International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

An IoT Machine Learning Approach for Visually Impaired People Walking Indoors and Outdoors

V. S. Saranya[1], Vijaya Krishna Sonthi[2], Dr. Prasanthi Boyapati[3], Boddu L. V. Siva Rama Krishna[4], Dr. Ganesh Naidu Ummadisetti[5], P. V. Naresh[6]

Submitted: 26/09/2023 **Revised**: 15/11/2023 **Accepted**: 27/11/2023

Abstract: This article describes the architecture and system design for assisting blind people in navigating freely inside an enclosed environment, such as the home or the outdoors. Thus, the proposed technology uses IoT technology and emerging techniques for machine learning to provide high-tech cane functionality that allows visually impaired navigators to walk independently. It also includes mobile applications to safeguard visually impaired persons and allow guardians to observe them. The proposed in this study system is intended to identify and classify any obstacles within a defined distance using machine learning. In this connection, an indoor and outdoor architecture on YOLO v3 is implemented for its detection technique, and multi-layer perceptron (MLP) neural network technology supports this framework. Based on the detection and classification, YOLO v3 and MLP are crucial for their accuracy.

Keywords: *Machine learning, Object Detection, Yolo, Visually Impaired People*

1. Introduction

According to the National Federation of the Blind (NFB) and the American Foundation for the Blind (AFB), the United States includes approximately 1.3 million blind persons, increasing the overall number of blind and visually impaired people to almost 10 million. There are around 100,000 students. Besides, about 160 million people worldwide have some form of visual issue, with 37 million being blind. Assistive aids were and will continue to be required. Blind or visually impaired users can choose from various navigation equipment and systems. An electronic vehicle (ETA) is a device [1] that converts environmental data from one sensory system to another. Typically, this data can be presented visually. Some solutions use a blind person's location, and orientation can be determined using Global Positioning System (GPS) technology [2].

However, these systems are appropriate for navigation systems due to direct availability to satellites, and additional components are required to increase resolution and proximity monitoring to prevent blind persons from crossing paths with other objects and endanger their existence in the process. Several scholars have suggested using robot dogs as, however, one feature: a combination. Furthermore, it incorporates technology like GPS to identify and avoid obstructions. These solutions are helpful. However, it can only be used outdoors, and misinterpreting requests by blind people or accuracy problems can seriously harm the user's health. Sensor technology constitutes one of the most critical aspects that may improve ETA reliability. A "smart cane" gadget is a type of ETA meant to be worn over a white cane to sense obstructions above the knee. This gadget is designed to assist visually impaired individuals in engaging in secure and effective independent travel by enhancing the availability of and access to specific categories of environmental information. Therefore, the accuracy of this sensor class is dependent on the signal strength and thus is affected by the reflectance and color of the object. An underwater sensor that aims at Infrared rays at obstacles and employs high-frequency sound waves rather than IR radiation is an example of an infrared (IR) sensor. The most generally used way to measure distances to objects is with laser rangefinders, which create laser wavelengths with the same objective [3]. Infrared sensor technology calculates the length by measuring signal strength. Since IR sensors offer a faster reaction time, a smaller range, and a more excellent resolution, they are better suited for sensing tiny distances [4, 5]. Visual aid technology is divided into three categories: visual improvement, replacement, and substitution [6-9]. In terms of functionality, the first two groups are almost identical. The

[1]*Assistant Professor, Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, Kattankulathur Campus, India*
[2]*Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.*
[3]*Assistant Professor, Department of Computer Science and Engineering, SRM University Andhra Pradesh,India.*
[4]*Assistant Professor, Department of Computer Science and Engineering, SRM University Andhra Pradesh,India.*
[5]*Department of CSBS, Assistant professor, B V Raju Institute Of Technology, Narsapur, Medak, Telangana, India.*
[6]*Assistant Professor, Department of Information Technology, Malla Reddy College of Engineering and Technology, Maisammaguda, Hyderabad, Telangana, India.*
** Corresponding Author Email: nareshgroup712@gmail.com, naresh.groups@gmail.com*

resulting image is analyzed and presented to a screening tool with visual magnification, and a visual replacement is performed using an alarm device capable of emitting vibration or voice with a small amount of information instead of sight. Vision replacement is more complicated than other surgeries because it involves medical technology. Based on visual aid technology, several walking aids have been created to solve the problem of blind people. White cane [10] is the only one that can supply location information and cannot determine the fastest way. The quickest way is indicated by guide dogs [11]. Guide-Cane [12] can sense the floor level ahead of users and difficulty moving sideways. Most walking helpers use obstacle detection and feedback signals [13,14].

