Research Article

Vision Navigator: A Smart and Intelligent Obstacle Recognition Model for Visually Impaired Users

Shubham Suman,[1] Sushruta Mishra [ID],[1] Kshira Sagar Sahoo [ID],[2] and Anand Nayyar [ID][3]

[1]*School of Computer Engineering, Kalinga Institute of Industrial Technology (Deemed to be University), Bhubaneswar, Odisha, India*
[2]*Department of Computer Science and Engineering, SRM University, Amravati, Andhra Pradesh, India*
[3]*Graduate School, Faculty of Information Technology, Duy Tan University, Da Nang, India*

Correspondence should be addressed to Anand Nayyar; anandnayyar@duytan.edu.vn

Vision impairment is a major challenge faced by humanity on a large scale throughout the world. Affected people find independently navigating and detecting obstacles extremely tedious. Thus, a potential solution for accurately detecting obstacles requires an integrated deployment of the Internet of Things and predictive analytics. This research introduces "Vision Navigator," a novel framework for assisting visually impaired users in obstacle analysis and tracking so that they can move independently. An intelligent stick named "Smart-fold Cane" and sensor-equipped shoes called "Smart-alert Walker" are the main constituents of our proposed model. For object detection and classification, the stick uses a single-shot detection (SSD) mechanism, which is followed by frame generation using the recurrent neural network (RNN) model. Smart-alert Walker is a lightweight shoe that acts as an emergency unit that notifies the user regarding the presence of any obstacle within a short distance range. This intelligent obstacle detection model using the SSD-RNN approach was deployed in real time and its performance was validated in indoor and outdoor environments. The SSD-RNN model computed an optimum accuracy of 95.06% and 87.68% indoors and outdoors, respectively. The model was also evaluated in the context of users' distance from obstacles. The proposed SSD-RNN model had an accuracy rate of 96.4% and 86.8% for close and distant obstacles, respectively, outperforming other models. Execution time for the SSD-RNN model was 4.82 s with the highest mean accuracy rate of 95.54% considering all common obstacles.

1. Introduction

In an extensive survey conducted worldwide, approximately 940 million people were detected to have a certain level of vision-related issues. Approximately 240 million among them suffered from extremely low vision, while approximately 39 million lacked complete vision [1]. Vision impairment, which restricts the ability to perceive, and is seldom curable, is a major global concern in recent times. Difficulties in normal movement, perceiving and detecting surrounding objects, and proper indoor and outdoor navigation are some of the basic complexities faced by affected people. Specifically, perceiving and identifying obstacles, and reacting to them in real time, is a challenge for them. These vision issues prevail mainly in underdeveloped nations because they are unable to afford the latest devices, which are often expensive. The problem is mainly dominant in old age people of densely populated underdeveloped countries [2]. Cataracts, refractive issues, glaucoma, retina problems, or age-based eye disorders are some common factors that lead to vision impairment [3, 4].

Navigation becomes a major concern for these people in unfamiliar environments. In addition to medical help, other services like awareness campaigns, regular rehabilitation programs, and social inclusion are used by these specially abled individuals. Many people use a white cane, where the cane's length is dependent on the touch sensation. However, its use is limited during traveling. Furthermore, its lack of flexibility results in cracks over regular use. In many scenarios, people use guide dogs to assist in

walking. These dogs alert the user of any potential obstacles in the way. However, these guide dogs may not always give accurate directions in crowded and complicated surroundings. GPS-enabled devices are also used by many people as assistive tools. These devices act as a navigation and orientation interface to locate the desired areas. Although they are effective in locating specific regions, they are not precise in avoiding and identifying obstacles. In many cases, an echolocation technique is used by visually disabled communities [5]. Here, sound echoes of mouth clicks are applied for the detection of obstacles in front of them. Short-distance range and leakage of data are the two main limitations of this approach. Other obstacle detection approaches, which are based on Quick Response codes and bar codes, are used to recognize various types of obstacles in crowded areas [6–8]. However, these approaches need a technologically advanced infrastructure and the support of a third person to accomplish tasks. Similarly, many devices exist to help users with visually disabled, but many of them are either used to detect obstacles through machine vision or used sensory modules like GPS and distance sensors. With recent advancements in science and technology, the lives of people with vision impairments can be improved and navigation can be made easier and more effective [9, 10]. Effective integration of sensor methodologies with computational vision and image processing can help in developing a real-time robust, cost-effective, and reliable model to assist users with vision impairments, thereby making them aware of any potential danger. The development of innovative modern technologies, such as the Internet of Things (IoT) and predictive analytics, has opened up possibilities for providing an interactive system to assist a person with vision concerns to independently navigate in all environments. The system or devices should be able to process data quickly, have a wide coverage area with enhanced detection of static and dynamic obstacles, and operate indoors and outdoors, depending on the needs of the user.

The major contributions of this article are as follows:

(1) This research analysis addresses this obstacle identification issue for users with major visual concerns and proposes a smart and intelligent obstacle recognition framework for them. In this research, a novel obstacle detection model named "Vision Navigator" is designed, which can assist users with vision concerns users to detect and recognize different types of obstacles in navigating indoors and outdoors.

(2) The integrated framework comprises an intelligent folded stick termed "Smart-fold Cane" and a pair of lightweight sneakers with two built-in ultrasonic sensors called "Smart-alert Walker." With the help of Smart-fold Cane, the distance between the user and obstacles can be conveniently computed, and water bodies on the way can be identified. In addition, the detected obstacle images can be captured and classified with notification to the user quickly.

(3) The second component, namely, Smart-alert Walker, is a pair of lightweight sneakers equipped with two ultrasonic sensors located at the front and left end of the sneakers. The sensors are emergency modules that determine the presence of any potential obstacle within a short range so that the user will be instantly informed about the danger to avoid any mishap.

(4) Smart-fold Cane and Smart-alert Walker combine to form the pillars of the proposed assistive obstacle detection model.

(5) The model was deployed in real time and its outcome was promising. The combined hybrid SSD-RNN model was successfully able to recognize, classify, and notify users about the presence of obstacles with excellent accuracy. Figure 1 depicts a skeleton model and the real-time usage of the proposed model.

This paper is organized as follows. Section 1 introduces the domain and addresses the importance of a new, efficient model for obstacle detection for people with vision concerns. Section 2 discusses the crucial background studies and related work by different researchers in the domain. Section 3 highlights the proposed working model in detail with the components and its functionalities discussed. Section 4 provides the results of deploying the model in a real-time scenario and analyzes the outcomes. Section 5 concludes and emphasizes the major inferences of the research.

2. Related Work

Detecting and carefully avoiding an obstacle at the right time is critical for visually impaired individuals. With the rise of modern technologies, a variety of working models have been developed and presented in this domain [11]. Many innovations have been proposed during the last two decades to help visually affected people effectively move around their surroundings. Therefore, numerous system models have appeared on the market. Some of them are wireless technologies. Visually impaired people face many challenges in their daily lives, but most of these challenges have been addressed through technological advancements [12]. In this section, extensive research is undertaken to describe the relevant work in the domain. The overall literature survey is partitioned into two sub-categories which are discussed below.

