



OPEN Intelligent feature fusion with dynamic graph convolutional recurrent network for robust object detection to assist individuals with disabilities in a smart IoT edge-cloud environment

Manal Abdullah Alohal¹✉, Fuhid Alanazi², Yaser Abdulaziz Alsahafi³ & Ishfaq Yaseen^{4,5}

Smart Internet of Things (IoT)-edge-cloud computing defines intelligent systems where IoT devices create data at the network's edge, which is then further processed and analyzed in local edge devices before transmission to the cloud for deeper insights and storage. Visual impairment, like blindness, has a deep effect on a person's psychological and cognitive functions. So, the use of assistive models can help mitigate the adverse effects and improve the quality of life for individuals who are blind. Much current research mainly concentrates on mobility, navigation, and object detection (OD) in smart devices and advanced technologies for visually challenged people. OD is a vital feature of computer vision that includes categorizing objects within an image, allowing applications like augmented reality, image retrieval, etc. Recently, deep learning (DL) models have emerged as an excellent technique for mining feature representation from data, primarily due to significant developments in OD. The DL model is well-trained with manifold images of objects that are highly applicable to visually impaired individuals. This paper presents an intelligent Feature Fusion with Dynamic Graph Convolutional Recurrent Network for Robust Object Detection (FFDGCRN-ROD) approach to assist individuals with disabilities. The paper aims to present an intelligent OD framework for individuals with disabilities utilizing a smart IoT edge cloud environment to enable monitoring and assistive decision-making. At first, the image pre-processing phase involves resizing, normalization, and image enhancement to eliminate the noise and enhance the image quality. For the OD process, the FFDGCRN-ROD approach employs the faster R-CNN to identify and locate specific targets within the images automatically. Furthermore, the fusion models, namely CapsNet, SqueezeNet, and Inceptionv3, are used for the feature extraction process. Finally, the FFDGCRN-ROD model implements the dynamic adaptive graph convolutional recurrent network (DA-GCRN) model to detect and classify objects for visually impaired people accurately. The experimental validation of the FFDGCRN-ROD methodology is performed under the Indoor OD dataset. The comparison analysis of the FFDGCRN-ROD methodology demonstrated a superior accuracy value of 99.65% over existing techniques.

Keywords Internet of things, Cloud computing, Object detection, Individuals with disabilities, Feature fusion, Deep learning

¹Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia. ²Department of Information Systems, Faculty of Computer and Information Systems, Islamic University of Madinah, Medina 42351, Saudi Arabia. ³Department of Information Technology, College of Computers and Information Technology, University of Jeddah, Jeddah 21493, Saudi Arabia. ⁴Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia. ⁵King Salman Centre for Disability Research, Riyadh 11614, Saudi Arabia. ✉email: maalohaly@pnu.edu.sa

Vision is the most crucial sense, and it is vital for day-to-day life activities. The World Health Organisation (WHO) reports that around 2.2 billion individuals suffer from loss of sight or impaired vision¹. For people who are visually challenged, assistive technologies have been developed to provide them with assistance. As they require independence and mobility, most of this technology is not available and is also expensive. Therefore, they necessitate economical, lightweight, versatile, and real-time applications that work with greater precision. Technological advancements enable systems to examine and view surroundings as individuals do, utilizing computer vision (CV)². For visually challenged individuals, CV provides visual data. OD is a vital task in a CV that enables the acquisition of intellectual capabilities³. OD mainly aims to detect and classify many objects in imagery and videos. Moreover, OD is considered a vital component in CV, tackling object classification in imagery, which allows applications such as content-based image retrieval (CBIR) and augmented reality⁴. Advancements in technologies have resulted in the growing acceptance of smart devices, like IoT devices, among the general public⁵.

This acceptance can be attributed to their capability to offer higher accessibility, efficacy, and functionality in several aspects of everyday living, and their incorporation with intelligent sensors, connectivity, and sophisticated analytics allows for real-time data assortment and transmission⁶. This has increased the count of devices linked to the network. As the data is mainly generated at the network edge, and demanding processing happens in central cloud environments, an adaptable connected framework is needed to bridge the gap, allowing edge and cloud environments to operate seamlessly together⁷. The complete employment of that continuum is anticipated to employ the assistance of latency-critical applications⁸. Furthermore, DL is presently being used in numerous applications connected to IoT, such as CV, where real-time processing is essential. For these applications, IoT devices produce extensive data that is examined by an artificial intelligence (AI) approach. Nevertheless, the more intricate the model is, the higher the computational resources needed for training and model inference⁹. This is due to the large number of parameters that create the model, as well as the millions of computations required for training and inference. Currently, the DL-driven methods are leveraged for OD to assist VIP¹⁰.

Key contributions

This paper proposes an intelligent Feature Fusion with Dynamic Graph Convolutional Recurrent Network for Robust Object Detection (FFDGCRN-ROD) model, designed to support disabled individuals within the smart IoT-edge-cloud environment. The main contributions of this paper are listed below:

- An intelligent FFDGCRN-ROD model is introduced to assist individuals with disabilities in a smart IoT-edge-cloud environment.
- To mitigate the noise and improve image quality, an advanced image pre-processing pipeline is developed by utilizing resizing, normalization, and enhancement. This facilitates improved accuracy in feature extraction. This process strengthens the overall performance in smart IoT edge applications.
- The Faster R-CNN model is integrated to enable reliable OD and precise localization of targets in real-time. It also improves detection speed and accuracy for intrinsic environments. This integration assists in the efficient detection of individuals with disabilities. The model is optimized for deployment in smart IoT edge cloud systems.
- A fusion of feature extraction models, namely CapsNet, SqueezeNet, and Inception V3, is employed for capturing multi-level and complementary feature representations. This integration also improves the model's capability to recognize diverse and complex patterns. It also enhances robustness and discriminative power in OD tasks. The approach ensures high accuracy in varying real-time IoT edge scenarios.
- The DA-GCRN methodology is employed for accurate classification and detection. The technique also facilitates efficient modelling of spatial-temporal dependencies in feature data. This method additionally strengthens the detection of subtle patterns and relationships. The network improves reliability and precision in smart edge-based environments.
- The FFDGCRN-ROD approach introduces a novel incorporation of CapsNet, SqueezeNet, and InceptionV3 models for enriched multi-level feature extraction. It utilizes DA-GCRN for capturing spatial-temporal dynamics for precise object classification. Integration with Faster R-CNN ensures real-time detection and localization. The end-to-end framework is optimized for smart IoT edge-cloud environments, targeting individuals with disabilities.
- Extensive experiments are demonstrating that the proposed model outperforms current advanced methodologies in terms of robustness and accuracy.

Organization of the work

This paper is structured to deliver a comprehensive understanding of the research. Section 2 inspects the literature relevant to the study. Section 3 details the methodological technique. Section 4 provides the outcomes of the experimental evaluations, showcasing the effectiveness and efficiency of the presented model. Ultimately, Sect. 5 summarises the research's significant findings and contributions.

Related works on indoor OD for individuals with disabilities

In¹¹, the authors presented an enhanced assistive technology for blind people and OD utilizing a Hiking optimizer algorithm (EATBP-ODHOA) method. The motive of this method is to design an efficient OD approach for VIP through sophisticated DL models. The image pre-processing step primarily uses an adaptive bilateral filtering (ABF) model. Moreover, the R-CNN methodology was utilized for detecting objects. In¹², the authors proposed an innovative OD model for disabled persons utilizing a state-of-the-art DL method with a sparrow search optimizer (ODSDP-ADLMSSO) model. This model aims to improve the OD technique for VIP. Moreover, the YOLO version 7 approach is employed to detect and classify objects from images. Kumar

et al.¹³ introduced a device with advanced technology that would allow VIPs to perform their own tasks rather than depending on others. For OD, this application leverages image processing methods, and for voice output, it utilizes speech synthesis. This technology focuses on providing VIP with real-time vocal or audio data regarding objects captured in their mobile cameras. Biradar and Gull¹⁴ presented a novel OD approach termed YOLOv8 alongside Spiking EfficientDet (Yv8SED), which enhances detection precision, minimizing incorrect classification with specific importance on small objects. The presented approach delivers extraordinary OD outcomes with decreased cost and time, which makes it a robust model for diverse OD missions. The segmentation procedure is enhanced by SegNet, which proficiently segments objects from images. Abidi et al.¹⁵ introduced an application, depending on the gradient support vector boosting-driven crossover golden jackal (GSB-CGJ) model. The introduced application aims to guide VIP, avoid accidents, and generate real-time feedback. Next, OD is effectively performed through the GSB-CGJ technique. Hebbar and Pullela¹⁶ provided a thorough analysis of developments in assistive technology personalized for VIPs with a particular emphasis on the function of the IoT in improving their quality of life. This study incorporates a detailed assessment of various studies and papers regarding wearable devices, navigation solutions, and IoT-based OD methods, highlighting the importance of user-centred design and possible advances in detection accuracy. Malla et al.¹⁷ focused on the VIPs, who encounter significant issues in social mobility because of neurological or physiological conditions that cause blindness. Despite a dependence on external support, there is an increasing preference for autonomy among these individuals. Given this, a groundbreaking device, SBS, is introduced to improve their mobility-related problems. The SBS is a cutting-edge adaptive device which is developed to solve everyday navigation problems faced by VIPs. Alagarsamy et al.¹⁸ designed a device that assists blind people in identifying various objects in their surroundings by employing the YOLO version 3 model integrated with R-CNN.

This includes various methodologies to design an application that rapidly detects diverse objects in VIP environments and guides them, utilizing audio output. CNN named YOLO, which identifies objects. This developed technique is more efficient and precise when compared to other models in detecting objects. Alazeb et al.¹⁹ presented a scene recognition framework by utilizing kernel convolution, UNet segmentation, discrete wavelet transform (DWT), and deep belief networks (DBN) incorporated with AlexNet for accurate multi-OD and labelling in robotic environments. Shawly and Alsheikhy²⁰ proposed an automated epilepsy prediction system using a novel attention module (NAM) integrated into a convolutional neural network (CNN), combined with the Fourier transform for feature extraction, principal component analysis (PCA) for dimensionality reduction, and an optimized stochastic gradient descent with the Adam optimizer to enhance learning from EEG signals. Xu et al.²¹ developed a Temporal-Frequency-Spatial Transformer (TFST) technique integrated with sub-domain knowledge alignment (SDKA) to improve the accuracy and generalization of major depressive disorder detection from EEG data by incorporating fine-grained frequency features and enhancing domain alignment. Bu et al.²² introduced OD-YOLO, an enhanced OD model based on You Only Look Once version 8 nano (YOLOv8n), integrating deformable convolutional networks, a hybrid attention transformer, a dynamic head, and the Optimised Intersection over Union (OIoU) loss for improving small OD in remote sensing images. Wang et al.²³ presented the YOLO-Deformable Convolution with Efficient Channel Attention and Bidirectional Feature Pyramid Network (YOLO-DBS) approach for improving underwater image pre-processing and enhancing real-time detection accuracy and efficiency. Lin and Leng²⁴ proposed a Scale Selection Network (SSN) with a landmark guided scale attention network and a reversible scale semantic flow preserving strategy for mitigating computational redundancy and accelerating multi-scale OD in remote sensing images. Apostolidis and Papakostas²⁵ analyzed the security, robustness, and performance of the YOLO OD models, including YOLOX variants, under adversarial attacks and image corruption to guide optimal model selection for real-world applications. Yuan et al.²⁶ developed a distributed Edge-Cloud collaborative framework for unmanned aerial vehicle (UAV) object detection using the Edge-Embedded Lightweight (E2L) algorithm with attention mechanism (AM) and a fuzzy neural network. Kim et al.²⁷ developed Cirrus, an end-to-end system with combinational point sampling for low-latency and energy-efficient 3D OD on outdoor Light Detection and Ranging (LiDAR) data, preserving accuracy while mitigating input size and processing overhead. Subramanian et al.²⁸ presented an audio-visual object detection system using TensorFlow Lite and FogBus on Raspberry Pi, incorporating multimodal data fusion with edge and fog computing to enable real-time, accurate scene description for visually impaired users.