These studies and approaches in blind or graphical impaired outdoor navigation frequently concentrate on determining distance and position. So, quite a technology is confined to knowing our crucial surroundings. This paper provides a unique approach using machine learning techniques to identify and distinguish distinct objects and barriers. Segmentation-based object identification uses pixels with segmentation color and class probability to find and recognize items. A single neural network evaluates the image to provide Segment predicate color for elements and class probability. As the detection pipeline uses a single neural network, you can directly optimize detection accuracy end-to-end. It continuously records the environment in front of the user and determines an object at an expected distance through object identification and image processing, giving the blind a verbal utterance of the object or obstacle in front. As the results show, visually impaired people can navigate better in the environment, and the proposed system will help them navigate unknown terrain and learn about potential hazards within the predicted distance. This paper will show a plan to allow blind or visually impaired people to navigate inside and outdoors independently.

2. Related Works and Background

If you are using Word, use either the Microsoft Equation Editor or the MathType add-on (http://www.mathtype.com) for equations in your paper (Insert | Object | Create New | Microsoft Equation or MathType Equation). “Float over text” should not be selected.

2.1. Equations

The scientific community has recommended various forms of assisted walking to help visually impaired people navigate and perform daily tasks. There have been many sensor-based systems suggested.

In [15], the author proposed manufacturing an Electronic Travel Aid (ETA) that is robust, reliable, and affordable. Blind and visually impaired persons use this ETA to navigate and distinguish indoor and outdoor objects. This ETA would be made up of a synergistic combination of ultrasonic sensors, a computer wand, and an object-detecting gadget. Four ultrasonic sensor nodes are used for navigation in an indoor environment. These nodes measure the closeness of surrounding objects, and if a thing is too close to another, the user is provided with vibrating feedback. The tip of the smart wand has a sensing function to dampen floors and stairs; if it does, the handle will begin to vibrate.

In [16], the author's main contribution is a review of the current state of vehicle design and an investigation of the following issues: (1) The significant design challenges that are presented by wearable travel aids, as well as the degrees to which various devices address these challenges; (2) Is there a connection between where and how you carry your travel gear and the design, features, and functionality of the travel accessory itself; (3) The limitations of currently available technologies, the absence of certain services, and the future paths of research, in particular as they relate to satisfying the requirements of prospective consumers.

In [17], the author demonstrated quick and secure electronic guidance for blind persons. Imagine an ultrasonic sensor-based obstacle detection system that automatically looks based on a USB camera. Sonar is used in the proposed method to identify impediments up to 300 centimeters away and gives feedback in the form of an audible signal to the user, informing them of the precise location of the obstruction. In addition, a USB webcam is coupled with the eBox 2300TM Embedded System to capture the user's field of view. This information is then utilized for determining the characteristics of the barrier and, more specifically, locating a human being in the context of this study. The identification of human presence is accomplished by recognizing faces and examining the textures of garments. These algorithms must be able to run on embedded systems despite significant limitations, the most important of which is a small image frame size (160 by 120 pixels) with a reduced number of faces, limited memory, and very little processing time available to meet real-time image processing requirements.

In [18], the authors provided an up-to-date and thorough summary of this study to provide developers with the tools needed to use the research's interdisciplinary nature. The approaches span the earliest "electronic travel aids" from early sensory substitutes or indoor/outdoor location research to more contemporary artificial vision technologies. Earlier methods would be concisely recounted and analyzed scientifically to achieve this purpose. After that, the concepts of user-centered design

are explained, and the key sources of criticism for previous methods are. In line with this, mobile phones and wearable with constructed cameras will be viewed as viable possibilities for enabling cutting-edge computer vision systems. This will allow for user placement and surveillance of the user's immediate community. Following that, these functions could be expanded further by utilizing distant services, which could lead to cloud services models and even environment monitoring through the usage of urban infrastructure. In [19], the authors suggested that electronic ones replace traditional travel aids, Helping 253 million blind people worldwide. Remarkably, most commercial products sold in today's market still operate on the same technology level as they were about 50 years ago. Advances in depth sensors and cameras could make a difference, even if there is competition in the industry.