2.1. Sensors Based Existing Works. Some significant existing works based on usage of smart sensors are presented. The usage of modern technologies for assisting users with vision impairments has generated positive reviews like the use of wearables [13], electronic traveling support [14], mobile assisting tools [15], machine vision–enabled models [16], haptics usage and design [17], substituting sensory equipment [18], and electronic navigation help [19]. With model-based state-feedback control, a smart cane and obstacle detection device for visually impaired people with multiple sensors has been developed. In Ref. [20], a white cane system composed of IC tags is designed to help visually disabled
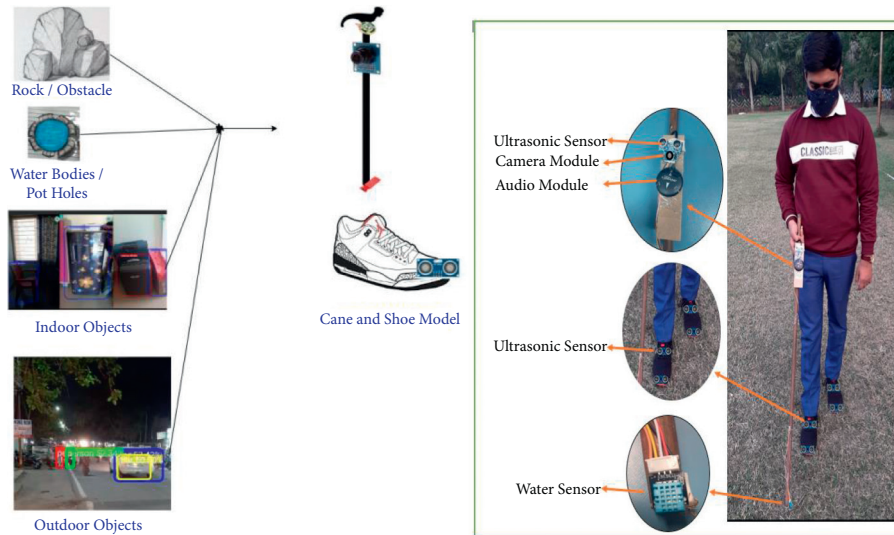
Figure 1: Skeleton prototype (left) and real deployment (right) of the devised model.

people walk comfortably in indoor areas. The authors of Ref. [21] proposed a system for determining TDOA using various signaling techniques. In Ref. [22], the authors introduced a nonlinear cost function for determining an object's indoor position by reducing the number of squares of a nonlinear cost function, such as least-squares algorithms. Other algorithms for calculating an object's indoor position include residual weighting and closest neighbor, which assess the location in comparison to reference points or base station coordinates [23]. Santos [24] proposed a module that uses a smartphone and an embedded device to communicate with public transit through wireless channels. Mutriara et al. [25] mentioned a tool that uses a GPS module to notify the position of a building that visually impaired users want to enter. Rehabilitative shoes and spectacles constituted another new method introduced by Abu-Faraj et al. [26]. The obstacles in front of users are sensed using ultrasonic transducers in this device. This method was used to determine the thickness of obstacles as well as the presence of any potholes in front of users. In 2018, Hu et al. proposed a method that relied on link models and assigned an equivalent amount of work based on its characteristics [27]. This method eliminates repetition and ensures precision to predetermined levels. It uses symbols instead of words because the objects are aligned in the 2D scale ratio. Al-Shehabi et al. [28] created a wearable navigation aid to help visually affected people reach their desired destination in a new environment. This module contains a Kinect chip, a tablet PC, a micro-controller, IMU, sensors, and vibration actuators for orienting the user in the next direction.

2.2. Computer Vision-Based Existing Works. In this subsection, some vital computer vision driven relevant works are discussed. In 2005, Chen and Yuille [29] published a cascaded model whose main purpose was to emphasize time complexities and their accuracy considering the various tests carried out by the greedy method; the model uses an algorithm that detects text from a cascade of images. In 2015, Wei et al. published a work in which the best output for image classification with a single label is achieved using a convolutional neural network (CNN) model [30]. In 2016, Zhang et al. presented a model that was used to detect trends in urban areas such as public streets and restaurants, as well as rainy environments [31]. This method categorized audio recordings, resulting in patterns. In 2015, Mekhalfi et al. [32] published a compressive sensing approach for vision-affected users. Here, objects were detected by grouping people using a camera in various indoor spaces. To determine the Euclidean distance and Gaussian method, this study used a multi-labeling strategy. It checks for the presence of a variety of artifacts related to the data collection. In Ref. [33], the authors provided a survey on electronic travel aids (ETA) for visually impaired navigation assistance. Various ETAs were discussed and compared in terms of their advantages and disadvantages. For activity recognition, a deep novel architecture for visually disabled people using a late combination of two parallel CNNs that outperform state-of-the-art methods was discussed [34]. Another approach proposed in Ref. [35] used CNN for object detection, followed by a recurrent neural network (RNN) and softmax classifier with intensity color thresholding for color recognition. In Ref. [36], researchers presented a visually impaired outdoor navigation assistant using a combination of machine vision and deep-learning techniques. The framework tracked objects without prior knowledge using a regression-based mechanism. It was able to handle sudden camera movements and uses you only look once (YOLO) for object recognition. In Ref. [37], a mobile app was designed to assist visually disabled people. It has two modes: offline and online based on the user's network access. Faster RCNN and YOLO are used in the online mode to produce predictions in stable conditions. In the offline mode, however, a feature recognition module based on Haar features and histogram of gradients serves this function. The ImageNet

dataset [38] has been used to develop a CNN for pretrained object recognition. Rajput et al. created a smart obstacle detector device to assist blind people in doing their jobs more conveniently and comfortably. A generic cane detection system was ineffective and precluded movement in blind people [39]. With the aid of a camera, the proposed device detected objects using a video processing method. This device used video processing to detect objects efficiently and quickly. Visual navigation aids were also developed by Thomas et al. [40]. To provide visual cues, this navigation aid was equipped with a wearable computer system with a see-through monitor, digital compass, and differential GPS.

2.3. Semantic Segmentation-Based Existing Works.
Pixel-by-pixel semantic segmentation is an effective method for detecting and identifying many classes of objects at the same time. There are various methods based on it used for assisting a visually impaired person. Deep learning pipelines have spurred the growth of semantic segmentation. FCNs which were proposed to convert CNNs, which were originally created for classification, to give pixel-wise classification outputs by making them completely convolutional, are a significant part of the literature. Another groundbreaking deep CNN architecture with a topologically symmetrical encoder-decoder design is SegNet. Instead of keeping all feature maps, SegNet up-samples the corresponding feature maps for the decoder using max-pooling indexes obtained from the encoder, drastically reducing memory and computational costs. ENet was offered as an efficient option for real-time semantic segmentation implementation. ENet was built with different bottleneck modules that may be utilized for either down-sampling or up-sampling images, based on ResNet views. ENet, unlike SegNet, has a larger encoder than its decoder because it is thought that the initial network layers should not immediately contribute to classification. ERFNet was created with the goal of increasing the accuracy/efficiency trade-off and making CNN-based segmentation appropriate for existing embedded hardware platforms. SQNet used parallel dilated convolutions and fused them as an element-wise sum to merge low-level knowledge from lower layers of the encoder, which helped with more precisely categorizing object outlines. By linking the encoder and the accompanying decoder, LinkNet attempted to obtain precise instance-level prediction without sacrificing processing time. In terms of pixel-exact categorization of tiny features, these architectures have outperformed ENet. PSPNet advocated using a decoder with max-pooling layers of various widths for large-scale scene parsing jobs in order to acquire various amounts of context in the last layers.

3. Materials and Methodology

At present, many people with vision-related concerns find challenges in navigation inside or outside of their homes. Thus, a portable and reliable smart device with intelligence functionality can assist them to independently navigate and move freely.

3.1. System Requirements.
Developing a smart and intelligent model for users with vision concerns requires systematic integration of various constituents. Table 1 highlights the vital hardware and software components used in configuration of the model.

Important constituents of the Vision Navigator include Arduino board, audio module, camera modules, water sensor, ultrasonic sensor, push button, and battery units. They are illustrated in a sample circuit diagram as shown in Figure 2.

3.2. Datasets Used.
The system will need some inputs to generate a model for the proposed system. We are using the following datasets.

3.2.1. MS COCO.
Microsoft Common Objects in Context (MS COCO) is an object detection dataset with 80 classes, 80,000 training images, and 40,000 validation images. The goal of this dataset is to determine the state of object recognition by placing different queries on object detection. This goal can be accomplished by gathering in a natural context various complex everyday scenes consisting of common objects from the environment. An object is labeled by optimization per instance to help locate objects precisely. The Single Shot MultiBox Detector (SSD) model uses this dataset for obstacle detection [41].

