Tech Science Press

IoT-Driven Optimal Lightweight RetinaNet-Based Object Detection for Visually Impaired People

Mesfer Alduhayyem[1,*], Mrim M. Alnfiai[2,3], Nabil Almalki[4], Fahd N. Al-Wesabi[5], Anwer Mustafa Hilal[6] and Manar Ahmed Hamza[6]

[1]Department of Computer Science, College of Sciences and Humanities-Aflaj, Prince Sattam bin Abdulaziz University, Saudi Arabia
[2]King Salman Center for Disability Research, Riyadh, Al-Hayāṯim 16273, Saudi Arabia
[3]Department of Information Technology, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif, 21944, Saudi Arabia
[4]Department of Special Education, College of Education, King Saud University, Riyadh, 12372, Saudi Arabia
[5]Department of Computer Science, College of Science & Arts, King Khaled University, Ar-Riyad 12372, Saudi Arabia
[6]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia
*Corresponding Author: Mesfer Alduhayyem. Email: m.alduhayyim@psau.edu.sa

Abstract: Visual impairment is one of the major problems among people of all age groups across the globe. Visually Impaired Persons (VIPs) require help from others to carry out their day-to-day tasks. Since they experience several problems in their daily lives, technical intervention can help them resolve the challenges. In this background, an automatic object detection tool is the need of the hour to empower VIPs with safe navigation. The recent advances in the Internet of Things (IoT) and Deep Learning (DL) techniques make it possible. The current study proposes IoT-assisted Transient Search Optimization with a Lightweight RetinaNet-based object detection (TSOLWR-ODVIP) model to help VIPs. The primary aim of the presented TSOLWR-ODVIP technique is to identify different objects surrounding VIPs and to convey the information via audio message to them. For data acquisition, IoT devices are used in this study. Then, the Lightweight RetinaNet (LWR) model is applied to detect objects accurately. Next, the TSO algorithm is employed for fine-tuning the hyperparameters involved in the LWR model. Finally, the Long Short-Term Memory (LSTM) model is exploited for classifying objects. The performance of the proposed TSOLWR-ODVIP technique was evaluated using a set of objects, and the results were examined under distinct aspects. The comparison study outcomes confirmed that the TSOLWR-ODVIP model could effectually detect and classify the objects, enhancing the quality of life of VIPs.

Keywords: Visually impaired people; deep learning; object detection; computer vision; long short-term memory; transient search optimization

1 Introduction

Visual impairment is one of the major health complications that affect humankind across the globe. Vision loss or vision impairment is characterized by loss of visual capacity or visual-sensing capability of an individual and is incurable by wearing spectacles [1]. In such cases, navigation in unknown locations is highly challenging for Visually Impaired Persons (VIPs) other than in their own house. Vision impairment occurs for multiple reasons, such as retinopathy, uncorrected refractive errors, glaucoma, age-related eye problems, diabetic trachoma, cataracts, unaddressed presbyopia and corneal opacity [2]. People use several aids for social inclusion, rehabilitation, education and work, excluding medical treatment. VIPs depend on their auditory perceptions and somatosensation—basically sound and braille—to obtain information from their environment. VIPs use assistive gadgets like canes to recognize their impediments [3]. However, nearly 28.22% of the global population are VIPs, whereas VIP-accessible facilities are not globally implemented. This leads to social discrimination problems since their activities have constraints. To be specific, VIPs cannot independently handle unpredictable circumstances outdoors. It limits their activities and movements indoors or in neighborhoods [4].

VIPs face important difficulties in their day-to-day activities: object detection and recognition, mobility/ navigation and safety, translation, identification of currency and textual information (signs, symbols) [5]. In literature, various methods, applications, systems and devices have been proposed, developed and validated in assistive technologies to facilitate VIPs in executing a task. These solutions usually consist of electronic gadgets equipped with microprocessors, cameras and sensors. These gadgets can take decisions and offer auditory or tactile feedback to the end-user, i.e., VIPs [6,7]. Though the existing object detection and recognition mechanisms are highly accurate, it lacks in presenting essential data and the associated attributes to track VIPs and ensure their safe movement. Since blind persons do not know the type of objects in their surroundings, such devices are highly helpful.

Moreover, a tracking system should be developed using which the VIPs' family members can monitor their movements [8]. Various authors have proposed numerous methods for developing assistive gadgets for VIPs. The following technologies are used in the prevailing gadgets such as low-energy Bluetooth beacons, vision-based sensors (camera), non-vision sensors (magnetic sensing, infrared, inertial and ultrasonic, etc.), and many more [9]. In the past, several research works attempted to address the following challenges: the nature and position of obstacles in a travelling path, understanding the environment, raising perception and easing VIPs in indoor and outdoor environments. Generally, advanced technologies have functions to choose routes and detect problems automatically [10]. Yet, there is a demand to develop and design intelligent systems that are helpful in object detection and recognition and enable the secure movement of VIPs.

The current study develops an IoT-assisted Transient Search Optimization with a Lightweight RetinaNet-based object detection (TSOLWR-ODVIP) model to help VIPs. For data acquisition, IoT devices are used in this study. Lightweight RetinaNet (LWR) model is applied to detect objects accurately. Next, the TSO algorithm is employed for fine-tuning the hyperparameters involved in the LWR model. Finally, the Long Short-Term Memory (LSTM) model is exploited for classifying objects. The performance of the proposed TSOLWR-ODVIP technique was validated using a set of objects, and the results were examined under distinct aspects.

