00	77.78	87.50	88.89
Average	**99.75**	**95.93**	**91.75**	**93.57**	**95.78**

Table 3. Object detection of EATBP-ODHOA model under 70:30 of TRAPA/TESPA.
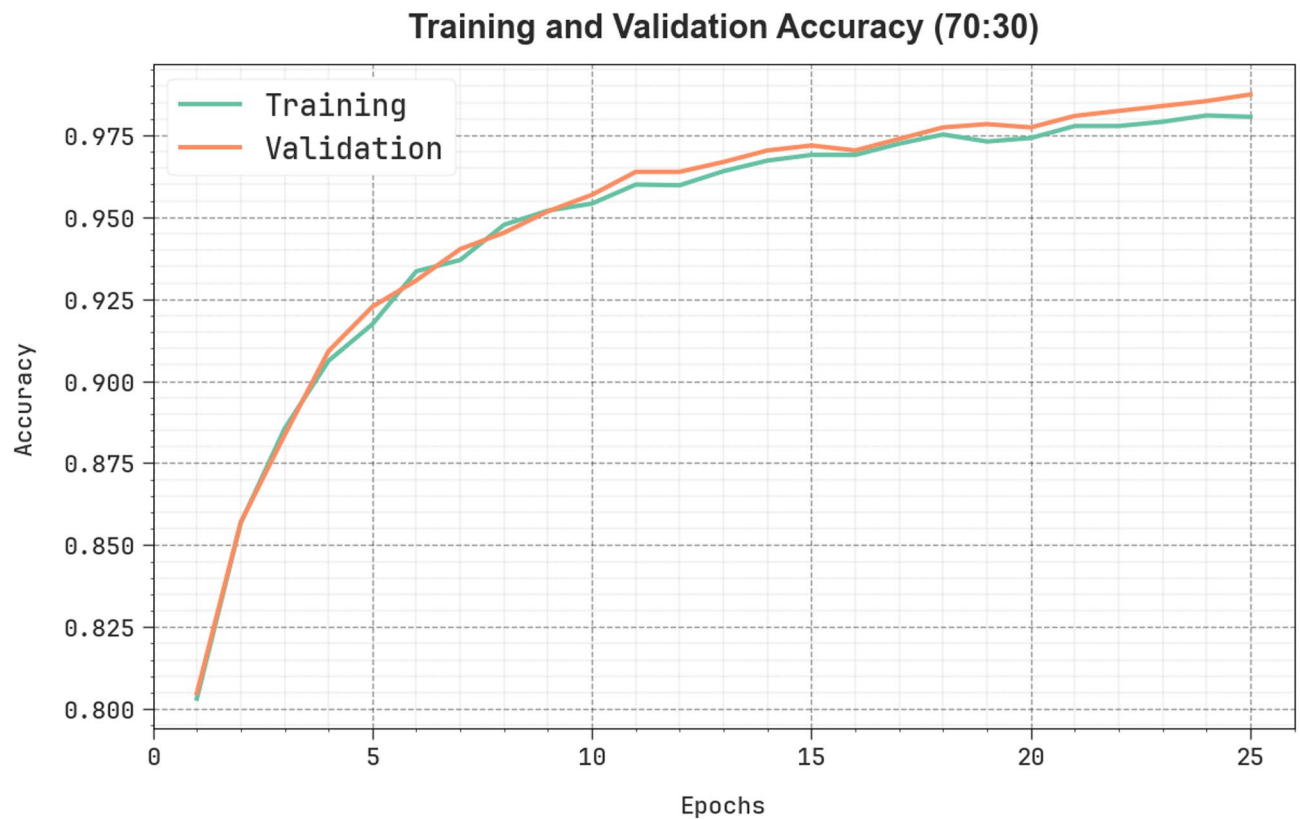
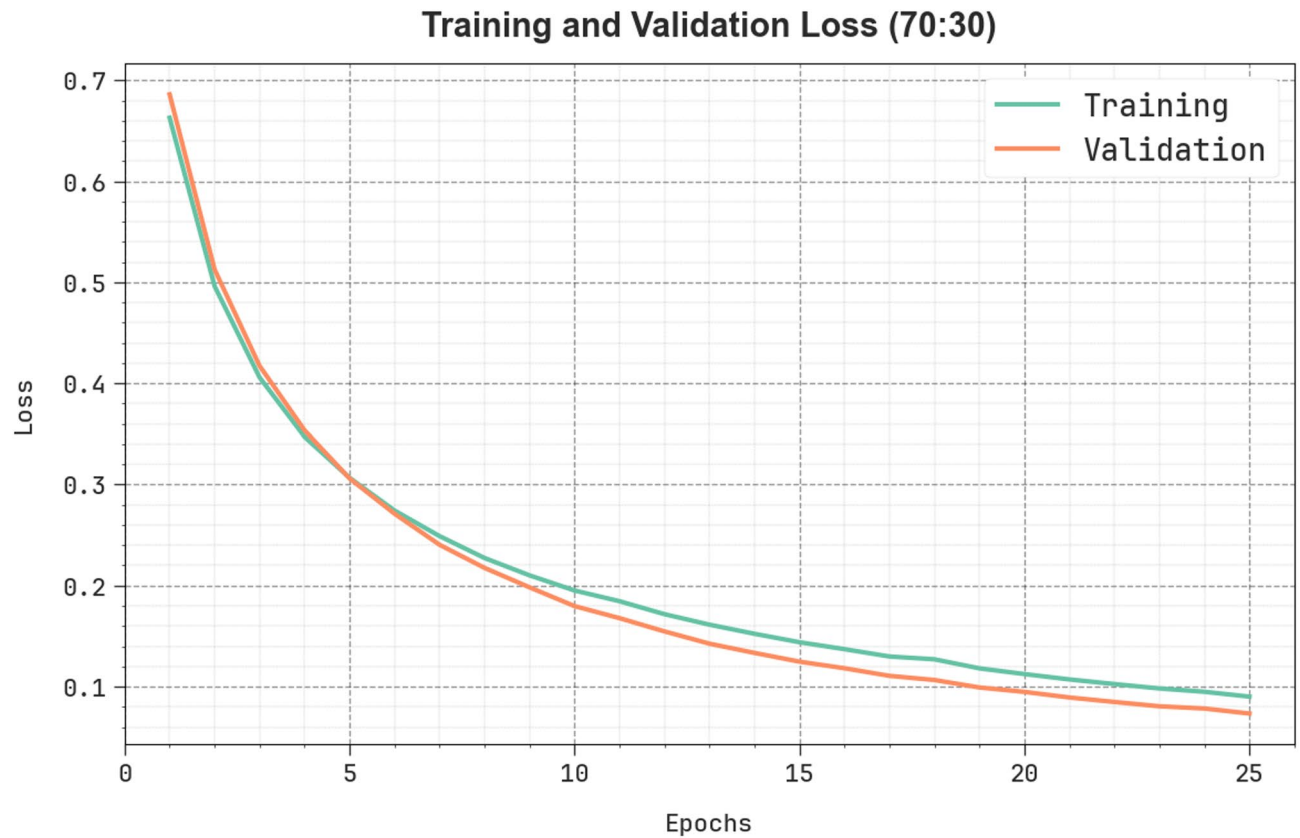**Fig. 6.** $Accu_y$ curve of the EATBP-ODHOA method under 70:30.