3.2.2. Flickr30k.
The Flickr30k dataset is a definitive source for the representation of phrase-based images. This work introduces Flickr30k Entities, which increases the 158k captions from which a single image is given different captions. The RNN model uses this dataset for suitable sentence framing from detected obstacle images [42].

3.3. SSD-RNN for Obstacle Recognition.
In this research, an SSD algorithm is used for detecting and classifying obstacles from the camera-captured images in the way of a person with vision concerns. Later, RNNs are used to generate an appropriate sentence for the detected obstacle by SSD, which is further communicated to users.

Using multibox [43], the SSD takes only one shot to detect multiple objects present in an image. The SSD has a substantially faster object detection algorithm with high accuracy. The SSD's high speed and precision when working with low-resolution images can be attributed to the following factors.

(i) Proposals of bounding boxes are no longer accepted.

(ii) For predicting object categories and offsets in bounding box locations, a progressively decreasing convolutional filter is used.

(iii) For object detection, multiple boxes or filters of various sizes and aspect ratios are used.

We add additional convolutional layers for detection to the base VGG network. The scale of the convolutional layers

TABLE 1: Hardware and software components of vision navigator.

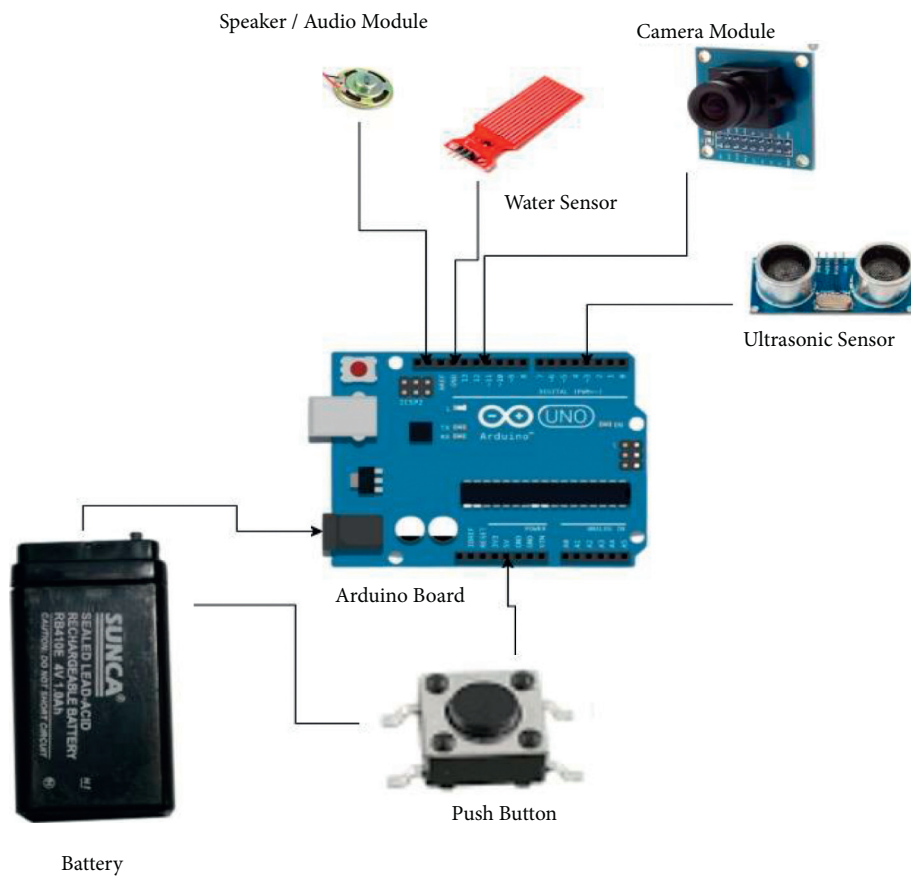| System Requirements	Description
Arduino board	Controls, processes, and generates all inputs and outputs. It receives the echo signals from the ultrasonic sensor that trigger it to take further actions and checks if the obstacle is there. It generates an immediate alert using a buzzer. It also generates a caption for the image captured by a camera and later converts that caption into speech that is played through an audio device.
Ultrasonic sensor	Determines the target obstacle distance by emission of ultrasonic sound waves, which maps the reflected sound into an electrical signal.
IR sensor	Detects infrared radiation that helps in sensing obstacles in surroundings.
Vibrator	Used to notify the user about any obstacle present.
Bluetooth module	Used to send and receive the signal through two devices in a wireless medium.
Push buttons	Used to switch on the microcontroller board.
Water sensor	Used to notify the user of any water body presence.
Audio module	Conveys image captions to the user in the form of audio. It receives an audio signal from Arduino once the caption for the image is converted into an audio format using a text-to-to-speech algorithm.
Camera modules	Acts as eyes for visually impaired people. Each time an ultrasonic sensor detects an obstacle, the camera modules capture the picture that is sent to the board for processing and caption generation.
Buzzers	Used in this system for an immediate alert. When the ultrasonic sensor detects an obstacle, it is triggered.
Plastic stick bodies	Acts as the outer texture body of the Smart-fold Cane.
Python	Programming interface to implement the model.
TensorFlow	Open-source machine learning platform.
Text-to-speech API	Application to map the obstacle text into speech to notify users.

FIGURE 2: Sample circuit diagram with vital constituents.

at the end of the base network decreases gradually, aiding in the detection of objects at a number of levels. Figure 3 shows a simple SSD model for obstacle detection.

3.3.1. Training of SSD. An input image with ground-truth bounding boxes for each object in the image is sent to the SSD. The function segmentation is based on the VGG-16
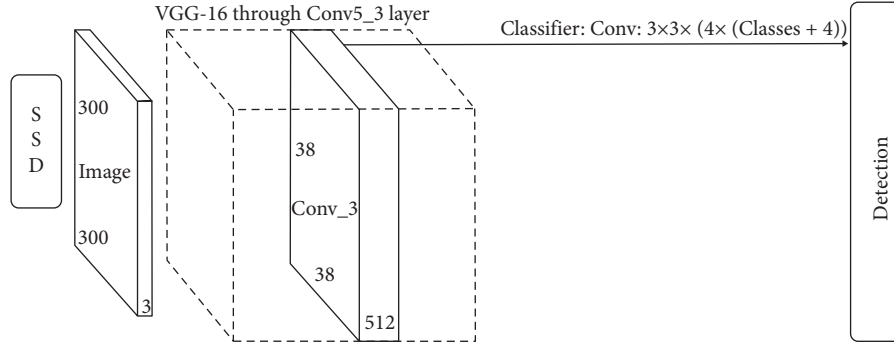
FIGURE 3: Illustration of an SSD model for obstacle detection.

base network. Convolution layers test boxes that comprise various aspect ratios that are present at each position in several feature maps in the form of various scales. Several default boxes of various sizes and aspect ratios are placed around the entire picture, aiding in evaluating the default box that most strongly matches the ground-truth bounding box containing objects.

3.3.2. Matching Strategy.

The default boxes are matched to the ground truth boxes during training in terms of aspect ratio, position, and size. The boxes that have the most overlap with the ground truth bounding boxes are chosen. The expected box and ground truth should have an intersection union greater than 0.5. The multibox chooses the expected box that has the most ground reality overlap. Each forecast is made up of the following parts:

(i) The offsets from the default box's middle section, as well as the height and width of the box.

(ii) All object types or classes have confidence. Class 0 is reserved for suggesting the object's absence.

3.3.3. Data Augmentation.

Shearing, zooming in and out, rotating, cropping, and other data augmentation techniques are used to manage a wide range of object sizes and shapes. Through data use, augmentation enhances the model's resilience to a wide range of input object sizes and shapes, aiding in improving the model's accuracy. The original opening input image is randomly sampled for each training sample. The operational step of an SSD model for obstacle detection as discussed is shown in Figure 4.