This research intends to develop a dependable, automated, and accessible buddy, allowing people to traverse known and unexpected terrains. In [20], the authors have devised a method for the visually handicapped or blind properly navigate the premises. The whole algorithm makes use of information from the Xbox Game Kinect 360. The gadget generates a three-dimensional model of the inside environment, calculates depth, and determines the relative angle and distance to barriers or individuals. Kinect is equipped with a color camera to capture environmental details in real-time and then process them accordingly. This helps ensure the accuracy of the tool.

The technology employed in each investigation, the gadget used, the functionality supplied, and the obstacle detection algorithm are all listed in the literature review. As can be seen, the suggested methodology employs combined sensor-based and device vision-based techniques while using a single data processing unit. As a result, it can identify and recognize obstacles, beep, and detect falls. It also allows guardians to check the system remotely. As a result, the suggested solution integrates IoT technology with unique machine learning methodologies to equip blind navigators with innovative cane capability that allows them to walk freely.

3. Materials and Methods

In this paper, by machine learning, the proposed system is meant to detect all obstacles within a certain distance by defining them as difficulties that the client does not intend to avoid, such as stairs or doors. Apart from these capabilities, the system was meant to be simple, small, light, and real-time, with power consumption being one of the key criteria. The proposed approach comprises software and hardware parts that depict the system's physical architecture, as seen in Figure 1.

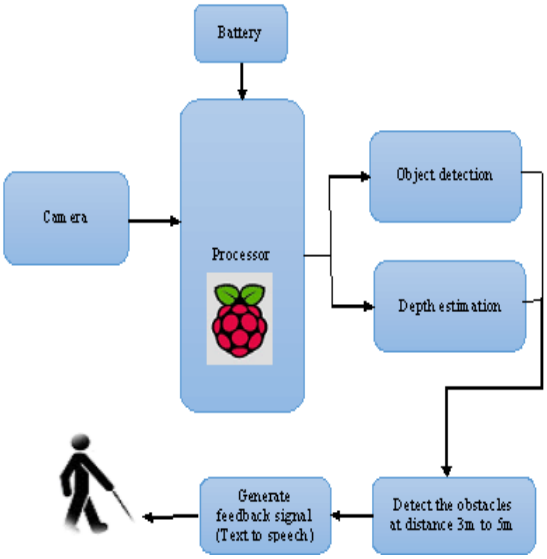


Fig 1. Proposed system framework

Cameras: The surrounding scene is captured using two high-quality 8-megapixel cameras. One camera connects through the camera's created serial interface, while the other connects via one of the peripheral computer device's USB ports.

Sound System: Produces an audible alert alerting cane users to obstacles ahead.

Dataset: This study used a freely available dataset to the public. There are 150 images in the data collection. These images are classified into two classes, indoor and outdoor: door (50 images), hollow pits (50), falling stairs (10), and upstairs (10). The other images (approximately 30) are from the author's neighborhood, classified with the same classifications indicated earlier (doors, stairs, and hollow pits), and included in the dataset.

Object Detection: With devices and minimum resource consumption, the suggested technique should be capable of recognizing an extensive collection of items and classifying them into specified item categories on the layers by each image. Furthermore, the job of sufficiently diverse items inside images typically entails giving bounding boxes and names for each object. This job differs from the segmentation and localization tasks in that it applies classification and localization to many objects rather than just a single dominating object. To address this, scholars have proposed several designs and frameworks. DeepLab [21], Fast R-CNN [22], and YOLO [23] are network topologies used in modern object detectors. For the YOLO approach, the image is divided into a GG grid. If the center of an item falls within a grid cell, that object will be detected. Each grid cell forecasts various sizes and forms of B-bounding units. The bounding box with the greatest IoU will be allocated to a detected object (Intersection over Union). A confidence level is also provided for each bounding box. (Union vs. Intersection). Every bounding box contains five descriptors: the

bounding box center (bx, by), this same box height (bh), the box widths (bw), and the boundary box level of confidence (pc), which is the likelihood that the element is located within the box. Besides that, each bounding box has C chances for all the classes, one for each observed object class. Every grid cell has B(5+C) descriptors, such as frames. To obtain bx, by, bw, and bh out from the output of the system, where tx, ty, tw, and th are the networks outputs, cx, and cy are really the grid's top-left positions, and pw and ph are indeed the box's anchors sizes. Because we predict center coordinates with a sigmoid, the outcome value will be between 0 and 1. When an object is detected, the camera also displays the object's relative position. The four predicted items after the grid cell's upper left corner are normalized to the dimensions of the functioning map cell. Bounding box size can also extract elements to the right, left, or in front of us.