2 Related Works

Liu et al. [11] modelled a fuzzy-enabled solution for vision challenges. This study leveraged a fuzzy-aided mechanism for the detection of targets that are poorly tracked with the help of response matrices of the samples. In poor tracking, the target tends to get relocated under the stored template. The solution was tested using an OTB100 data set, and the experimental outcomes reveal that the auxiliary solution

was effective for vision challenges. Alon et al. [12] utilized different gadgets for multiple purposes, such as a Pi camera as a capturing device, Raspberry Pi 4 as a microcontroller and a speaker for audio to declare an identified bill. EyeBill-PH was performed with an overall testing accuracy of 86.3%. Su et al. [13] formulated finger-worn gadgets—called Chinese FingerReader—that were realistically installed in VIPs to recognize conventional Chinese characters over micro Internet of Things (IoT) processors. The Chinese FingerReader, installed on the index finger, has small buttons and a camera. In this device, the small camera captures the images by finding the relative place of the index finger in the printed text. The buttons are implemented for VIPs to capture the images, whereas the audio output of the respective Chinese characters is provided to the VIPs in the form of a voice prompt.

In literature [14], the authors executed a new indoor object detector with the help of a Deep Convolutional Neural Network (CNN)-based structure. The structure was constructed based on a deep CNN called 'RetinaNet'. The proposed model was validated through several backbones such as VGGNet, ResNet, and DenseNet to enhance the processing time and detection performance. Bai et al. [15] introduced a wearable assistive device that empowers VIPs to navigate quickly and safely in strange ambiences and to recognize objects in outdoor and indoor environments. A lightweight CNN-related object recognition mechanism was formulated and deployed on the smartphone to increase the perception capability of VIPs and promote their navigation systems. Jiang et al. [16] innovatively used the image quality assessment method to select the images captured via vision sensors. This method could ensure the quality of the input scene for the final identification mechanism. Primarily, binocular vision sensors were used in this method to capture the images in a fixed frequency and choose the informative ones based on stereo image quality values. Then, the captured images were forwarded to the cloud for further computation processes. Specifically, the identification and automatic outcomes were obtained for all images from earlier stages. Then, big data-related CNN was used in this study.

3 The Proposed Object Detection Model

In this study, a new TSOLWR-ODVIP technique has been developed to detect and classify objects for VIPs. The major aim of the presented TSOLWR-ODVIP technique is to identify different objects surrounding the VIP and convey the information via an audio message.

3.1 Object Detection: LWR Model

In this study, the LWR model is applied to detect objects [17] accurately. RetinaNet structure encompasses Feature Pyramid Network (FPN), a detection backend and a backbone, as shown in Fig. 1. Initially, the images are managed using a backbone, which is generally a ResNet structure. It is important to note that though MobileNet efficiency is on par excellent with ResNet in terms of classification, it is not accurate enough to be observed as a corresponding replacement for ResNet. In this perspective, MobileNet is utilized as the backbone for detection tasks though it suffers from reduced accuracy in terms of classification outcomes. The primary purpose of MobileNet is that the confidence score of the MobileNet-based backbone can be decreased by trading off with low computation cost. As a result, it is not considered the desired choice as a backbone of a highly-accurate object detection network. Both backbone and the following FPN combine to form encoder and decoder networks. FPN has an advantage, i.e., it combines the characteristics of successive layers from the roughest to the finest level so that the features in different scales and levels are efficiently transmitted to the subsequent layer. Next, the multiple scale pyramid feature (P3-P7) feeds into the backend in which two detection subdivisions are applied for object classification and bounding box regression.
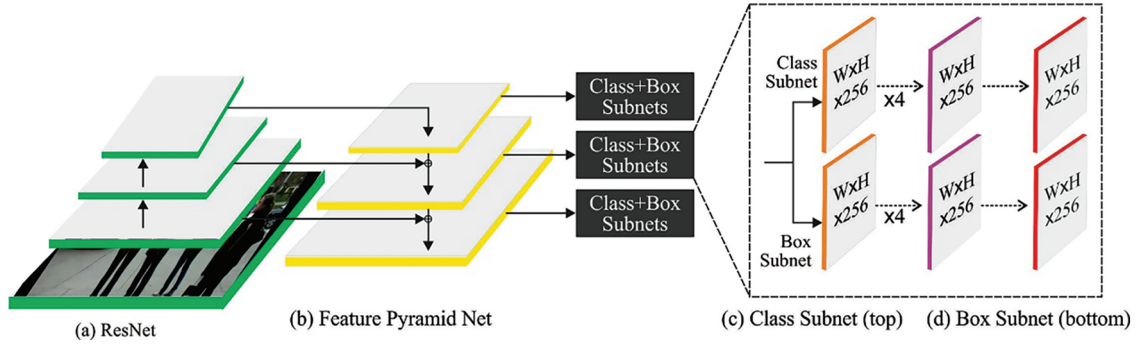
Figure 1: Structure of RetinaNet

Instinctively, the filter size is reduced to achieve FLOP reduction. Now, the D-blockv1 employs the MobileNet structure. A 1×1 convolution layer and a 3×3 depthwise (DW) convolutional layer replace one original layer. The D-block-v2 consecutively places 1×1 as well as 3×3 kernels. Based on YOLOv1, 3×3 kernels are replaced without presenting the residual blocks. In this study, the number of filters is fixed to a constant value across the layers. The D-block-v3 is highly aggressive because it substitutes each 3×3 convolution with a 1×1 convolution. In this scenario, the lightweight block experiences a decline in accuracy to trade off less computational cost. A lightweight detection block is a trade-off between low computation difficulty and a decline in accuracy. To indemnify the reduced accuracy, the fully-shared weight system is replaced with a partially-shared weight system in the new RetinaNet model. A partially-shared weight system primarily has two benefits. D3 has an independent weight parameter which learns additionally-tailored characteristics for the branch and can compensate for the reduced accuracy by low computation difficulty. For others, it allows the users not to touch the remaining network through heavy bottleneck blocks that are simply resolved.