RNNs are a form of neural network in which the output from the previous step is used as input in the current step [44–47]. RNNs have a "memory" that holds all details about the calculation. They use the same parameters for each input because they produce the same output by observing the performance on all inputs or hidden layers. RNNs make decisions based on historical data.

A network with one input layer, three hidden layers, and one output layer, as shown in Figure 5, is considered. Each layer, like other neural networks, has its own collection of weights and biases, such as (w_1, b_1) for hidden layer 1, (w_2, b_2) for the second hidden layer, and (w_3, b_3) for the third hidden layer. By giving the same weights and biases to the

layers, RNN transforms independent activation into dependent activation, minimizing the complexity of increasing parameters and memorizing each previous output by feeding each output into the next hidden layer. As a result, these three layers can be combined into a single recurrent layer with the same weights and bias as the hidden layers.

The following is the formula for calculating the current state:

$$h_t = f\left(h_t - 1, X_t\right), \tag{1}$$

where h_t: current state, X_t: input state, $h_t - 1$: previous state.

The formula for applying the activation function is as follows:

$$h_t = \tanh\left(W_{hh}h_t - 1 + W_{xh}X_t\right), \tag{2}$$

where W_{hh}: recurrent neuron weight, W_{xh}: input neuron weight.

The formula for calculating the output is as follows:

$$y_t = W_{hy}h_t, \tag{3}$$

where, y_t: output and w_{hy}: weight at the output layer.

3.3.4. Training through RNN

(i) The network receives a single time phase of the input.

(ii) Using the current input and the previous state, the current state is determined.

(iii) For the next time stage, the current h_t becomes $h_t - 1$.

(iv) As per the problem, as many time steps as can be repeated as necessary and combine the data from all previous states.

(v) The final current state is used to measure the performance after all time steps that have been completed.

(vi) The error is then calculated by comparing the output with the real output, i.e., the target output.

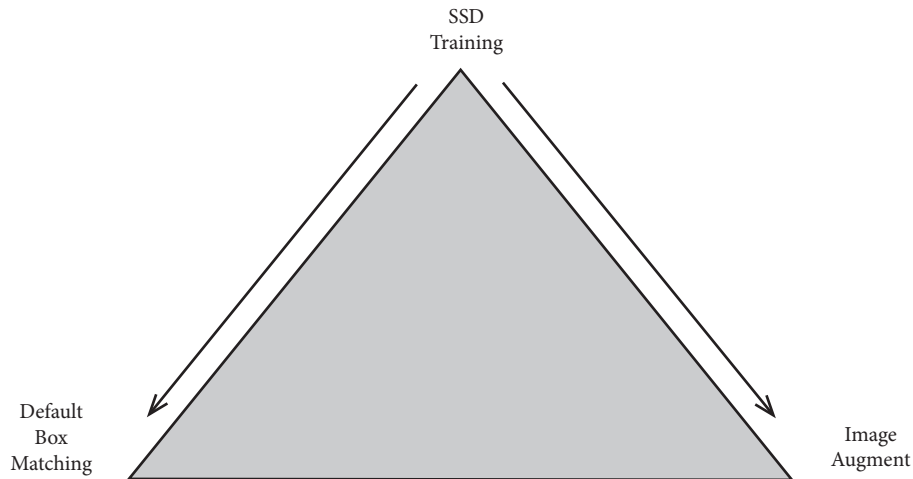(vii) The error is then back-propagated to the network, updating the weights and training the network (RNN).
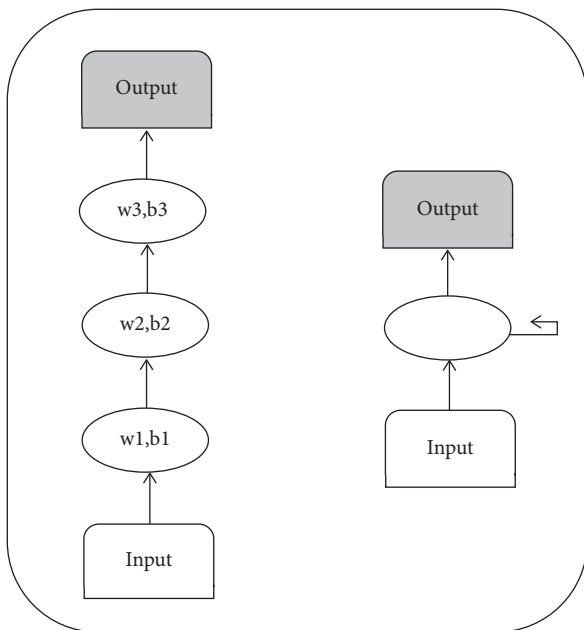
FIGURE 4: Functional steps of the SSD model.

FIGURE 5: Sample demonstration of the RNN model.

3.4. Proposed Model Workflow. The proposed Vision Navigator consists of a combination of a Smart-fold Cane with a Smart-alert Walker. The Smart-fold Cane is a stick comprising water sensors, camera modules, and an audio module. As opposed to a traditional stick, the Cane is a smart blind stick that will help people with vision concerns to an extent by allowing them to avoid obstacles while walking or going out, which may cause accidents. The stick is equipped with sensors and cameras to find the objects and give feedback alert messages to the user to avoid unnecessary accidents. The Smart-alert Walker has ultrasonic sensors mounted over it to find the obstacles in the path of the visually impaired person over a short-range distance. Thus, it acts as an emergency unit. The overall prototype of the device, which is laced with IoT sensors and enabled with predictive capabilities on an obstacle in front of the user, is

shown in Figure 6. The operation of the system is organized in such a manner that the person finds no difficulty in walking, reaching their destination.

The system model is activated when a visually impaired user uses it for navigation. An Arduino with an embedded Raspberry Pi camera is used as a single-board computer with Bluetooth facilities. It acts as the heart of the system because it controls, processes, and generates all inputs and outputs. The Smart-fold Cane has an ultrasonic sensor, water sensor, camera module, and audio module. The camera feed is taken live from the camera module. The camera module captures frames in real time and then the data are sent to the board where all frames are verified as per the model created to detect objects coming in front of the person. The SSD algorithm is used to detect potential obstacles such as animals, cars, and doors after training using the MS COCO dataset. The image feed is then validated against the trained deep learning model using the SSD model for object detection. Validated obstacles are transferred to RNN for sentence generation. The generated sentences interact with the Flickr30k dataset to frame appropriate sentences based on the matching image of the obstacle provided. The output of RNN is then forwarded to a text-to-speech application interface for vocal output. The audio module conveys the image captions to users in the form of audio alerts. It receives the audio signals from Arduino once the caption for the image is successfully converted into an audio format using a text-to-speech interface. This audio message reaches the user through an earpiece. An emergency alert provision is available using the Smart-alert Walker that is helpful in tracing any obstacle close to the user. Two ultrasonic sensors in Smart-alert Walker alerts for the presence of any obstacles that are too close, enhancing the accuracy of this system. These ultrasonic sensors are placed in such a way on the shoes that they can obtain data from obstacles that have height. The ultrasonic sensor feedback is then validated for its distance within 1 m and is sent to the Arduino board to verify whether the user will be collided with. Furthermore, water sensors are provided in the Smart-fold Cane. Various situations may arise in which the user may get trapped or face water bodies in their path if such bodies are not identified. The water sensor is able to sense any water body and
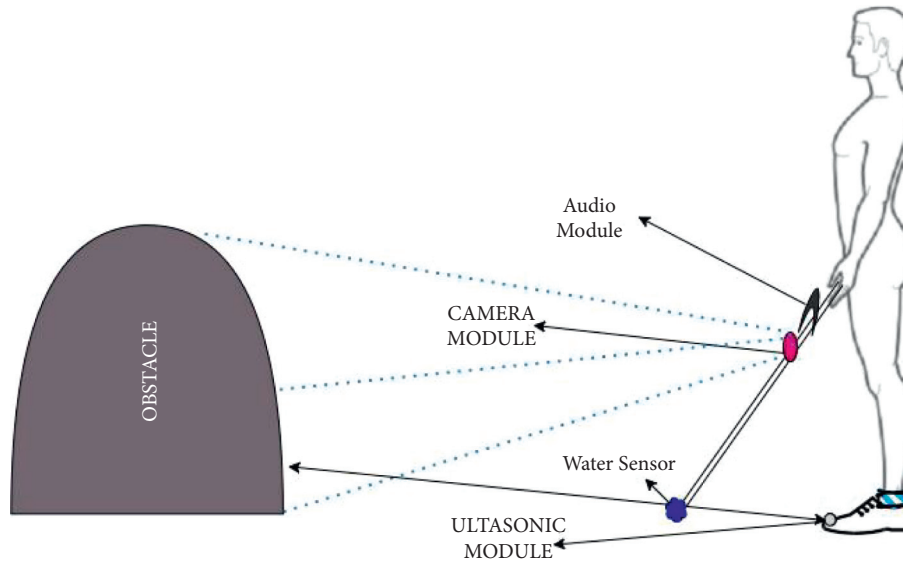
FIGURE 6: Prototyping of a user with Vision Navigator.