Figure 2 depicts a bounding box prediction example. The image split into 1313 grids, each with three border blocks predicted. For 30 classes, the feature height is 3 (5+30) and 255. Because each bounding box has a B frame, the complete image has a GGB bounding box. Several fields contain low-probability suggestions that can be promptly removed.

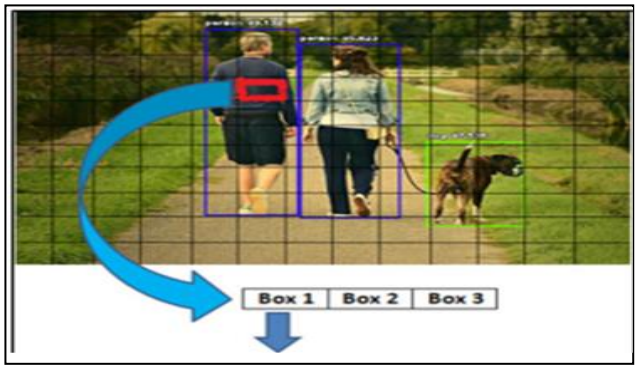


Fig 2. A border box prediction example

The remaining fields are addressed to "non-maximum suppression," eliminating duplicate elements while retaining the most accurate. YOLO v3 predicts on three different scales. The detection layer identifies elements in convolution layers of three sizes, rising by 32, 16, and 8 strides respectively. After the network detects the layers' feature maps in 32 strides, it samples the input image until it approaches the first detection layers. The surface is then upsampled by a factor of two and blended with the last layer's feature map using the relevant map—size object. Now another detection is done in increments of 16 per layer. The same up-sampling process is used again, and the final detection is done in 8-layer steps. As a result, YOLOv3 predicts more blocks than YOLOv2 for the same input image size.

Feature Extraction: The feature shape of the item has been extracted. For this, we used the Histogram Function Descriptor of Oriented Gradients (HOG)

Obstacle classification:

The procedure for classifying obstacles is divided into two parts: (i) preparing the dataset for training and testing and (ii) generating the classifier. It uses a deep learning approach to classify obstacles among available machine learning methods. The primary purpose of this stage is to develop a classifier that gives the most efficient obstacle classification given the abovementioned constraints. A classifier operates to evaluate the kind of barrier found in the image by using the visual features collected from the image. Second, we develop an indoor and outdoor training dataset (many labeled obstacle images for each class). Second, we use the training data to identify the most attractive features for obstacle detection and create an n-dimensional table. Finally, we choose the best deep-learning strategy for building the classifier.

Testing dataset: We create an important and representative corpus with examples that can reveal our skills. Images include all parameters of the environment (interior and exterior), ground (reflections, shadows, textures, and color contrast), and obstacle situations (people, vehicles, etc.). It should be noted that the collected images are indoors and outdoors. Divide all images obtained from the image dataset into three classes: people, vehicles, and others. The obtained data sets are utilized to build groups for training and testing, split randomly among training (80%) & testing (20%).

For every image in the datasets used for training and testing, feature vectors are created. The training set contains 11394 images, comprising 2006 images of humans, 1188 images of automobiles, and 8200 images. The test set contains 4623 images, including 708 images of humans, 405 images of automobiles, and 3510 various images.

Classifier generation: We create, test, and analyze three machine learning methods to determine the best deep learning technique for generating classifiers to evaluate the suggested strategy's efficacy. Examples of these approaches include multilevel perceptrons (MLPs), decision trees (DTs), nave Bayesian (NBs), and support vector machines (SVMs). The Accuracy Ratio (AR) and Percent Correct Classification (PCC) are utilized to evaluate the experimental outcomes of obstacle classification techniques.

Decision Tree (DT): A decision tree is a structure similar to a flowchart, where each inner node represents an attribute test, each branch represents a test result, and leaf nodes represent a class or class distribution. The ID3 method is a simple decision tree approach. Information acquisition is used as a separation criterion. ID3 evolved

into C4.5. They depend on the prize money. The CART algorithm uses the Gini coefficient as the test feature selection criterion. The characteristic with the most negligible Gini coefficient is chosen to illustrate the fact for each collection. The benefit of data classification is that it's simple to comprehend and evaluate.