Training and Validation Loss (70:30)

Fig. 7. Loss graph of the EATBP-ODHOA method below 70:30.

Model	$Accu_y$	$Prec_n$	$Reca_l$	$F_{Measure}$
DCASR	95.65	83.13	70.81	70.60
YOLOv5	97.95	76.72	71.46	73.13
CTF-Net	98.81	83.17	70.34	72.62
PointNet++	94.97	82.52	70.11	69.99
RSNet Model	97.26	76.15	71.88	74.37
MCNN	98.09	82.67	69.67	71.87
YOLOV3	91.40	74.22	70.86	74.11
SSD Model	93.42	80.92	69.67	69.82
Faster R-CNN	93.90	73.40	69.63	70.99
LCCP	97.04	80.11	70.34	72.35
EATBP-ODHOA	99.25	84.88	72.60	75.08

Table 4. Comparative outcomes of the EATBP-ODHOA model with existing techniques[19,20,35–37].
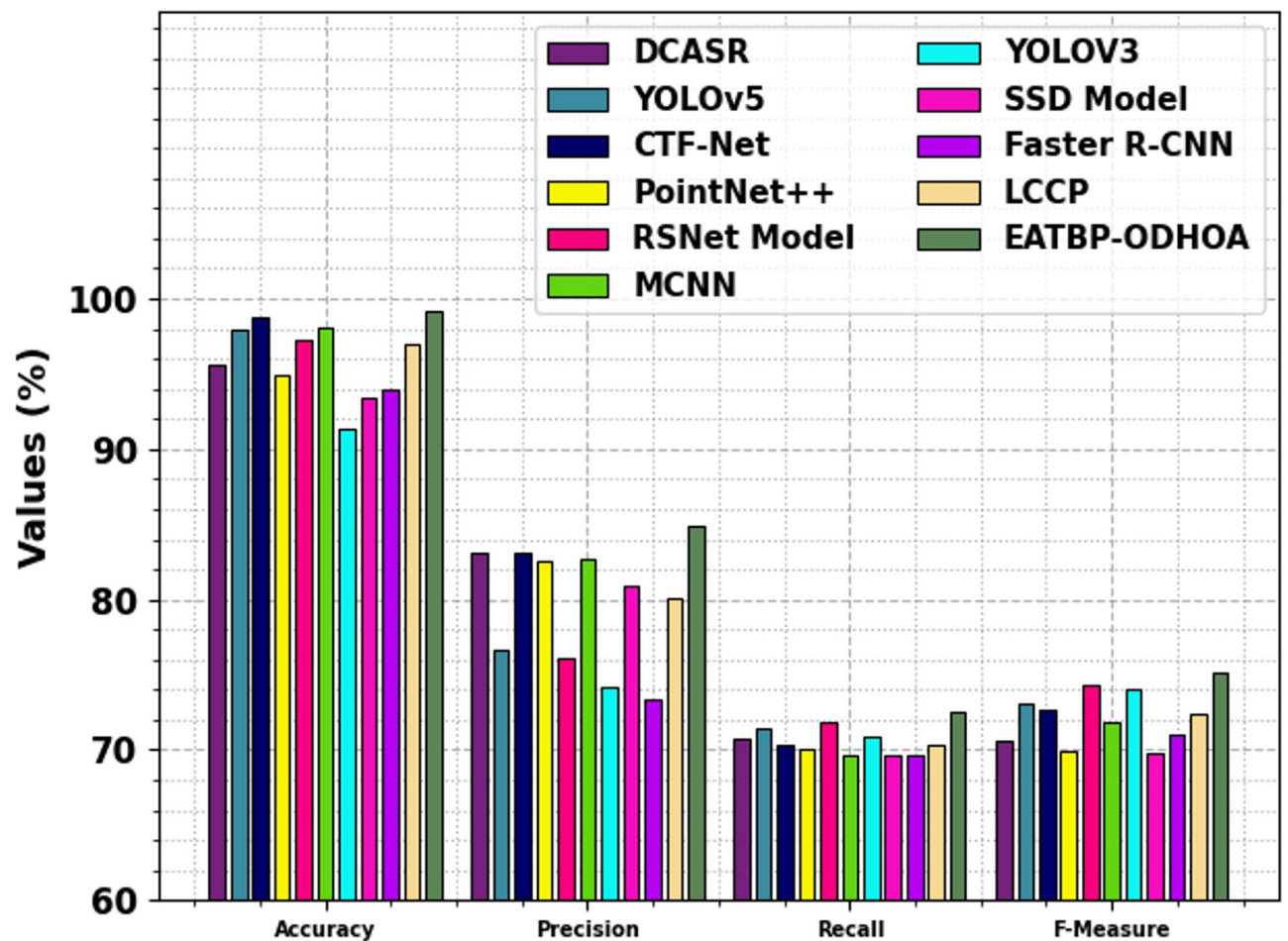
Fig. 8. Comparative analysis of the EATBP-ODHOA model with existing techniques.