Though the existing studies are efficient in the OD task for individuals with disabilities, they still exhibit various limitations. For instance, multiple models encounter challenges due to low lighting, high occlusion, and background clutter, affecting detection accuracy. The dependence on single-modal data (either visual or audio) restricts robustness in dynamic environments. Few methodologies illustrate high accuracy but are computationally intensive, making real-time deployment on low-power edge devices difficult. Furthermore, generalization poses a constraint due to limited adaptability across diverse environmental conditions and targets. Additionally, segmentation accuracy for small or overlapping objects remains suboptimal. The research gap lies in effectively balancing detection precision, computational efficiency, and multimodal integration to ensure real-time, on-device performance for practical OD in VIP applications.

Methodological approach

In this manuscript, an FFDGCRN-ROD technique is presented for individuals with disabilities. It concentrates on developing an intelligent OD framework for individuals with disabilities, employing smart technology to allow real-time monitoring and assistive decision-making. To achieve this, the study encompasses image pre-processing, OD process, feature representation, and classification processes. Figure 1 represents the general process of the FFDGCRN-ROD model.

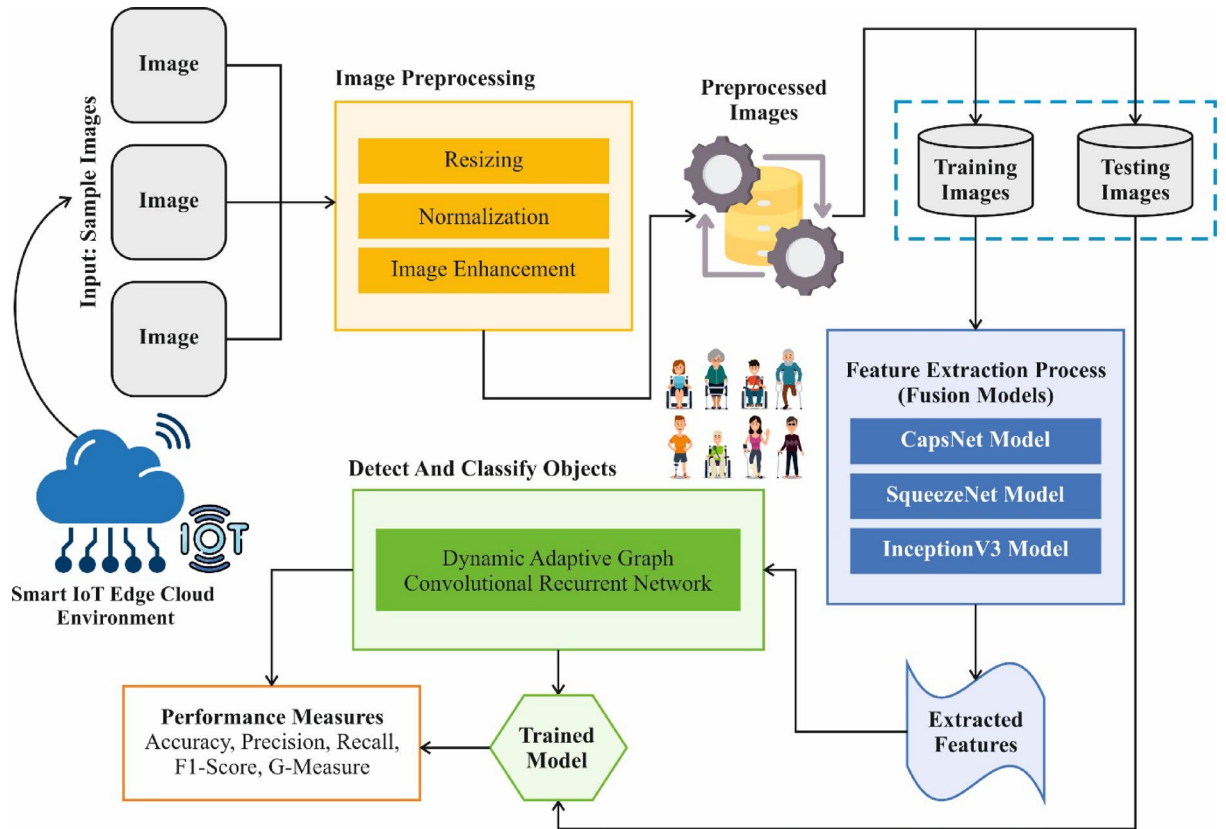


Fig. 1. General process of FFDGCRN-ROD technique.

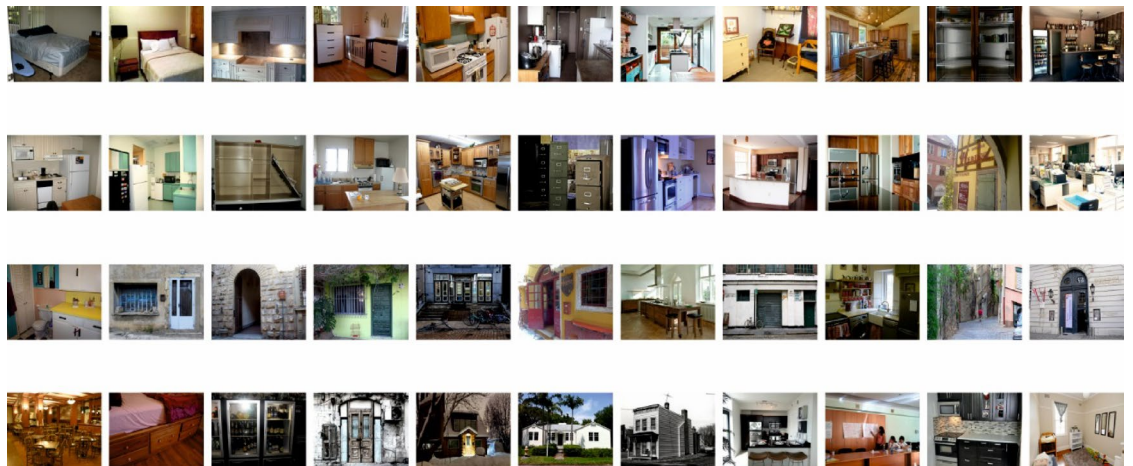


Fig. 2. Sample images.

Dataset description

This study utilizes a publicly available indoor OD dataset to assess the effectiveness of the proposed model. The dataset consists of 6,642 object instances distributed across multiple categories. Among them, Cabinet Door has the highest representation with 3,890 counts, followed by Refrigerator Door (879), Door (562), and Window (482). More minor but notable categories include Chair (223), Table (248), and Cabinet (208). Rarely occurring objects are Opened Door (90), Pole (36), and Couch (24). This varied distribution highlights the dominance of door-related objects while also including a mix of furniture and structural elements²⁹. Also, common augmentation techniques such as horizontal flipping, random rotation, and scaling were applied to enrich data diversity. These augmentations assist in improving the robustness of the model, specifically for underrepresented classes. Figure 2 depicts the sample images.

Pre-processing techniques

Primarily, the image pre-processing phase involves resizing, normalization, and image enhancement to remove noise and enhance image quality³⁰. This stage consists of a series of methods that process raw images to improve their quality for study. This may include processes such as noise reduction, normalization, contrast enhancement, and colour correction. The aim is to enhance the clarity of the image and make its attributes more critical for the following processing stages. Dual dissimilar pre-processing models are applied to the input images, based on the feature extraction methods, which emulate the requirement.

Image resizing

Resizing fine-tunes the image size, normalizes dimensions for ML methods, and maintains the aspect ratio until the alteration is projected. Here, histopathology images are re-scaled and standardized before being provided to the CNN method to detect and optimize, thereby optimizing processing time.

Normalization

It aids in stopping the above-mentioned issue. Normalization is the most effective and encompassing standardization method. A dataset will have a unit standard deviation and 0 mean through this complete after it has been normalized for all feature samples. Min-Max Normalization is used and then the pixel values of image I are scaled to the particular range, typically $[0,1]$. Here, I_{\min} and I_{\max} refers to the image's maximum and minimum values of pixel, consistently.

$$I_{nom} = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \quad (1)$$

Image enhancement

Image enhancement in ML improves image quality compared to methods like sharpening, contrast adjustment, and noise reduction, resulting in distinct features. These pre-processing stages are crucial for achieving a consistent outcome of the technique, particularly in creating images for feature extraction. CLAHE improves image contrast by using HE inside tiles, increasing local contrast and particularity, however, reducing noise amplification and maintaining image quality. It is streamlined using the succeeding expression for fine-tuning pixel intensity according to local HE:

$$Output(x, y) = \min \left(\max \left(\frac{Input(x, y) - Min_{local}}{Max_{local} - Min_{local}} \times Clip_{factor}, 0 \right), 255 \right) \quad (2)$$

Here, $(Input(x, y))$ refers to the new pixel value; however, (Min_{local}) and (Max_{local}) represent the maximum and minimum values. The 'Clip' feature limits contrast enhancement to the valid interval of 0 to 255.

Indoor OD process

For OD, the FFDGCRN-ROD approach utilizes the faster R-CNN for automatically recognizing and locating specific targets in the images³¹. The process of OD is implemented by the Faster R-CNN method for creating accurate region proposals. It is selected because of its higher precision and efficacy in identifying objects with changing scales and composite settings. Unlike traditional models or former R-CNN versions, it incorporates classification and region proposal. Its region proposal network (RPN) permits accurate and quick localization of objects. In comparison with methods like SSD or YOLO, Faster R-CNN presents better recognition outcomes, mainly in conditions requiring higher localization accuracy.

Its strength through different datasets makes it a better option for real-time applications. Faster R-CNN is one of the famous methods for its greater accuracy. It includes two components: RPN and a detection model. The fundamental source of Faster RCNN contains the following phases:

- 1) Base network: Use a pre-trained CNN to eliminate attributes from input images.
- 2) RPN: It creates candidate-directed regions by sliding a window across the feature maps, which are considered as anchor boxes.
- 3) Region of Interest (RoI) pooling: Divided every candidate's regions into permanent size sub-fields and mapped them on static size feature mapping.
- 4) Object classification network: Employ fully connected (FC) models for categorizing every candidate region, capturing an output of RoI Pooling as an input, and output the probabilities of every candidate region.
- 5) Bounding box regression: Apply these regressions to adjust the coordinates of the bounding box aims in every candidate region.

To perform OD, they use the RPN to create candidate regions. The RPN stage consists of two main formulations: one to calculate the coordinates of anchor boxes and the other to compute the overlap between bounding and ground truth anchor boxes.

$$\begin{aligned} x_{anchor} &= x_{center} - \frac{w_{anchor}}{2} \\ y_{anchor} &= y_{center} - \frac{h_{anchor}}{2} \\ w_{anchor} &= width_{anchor} \\ h_{anchor} &= height_{anchor} \end{aligned} \quad (3)$$

x_{anchor} and y_{anchor} are the anchor box. x_{center} and y_{center} represent the coordinates of the central point. w_{anchor} and h_{anchor} represent the height and width. $height_{anchor}$ and $width_{anchor}$ indicate the predefined height and width

$$L_{bbox} = \sum_i L_{smooth}(t_i - t'_i, 1_i \text{ is positive}) \quad (4)$$

Whereas t_i symbolizes the forecast bounding boxes offset, t'_i exemplifies the steady ground truth bounding boxes offset, and 1_i specifies the positive of an indicator.

$$F_{roi} = RoIpooling(F_{conv}, p) \quad (5)$$

F_{conv} symbolizes the convolution feature maps, and p signifies the input parameter. Pooling layer for RoI mapping, where $RoIs$ of different dimensions are mapped onto pre-determined size features.