Naïve Bayes (NB): Based on a probability model, the NB classifier gives the classification with the probability model to the feature set derived from the ROI. The posterior probabilities of a given class given a feature representation v are computed using Bayes' theorem v:

$$P(c_i|\vec{v}) = \frac{P(\vec{v}|c_i)P(c_i)}{P(\vec{v})} \quad (1)$$

This method works well if the properties are orthogonal. In practice, however, it works effectively without these assumptions. Because of the simplicity of the technique, short training sets yield excellent results. It also establishes soft decision limitations to avoid overtraining. Outliers (feature selections that do not reflect the category to which they correspond) may be avoided by building a probability model in practice. The random selection of a distribution model for calculating the probability P(x) results in performance constraints for complex multi-class constructions and the need for more flexibility in the decision boundary.

Support vector machines (SVM): Estimating the probability *P*(*x*) using an independent distribution system result in performance restrictions for complicated multi-class structures and the need for extra flexibility in the frame. If the characteristics are orthogonal, this technique works effectively. Nonetheless, it functions efficiently without these assumptions in reality. Short training sets produce outstanding outcomes due to the method's simplicity. Outliers (feature selections that do not represent the group toward which they belong) can be prevented by building a probability model in the application. It also uses soft decision constraints to avoid overtraining.

The SVM method seeks a decision function $f(\vec{v})$that minimizes the functional:

$$\min C \sum_{i}^{N} max\left(0,1 = y_i f(\vec{v}_i)\right)^2 + \|f\|_k \quad (2)$$

where N refers to the number of extracted features in total, $\|f\|_k$.

The positive definite function *K* defines as a norm in a replicating kernel Hilbert space, *H*, implying that its functional *f* is constrained. y_i is an abbreviation $y_i \in \{-1, 1\}$ (two-class problem). *C* is a parameter that defines the cost of mistakes and must be optimized. Many SVM models are constructed for the multiclass setup employing a one-against-one combination. Eventually, the dominant class is assigned.

Proposed method multi-layer perceptron (MLP): Using the suggested technique, the input layer of the last hidden neuron might take many different shapes. To pick the variables of all these training algorithms to solve the solution of constant equations once. It is assumed in this work that neurons in the MLP's hidden layer have sigmoid activation functions, and neurons in the output layer have linear activation functions *fout(x) = x*. Neurons with linear activation curves frequently populate this layer. These MLP formulations are supplied to clear up any confusion about MISO (multiple input, single output).

For networks with several outputs, a similar strategy is utilised. The function with the lowest cost is

$$E = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{F}(u_i) - d_i\right)^2 \quad (3)$$

Let *f (x)* represent the input vector in the final hidden nodes prior to using the suggested approach. Let *g1(x), g2(x),..., gm(x)* be the series' subsequent functions.

$$f\left(\frac{x}{2^h}\right), f\left(\frac{x}{2^{b-1}}\right), \dots f(x) \dots\dots, f(2^{k-1}x), f(2^k x) \quad (4)$$

$$f_k(x) = w_{k,1}g_1(x) + w_{k,1}g_2(x) + \cdots + w_{k,m}g_m(x) \quad (5)$$

$$Zw = d$$

$$w = \left[w_{1,1}, w_{1,2} \dots, w_{1,m}, w_{2,1}, w_{2,2}, \dots, w_{s,1}w_{s,2}, \dots, w_{s,m}, b\right]^T$$

$$d = [d_1, d_2, \dots, d_N]^T$$

$$Z = \left[\begin{pmatrix} z_1(u_1) & z_2(u_1) & \cdots & z_s(u_1) & 1 \\ z_1(u_1) & z_2(u_1) & \cdots & z_s(u_2) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ z_1(u_N) & z_2(u_N) & \cdots & z_s(u_N) & 1 \end{pmatrix}\right] \quad (6)$$

The parameter values of the new activation function minimizing the following equation determines the cost function (1):

$$w = (Z^T Z)^{-1} Z^T y \quad (7)$$

Numerical considerations should not be used to identify the parameters of the hidden neurons (4). Instead, they should be computed using proper numerical solutions for systems of ordinary differential equations (3). Several strategies for solving the Modified Least Square Problem (MLSP) arise when columns representing small vector elements are excluded from the ***Z*** matrix, which can be solved relatively quickly using QR decomposition. This can employ singular value decomposition. If some aspects of the vectors ***w*** are substantially smaller than others, you can decrease the total amount of elements in the sum of (4). Vector ***w*** may be generated for matrix ***Z*** inadequately

using shortened decomposition of singular values or rider analysis [13].