indicate its presence to the user. The feed is then sent to the Arduino board, which analyzes it and sends a command to the vibration and buzzer. Figure 7 presents the overall functionality of the model with different interlinked modules.

4. Results and Analysis

The proposed obstacle detection model Vision Navigator was designed as an assistive interface for users with vision concerns or individuals with fluctuating vision for precise navigation in indoor and outdoor surroundings. Apart from providing a dedicated sensory unit, the predictive capability of this model due to the integration of SSD and RNN is a distinguishing feature. The SSD technique was primarily used for object detection and RNN helped in mapping the detected obstacles with appropriate text generation. The effectiveness of the model in real-time scenarios was tested in indoor and outdoor environments. The combined SSD-RNN model used in the work was compared with existing models like Retina Net, Yolo Tiny, and Region-Based Convolutional Neural Networks (R-CNN).

Figure 8 shows the available obstacles in outdoor real-time roadside surroundings. The SSD model was able to successfully identify various moving and static obstacles in front of users. The developed detection model can observe cars and motorcycles present in the frame.

Figure 9 shows the available objects with labels in an indoor housing environment. The model detected a table fan, two different chairs, a suitcase, and a study table in front of the person, with their respective accuracy rates labeled.

An overall performance analysis was carried out with various common obstacles observed, as shown in Figure 10. Classification accuracy was the metric considered for evaluation. General obstacles were grouped into various distinct types: human, animal, vehicles, plastics, furniture, house, and others. Detection accuracy was done using the comparative models as discussed earlier. R-CNN and Yolo Tiny models generated an intermediate performance throughout the process with 92.98% and 91.35% mean accuracy, respectively. The SSD-RNN model computed the highest mean accuracy rate of 95.54% considering all obstacles.

An accuracy comparison analysis was undertaken using the SSD-RNN model with others in an indoor environment, as shown in Table 2. Various entities, including pets, humans, and furniture, were detected in the indoor scenario. Retina Net used 6 frames to produce a mean accuracy of 93.62%. Yolo Tiny gave 92.26% mean accuracy using 8 frames. R-CNN generated a mean accuracy of 91.48% taking 12 frames. The proposed SSD-RNN model computed the best accuracy of 95.06%. Overall, although Retina Net and Yolo Tiny models performed well on static obstacles like furniture, the SSD-RNN model was more efficient not only in detecting static obstacles but also in accurately detecting human beings.

Obstacle detection outdoors is more challenging because of the frequency of moving objects such as vehicles and human beings. Table 3 shows the results applied to different images in the outdoor environment. The Yolo Tiny model has a relatively low accuracy of 81.32% compared with other models. A very good accuracy rate of 86.64% was generated when implemented with the R-CNN model. Our SSD-RNN method gave an optimum accuracy of 87.68%.

On the basis of measuring the distance of obstacles, the proposed model is evaluated in terms of detection accuracy, as shown in Figure 11. Obstacles were categorized into "close" and "distant" types on the basis of the distance computed between the user and the obstacle. Obstacles within 1 m diameter were tagged as "close" and those beyond 1 m diameter were referred to as "distant." The analysis is applied to both detecting indoor and outdoor obstacles. As far as "close" obstacles are concerned, Retina Net and
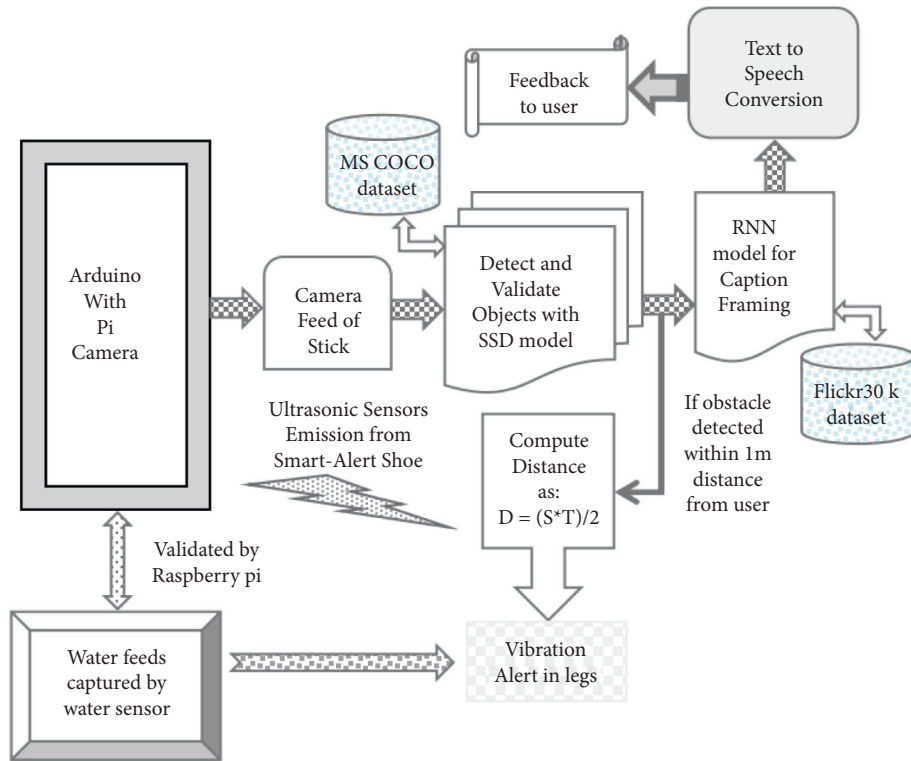
FIGURE 7: Functional units of Vision Navigator model.

FIGURE 8: Sample outdoor obstacle detection with the SSD model.

FIGURE 9: Sample indoor obstacle detection with the SSD model.