Method	ET (sec)
DCASR	25.88
YOLOv5	26.00
CTF-Net	17.91
PointNet++	29.17
RSNet Model	12.12
MCNN	22.50
YOLOV3	12.86
SSD Model	17.77
Faster R-CNN	24.51
LCCP	18.55
EATBP-ODHOA	10.87

Table 5. ET analysis of the EATBP-ODHOA approach with the existing methods.
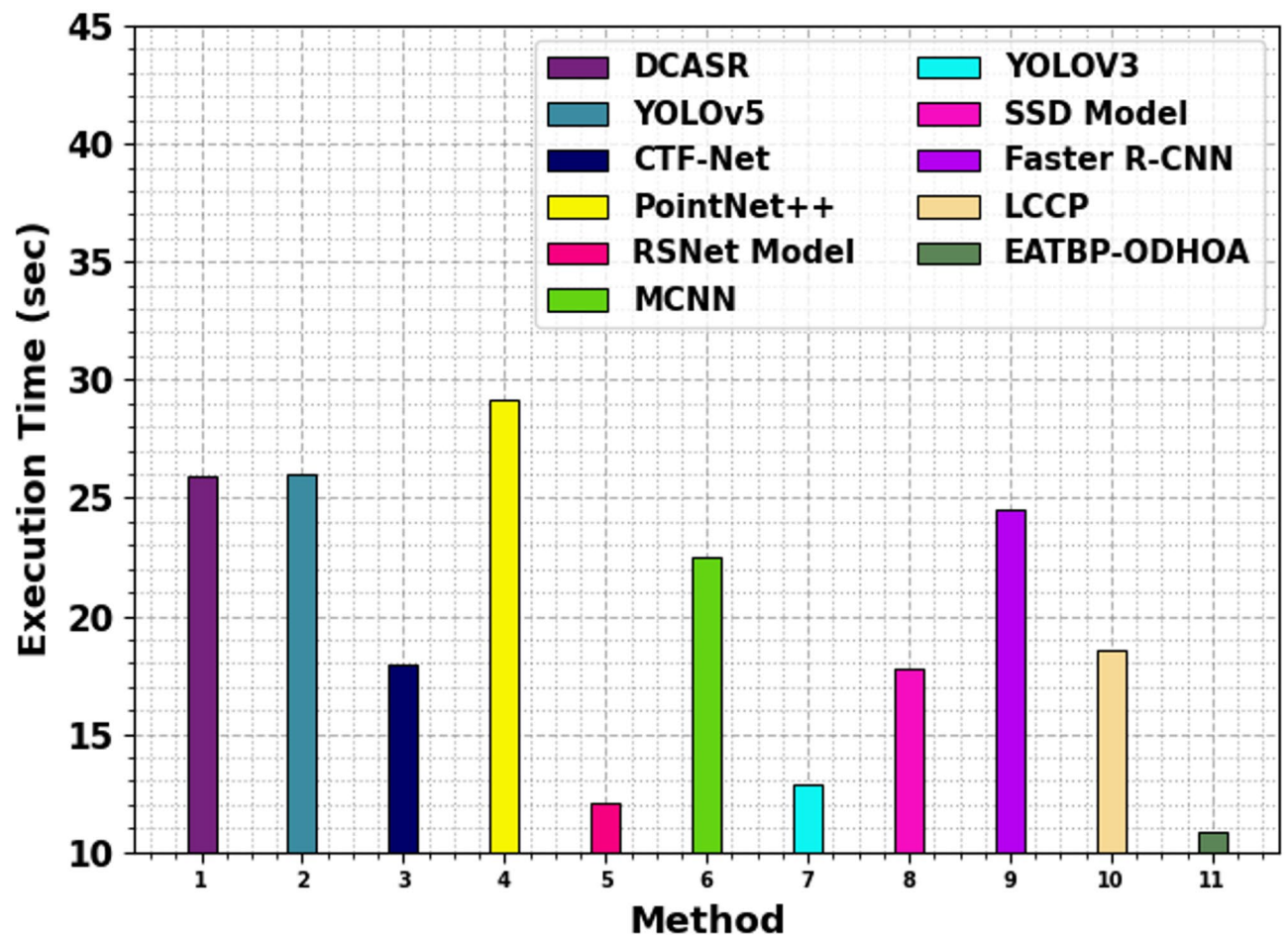
Fig. 9. ET outcome of EATBP-ODHOA approach with the existing methods.