This phase performs pooling procedures, which are equivalent to $RoIs$ of altered dimension, resulting in pre-determined size RoI characteristics.

$$F_{cls} = softmax(W_{cls} \cdot F_{roi} + b_{cls}) \quad (6)$$

The bounding box regression equation is given below:

$$F_{reg} = W_{reg} \cdot F_{roi} + b_{reg} \quad (7)$$

While, W_{cls} , b_{cls} , W_{reg} , and b_{reg} refer to learned parameters. The function of softmax has been applied to convert the output of object classification. By combining the detection network and RPN, this model achieves a correct OD level and has proved to be a significant performance improvement on many benchmark datasets.

Fusion of feature representation model

Next, the fusion models, namely CapsNet, SqueezeNet, and Inceptionv3, are utilized for the feature extraction. The fusion model is chosen for its complementary strengths in capturing spatial hierarchies and pose information, which helps recognize overlapping or deformed objects. The SqueezeNet is effective due to its lightweight architecture, facilitating efficient computation with fewer parameters, thus making it ideal for edge deployment. Inceptionv3 is chosen for its capability in handling diverse object scales through multi-level convolutions, improving feature richness. This integration enhances the robustness and generalization of the model across intrinsic scenes. They are trained using appropriate batch processing, epochs, categorical cross-entropy loss, dropout for regularisation, and weight initialization techniques to ensure efficient training and good convergence.

CapsNet architecture

In DL applications, CNNs have become a general method for analyzing image data. CNN is a particular kind of NN framework, which consists of FC neurons that are provided by specifically developed layers³². The layers included in this process are called flattening layers, convolutional layers, and pooling layers, respectively. Besides these typical layers, there are many specified layers intended to improve data handling in several methods. By incorporating these layers of CNN, image-based data is successfully examined, and classification tasks are implemented with enhanced precision.

CapsNet is an NN approach which uses capsule layers made up of groups of neurons. Unlike classic CNNs, these groupings inside capsules permit the detailed representation of dissimilar properties. Then, a CapsNet approach can well identify further details in image data. A standard CapsNet method improves the convolutional layer's outputs by processing them comprehensively over capsule layers. In this work, the CapsNet features have also been advanced with different changes, like dynamic routing, which contains class and primary capsule layers. These changes aid in computing added parameters, like the probability of the presence of particular features. In this work, CapsNet was used as the DL module. In comparison with classic CNNs, this model is a better option to analyze image data as it successfully takes into account the spatial hierarchy. In addition, this model has smaller parameters than classic CNNs, which decreases the overfitting hazard.

SqueezeNet architecture

SqueezeNet is a lightweight CNN designed to achieve greater precision with significantly fewer parameters, making it ideal for applications that require computing efficacy³³. It minimizes model size across three major approaches: 1x1 convolutions for parameter decline, Fire modules for effective feature extraction, and delayed downsampling for maintaining spatial data. These model selections enable SqueezeNet to maintain AlexNet-level precision while having 50 × smaller parameters, making it particularly suitable for tasks such as classification. The Fire module incorporates expand and squeeze layers. The layer of squeeze utilizes convolutions with kernel dimensions of 1 to reduce the input channel counts, resulting in a considerably lower computing cost. Then, the expanded layer incorporates convolutions with kernel dimensions of one and three, ensuring a balance between parameter efficacy and strong feature extraction. This framework guarantees that the model can seize fine-grained particulars that are important for precise classification.

Unlike conventional CNN structures that use pooling layers initially, this model delays these processes to deep layers. This assistance maintains higher-resolution spatial data that is important for classification, as more minor changes in colour, shape, size, and variation frequently identify the category. The system further utilizes GAP rather than FC layers, also reduces the number of parameters, and maintains strong feature representation. This leads to a method that is either computationally lightweight or able to remove discriminatory features essential for detection. By utilizing effective feature extraction, this model proves to be a perfectly suitable method for classification, offering the best balance between interpretability, precision, and speed. Figure 3 signifies the architecture of SqueezeNet.

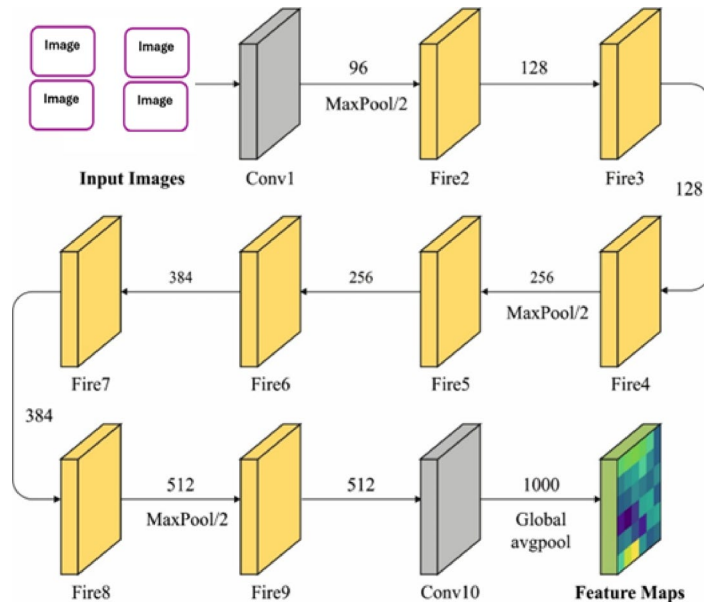


Fig. 3. SqueezeNet architecture.

Inceptionv3 architecture

Inception-v3 is a deep CNN image classifier which integrates numerous structural developments to improve precision and efficacy³⁴. A smaller convolution count, similar to dual 3×3 convolutions, completely substitutes large convolutions, such as 5×5 . Factorized convolution enables the model to reduce calculation costs without altering the receptive field. To further improve performance, asymmetric convolution, such as separating an $n \times n$ matrix into $1 \times n$ and $n \times 1$, decreases the total parameter counts. Other prominent features include an auxiliary classifier, a smaller network inserted at layers in the centre, which serves as a regularizer to mitigate the problem of gradient vanishing and assist in convergence improvement after training. For training, the system utilizes categorical loss of cross-entropy that can be specified by:

$$L = - \sum_{i=1}^N y_i \log(x_j) \quad (8)$$

Whereas y_i denotes real classes, and x_i means predicted possibility for all classes I . To prevent the method from being overly confident in its predictions, label smoothing is applied substantially. It fine-tunes the labels of ground truth by distributing a smaller likelihood mass to inaccurate class labels.

$$y'_i = y_i (1 - \epsilon) + \frac{\epsilon}{N} \quad (9)$$

Here, ϵ (epsilon) assists in avoiding overfitting by making the method less reliable for this prediction. These optimizations make inceptionv3 a very effective method for composite image classification tasks.

Classification with DA-GCRN

Lastly, the FFDGCRN-ROD method employs the DA-GCRN technique to precisely detect and classify objects for visually impaired people³⁵. The spatial and temporal dependencies in sequential data are effectively captured by this method, which is considered crucial for accurate OD and classification in dynamic environments. The domain adaptation capability across diverse subjects and conditions assists the generalization of the model. Unlike conventional models, this methodology demonstrates excellence in efficiently handling discrepancies in data distribution, thus improving robustness and mitigating errors in real-world scenarios. Additionally, the contextual relationships are effectively learned by the recurrent structure of the model, thereby enhancing detection accuracy and reliability in assistive applications.

This work proposes a DA-GCRN method for spatio-temporal prediction, and it is based on a sequence-to-sequence structure, using the encoder-decoder framework. Both the decoder and encoder are comprised of adaptive graph convolutional recurrent units (AGCRUs). The AGCRUs incorporate DAGC processes with the GRU, which is a simpler version of LSTM, considerably increasing computational efficacy while maintaining performance. The introduction of the DAGC substitutes the multi-layer perceptron layer in the new GRU, expanding the ability of the model to complex model dependences and spatio-temporal relationships in dynamic time-series.

The AGCRU becomes the primary element for spatio-temporal modelling, successfully capturing temporal and spatial dependences in the input graph topology. The key to adaptive graph convolution rests in how it uses the dynamic graph framework \tilde{P} , which characterizes spatial correlations and is described as demonstrated:

$$H = \sigma (X \star_G \Theta) = \sigma \left(\sum_{k=0}^K \tilde{P}^k X W_k \right) \quad (10)$$

Whereas $X \in \mathbb{R}^{N \times C}$ and $H \in \mathbb{R}^{N \times h}$ characterize the output and input of the GC operation (\star_G), correspondingly, $\tilde{\Theta}_k$ and $W_k \in \mathbb{R}^{N \times C \times h}$ represent the k th order kernel parameter, σ means activation function, and $\tilde{P} \in \mathbb{R}^{N \times N}$ characterizes the dynamic topological relationship among the nodes in graph G . The computation procedure of the AGCRU, made by incorporating adaptive GC with GRU components, is as demonstrated:

$$u_t = \text{Sigmoid}([X_t, H_{t-1}] \star_G \Theta_u + b_u) \quad (11)$$

$$r_t = \text{Sigmoid}([X_t, H_{t-1}] \star_G \Theta_r + b_r) \quad (12)$$

$$C_t = \text{Tanh}([X_t, (r_t \odot H_{t-1})] \star_G \Theta_C + b_C) \quad (13)$$

$$H_t = u_t \odot H_{t-1} + (1 - u_t) \odot C_t \quad (14)$$

Here, C , u , and r characterize the candidate state, update gate, and reset gate in the AGCRU component, respectively. \odot signifies the product of Hadamard, while $\Theta_{\{u,r,C\}} \in \mathbb{R}^{K \times (C+h) \times h}$ characterizes the learnable gate parameters. $[X_t, H_{t-1}]$ and H_t represent input and output at time step t . Moreover, to the observation X_t , the AGCRU needs auxiliary input from the dynamic topological framework \tilde{P} .

Evaluation metrics

Accuracy, precision, recall, F1-score, and G-measure are some of the traditional performance measures used to verify. It also allows for investigating the outcomes of predictions using the confusion matrix, where TP and TN signify true positive and negative, and FP and FN signify false positive and negative. These metrics assess the model's ability to generalize across diverse image data types while maintaining patient privacy in a DL setting.

Accuracy

A prevalent statistic to assess how precise a classifier model is in terms of accuracy. In the below-mentioned Eq. (5), it is calculated as the ratio of clearly marked examples to every occurrence in the dataset.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (15)$$

Precision

Precision is described as the ratio of properly classified images to every categorized image. The precision (Pr) might be stated utilizing Eq. (6):

$$\text{Precision} = \frac{TP}{TP + FR} \times 100 \quad (16)$$

Recall

Recall determines how many of all images in the dataset were correctly identified. Equation (17) offers a brief formulation that outlines it as shown:

$$\text{Recall} = \frac{TP}{TP + FN} \times 100 \quad (17)$$

F1-score

A maximum F-score proposes that the model has a successful predictive ability, as it characterizes the harmonic means of precision and recall. Precision and recall are still insufficient metrics to evaluate the performance. Based on Eq. (8), the F -score is established as demonstrated in Eq. (18).

$$F1 - \text{score} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (18)$$

G-measure

It is a performance metric that balances recall and precision by capturing their geometric mean. It offers a single score that imitates the tradeoff between these two measures, ensuring neither is neglected.

$$G - \text{measure} = \sqrt{\text{precision} * \text{recall}} \quad (19)$$

Results and discussion

The investigational study of the FFDGCRN-ROD model is inspected under the Indoor OD dataset. The technique is simulated using Python 3.6.5 on a PC with an i5-8600k, 250GB SSD, GeForce 1050Ti 4GB, 16GB

RAM, and 1 TB HDD. Parameters include a learning rate of 0.01, ReLU activation, 50 epochs, 0.5 dropout, and a batch size of 5.