Below is a diagram of the proposed method.

1) Transform all neurons in the final hidden layer's activation function to the function described in (4).

(2) In the output layer, set the frequency of all neurons to 1.

(3) Use the right numerical approach to solve the standard equation.

(4) Analysis of the cost function's value. (3)

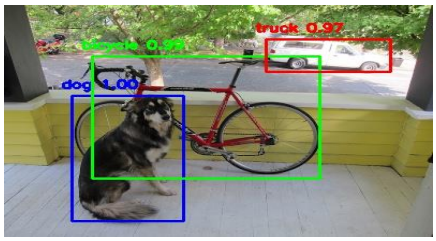
(5) The column in matrix Z corresponds to the lowest number of w and the capacity to eliminate the MLSP solution as a result of this elimination. Cost function values are recalculated. Step 5 should be repeated if the functional form changes considerably. Nevertheless, the modifications to this section have been reversed.

The equation (5) system may be solved significantly quicker using nonlinear optimization approaches than learning MLP. f(x) corresponds to a particular equation in the series g1(x), g2(x) ,..., gm (x). As a result, the updated system must have, at minimum, the very same approximate function as MLP before implementing the suggested technique. The next section demonstrates how the suggested technique enhanced the efficiency of an MLP trained with the Levenberg-Marquardt method.

4. Results And Discussions

The training process involved feeding a set of input data into the model. Multilevel Perceptron (MLPs), Decision Trees (DTs), Nave Bayes (NBs), and Support Vector Machines (SVMs) were used in this study.

(a)



(b)

(c)

Fig 3. Types of objects (a) Sample image (b) indoor (c) outdoor

We built and evaluated the results of many object identification algorithms to find the best one, as seen in Table 1. All of these techniques have been tested on a variety of photos. The table illustrates the amount of confidence in one of these photos. Table 1 compares YOLO v3 against two other DeepLab object identification methods, R-CNN and YOLO when applied to the same picture. Regardless of the item, YOLO v3 gives the maximum accuracy.

Table 1. A review of several object detection models

Detection models	Object 1 assurance	Object 2 assurance
DeepLab	97.35	98.51
R-CNN	96.47	97.77
YOLO	98.11	98.74
YOLO v3	99.32	99.74

Figure 4 shows that the proposed YOLO v3 achieves the highest confidence levels for two objects at 99.32% and 99.74%. On the other hand, R-CNN reaches the poorest confidence levels for two objects at 96.47% and 97.77%, respectively.

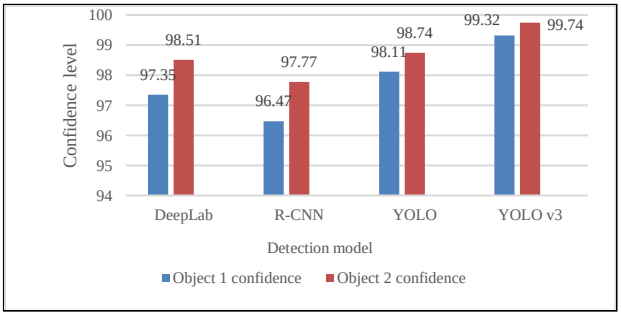


Fig 4. Object detection models

Table 2 shows the results applied to various indoor images. "YOLO v3" has the highest indoor accuracy compared to "Yolo 2". To better clarify the algorithm model, we measured the accuracy based on the measurement parameter of all images.

Table 2. Comparison of indoor detection objects on different images

Type	Algorithm	Object	Accuracy
Indoor	YOLO 2	Table	98.62
		Computer	98.88
		Dog	98.24
		Sofa	98.77
	YOLO v3	Table	99.10
		Computer	98.98
		Dog	99.11
		Sofa	99.32

Figure 5 shows the results applied to various indoor images. "YOLO v3" has the highest accuracy of 99.10%, 98.98%, 99.11%, and 99.32%, each of the four attributes in the indoor environment compared to the 'Yolo 2'.