3.2 Hyperparameter Tuning: TSO Algorithm

In this stage, the TSO algorithm is employed for fine-tuning the hyperparameters involved in the LWR model [18]. TSO procedure is modelled as follows; 1) initialize the search agents between the upper and lower limits of the search area; 2) identify the optimal solution (Exploration), and 3) obtain the optimal solution or steady-state (Exploitation) phase. At first, the initialization of the searching agent is arbitrarily produced. Then, the exploration behaviour of TSO is followed based on the oscillation of the second order RLC circuit nearby 0. But the exploitation of TSO occurs based on exponential decay of the initial order release. An arbitrary value r_1 is applied to create a balance between the exploitation ($r_1 < 0.5$) and exploration $(r_1 \geq 0.5)$ phases. The optimal solution (y_l^*) mimics the final value $(x())$ and steady state of the electric circuit as well $B_1 = B_2 = |Y_l - C_1 . Y_l^*|$.

$$y = lb + rand \times (ub - lb) \tag{1}$$

$$Y_{l+1} = \begin{cases} Y_l^* + (Y_l - C_1 Y_l^*)e^{-T} \\ Y_{l*} + e^{-T}[\cos(2\pi T) + \sin(2\pi T)]|Y_l - C_1 . Y_{l*}| \end{cases} \quad r_1 < 0.5 \tag{2}$$

$$r_1 \geqq 0.5 \tag{3}$$

$$T = 2 \times z \times r_2 - z \tag{4}$$

$$C_1 = k \times z \times r_3 + 1 \tag{5}$$

$$z = 2 - 2(^l/_{L_{ma}}) \tag{6}$$

In this expression, *lb* and *ub* indicate the lower and upper limits, respectively, *rand* denotes a uniformly-distributed arbitrary value, z denotes a parameter in the range of [2, 0], T and C_1 denote the arbitrary coefficients, respectively, r_1, r_2, and r_3 denote the uniformly-distributed arbitrary values $\varepsilon[0, 1]$, Y_l indicates the location of the searching agent, Y_l^* represents the optimal location, l shows the iteration value, k refers to a constant value ($k = 0,\ 1,\ 2$) and L_{max} denotes the maximal iteration count. Here, the balance between the exploration and exploitation phases is understood as the coefficient T that differs in the range of [−2, 2]. The exploitation procedure is accomplished if $\tau > 0$. On the other hand, the exploration procedure is accomplished as well. It is to be noted that the transient response is initiated by a higher response value which then dumps into a minimum value if $\tau > 0$. Afterwards, it oscillates again and moves towards the higher value if $\tau < 0$. The approach is not complicated and can be applied to update the location and create a balance between the exploitation and exploration phases.

Algorithm 1: Pseudocode of TSO algorithm

Initialization of the population and the optimal position $Y_l,\ Y_{l^*}$

Assess the cost function of the population

while $l < L_{max}$

    Upgrade the T and Values

    do every population y_l

        Upgrade the location of the population

    end do

    Evaluate the cost function of each novel population

    Upgrade the optimal value if the current cost function is lesser than the preceding optimal cost function

    $l = l + 1$

end while

output the optimal value Y_l^*

3.3 Object Classification: LSTM Model

Finally, the LSTM model is exploited for classifying objects [19]. Recurrent Neural Network (RNN) is a new Artificial Intelligence technique applied in real-time applications. A conventional RNN is generally employed to forecast the trained temporal dataset. However, it faces many challenges in terms of dealing with gradient explosion datasets. To resolve these issues, the LSTM method was developed. The LSTM method applies a memory function to substitute the hidden RNN unit. Fig. 2 shows the architecture of the LSTM mechanism which detects intrusions in IoT data. The LSTM technique has three major gates: output, forget, and input.
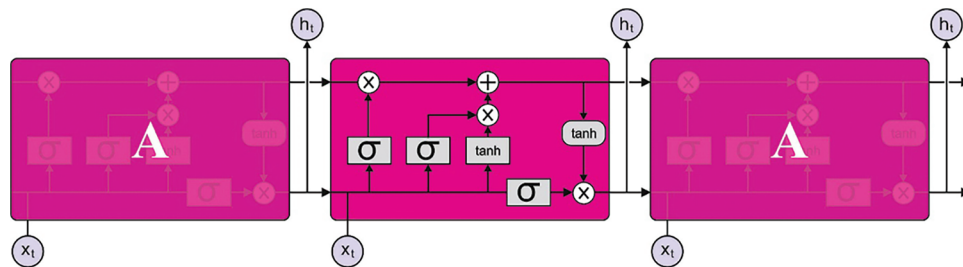
Figure 2: LSTM model

The forget gate identifies the forgotten datasets, whereas h_t refers to the input dataset, and the output gate lies in the range of [0, 1]. If the value is 1, the output is 'completely retained', whereas 0 denotes 'completely discarded'. The present state is characterized by c_t:

$$h_t = \text{sigma} \left(W_{xt} + Uh_{t-1} + b^{(h)} \right) \tag{7}$$

$$f_t = \text{sigma} \left(W^{(f)} + X_t + U^{(f)}h_{t-1} + b^{(f)} \right) \tag{8}$$

Here, h_t indicates the trained input dataset, whereas the input to the preceding cell is characterized by h_{t-1}. The forget gate is specified as f_t, whereas the crucial parameter of the LSTM is denoted by weight $W^{(f)}$ and $b^{(f)}$ denotes the bias. The data is updated at the input gate using two functions such as *sigma* and *tanh*. The sigma function is applied to detail the data required for upgradation, whereas the *tanh* function produces the data for the updation process.

$$i_t = \text{sigma} \left(W^{(i)} + X_t + U^{(i)}h_{t-1} + b^{(i)} \right),$$

$$m_t = \tanh \left(W^{(m)} + X_t + U^{(m)}h_{t-1} + b^{(m)} \right) \tag{9}$$

$$c_t = i_t \cdot m_t + f_t \cdot c_{t-1}.$$

The cell state c_{t-1} from the preceding cell is used for updating with the help of cell state c_t. The novel dataset should be discarded, whereas $f_t . c_{t-1}$ and $i_t . m_t$ is integrated to obtain the following cell state.

$$o_t = \text{sigma} \left(W^{(0)} + X_t + h_{t-1} + b^{(0)} \right) \tag{10}$$

$$h_t = 0_t \cdot \tanh (c_t) \tag{11}$$

In Eq. (11), 0_t denotes the output gate. In contrast, the weight vector of the Neural Network (NN) is signified by U and V. The *sigma* function is used to find the data that would be the output, whereas *tanh* is applied to present the cell state and declare the concluding output.