Data availability
The data that support the findings of this study are openly available at https://www.kaggle.com/datasets/thepbordin/indoor-object-detection, reference number[34].

References
1. Zou, Z., Chen, K., Shi, Z., Guo, Y. & Ye, J. Object detection in 20 years: A survey. *Proceedings of the IEEE*, *111*(3), pp.257–276. (2023).
2. Dhillon, A. & Verma, G. K. Convolutional neural network: a review of models, methodologies and applications to object detection. *Progress Artif. Intell.* **9** (2), 85–112 (2020).
3. Liu, L. et al. Deep learning for generic object detection: A survey. *Int. J. Comput. Vision.* **128**, 261–318 (2020).
4. Sharma, V. K. & Mir, R. N. A comprehensive and systematic look up into deep learning based object detection techniques: A review. *Computer Science Review*, *38*, p.100301. (2020).
5. Pareek, P. K. Pixel level image fusion in moving objection detection and tracking with machine learning. *J. Fusion: Pract. Appl.* **2** (1), 42–60 (2020).
6. Murthy, C. B., Hashmi, M. F., Bokde, N. D. & Geem, Z. W. Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A comprehensive review. *Applied sciences*, *10*(9), p.3280. (2020).
7. Ariji, Y. et al. Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral surgery, oral medicine, oral pathology and oral radiology*, *128*(4), pp.424–430. (2019).
8. Kumar, A., Zhang, Z. J. & Lyu, H. Object detection in real time based on improved single shot multi-box detector algorithm. *EURASIP Journal on Wireless Communications and Networking*, *2020*(1), p.204. (2020).
9. Akyon, F. C., Altinuc, S. O. & Temizel, A. October. Slicing aided hyper inference and fine-tuning for small object detection. In *2022 IEEE international conference on image processing (ICIP)* (pp. 966–970). IEEE. (2022).
10. Madhuri, A. & Umadevi, T. Role of context in visual Language models for object recognition and detection in irregular scene images. *Fusion: Pract. & Applications*, **15**(1). (2024).
11. Aniedu, A. N., Nwokoye, S. C., Okafor, C. S., Anyanwu, K. & Isizoh, A. N. Enhanced AI-Based navigation system for the visually impaired. *Inform: Jurnal Ilmiah Bidang Teknologi Informasi Dan. Komunikasi*. **10** (1), 16–20 (2025).
12. Arsalwad, G. et al. YOLOInsight: artificial Intelligence-Powered assistive device for visually impaired using internet of things and Real-Time object detection. *Cureus*, **1**(1). (2024).