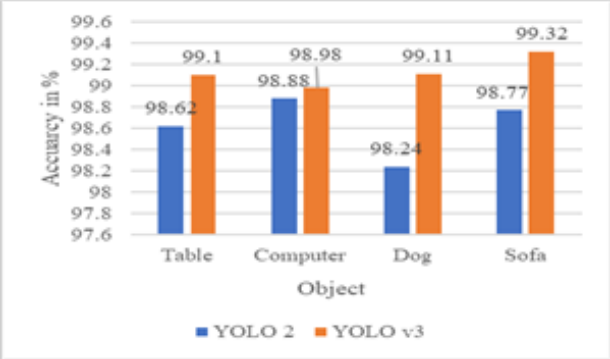


Fig 5. Performance comparison of YOLO 2 and YOLO v3 indoor detection objects.

Table 3. Comparison of outdoor detection objects on different images

Type	Algorithm	Object	Accuracy
Outdoor	YOLO 2	Person A	97.45
		Car	97.22
		Truck	98.35
		Bike	98.55
	YOLO v3	Person A	99.23
		Car	98.75
		Truck	99.34
		Bike	99.44

Figure 6 shows the results applied to various indoor images. "YOLO v3" has the highest accuracy of 99.23%, 98.75%, 99.34%, and 99.44% of each of the four attributes in the indoor environment compared to the 'Yolo 2'

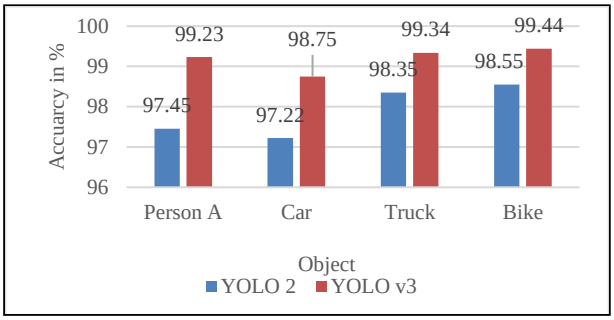


Fig 6. Outside detecting object comparison of the performance of YOLO 2 and YOLO v3.

The proposed MLP algorithm's performance is measured using evaluation criteria such as accuracy, recall, and accuracy. The amount of correct this about an identical class is called accuracy. A recall is the correct figure of suggestions produced across all categories in the data set. Model accuracy is the ability to identify the optimal model using trained data.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$Precision = \frac{No.of\ corrected\ predictions}{Total\ no.of\ predictions} \quad (10)$$

Table 4. Comparison of indoor and outdoor classification objects on different images

Classifier	Accuracy	Precision	Recall
DT	98.47	98.88	98.74
NB	97.35	97.44	98.10
SVM	98.75	98.90	98.88
MLP (Proposed work)	99.10	99.32	99.21

Figure 7 scores metric results for several architectures like DT, NB, SVM, and MLP are displayed. According to the experimental data, MLP has the maximum average accuracy of 99.32% & NB has the poorest average accuracy of 97.44%. With a recall rating of 98. 10%, NB has the most negligible recall value. The MLP architecture provides an overall accuracy of 99.10% compared to other network designs. Empirical evidence indicates that MLP models built with the YOLO v3 network design outperform image prediction.

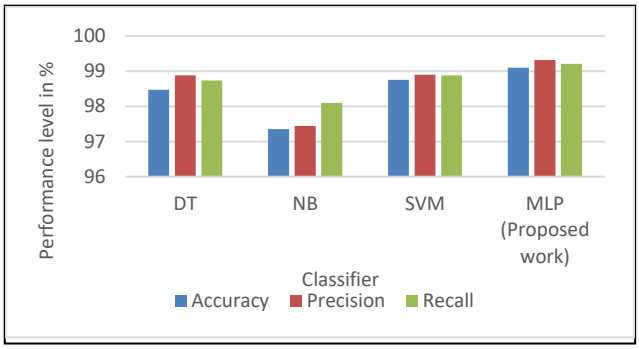


Fig 7. Evaluation of DT, NB, SVM, and Suggested Model Performance

5 Conclusions

This article presents the latest assistive technologies for the visually impaired in computer vision, embedded systems, and mobile platforms. The purpose of the system under development is to generate sound signals and vibrations in the presence of obstacles to the internal and external environment. The technology uses machine learning to identify impediments and warn users of their features. The suggested system's prototype was developed and tested. The obstacle detection module outperforms several modules in the literature.

References

[1] World Report on Vision. World Health Organization. 2019. Available online: https://www.who.int/publications/i/item/9789241516570 (accessed on 7 May 2022).