Figure 4 shows the classifier outcome of FFDGCRN-ROD methodology on 80:20. Figure 4a and b demonstrate the confusion matrices with accurate detection of each class. Figure 4c exhibits 80% of the training phase (TRAPH) for several classes at diverse metrics. Figure 4d illustrates 20% of the testing phase (TESPH) for several classes on dissimilar metrics.

Table 1; Fig. 5 display the OD of the FFDGCRN-ROD method at 80%:20%. On 80% TRAPH, the FFDGCRN-ROD model attains an $accu_y$, $prec_n$, $recal$, $F1_{Score}$, and $G_{Measure}$ of 99.57%, 93.57%, 88.13%, 90.09%, and 90.46%, respectively. Likewise, at 20% TESPH, the FFDGCRN-ROD model attains an $accu_y$, $prec_n$, $recal$, $F1_{Score}$, and $G_{Measure}$ of 99.65%, 93.85%, 92.25%, 92.85%, and 92.95%, respectively.

Figure 6 describes the training (TRAN) and validation (VALD) accuracy of the FFDGCRN-ROD model on 80:20 over 200 epochs. Both curves progressively surge and steadily converge, denoting that the method is learning effectively. The VALD accuracy consistently exceeds the TRAN accuracy, indicating that the methodology is not over-fitting and is better at generalizing to unseen data. The variations in accuracy are due to the task's intricacy, but general upward tendencies reveal robust performance and strength of the model in differentiating between normal and attack data.

Figure 7 describes the TRAN and VALD loss of the FFDGCRN-ROD approach at 80:20 over 200 epochs. Both curves depict a persistent downward trend, denoting that the model is efficiently reducing error at learning. The VALD loss remains minimally lesser than the training loss across most epochs, signifying better generality and no signs of over-fitting. While some fluctuations are observed, it is becoming gradually reliable and steady at differentiating between normal and attack data as training emerges.

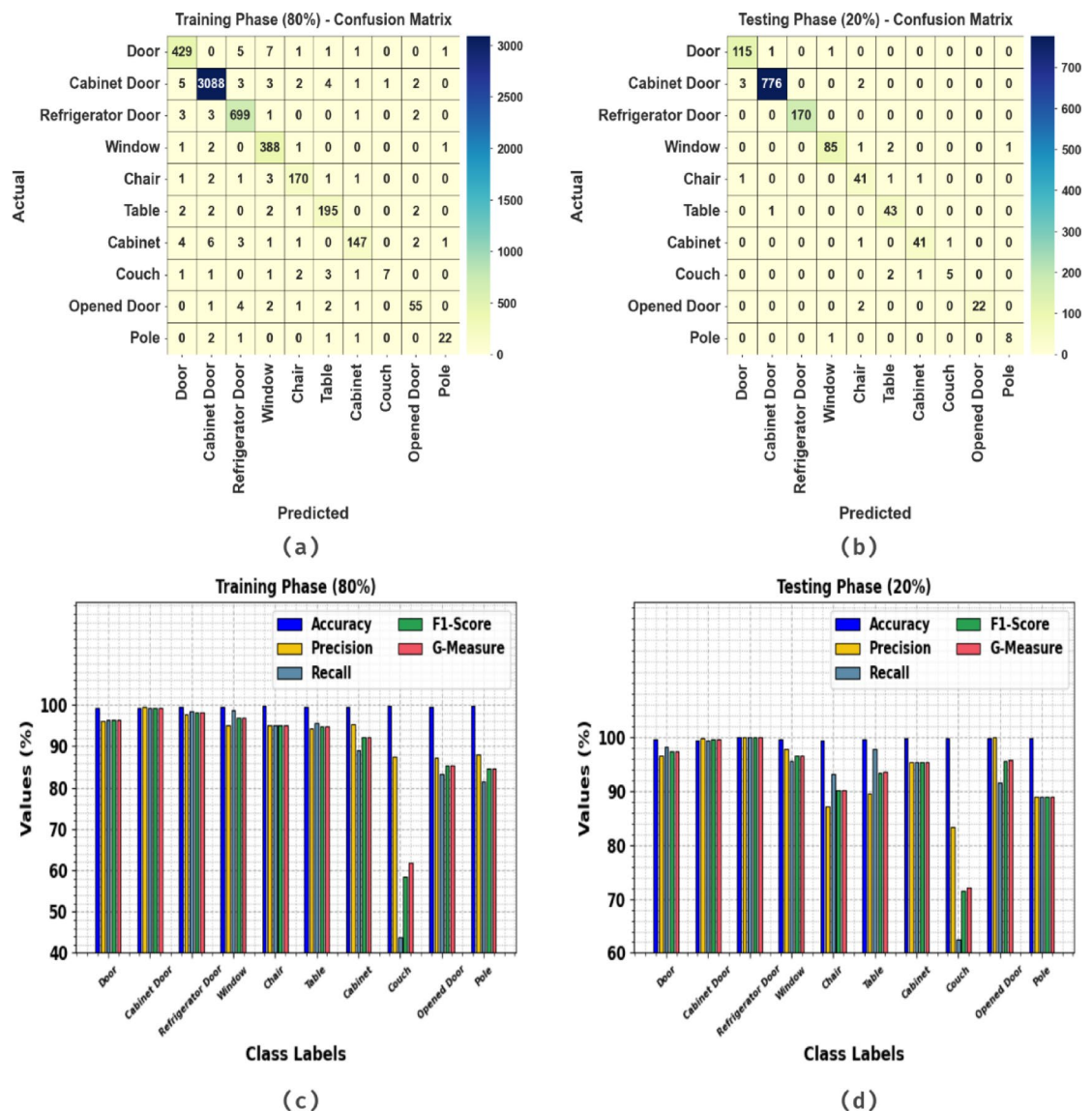


Fig. 4. 80:20 of (a, b) confusion matrix, and (c, d) TRAPH/TESPH.

Class Labels	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_t</i>	<i>F1_{Score}</i>	<i>G_{Measure}</i>
TRAPH (80%)					
Door	99.38	96.19	96.40	96.30	96.30
Cabinet Door	99.25	99.39	99.32	99.36	99.36
Refrigerator Door	99.49	97.63	98.59	98.11	98.11
Window	99.53	95.10	98.73	96.88	96.90
Chair	99.66	94.97	94.97	94.97	94.97
Table	99.60	94.20	95.59	94.89	94.89
Cabinet	99.53	95.45	89.09	92.16	92.22
Couch	99.81	87.50	43.75	58.33	61.87
Opened Door	99.64	87.30	83.33	85.27	85.29
Pole	99.85	88.00	81.48	84.62	84.68
Average	99.57	93.57	88.13	90.09	90.46
TESPH (20%)					
Door	99.55	96.64	98.29	97.46	97.46
Cabinet Door	99.47	99.74	99.36	99.55	99.55
Refrigerator Door	100.00	100.00	100.00	100.00	100.00
Window	99.55	97.70	95.51	96.59	96.60
Chair	99.32	87.23	93.18	90.11	90.16
Table	99.55	89.58	97.73	93.48	93.57
Cabinet	99.70	95.35	95.35	95.35	95.35
Couch	99.70	83.33	62.50	71.43	72.17
Opened Door	99.85	100.00	91.67	95.65	95.74
Pole	99.85	88.89	88.89	88.89	88.89
Average	99.65	93.85	92.25	92.85	92.95

Table 1. OD of FFDGCRN-ROD method on 80:20.

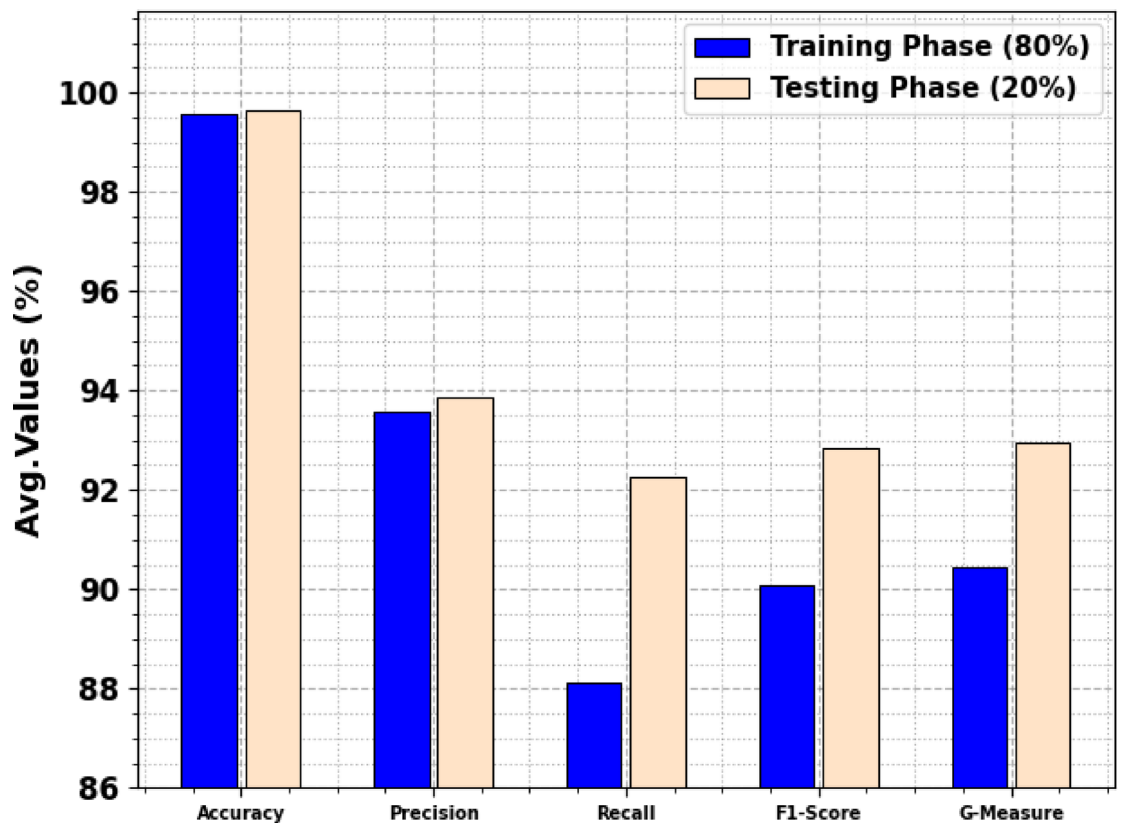


Fig. 5. Average values of FFDGCRN-ROD method under 80:20.

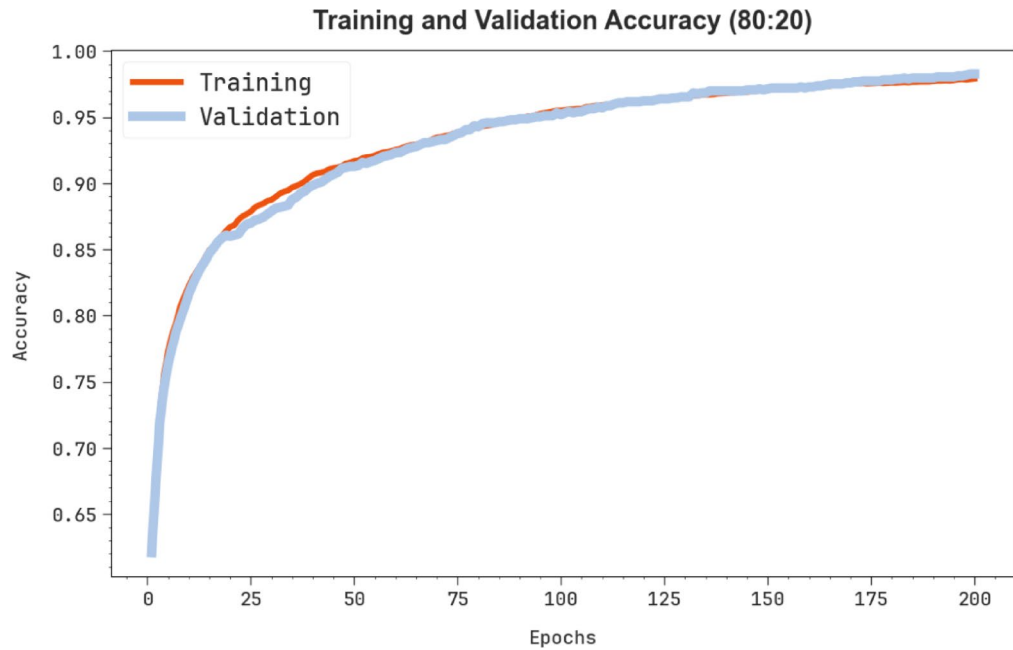


Fig. 6. $Accu_y$ curve of FFDGCRN-ROD technique under 80:20.