4 Experimental Validation

The current section experimentally validates the proposed TSOLWR-ODVIP model using a dataset that is composed of 1,600 samples. The dataset has a total of eight distinct classes. Table 1 provides the details regarding the dataset. The proposed model was simulated in Python 3.6.5 tool, whereas the PC configurations are as follows; i5-8600 k, GeForce 1050Ti 4 GB, 16 GB RAM, 250 GB SSD, and 1 TB HDD. The parameter settings are given herewith: learning rate: 0.01, dropout: 0.5, batch size: 5, epoch count: 50, and activation: ReLU.

Table 1: Dataset details

Labels	Class	No. of samples
C-1	Wrist watch	200
C-2	Dog	200
C-3	Stop sign	200
C-4	Person	200
C-5	Stairs	200
C-6	Chair	200

(Continued)

Table 1 (continued)

Labels	Class	No. of samples
C-7	Table	200
C-8	Washroom	200
Total number of samples		**1600**

Fig. 3 illustrates the confusion matrices offered by the TSOLWR-ODVIP model upon the entire Training (TR) and Testing (TS) datasets. The results demonstrate that the proposed TSOLWR-ODVIP model recognized all the class labels proficiently.

Table 2 provides the comprehensive object detection outcomes achieved by the proposed TSOLWR-ODVIP model. Fig. 4 portrays the results accomplished by the TSOLWR-ODVIP model on the entire dataset under distinct classes. TSOLWR-ODVIP model categorized C-1 samples with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, and Mathew Correlation Coefficient (MCC) values such as 99.62%, 97.09%, 100%, 98.52%, and 98.32% respectively. Meanwhile, the proposed TSOLWR-ODVIP method classified C-2 samples with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, and MCC values such as 99.19%, 93.90%, 100%, 96.85%, and 96.45% correspondingly. Ultimately, the presented TSOLWR-ODVIP technique categorized C-3 samples with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, and MCC values such as 99.25%, 99.47%, 94.50%, 96.92%, and 96.54% correspondingly.

Fig. 5 shows the outcomes accomplished by the proposed TSOLWR-ODVIP approach on dissimilar classes of 70% TR data. The presented TSOLWR-ODVIP system classified C-1 samples with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, and MCC values such as 99.55%, 96.43%, 100%, 98.18%, and 97.95% correspondingly. Meanwhile, the TSOLWR-ODVIP method categorized C-2 samples with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, and MCC values such as 99.11%, 93.42%, 100%, 96.60%, and 96.16% correspondingly. Eventually, the proposed TSOLWR-ODVIP algorithm classified C-3 samples with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, and MCC values such as 99.20%, 100%, 92.97%, 96.36%, and 95.99% correspondingly.

Fig. 6 reviews the outcomes achieved by TSOLWR-ODVIP algorithm on dissimilar classes of 30% TS data. The proposed TSOLWR-ODVIP approach categorized C-1 samples with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, and MCC values such as 99.79%, 98.48%, 100%, 99.24%, and 99.12% correspondingly. Meanwhile, the TSOLWR-ODVIP system classified C-2 samples with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, and MCC values such as 99.38%, 95.08%, 100%, 97.48%, and 97.16% correspondingly. Eventually, the proposed TSOLWR-ODVIP technique categorized C-3 samples with $accu_y$, $prec_n$, $reca_l$, $F1_{score}$, and MCC values such as 99.38%, 98.59%, 97.22%, 97.90%, and 97.54% correspondingly.

Fig. 7 provides the ROC curve examination results of the TSOLWR-ODVIP model. The figure demonstrates that the proposed TSOLWR-ODVIP model accomplished enhanced ROC values under all the classes.

At last, Table 3 provides an overview of the detailed comparative analysis results achieved by the proposed TSOLWR-ODVIP model and other existing DL models. Fig. 8 portrays the comparative analysis outcomes offered by the TSOLWR-ODVIP model and other DL models in terms of $accu_y$. The figure implies that AlexNet, VGG-16 and VGG-19 models reported the least $accu_y$ values such as 85.89%, 86.50%, and 84.43% respectively. At the same time, GoogleNet model achieved a slightly enhanced $accu_y$ of 88.77%. Next, SSD-MobileNet and YOLO-v3 models attained considerable outcomes

