Section A-Research paper

ISSN 2063-5346

Empowering computer vision to visually impaired people for satisfactory movements through Artificial Intelligence & Machine Learning with the help of nano-chemical concepts

[1] Dr. Suhasini V.K., [2] Dr. K.N. Vijayakumar, [3] Pavan Vilas Rayar, [4] Dr. V. Venkataramanan, [5] Dr. Pavithra G., [6] Dr. T.C.Manjunath*

[1] Professor & Principal, MCA Dept., Bharati Vidyapeeth's Institute of Management & Tech. Sector 8, CBD Belapur, Navi Mumbai, Mumbai-400614, Maharashtra State, India
[2] Professor & HOD, Dept. of Mechancial Engg., No. U-15, JVPD Scheme, Opp.Cooper Hospital, DJ Sanghvi College of Engineering, Bhaktivedanta Swami Marg, Vile Parle (West), Mumbai-400 056, Maharasthra, India
[3] Assistant Professor, Dept. of Mechancial Engg., No. U-15, JVPD Scheme, Opp.Cooper Hospital, DJ Sanghvi College of Engineering, Bhaktivedanta Swami Marg, Vile Parle (West), Mumbai-400 056, Maharasthra, India
[4] Assistant Professor, Dept. of Electronics & Telecommunication Engg., JVPD Scheme, Opp.Cooper Hospital, DJ Sanghvi College of Engineering, Bhaktivedanta Swami Marg, Vile Parle (West), Mumbai-400 056, Maharasthra, India
[5]Associate Prof., Electronics & Communication Engg. Dept., Dayananda Sagar College of Engineering, Bangalore, Karnataka
[6] Professor & HOD, Electronics & Communication Engg. Dept., Dayananda Sagar College of Engineering, Bangalore, Karnataka
*Corresponding author : tcmanju@iitbombay.org

doi: 10.31838/ecb/2023.12.si6.295

Abstract

This paper deals with the empowering of vision to visually impaired people for satisfactory movements through Artificial Intelligence & Machine Learning algorithms. Our research paper approaches on how Artificial Intelligence can enhance the regular life of the vision impaired individuals. They suffer a lot when they are submitted with unexpected situations which they might not be aware of. Visually impaired individuals face unique challenges in navigating and interacting with the world around them, which can lead to social exclusion and a reduced quality of life. Fortunately, advances in technology have made it possible for visually impaired individuals to live more independently and fully participate in society. For example, there are now a wide range of assistive technologies available that can help visually impaired individuals to access information in their surroundings. These can include text-to-speech software, braille displays, screen readers, and other tools that allow individuals to interact with digital and physical information in new ways. Additionally, there are many organizations and resources available that can provide training and support to visually impaired individuals who are seeking to live more independently. It is important to continue to develop and improve these technologies and resources in order to ensure that visually impaired individuals have equal access to information and opportunities. Additionally, society as a whole must work to become more inclusive and welcoming of individuals with disabilities, recognizing the value and contributions that they can make to our communities. With the help of the advanced technology such as object-detection, facial recognition, emotion recognition, currency recognition, web-automation, screen-reader and emergency calling features now the visually impaired can lead a life similar to a normal individual which in turn enables them to adapt and react to the environment in an effective manner. All the above-mentioned technologies are implemented through the use of machine learning, deep learning, speech-to-text and Selenium web automation frameworks.

Keywords Artificial intelligence, Object detection, Emotion Intelligence, Currency Recognition, Facial Recognition, Machine Learning. They should be separated by commas. Keywords (except the first one) start with small letters and the last one ends with a dot.

ISSN 2063-5346

1 Introduction

The study and development of computer systems that can see, reason, and act are known as artificial intelligence (AI). The intelligence must be represented by learning, thinking, making decision, solving problems. With the help of this advanced technology the visually impaired can now lead a life similar to normal individuals. Detecting instances of semantic objects of a specific class (such as people, buildings, or cars) in digital photos and videos is the task of object detection, a branch of computer vision and image processing, which enables the visually impaired to be completely aware of the surrounding that they are in [1]. A facial recognition system is a piece of technology that can compare a human face in a digital photo or video frame to a database of faces. This technology combined with object detection enables the visually impaired to not only detect the presence of an individual in the frame but also enables to identify the individual as well [2].