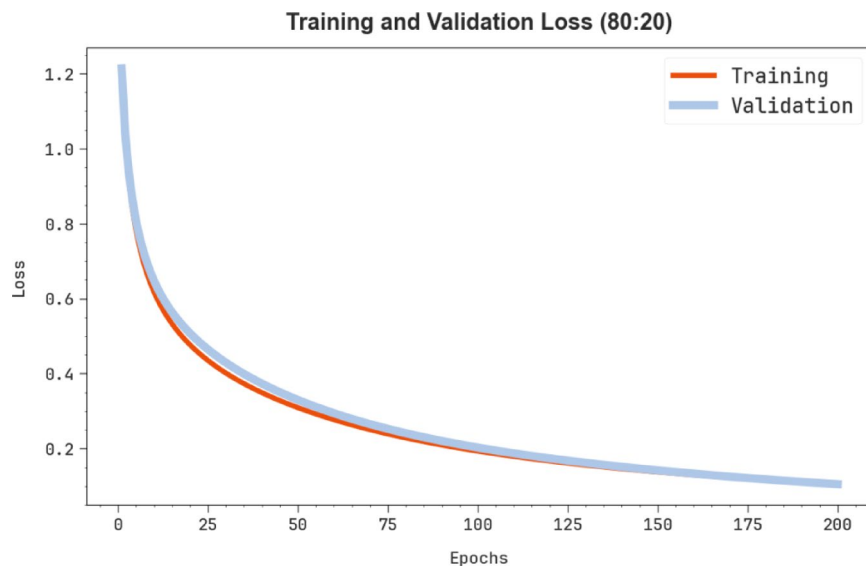


Fig. 7. Loss curve of FFDGCRN-ROD technique on 80:20.

In Fig. 8, the precision-recall (PR) curve analysis of the FFDGCRN-ROD approach on 80:20 provides insights into its performance by charting Precision against Recall for each class. The figure demonstrates that the FFDGCRN-ROD approach consistently yields increased PR values across multiple classes. The constant improvement in every class describes the proficiency of the FFDGCRN-ROD technique.

In Fig. 9, the ROC curve of the FFDGCRN-ROD approach on 80:20 is examined. The outcomes suggest that the FFDGCRN-ROD approach yields increased ROC values for each class, demonstrating a significant ability to differentiate between them. This persistent pattern over several classes implies the efficacious performance of the FFDGCRN-ROD model on class prediction.

Figure 10 displays the classifier outcome of the FFDGCRN-ROD technique on a 70:30 split. Figure 10a and b exhibit the confusion matrices with perfect recognition of every class. Figure 10c represents 70% of the TRAPH for numerous class labels under dissimilar metrics. Figure 10d illustrates 30% of the TESP for numerous classes under dissimilar metrics.

Table 2; Fig. 11 portray the OD of the FFDGCRN-ROD technique under 70%:30%. On 70% TRAPH, the FFDGCRN-ROD model attains an $accu_y$, $prec_n$, $reca_l$, $F1_{Score}$, and $G_{Measure}$ of 99.57%, 93.89%, 86.15%,

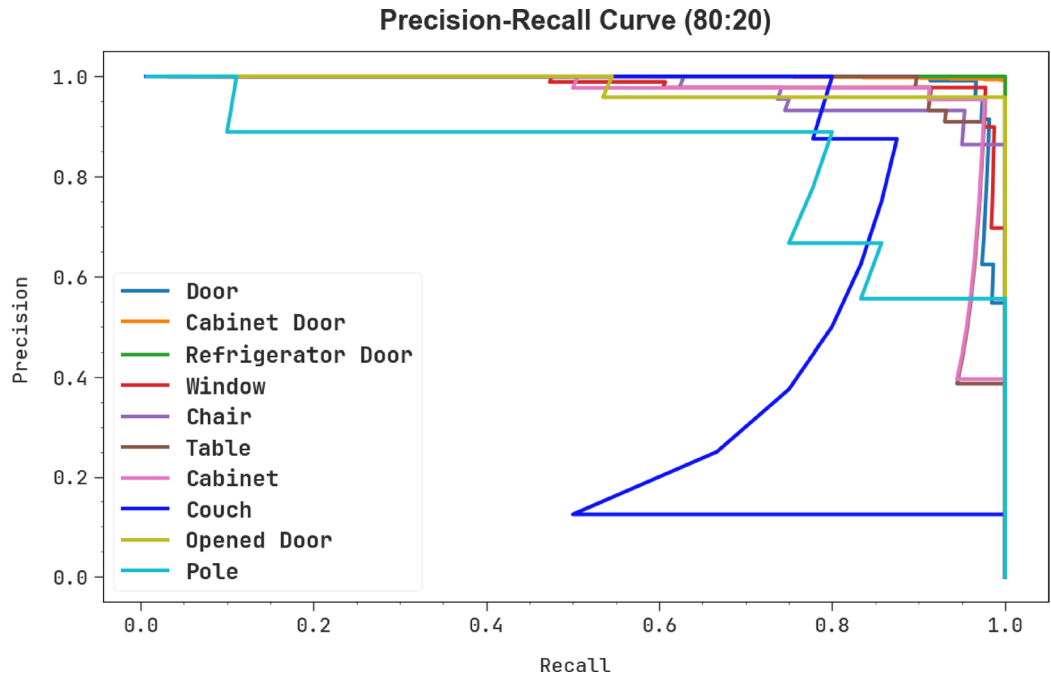


Fig. 8. PR curve of FFDGCRN-ROD technique under 80:20.

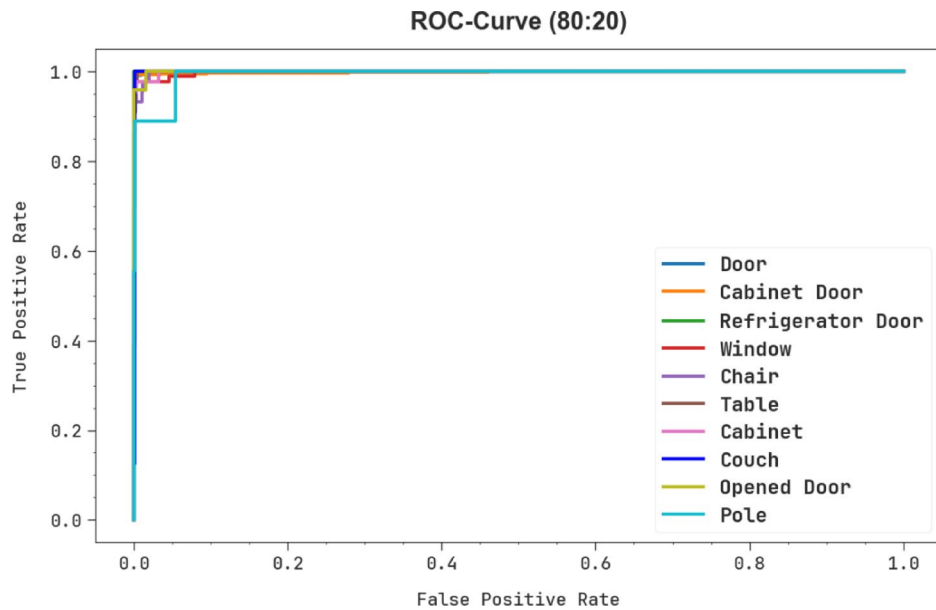


Fig. 9. ROC curve of FFDGCRN-ROD technique under 80:20.

88.56%, and 89.24%, respectively. Further, at 30% TESP, the FFDGCRN-ROD model attains an $accu_y$, $prec_n$, $recall$, $F1Score$, and $G_{Measure}$ of 99.63%, 96.33%, 91.49%, 93.59%, and 93.75%, respectively.

Figure 12 portrays the TRAN and VALD accuracy of the FFDGCRN-ROD methodology under 70:30 over 200 epochs. Both curves steadily increase and progressively converge, indicating that the model is learning effectively. The VALD accuracy consistently remains slightly higher than the TRAN accuracy, indicating that the model is not over-fitting and is generalizing better to unseen data. The fluctuations in accuracy are due to the complication of the task. Still, the overall upward trend reveals robust performance and stability of the model in discriminating between normal and attack data.

Figure 13 exemplifies the TRAN and VALD loss of the FFDGCRN-ROD methodology on 70:30 over 200 epochs. Both curves show steady downward tendencies, indicating that the model is excellently minimizing error during learning. The VALD loss remains minimally lower than the training loss across most epochs,

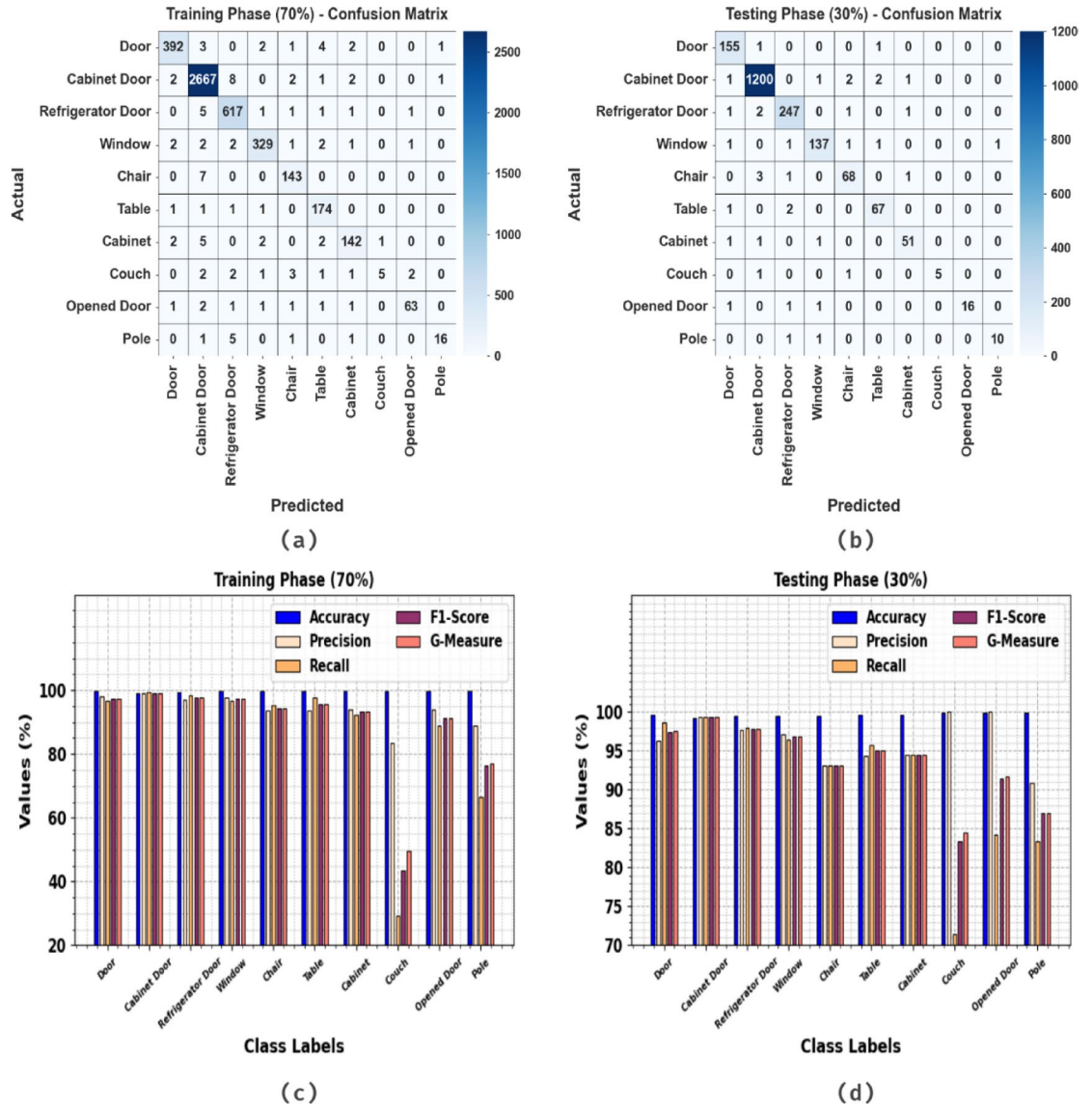


Fig. 10. 70:30 of (a, b) confusion matrix, and (c, d) TRAPH/TESPH.

implying good generality and no signs of over-fitting. Although some fluctuations are observed, it is becoming progressively stable and reliable at differentiating between normal and attack data as training progresses.