with $accu_y$ values such as 98.85% and 95.55%, respectively. But, the proposed TSOLWR-ODVIP model attained effectual outcomes with an increased $accu_y$ of 99.69%.
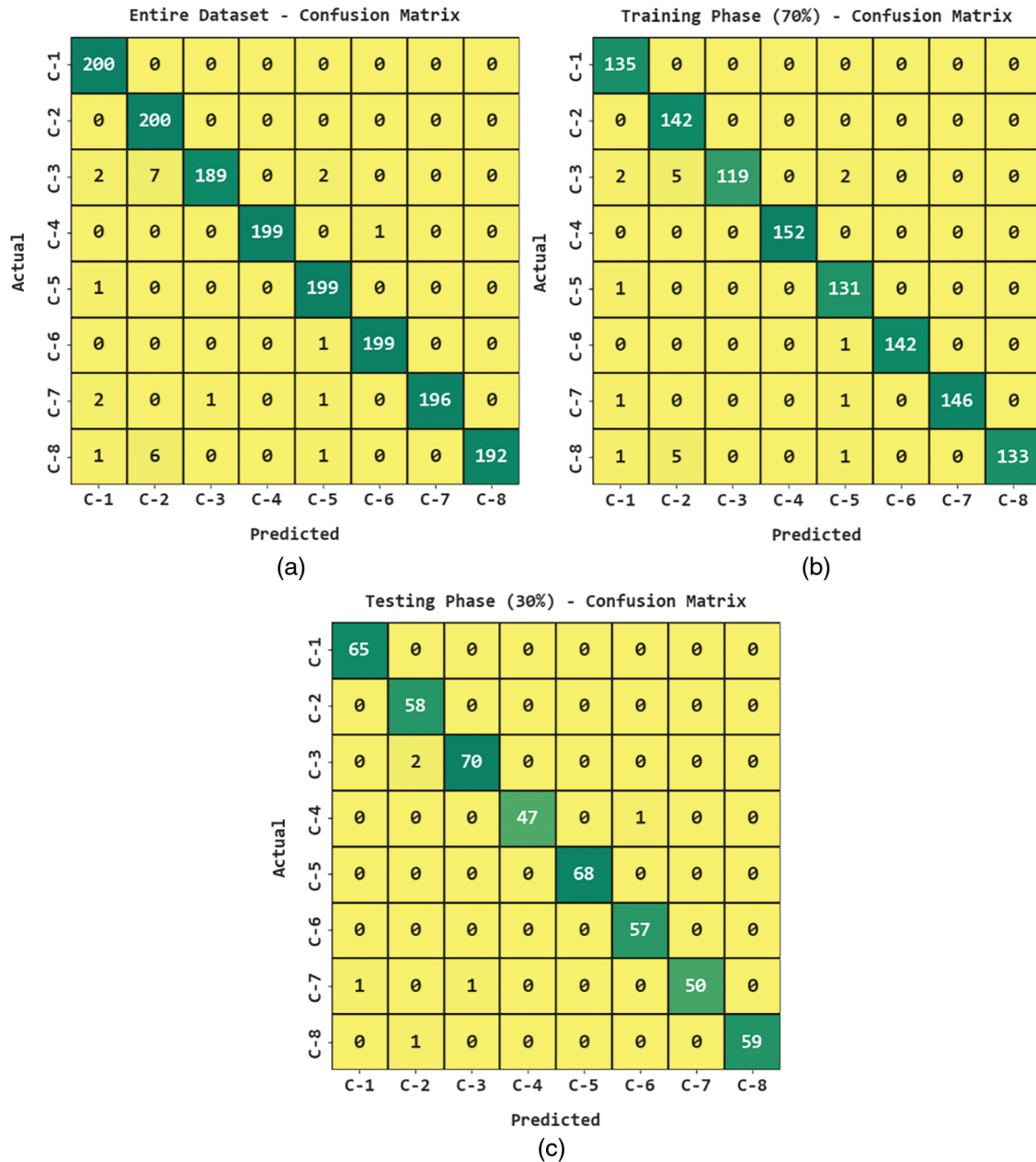
Figure 3: Confusion matrices of TSOLWR-ODVIP model (a) entire dataset, (b) 70% of TR data and (c) 30% of TS data

Table 2: Overall results of TSOLWR-ODVIP model

Labels	Accuracy	Precision	Recall	F1-Score	MCC
Entire dataset					
C-1	99.62	97.09	100.00	98.52	98.32
C-2	99.19	93.90	100.00	96.85	96.45
C-3	99.25	99.47	94.50	96.92	96.54
C-4	99.94	100.00	99.50	99.75	99.71
C-5	99.62	97.55	99.50	98.51	98.31
C-6	99.88	99.50	99.50	99.50	99.43
C-7	99.75	100.00	98.00	98.99	98.85
C-8	99.50	100.00	96.00	97.96	97.70
Average	**99.59**	**98.44**	**98.38**	**98.38**	**98.16**
Training phase (70%)					
C-1	99.55	96.43	100.00	98.18	97.95
C-2	99.11	93.42	100.00	96.60	96.16
C-3	99.20	100.00	92.97	96.36	95.99
C-4	100.00	100.00	100.00	100.00	100.00
C-5	99.46	96.32	99.24	97.76	97.47
C-6	99.91	100.00	99.30	99.65	99.60
C-7	99.82	100.00	98.65	99.32	99.22
C-8	99.38	100.00	95.00	97.44	97.12
Average	**99.55**	**98.27**	**98.15**	**98.16**	**97.94**
Testing phase (30%)					
C-1	99.79	98.48	100.00	99.24	99.12
C-2	99.38	95.08	100.00	97.48	97.16
C-3	99.38	98.59	97.22	97.90	97.54
C-4	99.79	100.00	97.92	98.95	98.84
C-5	100.00	100.00	100.00	100.00	100.00
C-6	99.79	98.28	100.00	99.13	99.02
C-7	99.58	100.00	96.15	98.04	97.83
C-8	99.79	100.00	98.33	99.16	99.05
Average	**99.69**	**98.80**	**98.70**	**98.74**	**98.57**

Fig. 9 shows the comparative analysis results of the proposed TSOLWR-ODVIP and other DL techniques in terms of $prec_n$. The figure infers that AlexNet, VGG-16, and VGG-19 approaches reported the least $prec_n$ values such as 86.54%, 85.70%, and 84.06%, correspondingly. Simultaneously, the GoogleNet algorithm accomplished a slightly improved outcome with a $prec_n$ of 86.80%. SSD-MobileNet and YOLO-v3 techniques accomplished significant outcomes with $prec_n$ values such as

98.33% and 95.16%, correspondingly. But, the proposed TSOLWR-ODVIP method obtained efficient results with an improved $prec_n$ of 98.80%.
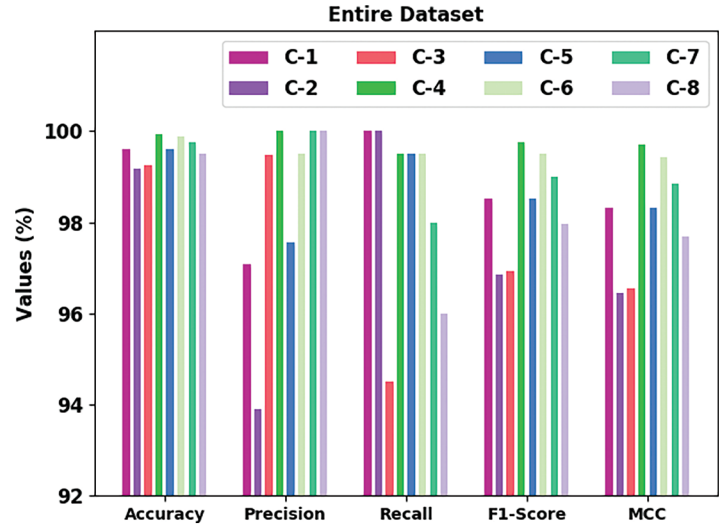