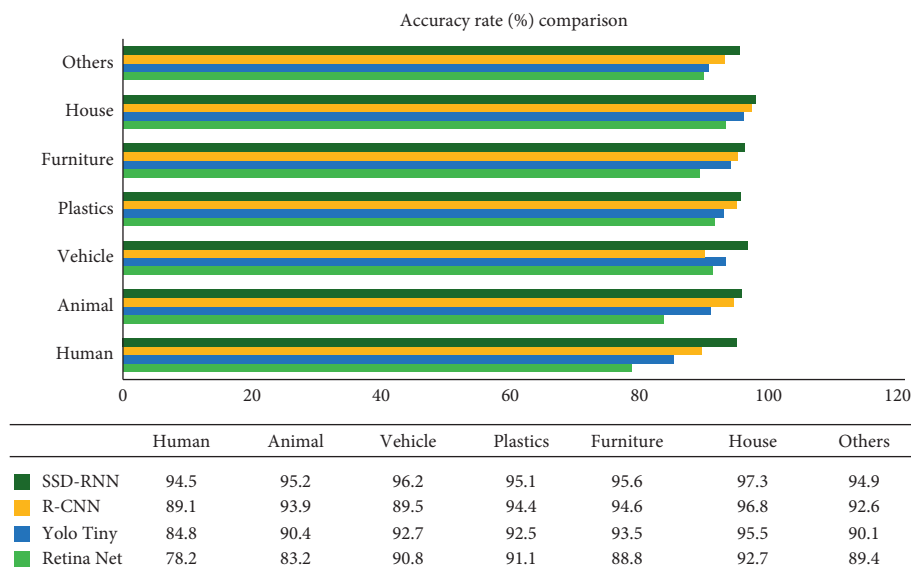|	Human	Animal	Vehicle	Plastics	Furniture	House	Others
SSD-RNN	94.5	95.2	96.2	95.1	95.6	97.3	94.9
R-CNN	89.1	93.9	89.5	94.4	94.6	96.8	92.6
Yolo Tiny	84.8	90.4	92.7	92.5	93.5	95.5	90.1
Retina Net	78.2	83.2	90.8	91.1	88.8	92.7	89.4

FIGURE 10: Detection accuracy comparison of different obstacles using different deep learning models.

R-CNN models gave good performance with accuracy rates of 94.5% and 93.8%, respectively. In terms of "distant" obstacle identification, a visible dip in the performance of all models was observed. Still, among all methods, Retina Net performed reasonably well with 83.9% accuracy. With an accuracy rate of 96.4% and 86.8%, respectively, in "close" and "distant" categories, the proposed SSD-RNN model outperformed the other models.

Latency delay is an important parameter to determine the efficiency of an obstacle detection model in real time. An implementation analysis was performed in this context using the discussed algorithmic models for comparison purposes, as depicted in Figure 12. Retina Net and R-CNN exhibited a consistent performance, but a slight delay in executing obstacle detecting was observed with Yolo Tiny. The SSD-RNN model was comparatively faster in providing the desired result. The recorded latency time for Retina Net, Yolo Tiny, R-CNN, and SSD-RNN models are 6.35, 8.54, 6.96, and 4.82 s, respectively.

A pilot study was conducted to test the effectiveness of the model in a real-time environment. The system was formulated at the time when the world was going through the COVID-19 outbreak. So, we met few people who agreed to test the system. The motto behind the development of this device was to bring a smile on the faces of such specially abled beings. The system application was tested over a group of specially abled people at School of Hope, which is a school for such kind of specially abled people. Authors recorded the testing process by validating it on 10 people and asked the feedback and recorded the confidence quotient of users. Table 4 highlights the metrics of evaluation and the generated outcome. As noted in Table 4, the confidence level of the majority of users is "High," which validates the effectiveness of the model.

TABLE 2: Accuracy rate analysis of observed obstacles with varying obstacle detection models in indoor settings.

Environment	Model	Obstacle	Accuracy (%)	Mean Accuracy (%)
Indoor	Retina Net	Cat	92.5	
		Human-1	93.2	
		Table-1	92.6	93.62
		Door	94.6	
		Chair-1	95.2	
Indoor	Yolo Tiny	Dog	91.7	
		Human-2	90.4	
		Sofa	92.2	92.26
		Door-2	94.1	
		Chair-2	92.9	
Indoor	R-CNN	Human-3	91.2	
		Bag	90.4	
		Dustbin	89.3	91.48
		Board	93.7	
		Table-2	92.8	
Indoor	SSD-RNN	Chair-1	95.6	
		Sofa	95.9	
		Human-4	93.9	95.06
		Clothes	95.1	
		Human-5	94.8	

TABLE 3: Accuracy rate analysis of observed obstacles with varying obstacle detection models in outdoor settings.

Environment	Model	Obstacle	Accuracy (%)	Mean accuracy (%)
Outdoor	Retina Net	Bus-1	82.1	
		Human-1	70.2	
		Car-1	91.6	84.76
		Tree	90.5	
		Cow	89.4	
Outdoor	Yolo Tiny	Car-2	83.6	
		Human-2	81.1	
		Human-3	82.2	81.32
		Truck	78.5	
		Dog	81.2	
Outdoor	R-CNN	Human-4	86.3	
		Cycle	90.2	
		Human-5	88.1	86.64
		Bus-2	88.5	
		Human-6	80.1	
Outdoor	SSD-RNN	Human-7	85.6	
		Car-3	87.2	
		Human-8	88.9	87.68
		Shop	90.1	
		Human-9	86.8	

Besides some classical models, the developed SSD-RNN model was also compared with some state-of-the-art models like FCOS, DETR, YOLO4, and EfficientDet. The accuracy and latency were computed and analyzed, as shown in Figure 13. It was found that the SSD-RNN model used in the research generated an optimum outcome with 95.5%
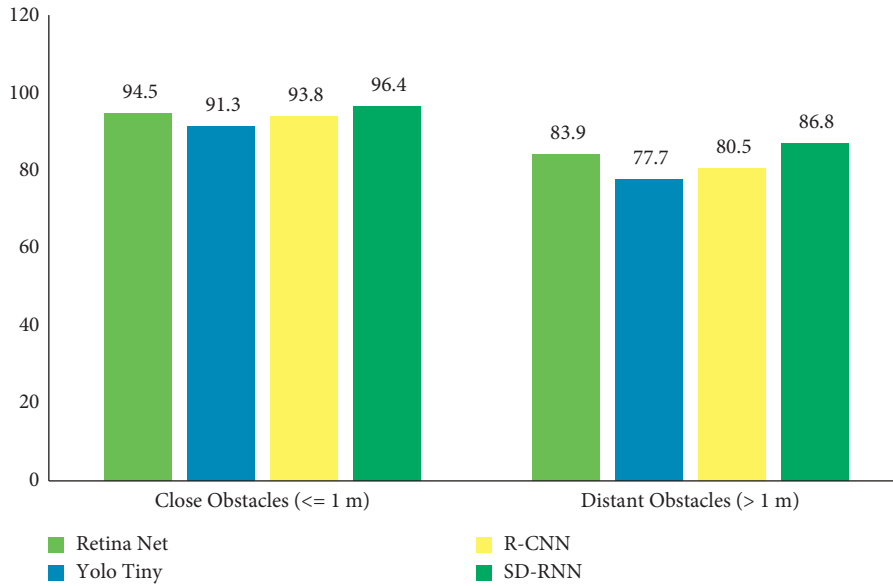
Figure 11: Detection accuracy comparison of close and distant obstacles using different obstacle detection models.
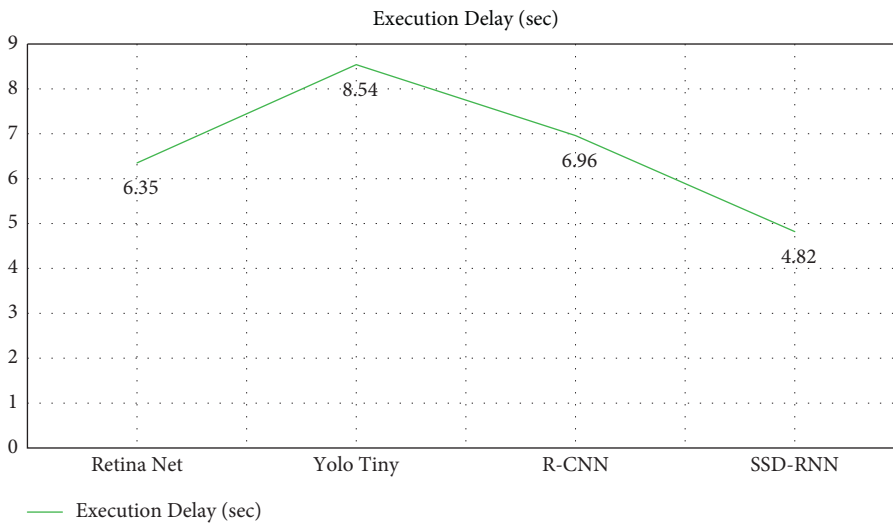

Figure 12: Execution latency delay comparison with different obstacle detection models.