A emotion recognition system is a technology that analyses facial expressions from both static images and videos in order to reveal information on one's emotional state. This could help the visually impaired to know how the other individuals present in the frame feel and therefore react accordingly to it. According to research, those with disabilities are 31% more likely to access the internet than those with visual impairments. The main issue in creating dependable software is to use the fewest number of keystrokes feasible and to deliver a whole experience using speech alone. This is made possible through the use of Web-automation which could automate the process of internet surfing with the inputs provided in the form of voice commands by the visually impaired [3].

A Currency-recognizing module could be a new tech that is capable of automatically recognizing the monetary value of the currency and help to convert it to other currencies without any human intervention. This enables the visually impaired to be assured of transactions performed in day to day activities. A Screen reader is a software that reads the text is displayed on the computer screen into a form that a visually impaired user can process. All these technologies are combined together with the input received from the Bluetooth camera placed in the hat of the visually impaired to help them lead a life similar to normal individuals [4].

2 Research Objectives

The research objectives that are mentioned in this section is sub-divided into 6 streams, viz., this paper addresses the important research objectives such as [5]

a) To develop an object detection embedded with facial recognition software to aid the visually impaired adopt to the environment effectively [6].
b) To Develop an Emotion detection model that could identify the emotions of individuals present in the frame of the visually impaired [7].
c) To develop a currency recognition software that could help in assuring the transactions done by the visually impaired persons [8].
d) To automate the process of web-surfing by taking voice commands from visually impaired which makes browsing more effective and efficient for them [9].
e) To develop a software that could inform to the registered phone number whenever there is a panic situation faced by the visually impaired [10].

3 Literature Survey / Review of Literatures

The technology developed involves only detection of obstacles to alert the individual but does not provide any information related to the type of the object or any information regarding the environment. To overcome this we have chosen to implement object detection with embedded facial recognition which provides the type of objects present in the environment and the identity of the individual present in the frame which enables the visually impaired individuals to effectively and efficiently to interact and adopt to the environment [11]. The designed model only detects if an individual is present in the frame and does not provide any information regarding the identity of the individual to overcome this we have embedded facial recognition to object detection which also identifies the individual hidden in the frame provided the individual's image or picture is present in our database [12].

They had automated the web browsing process of Wikipedia and Gmail as a part of extension of work done we wish to automate shopping experience of visually impaired in websites such as Amazon which operates by receiving the voice commands and performing the required operations without any human intervention [13. They have implemented object detection, Emotion intelligence software through a Android application as a part of

ISSN 2063-5346

improving the work we wish to add the features of currency detection which would assure the visually impaired individuals about the transactions performed [14]. In addition to existing technologies we wish to also aim to include Emergency calling function through voice command which could help the visual impaired to contact the required individuals during the panic situations [15].

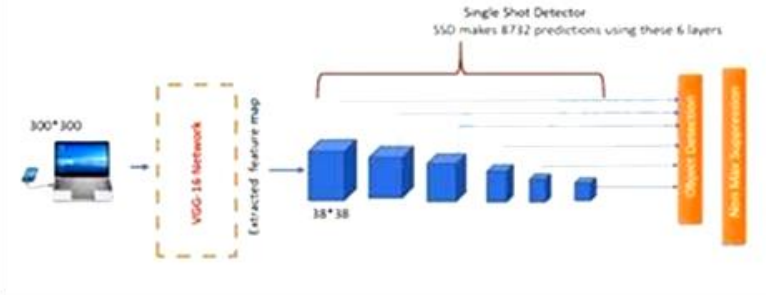

Fig. 1. Block diagrammatic module of the proposed model

4 Object-detection using Mobinet-SSD architecture

The dataset considered for our research is the coco dataset which consists the data pertaining to 81 different objects which is provided as input to the training network. The architecture chosen for real time implementation of object detection is Mobinet-Single shot detectors. The various steps that are included are as listed diagrammatically as shown in the Fig. 1 in the form of an aarchitecture of Mobinet, i.e., as an single shot detector [16].

5 Architecture of a Mobinet Structure

The dataset considered for our research is the coco dataset, which involves 3 steps, viz. [17].

Step1: To train the network-During the training process training image is an image consisting of ground truth boxes for each object to be trained which is applied as an input to the VGG-16 network which consists of convolution layers of 3*3 filters with a stride 1 and always having the same padding and a Maxpooling layer of 2*2 filter with a stride of 2 which outputs the respective extracted feature maps for each object [18].