Figure 4: Overall results of the TSOLWR-ODVIP model on the entire dataset
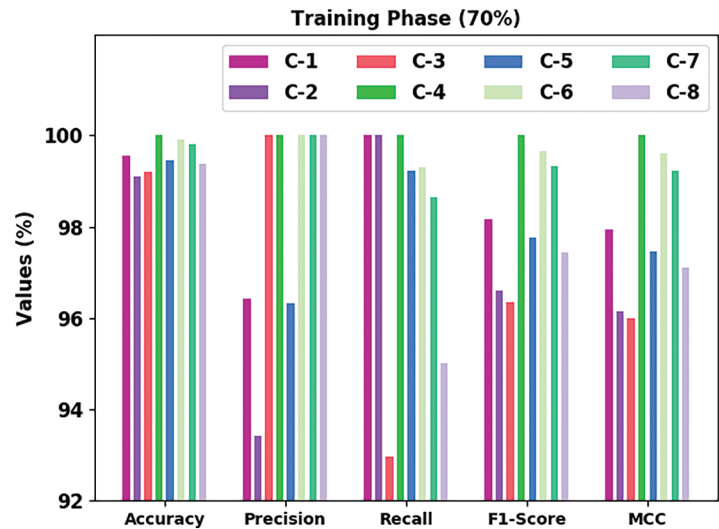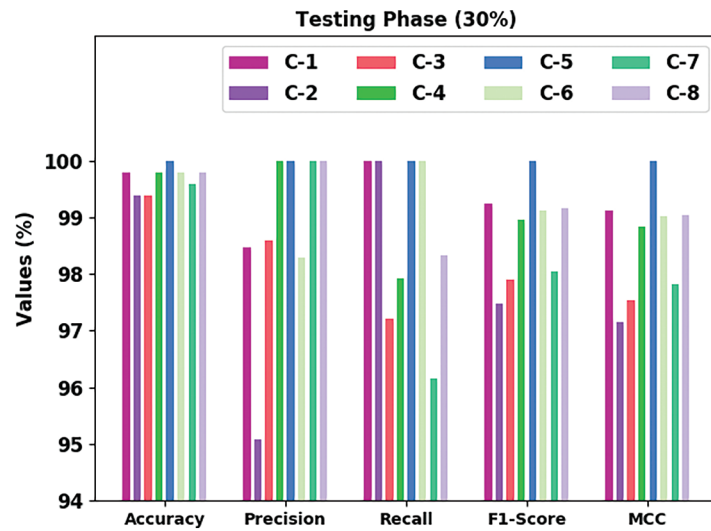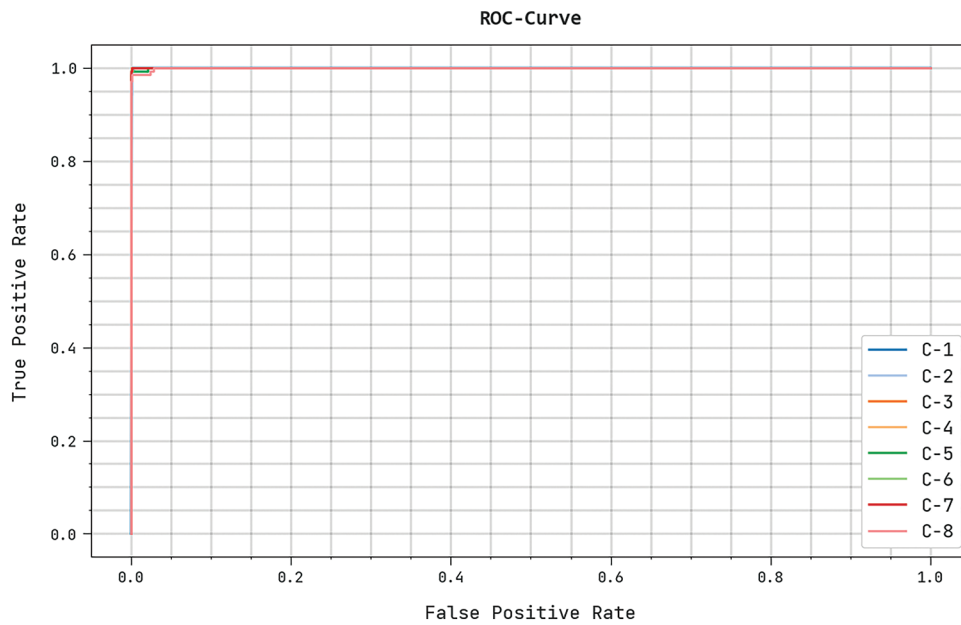

Figure 5: Overall results of TSOLWR-ODVIP model on 70% of TR dataset

Fig. 10 shows the comparative analysis outcomes of TSOLWR-ODVIP and other DL methods in terms of $reca_l$. The figure infers that AlexNet, VGG-16, and VGG-19 techniques reported the least $reca_l$ values such as 84.81%, 87.92%, and 87.33% correspondingly. On the other hand, GoogleNet approach achieved an improved outcome of 89.05% $reca_l$ Followed by, SSD-MobileNet and YOLO-v3 algorithms achieved substantial outcomes with $reca_l$ values being 98.17% and 95.77%, correspondingly. But, the proposed TSOLWR-ODVIP system attained efficient results with an improved $reca_l$ of 98.70%.