In Fig. 14, the precision-recall (PR) curve analysis of the FFDGCRN-ROD model on 70:30 provides insights into its outcomes by charting Precision against Recall for every class. The figure shows that the FFDGCRN-ROD model consistently achieves higher PR values across multiple classes. The persistent improvement in each class reveals the efficacy of the FFDGCRN-ROD in the classification process.

In Fig. 15, the ROC curve of the FFDGCRN-ROD approach on 70:30 is inspected. The results indicate that the FFDGCRN-ROD model achieves increased ROC values across all classes, demonstrating a significant ability to distinguish between them. This dependable pattern of maximal ROC values on multiple classes implies the efficacious outcomes of the FFDGCRN-ROD model on class prediction.

Table 3; Fig. 16 depict the comparative analysis of the FFDGCRN-ROD model with existing models^{20,21,36,37}. The table values indicate that the methodologies, including NAM, PCA, TFST, Yolo-V8, Yolo-V5 large, MR-CNN, YOLO-inception, CADNet, DICSSD300, and AttenRetina, achieved the minimum performance across various metrics. At the same time, the FFDGCRN-ROD model has obtained maximum performance with *accu_y*, *prec_n*, *reca_l* and *F1_{score}* of 99.65, 93.85%, 92.25%, and 92.85%, respectively.

In Table 4; Fig. 17, the time complexity (TC) of the FFDGCRN-ROD methodology is proven with the present approaches. The outcomes indicate that the FFDGCRN-ROD model accomplishes better performance with a TC of 11.95 s.

Among the existing methodologies, NAM, PCA, TFST, Yolo-V8, Yolo-V5 large, MR-CNN, YOLO-inception, CADNet, DICSSD300, and AttenRetina achieved the highest TCs of 20.09 s, 22.45 s, 19.34 s, 18.46 s, 30.18 s,

Class Labels	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_t</i>	<i>F1_{Score}</i>	<i>G_{Measure}</i>
TRAPH (70%)					
Door	99.55	98.00	96.79	97.39	97.39
Cabinet Door	99.05	98.96	99.40	99.18	99.18
Refrigerator Door	99.38	97.01	98.41	97.70	97.71
Window	99.59	97.63	96.76	97.19	97.19
Chair	99.63	93.46	95.33	94.39	94.39
Table	99.66	93.55	97.75	95.60	95.63
Cabinet	99.55	94.04	92.21	93.11	93.12
Couch	99.72	83.33	29.41	43.48	49.51
Opened Door	99.74	94.03	88.73	91.30	91.34
Pole	99.78	88.89	66.67	76.19	76.98
Average	99.57	93.89	86.15	88.56	89.24
TESPH (30%)					
Door	99.60	96.27	98.73	97.48	97.49
Cabinet Door	99.25	99.34	99.42	99.38	99.38
Refrigerator Door	99.45	97.63	98.02	97.82	97.82
Window	99.55	97.16	96.48	96.82	96.82
Chair	99.50	93.15	93.15	93.15	93.15
Table	99.65	94.37	95.71	95.04	95.04
Cabinet	99.70	94.44	94.44	94.44	94.44
Couch	99.90	100.00	71.43	83.33	84.52
Opened Door	99.85	100.00	84.21	91.43	91.77
Pole	99.85	90.91	83.33	86.96	87.04
Average	99.63	96.33	91.49	93.59	93.75

Table 2. OD of FFDGCRN-ROD methodology under 70:30.

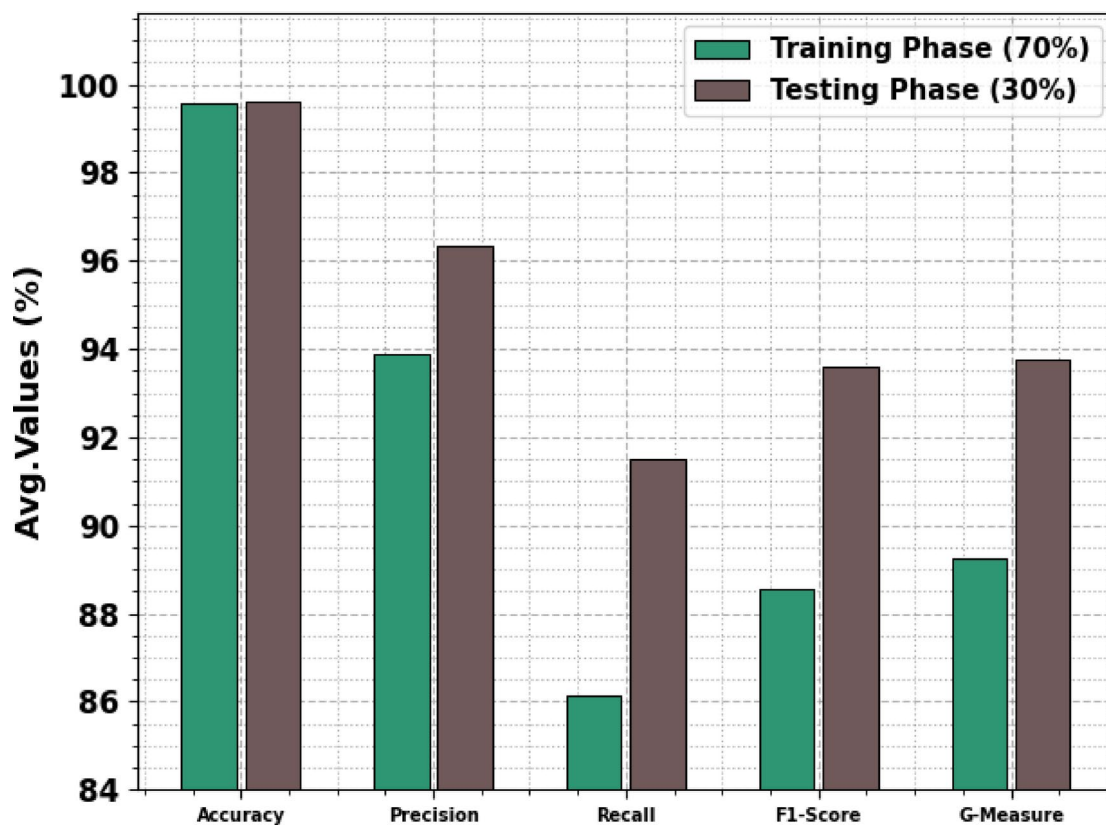


Fig. 11. Average values of FFDGCRN-ROD methodology under 70:30.

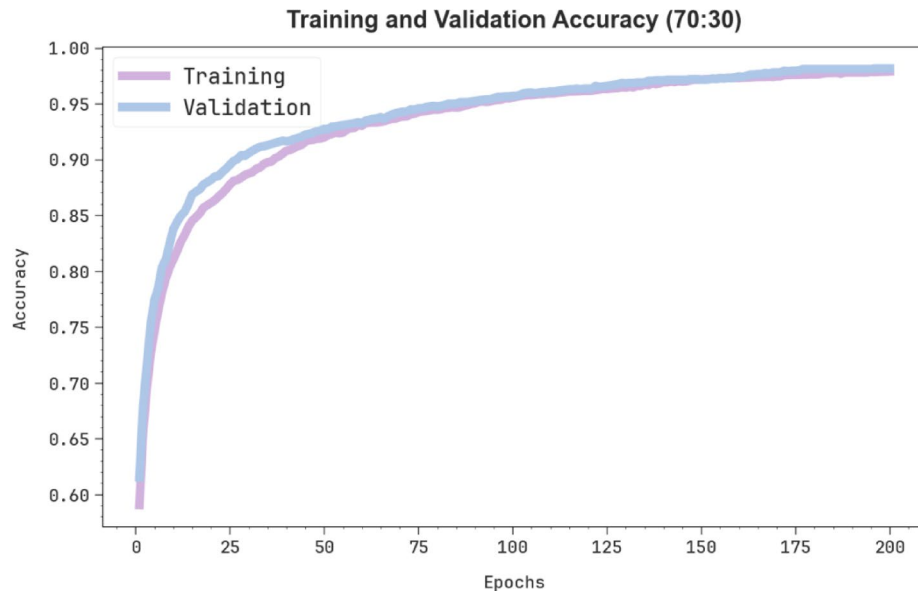


Fig. 12. $Accu_y$ curve of FFDGCRN-ROD technique under 70:30.

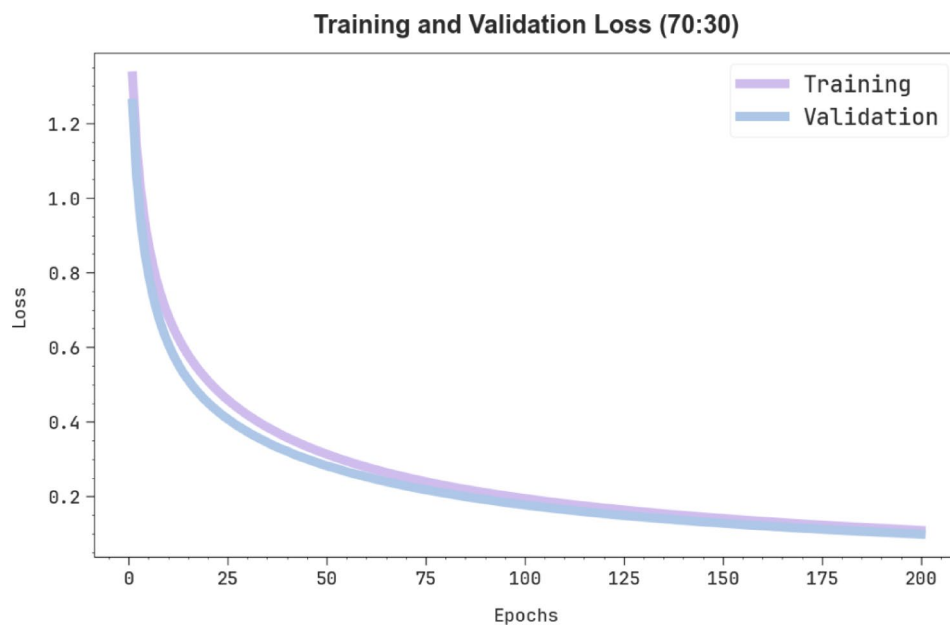


Fig. 13. Loss curve of FFDGCRN-ROD technique under 70:30.