Table 4: Confidence level analysis of Vision Navigator using pilot study.

User	Visually challenged degree	Intelligence level of Smart Fold Cane	Efficiency level of Smart Alert Walker	Accuracy level of SSD-RNN-based Object Detection	Overall Confidence level
P1	High	High	High	High	High
P2	High	Low	High	High	High
P3	Moderate	—	Low	High	High
P4	High	High	Moderate	Moderate	Moderate
P5	Moderate	High	Moderate	High	High
P6	Moderate	High	High	Low	High
P7	High	Moderate	High	High	High
P8	Low	High	High	Moderate	Moderate
P9	High	High	High	High	High
P10	High	Low	High	Moderate	Moderate

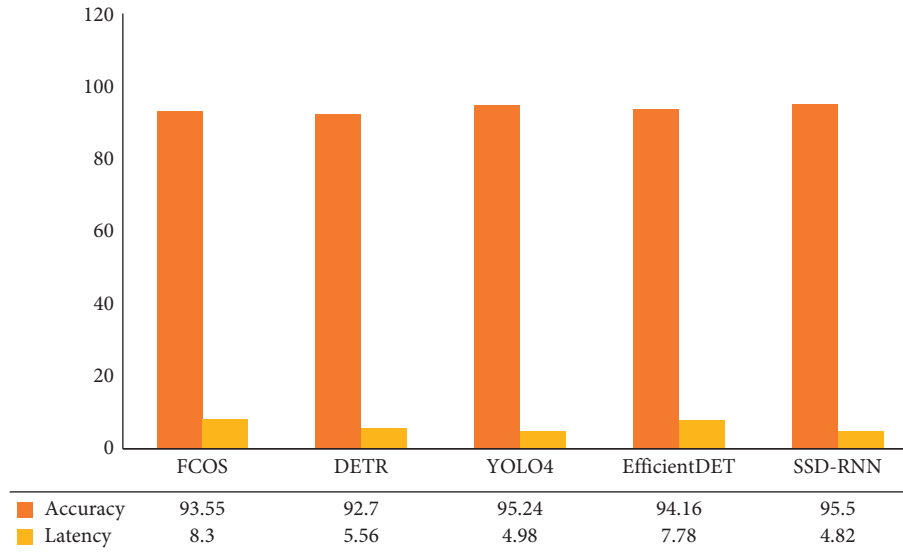|	FCOS	DETR	YOLO4	EfficientDET	SSD-RNN
■ Accuracy	93.55	92.7	95.24	94.16	95.5
■ Latency	8.3	5.56	4.98	7.78	4.82

FIGURE 13: Performance comparison of the proposed SSD-RNN model with other state-of-the-art approaches.

accuracy and a mean latency of 4.82 seconds. The performance of other counterparts was equally good but slightly less compared to the SSD-RNN model for obstacle detection.

5. Conclusion

Proper navigation is important for any individual with vision concerns for detecting and avoiding potential obstacles. With advancements in technology, hybrid models can be designed to serve the purpose. In this work, Vision Navigator, a smart framework with intelligence capability using obstacle detection, classification, and notification to the user in real time has been presented to assist the visually impaired community. Smart-fold Cane and Smart-alert Walker are the sub-constituents of the model. Smart-fold Cane is a lightweight stick that has in-built sensors and cameras that are responsible for obstacle image capture and detection. The SSD algorithm is used for obstacle recognition, while the RNN model maps the detected obstacle into the text form. Water pits are also detected through water sensors embedded into the stick. An alert notification is sent to the user through the audio module. The Smart-alert Walker, which is a pair of sneakers, is equipped with ultrasonic sensors that act as an emergency unit and alert the user if any obstacle is present at a close distance from the user. The SSD-RNN obstacle recognition hybrid framework was evaluated indoors and outdoors with other relevant models, namely, Retina Net, Yolo Tiny, and R-CNN. The SSD-RNN model gave an optimum performance, generating an accuracy of 95.06% and 87.68% indoors and outdoors, respectively. It also recorded an accuracy of 96.4% and 86.8% when used for close and distant obstacle detection, respectively. A minimum latency delay of 4.82 s was computed using the SSD-RNN model. An overall 95.54% accuracy with the model for common obstacles was noted. Thus, the designed model is fairly easy to deploy and use, making it a more generic framework compared with other models and well-equipped with all vital features. The proposed system can constructively serve visually affected users for proper navigation.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] WHO, *Universal Eye Health: A Global Action Plan 2014-2019*, WHO, Switzerland, 2014.

[2] "Global trends in the magnitude of blindness and visual impairment," 2020, https://www.who.int/blindness/causes/trends/en/.

[3] WHO, "Blindness and vision impairment," 2020, https://www.who.int/news-room/fact-sheets/detail/blindness-andvisual-impairment.

[4] L. Thaler, S. R. Arnott, and M. A. Goodale, "Neural correlates of natural human echolocation in early and late blind echolocation experts," *PLoS One*, vol. 6, no. 5, p. e20162, 2011.

[5] D. López-De-Ipiña, T. Lorido, and U. B. López, "Enabling accessible shopping for visually impaired people through mobile technologies," in *Proceedings of the 9th International Conference on Smart Homes and Health Telematics*, pp. 266–270, Montreal, QC, Canada, June 2011.

[6] S. Dambhare and A. Sakhare, "Smart stick for blind: obstacle detection, artificial vision and real-time assistance via GPS," in *Proceedings of the 2nd National Conference on Information and Communication Technology*, pp. 23-24, Chennai, India, December 2011.

[7] Z. O. Abu-Faraj, E. Jabbour, P. Ibrahim, and A. Ghaoui, "Design and development of a prototype rehabilitative shoes and spectacles for the blind," in *Proceedings of the 2012 5th International Conference on BioMedical Engineering and Informatics*, pp. 795–799, Chongqing, China, October 2012.

[8] P. Wawrzyniak and P. Korbel, "Wireless indoor positioning system for the visually impaired," in *Proceedings of the Federated Conference on Computer Science and Information Systems (FedCSIS)*, pp. 907–910, Krakow, Poland, September 2013.

[9] C. Ray, H. Kumar Tripathy, and S. Mishra, "Assessment of autistic disorder using machine learning approach," in *Proceedings of the International Conference on Intelligent Computing and Communication*, pp. 209–219, Springer, Bengaluru, India, June 2019.

[10] H. K. Tripathy, P. K. Mallick, and S. Mishra, "Application and evaluation of classification model to detect autistic spectrum disorders in children," *International Journal of Computer Applications in Technology*, vol. 65, no. 4, pp. 368–377, 2021.

[11] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "A stereo image processing system for visually impaired," *International Journal of Signal Processing*, vol. 2, pp. 136–145, 2006.

[12] J. Mercado, G. Chu, E. J. Imperial, K. G. Monje, R. M. Pabustan, and A. Silverio, "Smart cane: instrumentation of a quad cane with audio-feedback monitoring system for partial weight-bearing support," in *Proceedings of the 2014 IEEE International Symposium on Bioelectronics and Bioinformatics (ISBB)*, pp. 1–4, Chung Li, Taiwan, April 2014.

[13] A. Dutta, C. Misra, R. K. Barik, and S. Mishra, "Enhancing mist assisted cloud computing toward secure and scalable architecture for smart healthcare," in *Advances in Communication and Computational Technology*, G. Hura, A. Singh, and L. Siong Hoe, Eds., vol. 668, Singapore, Springer, 2021.

[14] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: a survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 1, pp. 25–35, 2010.

[15] S. Mishra, H. Thakkar, P. K. Mallick, P. Tiwari, and A. Alamri, "A sustainable IoHT based computationally intelligent healthcare monitoring system for lung cancer risk detection," *Sustainable Cities and Society*, vol. 73, p. 103079, 2021.