Figure 6: Overall results of the TSOLWR-ODVIP model on 30% of testing dataset

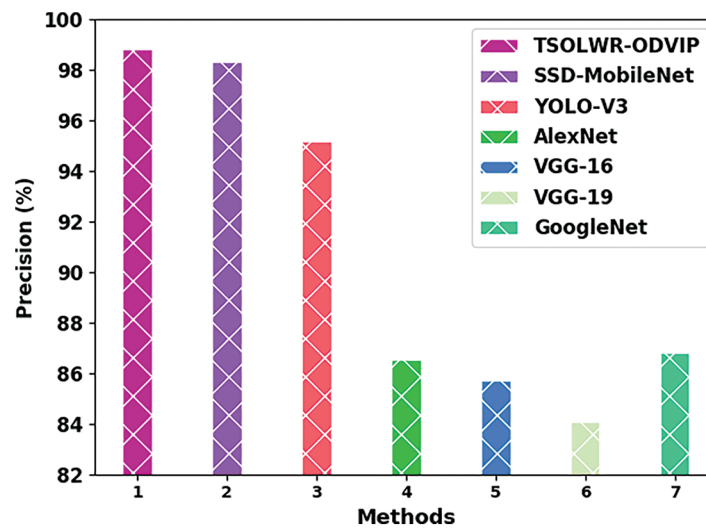**Figure 7:** ROC analysis results of the TSOLWR-ODVIP model

Table 3: Comparative analysis results of the TSOLWR-ODVIP model and other existing models

Methods	Accuracy	Precision	Recall	F1-Score
TSOLWR-ODVIP	99.69	98.80	98.70	98.74
SSD-MobileNet	98.85	98.33	98.17	98.38
YOLO-V3	95.55	95.16	95.77	95.80

(Continued)

Table 3 (continued)

Methods	Accuracy	Precision	Recall	F1-Score
AlexNet	85.89	86.54	84.81	89.53
VGG-16	86.50	85.70	87.92	89.11
VGG-19	84.43	84.06	87.33	87.22
GoogleNet	88.77	86.80	89.05	86.66
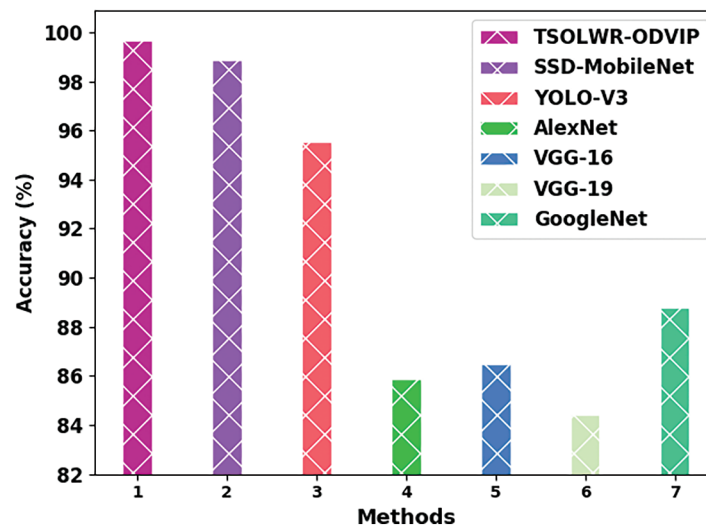
Figure 8: Comparative $accu_y$ examination outcomes of TSOLWR-ODVIP model

Figure 9: Comparative $prec_n$ examination results of the TSOLWR-ODVIP model and other models
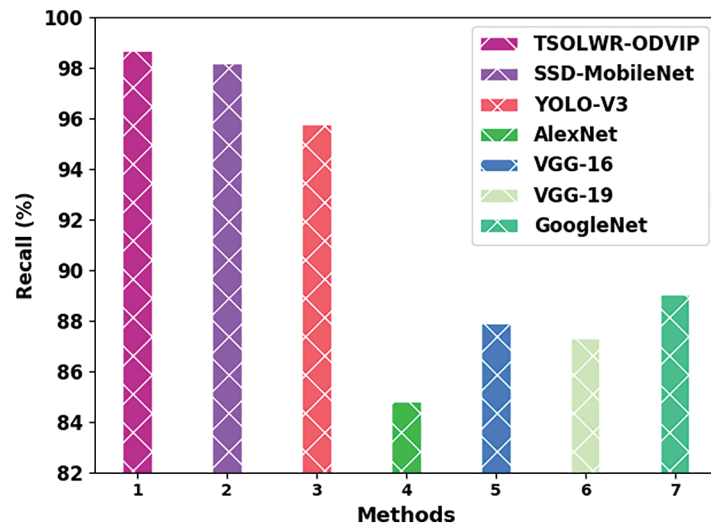
Figure 10: Comparative *reca$_l$* examination outcomes of TSOLWR-ODVIP model and other techniques

Fig. 11 portrays the comparative analysis results, in terms of $F1_{score}$ values, offered by TSOLWR-ODVIP and other DL approaches. The figure shows that AlexNet, VGG-16 and VGG-19 systems accomplished the least $F1_{score}$ values, such as 89.53%, 89.11% and 87.22% correspondingly. GoogleNet algorithm demonstrated a good $F1_{score}$ of 86.66%. SSD-MobileNet and YOLO-v3 techniques too accomplished significant results, with its $F1_{score}$ values being 98.38% and 95.80%, correspondingly. But, the proposed TSOLWR-ODVIP method achieved effective outcomes with a high $F1_{score}$ of 98.74%. Therefore, the experimental values confirm the improved outcomes of TSOLWR-ODVIP approach compared to other existing models.
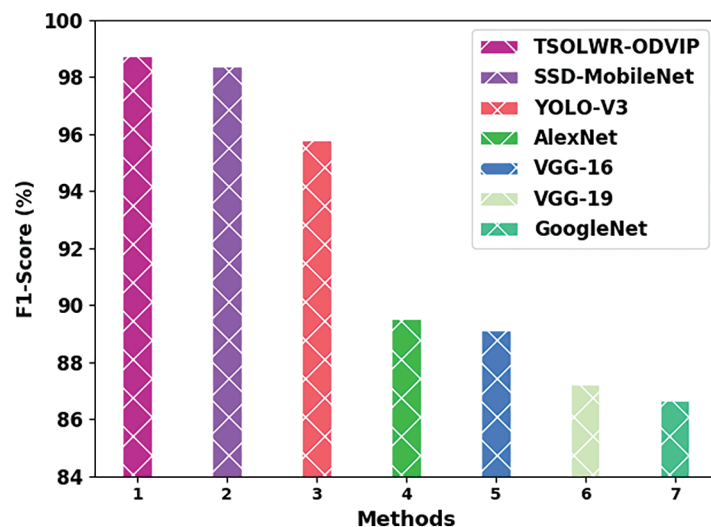
Figure 11: Comparative $F1_{score}$ examination results of TSOLWR-ODVIP model and other techniques