29.40 s, 25.71 s, 25.08 s, 26.81 s, and 28.66 s, respectively. These results concluded the betterment of the proposed model over other recent studies.

Table 5 specifies the ablation study of the FFDGCRN-ROD methodology. The base DA-GCRN model achieved an $accu_y$ of 97.36%, $prec_n$ of 90.94%, $reca_l$ of 89.81%, and $F1_{Score}$ of 90.00%. Integrating CapsNet model enhanced the outcomes slightly, achieving an $accu_y$ of 97.90%, $prec_n$ of 91.67%, $reca_l$ of 90.33%, and $F1_{Score}$ of 90.70%. With SqueezeNet, the model illustrated further improvement, achieving an $accu_y$ of 98.47%, $prec_n$ of 92.28%, $reca_l$ of 90.94%, and $F1_{Score}$ of 91.47%. Adding InceptionV3 provided even better results, with an $accu_y$ of 99.05%, $prec_n$ of 93.05%, $reca_l$ of 91.60%, and $F1_{Score}$ of 92.27%. The overall model portrayed optimum performance, achieving an $accu_y$ of 99.65%, $prec_n$ of 93.85%, $reca_l$ of 92.25%, and $F1_{Score}$ of 92.85%, highlighting the efficiency of intelligent feature fusion in improving robust object detection.

Conclusion

This paper proposes the FFDGCRN-ROD method for OD among individuals with disabilities in the smart IoT-edge-cloud environment. This study has developed an intelligent OD framework for individuals with disabilities

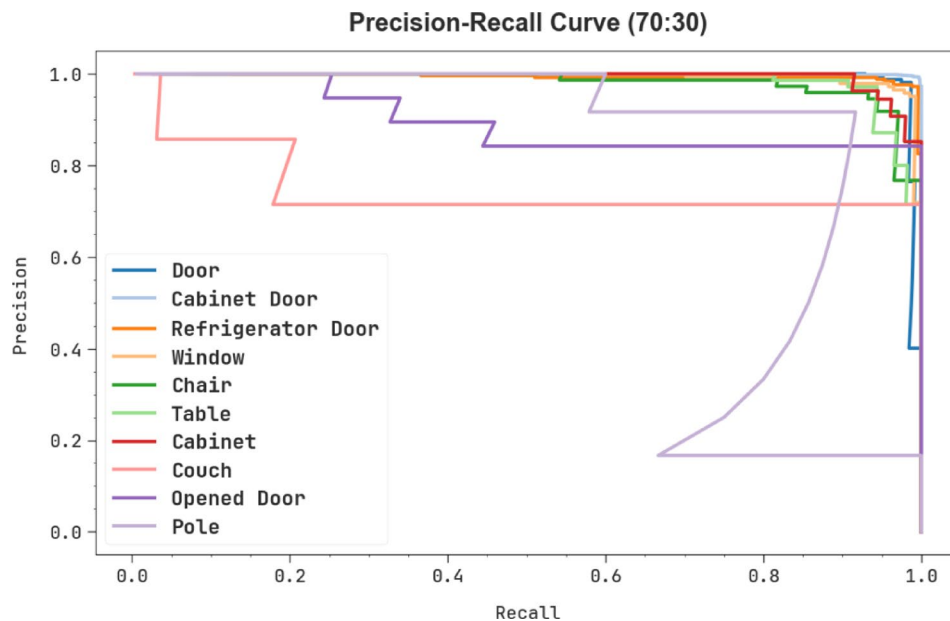


Fig. 14. PR curve of FFDGCRN-ROD technique on 70:30.

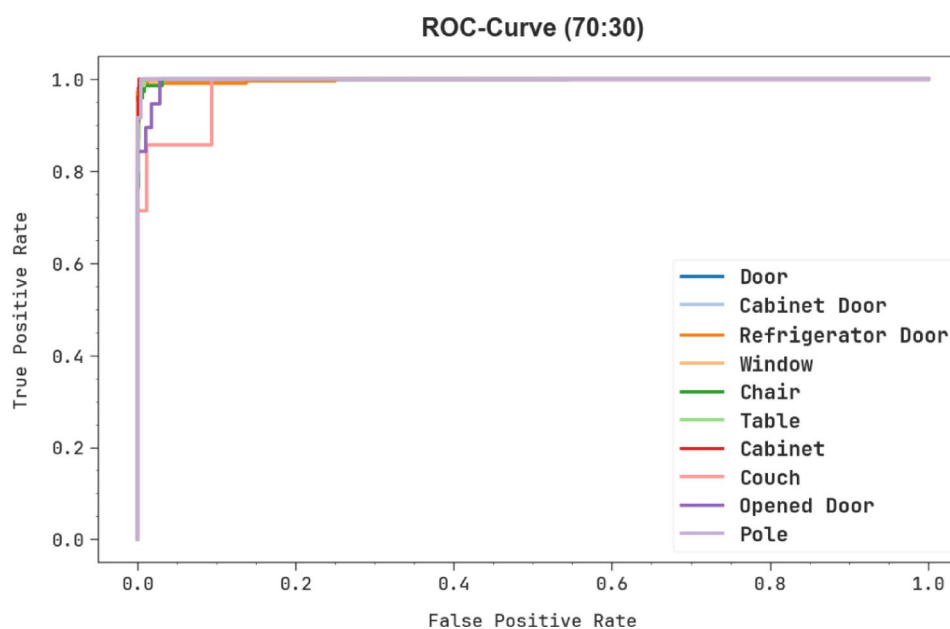


Fig. 15. ROC curve of FFDGCRN-ROD technique under 70:30.

utilizing smart technology to allow real-time monitoring and assistive decision-making. Primarily, the image pre-processing step encompasses resizing, normalization, and image enhancement to eliminate noise and improve image quality. For the OD process, the FFDGCRN-ROD method utilizes the faster R-CNN technique for automatically identifying and locating specific targets in the imagery. Furthermore, the fusion models, such as CapsNet, SqueezeNet, and Inceptionv3, are utilized for feature extraction. Finally, the FFDGCRN-ROD methodology implements the DA-GCRN technique to detect and classify objects for visually impaired individuals precisely. The comparison analysis of the FFDGCRN-ROD methodology demonstrated a superior accuracy value of 99.65% over existing methods when examined under the Indoor OD dataset. The limitations of the FFDGCRN-ROD methodology include high computational complexity and the requirement for large labelled datasets. The deployment of the model on resource-constrained devices may be affected, and the technique may also encounter threats in adapting to diverse real-world environments with varying lighting and occlusion conditions, affecting detection accuracy. The research gap is in improving lightweight, adaptive models that maintain high accuracy while being efficient for on-device processing. Real-world deployment challenges

Model	$Accu_y$	$Prec_n$	$Recal$	$F1_{Score}$
NAM	89.72	88.49	88.36	91.50
PCA	91.96	92.84	90.60	87.18
TFST	91.74	89.48	86.19	86.50
Yolo-V8	89.07	87.89	87.68	91.79
Yolo-V5 large	91.26	93.15	91.91	86.48
MR-CNN	91.10	88.72	85.60	85.97
YOLO-inception	88.58	89.03	89.95	87.79
CADNet	90.08	88.65	85.89	88.36
DICSSD300	97.81	88.51	90.28	86.85
AttenRetina	93.39	92.35	86.77	92.59
FFDGCRN-ROD	99.65	93.85	92.25	92.85

Table 3. Comparative study of FFDGCRN-ROD model with existing techniques.

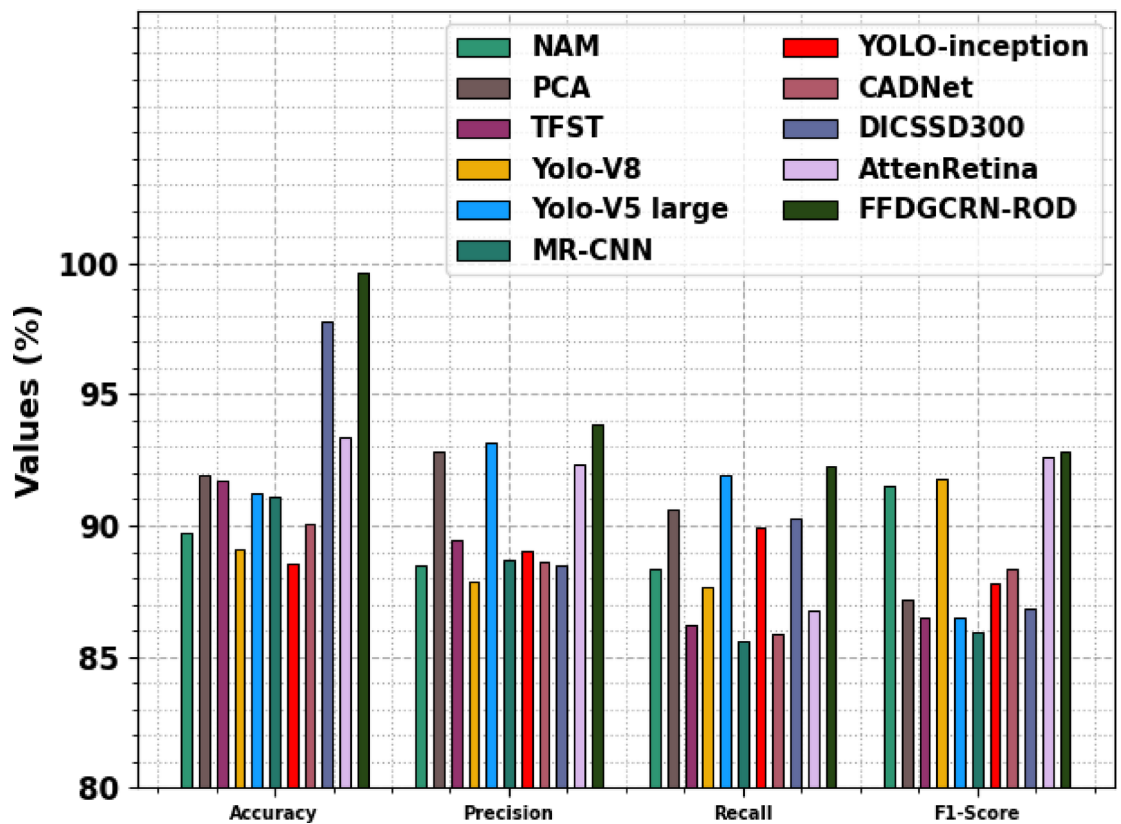


Fig. 16. Overall performance of FFDGCRN-ROD model with existing techniques.

include ensuring low latency, managing power consumption, and maintaining robustness against noise and environmental variability. Addressing these issues is significant for practical assistive technology applications.

Approach	TC (sec)
NAM	20.09
PCA	22.45
TFST	19.34
Yolo-V8	18.46
Yolo-V5 large	30.18
MR-CNN	29.40
YOLO-inception	25.71
CADNet	25.08
DICSSD300	26.81
AttenRetina	28.66
FFDGCRN-ROD	11.95

Table 4. TC outcome of FFDGCRN-ROD approach with existing methods.