Step2: The output of the VGG-16 network which is the extracted feature maps are further applied as an input to the progressively decreasing six convolutional layers to obtain 8732 bounding boxes for each object detected with respect to the ground truth box of the training image [19]. Step3: The output of these progressively decreasing convolutional layer(which are the 8732 bounding boxes of different aspect ratio's)are reduced to a single bounding box for each object which is the best fit for the provided training image by using a technique called non max suppression which is based on intersection over union concepts [20].

6 Facial recognition using Histogram of Gradients (HOG) approach

A facial recognition system is a piece of technology that can compare a human face from a digital image or a video frame to a database of faces. It primarily consists of four steps, including Step 1: Identifying the faces in the photo We'll begin by converting our image to black and white in order to detect faces in it. We want to look at the pixels that are right next to every single pixel. To determine how dark a pixel is in comparison to its immediate surroundings is our objective. Then, in order to obtain the histogram representation of the image, we want to draw an arrow indicating the direction in which the image is becoming darker [21].

Posing and projecting faces in step two In this phase, we'll aim to warp each image so that the mouth and eyes are consistently in the same location. We'll employ a face landmark estimation technique to accomplish this. The basic plan is to identify 68 distinct places on each face (known as landmarks), such as the top of the chin, the outside corner of each eye, the inner corner of each brow, etc. Then, using machine learning, we'll train an algorithm to be able to locate these 68 precise locations on any face. We'll just rotate, scale, and shear the image with the knowledge of where the eyes and mouth are to get the greatest possible centering [22]. The machine learning system will be trained to provide 128 measurements for each face in Step 3: Encoding the faces that have been detected. By examining three facial photos simultaneously, the training procedure operates. [23]

- Load a practise face image of a well-known individual.

ISSN 2063-5346

- Add an additional image of the same well-known individual.
- Upload a photo of a completely different individual.

The main aim is here to increase the difference between 2 unlike individuals and decrease the difference between the like individuals. This process is repeated foe every face detected. Step4:To compare the encodings of the input image with the image present in the database to identify the name of the person-To do this we use a SVM classifier which compares the two images and provides the closest match, thus the name of the person. The parts of the facial image that could be used for the detection process is displayed in Fig. 2 [24].

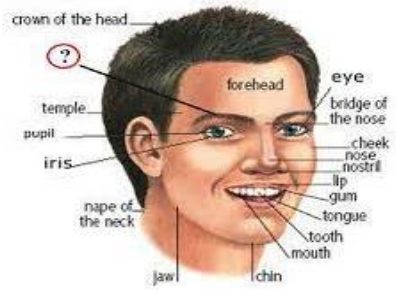

Fig. 2. Finding the parts of the facial image

Fig. 3.

7 Emotional Intelligence Usage in Design Process

The training process is done by using the FER2013 dataset. Steps involved in emotion recognition process using landmarks are shown in the Fig. 3. The three main steps involved are [25]

- Face detection
- Feature Extraction
- Feature Classification

Fig. 4. Steps involved in the emotional intelligence

Step 1 : Facial Detection - Here for the 1st step the same model trained for facial recognition is utilized which identifies the faces using histogram of gradient approach .

Step 2 : Feature Extraction - By separating the landmarks from the detected faces, this is accomplished. The key features on photographs of the human face are known as facial landmarks. The (x,y) coordinates of the points on the image serve as their definition. These points serve as markers for important facial features such the eyes, brows, nose, mouth, and jawline. The Dlib's pre-trained Facial Landmark Detection Model, which recognises 68 2-Dimensional points on the human face, was the facial landmark model we employed.

Step 3 : Feature Classification - A list of softmax scores for each of the seven kinds of emotions is produced by the model from step. The user can choose the primary parametric processes by reading out the entire search result and selecting the emotion with the highest score that was determined by calculating the Euclidian distance.

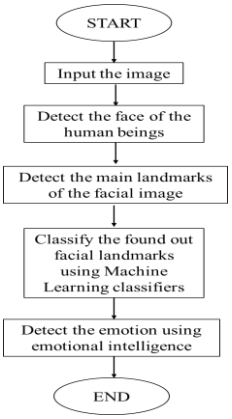

Fig. 5. Steps involved in emotion recognition process using landmarks

8 Currency Recongtions

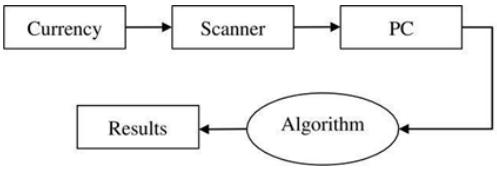

Fig. 6. Steps invovled in the recogntion of currency

The steps involved in the currency recognition are as mentioned as follows one below the other in successions. A Currency recognition system is a technology capable of automatically recognizing the monetary value of the currency and help to convert it to other currencies without any human interventions. Fig 5 gives the steps involved in the recognition of the currency.