[2] Bourne, R.; Steinmetz, J.D.; Flaxman, S.; Briant, P.S.; Taylor, H.R.; Casson, R.; Bikbov, M.; Bottone, M.; Braithwaite, T.; Bron, A.M.; et al. Trends in prevalence of blindness and distance and near vision impairment over 30 years: An analysis for the Global Burden of Disease Study. Lancet Glob. Health 2021, 9, e130–e143.

[3] Wafa Elmannai and Khaled Elleithy. Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions. Sensors, 17(3), 2017

[4] A. S. Al-Fahoum, H. B. Al-Hmoud, and A. A. Al-Fraihat. A smart infrared microcontroller-based blind guidance system. Active and Passive Electronic Components, 2013(726480), 2013.

[5] B. Mustapha, A. Zayegh, and R. K. Begg. Ultrasonic and infrared sensors performance in a wireless obstacle detection system. In 2013 1st International Conference on Artificial Intelligence, Modelling and Simulation, pages 487–492, Dec 2013.

[6] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: A survey," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 1, pp. 25–35, Jan. 2010.

[7] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," Sensors, vol. 17, no. 3, pp. 565–606, Mar. 2017.

[8] A. Bhowmick and S. M. Hazarika, "An insight into assistive technology for the visually impaired and blind people: State-of-the-art and future trends," J. Multimodal User Interfaces, vol. 11, no. 2, pp. 149–172, Jan. 2017.

[9] H. Fernandes, P. Costa, V. Filipe, H. Paredes, and J. Barroso, "A review of assistive spatial orientation and navigation technologies for the visually impaired," Univ. Access Inf. Soc., vol. 2017, pp. 1–14, Aug. 2017.

[10] H. Zhang and C. Ye, "An indoor wayfinding system based on geometric features aided graph SLAM for the visually impaired," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 25, no. 9, pp. 1592–1604, Sep. 2017.

[11] R. Jafri, R. L. Campos, S. A. Ali, and H. R. Arabnia, "Visual and infrared sensor data-based obstacle detection for the visually impaired using the Google project tango tablet development kit and the unity engine," IEEE Access, vol. 6, pp. 443–454, 2018.

[12] I. Ulrich and J. Borenstein, "The GuideCane-applying mobile robot technologies to assist the visually impaired," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 31, no. 2, pp. 131–136, Mar. 2001.

[13] F. Penizzotto, E. Slawinski, and V. Mut, "Laser radar based autonomous mobile robot guidance system for olive groves navigation," IEEE Latin Amer. Trans., vol. 13, no. 5, pp. 1303–1312, May 2015.

[14] Y. H. Lee and G. Medioni, "Wearable RGBD indoor navigation system for the blind," in Proc. ECCV Workshops, Zürich, Switzerland, 2014, pp. 493–508.

[15] Tirlangi, Ram & Sankar, Ch. (2016). Electronic Travel Aid for Visually Impaired People based on Computer Vision and Sensor Nodes using Raspberry Pi. Indian Journal of Science and Technology. 9. 10.17485/ijst/2015/v8i1/106850.

[16] Ackland, P.; Resnikoff, S.; Bourne, R. World blindness and visual impairment: Despite many successes, the problem is growing. Community Eye Health 2017, 30, 71.

[17] Kumar, A. & Patra, Rusha & Mahadevappa, Manjunatha & Mukhopadhyay, Jimut & Majumdar,

Arun. (2011). An electronic travel aid for navigation of visually impaired persons. 2011 3rd International Conference on Communication Systems and Networks, COMSNETS 2011. 1 - 5. 10.1109/COMSNETS.2011.5716517.

[18] Haigh A., Brown D.J., Meijer P., Proulx M.J. How well do you see what you hear? The acuity of visual-to-auditory sensory substitution. Front. Psychol. 2013;4 doi: 10.3389/fpsyg.2013.00330.

[19] Vance Landford, "Electronic travel aids ETAs, past and present," slides, TAER April 2004.

[20] A. Awada, Y. B. Issa, C. Ghannam, J. Tekli, and R. Chbeir, "Towards Digital Image Accessibility for Blind Users Via Vibrating Touch Screen: A Feasibility Test Protocol," presented at the Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on, 2012, pp. 547–554.

[21] L.-C. Chen, G. Papandreou, S. Member, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs." arXiv:1606.00915v2, 2017.

[22] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages91–99, 2015.

[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You Only Look Once: Unified, real-time object detection. arXiv:1506.02640, 2015.