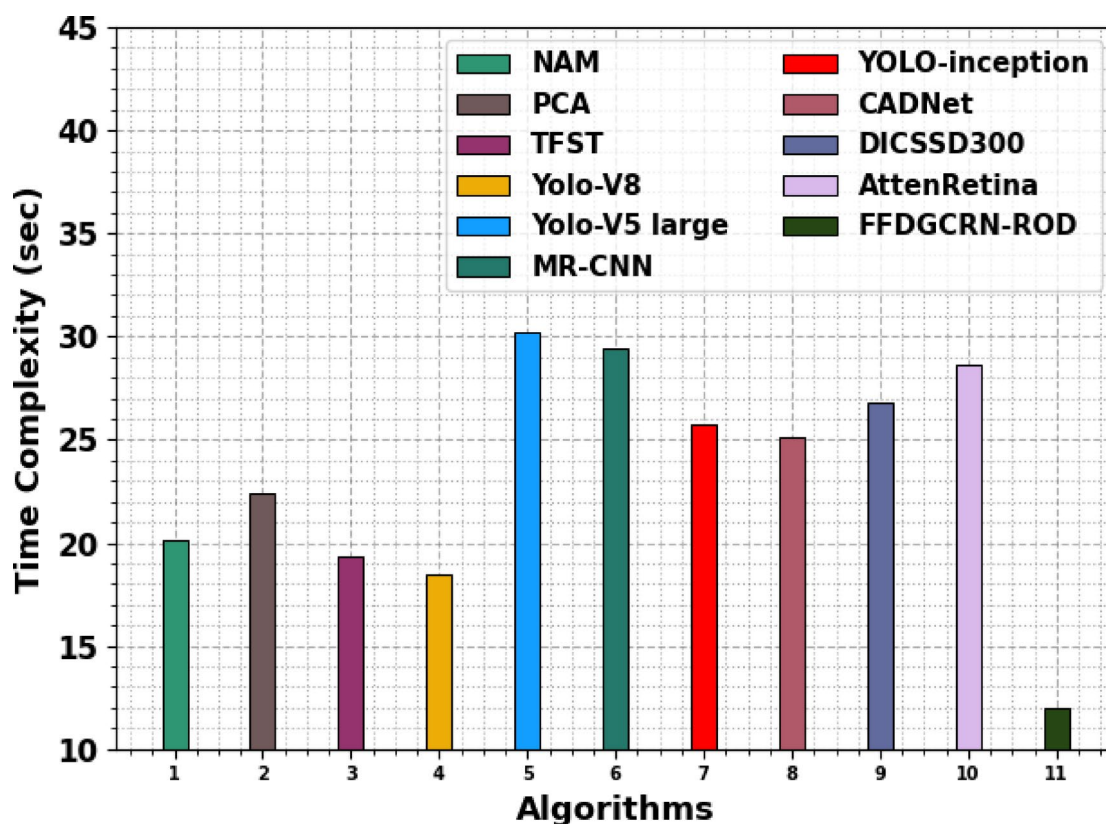


Fig. 17. TC outcome of FFDGCRN-ROD approach with existing methods.

Model	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal</i>	<i>F1Score</i>
DA-GCRN	97.36	90.94	89.81	90.00
DA-GCRN + CapsNet (with CapsNet without SqueezeNet and InceptionV3)	97.90	91.67	90.33	90.70
DA-GCRN + SqueezeNet (with SqueezeNet without CapsNet and InceptionV3)	98.47	92.28	90.94	91.47
DA-GCRN + InceptionV3 (with InceptionV3 without CapsNet and SqueezeNet)	99.05	93.05	91.60	92.27
FFDGCRN-ROD (DA-GCRN with all Feature Extraction)	99.65	93.85	92.25	92.85

Table 5. Performance comparison through ablation study of diverse components integrated with the DA-GCRN model.

Data availability

The data that support the findings of this study are openly available in the Kaggle repository at [<https://www.kaggle.com/datasets/thebordin/indoor-object-detection>] (<https://www.kaggle.com/datasets/thebordin/indoor-object-detection>), reference number²⁹.

Received: 2 September 2025; Accepted: 17 October 2025

Published online: 21 November 2025

References

- Gkonis, P., Giannopoulos, A., Trakadas, P., Masip-Bruin, X. & D'Andria, F. A survey on IoT-edge-cloud continuum systems: Status, challenges, use cases, and open issues. *Future Internet*, 15(12), p.383. (2023).
- Schneider, P. & Khafa, F. *Anomaly Detection and Complex Event Processing Over Iot Data Streams: with Application To EHealth and Patient Data Monitoring* (Academic, 2022).
- Bibri, S. E. & Jagatheesaperumal, S. K. Harnessing the potential of the metaverse and artificial intelligence for the internet of City things: Cost-effective XRReality and synergistic AIoT technologies. *Smart Cities*. 6 (5), 2397–2429 (2023).
- Okafor, C. & Bello, A. Designing a semantic analysis framework for intelligent learning by reading systems using advanced text comprehension techniques. *Stud. Knowl. Discovery Intell. Syst. Distrib. Analytics*. 14 (11), 13–25 (2024).
- Okokpujie, I. P., Kinggs, K. O. A., Adetunla, A., Okokpujie, K. & Awomoyi, M. E. September. Development of IoT based controlled bench power generation system. in world conference on information systems for business management (pp. 601–614). Singapore: Springer Nature Singapore. (2023).
- Lupi3n, M., Sanjuan, J. F., Gonz3lez-Ruiz, V. & Ortigosa, P. M. November. empowering the disabled: object detection and recognition for enhanced autonomy. in international conference on ubiquitous computing and ambient intelligence (pp. 275–287). Cham: Springer Nature Switzerland. (2024).
- Hussan, M. I., Saidulu, D., Anitha, P. T., Manikandan, A. & Naresh, P. Object detection and recognition in real time using deep learning for visually impaired people. *Int. J. Electr. Electron. Res.* 10 (2), 80–86 (2022).
- 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* Vol. 10, pp.13428–13441 (IEEE access, 2022).
- Alruwaili, M. et al. Deep learning-based YOLO models for the detection of people with disabilities. *IEEE Access*. 12, 2543–2566 (2023).
- Sharma, A., Vats, A., Shiv Shankar, D. & Surinder, K. Artificial intelligence enabled virtual sixth sense application for the disabled. *Thought* 4, 10 (2021).
- Adam, M. M. S., Aljehane, N. O., Alzahrani, M. Y. & Al Zanin, S. Leveraging assistive technology for visually impaired people through optimal deep transfer learning based object detection model. *Scientific Reports*, 15(1), p.30113. (2025).
- Obayya, M., Al-Wesabi, F. N., Alshammeri, M. & Iskandar, H. G. An intelligent optimized object detection system for disabled people using advanced deep learning models with optimization algorithm. *Scientific Reports*, 15(1), p.16514. (2025).
- Kumar, J. R., Reddy, M. B., Srikanth, M. & Giri, D. R. Automated object recognition with IoT for visually impaired users. In *Algorithms in Advanced Artificial Intelligence*, 1st Edition (82–87). CRC. (2024).
- Biradar, V. & Gull, K. C. *Enhancing Object Detection for Visually Impaired Integrating YOLOv8 with Spiking EfficientDet* pp.1–29 (Multimedia Tools and Applications, 2025).
- Abidi, M. H., Alkhalefah, H. & Siddiquee, A. N. Enhancing navigation and object recognition for visually impaired individuals: a gradient support vector boosting-based crossover golden jackal algorithm approach. *Journal of Disability Research*, 3(5), p.20240057. (2024).
- Hebbar, M. K. & Pulela, P. K. November. Object recognition system for the visually impaired: leveraging iot and remote server integration. In 2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM) (pp. 1–7). IEEE. (2023).
- Malla, S., Sahu, P. K., Patnaik, S. & Biswal, A. K. Obstacle detection and assistance for visually impaired individuals using an iot-enabled smart blind stick. *Revue d'Intelligence Artificielle* 37(3), 783–794 (2023).
- Alagarsamy, S. et al. January. An real time object detection method for visually impaired using machine learning. In 2023 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1–6). IEEE (2023).
- Alazeb, A. et al. Remote intelligent perception system for multi-object detection. *Frontiers in Neurorobotics*, 18, p.1398703. (2024).
- Shawly, T. & Alsheikhy, A. A. Eeg-based detection of epileptic seizures in patients with disabilities using a novel attention-driven deep learning framework with SHAP interpretability. *Egyptian Informatics Journal*, 31, p.100734. (2025).
- Xu, C. Y. et al. *Major Depressive Disorder Detection Via temporal-frequency-spatial Transformer with sub-domain Knowledge Alignment Using EEG* p.107965 (Neural Networks, 2025).
- Bu, Y., Ye, H., Tie, Z., Chen, Y. & Zhang, D. OD-YOLO: Robust small object detection model in remote sensing image with a novel multi-scale feature fusion. *Sensors*, 24(11), p.3596. (2024).
- Wang, X., Song, X., Li, Z. & Wang, H. YOLO-DBS: efficient target detection in complex underwater scene images based on improved YOLOv8. *J. Ocean. Univ. China*. 24 (4), 979–992 (2025).
- Lin, Z. & Leng, B. Ssn: Scale selection network for multi-scale object detection in remote sensing images. *Remote Sensing*, 16(19), p.3697. (2024).
- Apostolidis, K. D. & Papakostas, G. A. Delving into YOLO object detection models: Insights into Adversarial Robustness. *Electronics*, 14(8), p.1624. (2025).
- Yuan, Y. et al. Edge-cloud collaborative UAV object detection: Edge-embedded lightweight algorithm design and task offloading using fuzzy neural network. *IEEE Trans. Cloud Comput.* 12 (1), 306–318 (2024).
- Kim, J., Yi, J., Kim, W., Shin, S. & Lee, Y. May. Combinational point sampling for fast and accurate on-device LiDAR 3D object detection. In IEEE INFOCOM 2025-IEEE Conference on Computer Communications (pp. 1–10). IEEE. (2025).
- Subramanian, R. R., Ravikiran, L., Teja, K. V. P., Reddy, K. V. & Chowdary, K. A. November. Enhancing object detection through auditory-visual fusion on raspberry Pi and FogBus. In 2024 5th International Conference on Data Intelligence and Cognitive Informatics (ICDIC) (pp. 342–347). IEEE. (2024).
- <https://www.kaggle.com/datasets/thebordin/indoor-object-detection>
- Vittal, M. V. R. Advanced colon cancer detection: integrating context-aware multi-image fusion (Camif) in a multi-stage framework. *Egypt. Inf. J.* 29, 100609 (2025).
- Khadidos, A. O. & Yafoz, A. An intelligent object detection and classification framework for assisting visually challenged persons using deep learning and improved crow search optimization. *Scientific Reports*, 15(1), p.29822. (2025).
- Rodríguez-Aguilar, R., Jose-Antonio, M. S. & Utku, K. Federated learning based on an internet of medical things framework for a secure brain tumor diagnostic system: a capsule networks application. *Mathematics*, 13(15), p.2393. (2025).
- Ajay, A. et al. *LeafVisionNet: A Deep learning-based Approach for the Classification of Black Gram Leaf Disease Using attention-driven and semi-local Feature Boosted SqueezeNet* p.101245 (Smart Agricultural Technology, 2025).
- Chou, C. K. et al. Integrating AI with advanced hyperspectral imaging for enhanced classification of selected gastrointestinal diseases. *Bioengineering*, 12(8), p.852. (2025).

35. Peng, G., Liu, Y., Xiao, C., Du, W. & Xiao, C. A dynamic adaptive graph convolutional recurrent network model for efficient mid-short term prediction of global sea surface salinity. *International Journal of Digital Earth*, 18(1), p.2548008. (2025).
36. Nimma, D. et al. Object detection in real-time video surveillance using attention based transformer-YOLOv8 model. *Alexandria Eng. J.* **118**, 482–495 (2025).
37. Liu, G. et al. Object detection algorithm for autonomous driving: design and real-time performance analysis of attenretina model. *Alexandria Eng. J.* **123**, 392–402 (2025).

Acknowledgements

The authors extend their appreciation to the King Salman center For Disability Research for funding this work through Research Group no KSRG-2024-094.

Author contributions

Manal Abdullah Alohal: Conceptualization, methodology, validation, investigation, writing—original draft preparation, Fuhid Alanazi: Conceptualization, methodology, writing—original draft preparation, writing—review and editing Yaser Abdulaziz Alshafi: methodology, validation, writing—original draft preparation Ishfaq Yaseen: software, visualization, 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.A.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.

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