9 Web Automations

Web-Automation is a process of automating the web browsing through the use of web-automation frameworks like Selenium through which the visually impaired can browse through the internet in a hassle free and efficient manner. This module consists of a Python script that uses Selenium and Beautiful Soup to automate the website. Through the speech-to-text and text-to-speech interfaces, the user can search for any query, and Selenium is used to search the recognised query. The Beautiful Soup module of Python is used to scrape the contents of the web page and feed back the search results to the user. In comparison to the straightforward 4-layer CNN, the test accuracy reached 63.2 percent after 50 epochs thanks to the indexing of the search results, which provides quick access to the website in accordance with the user's preferences and saves time. The block diagram of the selenium stacks-based web automation process is shown in Fig. 6.

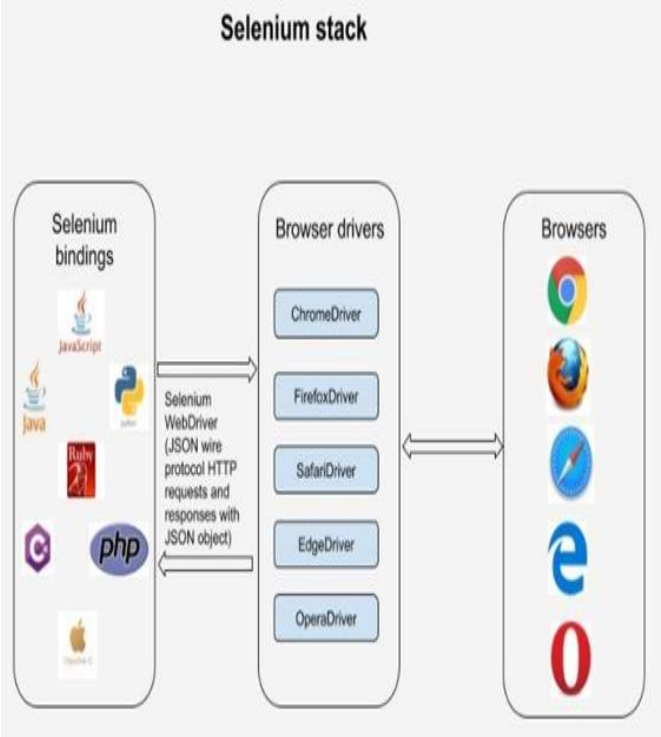

Fig. 7. Web automation using selenium stacks

Successfully implemented a currency recognition model that could aid the visually impaired in their daily transactions. The currency recognition model developed using random forest algorithm, Scikit-learning, and openCV can identify the denominations of 10, 20, 50, 100, 200, 500, 2000 of Indian currency. Implemented a web-automation framework for applications of Gmail and shopping website amazon using the selenium automation framework that receives the input in the form of voice commands, and sends the mail to the required participant in the case of email automation and finds a product of user desire and place the order in the case of automation of shopping websites such as Amazon. Implemented an emergency calling software using the Twilio model of py-

thon, that could call to the registered phone number whenever the visually impaired feels he is in a panic situation and would need help.

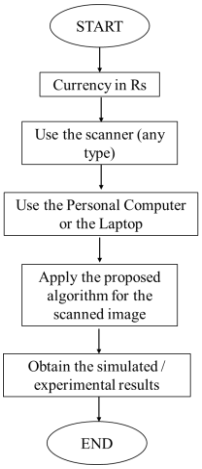

Fig. 8. Steps that are involved in the recognition of the currency detection processes

10 Steps in Image Processing Concepts

The steps used in the imaging process for the detection process could be exaggerated as follows. Fig. 7. gives the steps that are involved in the recognition of the currency detection processes.