13. Yu, X. & Saniie, J. Visual Impairment Spatial Awareness System for Indoor Navigation and Daily Activities. *Journal of Imaging*, *11*(1), p.9. (2025).
14. Baig, M. S. A. et al. AI-based Wearable Vision Assistance System for the Visually Impaired: Integrating Real-Time Object Recognition and Contextual Understanding Using Large Vision-Language Models. *arXiv preprint arXiv:2412.20059*. (2024).
15. Sabit, H. Artifical Intelligence-Based Smart Security System Using Internet of Things for Smart Home Applications. *Electronics*, *14*(3), p.608. (2025).
16. Joshi, R. C., Singh, N., Sharma, A. K., Burget, R. & Dutta, M. K. AI-SenseVision: a low-cost artificial-intelligence-based robust and real-time assistance for visually impaired people. *IEEE Trans. Human-Machine Systems* (2024).
17. Sophia, N. A. et al. Li-Fi-enabled smart indoor navigation system with obstacle detection and voice assistance for visually impaired individuals. In Challenges in Information, Communication and Computing Technology (141–146). CRC. (2025).
18. Naz, S. & Jabeen, F. Towards improved assistive technologies: classification and evaluation of object detection techniques for users with visual impairments. *VAWKUM Trans. Comput. Sci.* **12** (2), 165–177 (2024).
19. Malkocoglu, A. B. V. & Samli, R. A novel model for higher performance object detection with deep channel attention super resolution. *Engineering Science and Technology, an International Journal*, *64*, p.102003. (2025).
20. Zhang, D., Wang, C., Wang, H., Fu, Q. & Li, Z. An effective CNN and Transformer fusion network for camouflaged object detection. *Computer Vision and Image Understanding*, p.104431. (2025).
21. Zin, T. T. et al. Real-time action recognition system for elderly people using stereo depth camera. *Sensors*, *21*(17), p.5895. (2021).
22. Khadidos, A. O. & Yafoz, A. Leveraging retinanet based object detection model for assisting visually impaired individuals with metaheuristic optimization algorithm. *Scientific Reports*, *15*(1), p.15979. (2025).
23. Aung, M. M., Maneetham, D., Crisnapati, P. N. & Thwe, Y. Enhancing object recognition for visually impaired individuals using computer vision. *Int. J. Eng. Trends Technol.* **72** (4), 297–305 (2024).
24. Alsolai, H., Al-Wesabi, F. N., Motwakel, A. & Drar, S. Sine Cosine Optimization With Deep Learning–Driven Object Detector for Visually Impaired Persons. *Journal of Mathematics*, *2025*(1), p.5071695. (2025).
25. Bhandari, A. et al. July. Using Transfer Learning to Refine Object Detection Models for Blind and Low Vision Users. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1–4). IEEE. (2024).
26. Venkatkrishnan, G. R., Jeya, R., Ramyalakshmi, G., Sindhu, S. & Sreenithi, K. Real-Time object detection for the visually impaired using Yolov8 and NLP on Iot devices. *Int. J. Environ. Sci.* **11** (8s), 907–914 (2025).
27. Srivastava, P. K. et al. August. A Portable Object Detection System for Visually Impaired Individuals in Outdoor Environments. In *2024 1st International Conference on Advanced Computing and Emerging Technologies (ACET)* (pp. 1–5). IEEE. (2024).
28. Guo, W., Lu, P., Peng, X. & Zhao, Z. Learnable adaptive bilateral filter for improved generalization in single image Super-Resolution. *Pattern Recognition*, p.111396. (2025).
29. Barberena, X., Saiz, F. A. & Barandiaran, I. Handling Drift in Industrial Defect Detection Through MMD-Based Domain Adaptation. *Proceedings Copyright*, *420*, p.429.
30. Yousra, K. A. T. E. B. *INTELLIGENT MONITORING OF AN INDUSTRIAL SYSTEM USING IMAGE CLASSIFICATION* (Doctoral dissertation, MINISTRY OF HIGHER EDUCATION). (2025).
31. Jahan, I. et al. Deep learning and vision transformers-based framework for breast cancer and subtype identification. *Neural Comput. Applications*, pp.1–20. (2025).
32. Wang, S., Omar, K. S., Miranda, F. & Bhatt, T. Automatic gait EVENT detection in older adults during perturbed walking. *Journal of NeuroEngineering and Rehabilitation*, *22*(1), p.40. (2025).
33. Duan, X. Settlement prediction of Nanjing metro line 10 with HOA-VMD-LSTM. *Measurement* **244**, 116477 (2025).
34. https://www.kaggle.com/datasets/thepbordin/indoor-object-detection
35. Luo, N., Wang, Q., Wei, Q. & Jing, C. Object-level segmentation of indoor point clouds by the convexity of adjacent object regions. *IEEE Access*. **7**, 171934–171949 (2019).
36. Qin, P., Shen, W. & Zeng, J. DSCA-Net: indoor head detection network using Dual-Stream information and channel attention. *Chin. J. Electron.* **29** (6), 1102–1109 (2020).
37. Lv, Y., Fang, Y., Chi, W., Chen, G. & Sun, L. Object detection for sweeping robots in home scenes (ODSR-IHS): A novel benchmark dataset. *IEEE Access*. **9**, 17820–17828 (2021).

Acknowledgements

Author contributions

Mahir Mohammed Sharif Adam: Conceptualization, methodology, validation, investigation, writing—original draft preparation, fundingNojood O Aljehane: Conceptualization, methodology, writing—original draft preparation, writing—review and editingMohammed Yahya Alzahrani: methodology, validation, writing—original draft preparationSamah Al Zanin: software, validation, data curation, writing—review and editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.M.S.A.

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.