5 Conclusion

In this study, a new TSOLWR-ODVIP technique has been developed to detect and classify objects to help VIPs. The major aim of the presented TSOLWR-ODVIP technique is to identify different objects surrounding the VIPs and convey the information to them via audio messages. For data acquisition, IoT devices are used in this study. Followed by LWR model is applied for the accurate detection of the objects. Then, the TSO algorithm is employed for fine-tuning the hyperparameters involved in LWR model. At last, the LSTM model is exploited for the classification of objects. The performance of the proposed TSOLWR-ODVIP technique was validated using a set of objects and the results were examined under distinct aspects. The proposed model achieved a high accuracy of 99.69%. The comparison study outcomes confirmed that the TSOLWR-ODVIP model detects and effectually classifies the objects. So, it can be exploited to enhance the quality of life of VIPs. In future, hybrid DL models can be used to boost the detection efficacy of the TSOLWR-ODVIP technique.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] R. R. A. Bourne, S. R. Flaxman and T. Braithwaite, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: A systematic review and meta-analysis," *The Lancet Global Health*, vol. 5, no. 9, pp. 888–897, 2017.

[2] E. Cardillo, V. Di Mattia, G. Manfredi, P. Russo, A. De Leo *et al.* "An electromagnetic sensor prototype to assist visually impaired and blind people in autonomous walking," *IEEE Sensors Journal*, vol. 18, no. 6, pp. 2568–2576, 2018.

[3] O. Arslan and D. E. Koditschek, "Sensor-based reactive navigation in unknown convex sphere worlds," *The International Journal of Robotics Research*, vol. 38, no. 2–3, pp. 196–223, 2019.

[4] S. Caraiman, O. Zvoristeanu, A. Burlacu and P. Herghelegiu, "Stereo vision based sensory substitution for the visually impaired," *Sensors*, vol. 19, no. 12, pp. 2771, 2019.

[5] R. A. Z. Daou, J. Chehade, G. A. Haydar, A. Hayek, J. Boercsoek *et al.* "Design and implementation of smart shoes for blind and visually impaired people for more secure movements," in *32nd Int. Conf. on Microelectronics (ICM)*, Aqaba, Jordan, pp. 1–6, 2020.

[6] S. Shaikh, "Assistive object recognition system for visually impaired," *International Journal of Engineering and Technical Research*, vol. V9, no. 9, pp. 736–140, 2020.

[7] V. V. Meshram, K. Patil, V. A. Meshram and F. C. Shu, "An astute assistive device for mobility and object recognition for visually impaired people," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 5, pp. 449–460, 2019.

[8] M. M. Islam, M. S. Sadi, K. Z. Zamli and M. M. Ahmed, "Developing walking assistants for visually impaired people: A review," *IEEE Sensors Journal*, vol. 19, no. 8, pp. 2814–2828, 2019.

[9] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar *et al.* "A survey of modern deep learning based object detection models," *Digital Signal Processing*, vol. 126, pp. 103514, 2022.

[10] P. M. Kumar, U. Gandhi, R. Varatharajan, G. Manogaran and T. Vadivel "Intelligent face recognition and navigation system using neural learning for smart security in internet of things," *Cluster Computing*, vol. 22, no. S4, pp. 7733–7744, 2019.

[11] S. Liu, X. Liu, S. Wang and K. Muhammad, "Fuzzy-aided solution for out-of-view challenge in visual tracking under IoT-assisted complex environment," *Neural Computing and Applications*, vol. 33, no. 4, pp. 1055–1065, 2021.

[12] A. S. Alon, R. M. Dellosa, N. U. Pilueta, H. D. Grimaldo and E. T. Manansala, "EyeBill-PH: A machine vision of assistive philippine bill recognition device for visually impaired," in *11th IEEE Control and System Graduate Research Colloquium (ICSGRC)*, Shah Alam, Malaysia, pp. 312–317, 2020.

[13] Y. S. Su, C. H. Chou, Y. L. Chu and Z. Y. Yang, "A Finger-worn device for exploring Chinese printed text with using cnn algorithm on a micro iot processor," *IEEE Access*, vol. 7, pp. 116529–116, 2019.

[14] M. Afif, R. Ayachi, Y. Said, E. Pissaloux and M. Atri, "An evaluation of retinanet on indoor object detection for blind and visually impaired persons assistance navigation," *Neural Processing Letters*, vol. 51, no. 3, pp. 2265–2279, 2020.

[15] J. Bai, Z. Liu, Y. Lin, Y. Li and S. Lian, "Wearable travel aid for environment perception and navigation of visually impaired people," *Electronics*, vol. 8, no. 6, pp. 1–27, 2019.

[16] B. Jiang, J. Yang, Z. Lv and H. Song, "Wearable vision assistance system based on binocular sensors for visually impaired users," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1375–1383, 2019.

[17] Y. Li, A. Dua and F. Ren, "Lightweight retinanet for object detection on edge devices," in *IEEE 6th World Forum on Internet of Things (WF-IoT)*, New Orleans, LA, USA, pp. 1–6, 2020.

[18] M. H. Qais, H. M. Hasanien and S. Alghuwainem, "Transient search optimization: A new meta-heuristic optimization algorithm," *Applied Intelligence*, vol. 50, no. 11, pp. 3926–3941, 2020.

[19] F. Karim, S. Majumdar, H. Darabi and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1, 2018.