[16] J. R. Terven, J. Salas, and B. Raducanu, "New opportunities for computer vision-based assistive technology systems for the visually impaired," *Computer*, vol. 47, no. 4, pp. 52–58, 2014.

[17] V. Levesque, "Blindness, technology and haptics," Technical Report, pp. 19–21, McGill University, UK, 2005.

[18] S. Maidenbaum, S. Hanassy, S. Abboud et al., "The "EyeCane", a new electronic travel aid for the blind: technology, behavior & swift learning," *Restorative Neurology and Neuroscience*, vol. 32, no. 6, pp. 813–824, 2014b.

[19] U. R. Roentgen, G. J. Gelderblom, M. Soede, and L. P. de Witte, "Inventory of electronic mobility aids for persons with visual impairments: a literature review," *Journal of Visual Impairment & Blindness*, vol. 102, no. 11, pp. 702–724, 2008.

[20] D. Baldwin, "Wayfinding technology: a road map to the future," *Journal of Visual Impairment & Blindness*, vol. 97, no. 10, pp. 612–620, 2003.

[21] A. Pereira, N. Nunes, D. Vieira, N. Costa, H. Fernandes, and J. Barroso, "Blind Guide: an ultrasound sensor-based body area network for guiding blind people," *Procedia Computer Science*, vol. 67, pp. 403–408, 2015.

[22] M. C. Domingo, "An overview of the Internet of Things for people with disabilities," *Journal of Network and Computer Applications*, vol. 35, no. 2, pp. 584–596, 2012.

[23] M. Kanaan and K. Pahlavan, "A comparison of wireless geolocation algorithms in the indoor environment," in *Proceedings of the 2004 IEEE Wireless Communications and Networking Conference (IEEE Cat. No.04TH8733)*, vol. 1, pp. 177–182, Atlanta, GA, USA, March 2004.

[24] E. A. B. Santos, "Design of an interactive system for city bus transport and visually impaired people using wireless communication, smartphone and embedded system," in *Proceedings of the 2015 SBMO/IEEE MTT-S International Microwave and Optoelectronics Conference (IMOC)*, pp. 1–5, Porto de Galinhas, Brazil, November 2015.

[25] G. A. Mutiara, G. I. Hapsari, and R. Rijalul, "Smart guide extension for blind cane," in *Proceedings of the 2016 4th International Conference on Information and Communication Technology (ICoICT)*, pp. 1–6, Bandung, Indonesia, May 2016.

[26] Z. O. Abu-Faraj, E. Jabbour, P. Ibrahim, and A. Ghaoui, "Design and development of a prototype rehabilitative shoes and spectacles for the blind," in *Proceedings of the 2012 5th International Conference on BioMedical Engineering and Informatics*, p. 795, Chongqing, China, October 2012.

[27] H. Hu, J. Gu, Z. Zhang, J. Dai, and Y. Wei, "Relation networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3588–3597, Salt Lake City, UT, USA, June 2018.

[28] M. M. Al-Shehabi, M. Mir, A. M. Ali, and A. M. Ali, "An obstacle detection and guidance system for mobility of visually impaired in unfamiliar indoor environments," *International Journal of Computer and Electrical Engineering*, vol. 6, no. 4, pp. 337–341, 2014.

[29] X. Chen and A. L. Yuille, "A time-efficient cascade for real-time object detection: with applications for the visually impaired," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops*, p. 28, San Diego, CA, USA, September 2005.

[30] Y. Wei, W. Xia, M. Lin et al., "A flexible CNN framework for multi-label image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 9, pp. 1901–1907, 2015.

[31] L. Zhang, M. Towsey, J. Xie, J. Zhang, and P. Roe, "Using multi-label classification for acoustic pattern detection and assisting bird species surveys," *Applied Acoustics*, vol. 110, pp. 91–9810, 2016.

[32] M. L. Mekhalfi, F. Melgani, Y. Bazi, and N. Alajlan, "A compressive sensing approach to describe indoor scenes for blind people," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 7, pp. 1246–1257, 2014.

[33] R. Tapu, B. Mocanu, and E. Tapu, "A survey on wearable devices used to assist the visual impaired user navigation in outdoor environments," in *Proceedings of the 2014 11th International Symposium on Electronics and Telecommunications (ISETC)*, pp. 1–4, IEEE, Timisoara, Romania, November 2014.

[34] J. Monteiro, J. P. Aires, R. Granada, R. C. Barros, and F. Meneguzzi, "Virtual guide dog: an application to support visually-impaired people through deep convolutional neural networks," in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 2267–2274, IEEE, Anchorage, AK, USA, May 2017.

[35] R. Kumar and S. Meher, "A novel method for visually impaired using object recognition," in *Proceedings of the 2015 International Conference on Communications and Signal Processing (ICCSP)*, pp. 0772–0776, IEEE, Melmaruvathur, India, April 2015.

[36] R. Tapu, B. Mocanu, and T. Zaharia, "Seeing without sight-an automatic cognition system dedicated to blind and visually impaired people," in *Proceedings of the IEEE International*

Conference on Computer Vision Workshops*, pp. 1452–1459, Venice, Italy, October 2017.

[37] B.-S. Lin, C.-C. Lee, and P.-Y. Chiang, "Simple smartphone-based guiding system for visually impaired people," *Sensors*, vol. 17, no. 6, p. 1371, 2017.

[38] K. Potdar, C. D. Pai, and S. Akolkar, "A convolutional neural network based live object recognition system as blind aid," 2018, http://arXiv.org/abs/1811.10399.

[39] D. Daniyal, F. Ahmed, H. Ahmed, E. Z. A. Shaikh, and A. Shamshad, "Smart obstacle detector for blind person," *Journal of Biomedical Engineering and Medical Imaging*, vol. 1, no. 3, pp. 31–40, 2014.

[40] B. Thomas, V. Demczuk, W. Piekarski, D. Hepworth, and B. Gunther, "Wearable computer system with augmented reality to support terrestrial navigation," in *Proceedings of the 2nd Int. Symp. On Wearable Computers*, pp. 168–171, Pittsburgh, PA, USA, October 1998.

[41] T.-Yi Lin, M. Maire, S. Belongie et al., "Microsoft COCO: common objects in context," 2015, http://arXiv.org/abs/1405.0312v3.

[42] B. A. Plummer, L. Wang, C. M. Cervantes, C. C. Juan, H. Julia, and S. Lazebnik, "Flickr30k entities: collecting region-toPhrase correspondences for richer image-toSentence models," 2019, http://arXiv.org/abs/1505.04870v4.

[43] A. Arora, A. Grover, R. Chugh, and S. S. Reka, "Real time multi object detection for blind using single shot multibox detector," *Wireless Personal Communications*, vol. 107, no. 1, pp. 651–661, 2019.

[44] G. Ning, Z. Zhang, C. Huang, Z. He, X. Ren, and H. Wang, "Spatially supervised recurrent convolutional neural networks for visual object tracking," 2016, http://arXiv.org/abs/1607.05781.

[45] H. K. Tripathy, S. Mishra, S. Suman, A. Nayyar, and K. S. Sahoo, "Smart COVID-shield: an IoT driven reliable and automated prototype model for COVID-19 symptoms tracking," *Computing*, pp. 1–22, 2022.

[46] R. K. Rout, S. S. Hassan, S. Sheikh, S. Umer, K. S. Sahoo, and A. H. Gandomi, "Feature-extraction and analysis based on spatial distribution of amino acids for SARS-CoV-2 Protein sequences," *Computers in Biology and Medicine*, vol. 141, p. 105024, 2021.

[47] S. K. Sahu, D. P. Mohapatra, J. K. Rout, K. S. Sahoo, Q.-V. Pham, and N.-N. Dao, "A LSTM-FCNN based multi-class intrusion detection using scalable framework," *Computers & Electrical Engineering*, vol. 99, p. 107720, 2022.