Step 1 : Image pre-processing - The aim of image preprocessing is to suppress undesired distortions or enhance some image features that are important for further processing or analysis. In our work, image pre-processing includes these parts:

- Image adjusting
- Image smoothening (removing noise)

Step 2 : Edge detection and ROI determination - Edge detection is used in image analysis for finding region boundaries .For boundary detection, we require a binary image, which has only 2 colors, black and white. All we do in this process is simply, separate the background and the foreground, and separate the ROI

Step 3 : Extract the required information from ROI - We extract the HSV values of all the pixels and take average of their H, S, V features and again compare them with the values from the database. We make use of Euclidian distance to compare between the target HSV and ideal HSV to display the value of the currency which has the least Euclidian distance.

11 Camera Module used for taking the input from the visually impaired environments

We plan to make use of a wireless camera / webcam to take the input from the surrounding which is utilized as inputs for various applications such as object-detection, emotional recognition, currency-recognition and text to speech converter . Here the results are provided as voice output through the Bluetooth earphones attached to the Hat. Fig. 8 gives the CCD camera module embedded in a hat to take the input wirelessly modes.

Fig. 9. Camera module embedded in a hat to take the input wirelessly

ISSN 2063-5346

12 Results & Discussions

Here, in this paper, we have successfully implemented an object detection model embedded with facial recognition using AI ML concepts. Fig. 9 gives the detection of Indian Currencies of various denominations.

Fig. 10. Detection of Indian Currencies

In this work, we have succesfully implemented a object detection model embedded with facial recognition system. The model developed using the coco data set and the Mobinet-SSD architecture is capable of identifying a total of 81 different objects present in the frame of the visually impaired. The facial recognition model embedded with object detection not only helps to indicate a presence of an individual but also help in identifying who the individual is provided their data is stored in the database. In the above shown result the module is successful in recognizing multiple objects –person, Cell phone, Zebra with a great accuracy. The trained model is tested for various other different inputs and results obtained were correct and accurate as shwon in the Fig. 10 & 11 respectively.

Fig. 11. Sample output of the object detection model

In the next module of the simulation process, we have successfully implemented the emotional intelligence model. The .model developed using the FER-2013 dataset is capable of identifying 7 types of emotions angry. The plot of model accuracy and model loss is as shown in the graph above. This implementation by default detects emotions on all faces in the webcam feed. With a simple 4-layer CNN, the test accuracy reached 63.2% in 50 epochs as shwon in the Fig.12.

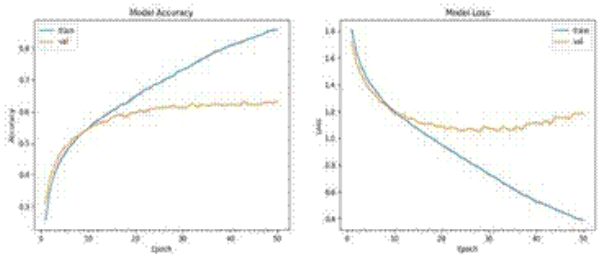

Fig. 12. Plot of the model accuracy & the model loss

In the 3rd module, we have successfully implemented a currency recognition model that could aid the visually impaired in their daily transactions. The currency recognition model developed using random forest algorithm , Scikit-learning, and opencv can identify the denominations of 10, 20, 50, 100, 200, 500, 2000 of Indian currency. In the 4th module, we have implemented a Web-automation framework for applications of Gmail and shopping website amazon using the selenium automation framework that receives the input in the form of voice commands , and sends the mail to the required participant in the case of email automation and finds a product of user desire and place the order in the case of automation of shopping websites such as Amazon.

Finally, in the last module, we have implemented a emergency calling software using the Twilio model of python, that could call to the registered phone number whenever the visually impaired feels he is in a panic situation and would need help as shwon in the Fig. 13.

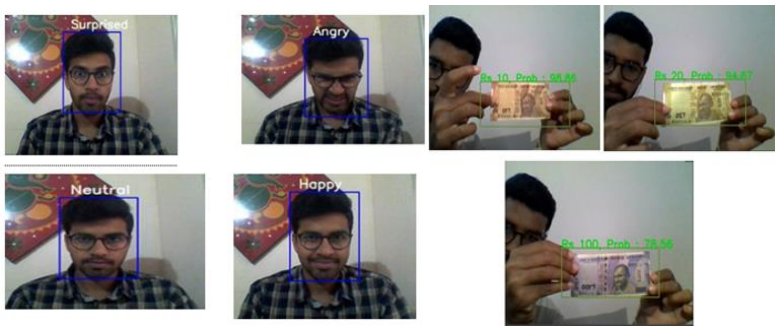

Fig. 13. Sample output of the currency recognition model

13 Conclusive Remarks

Artificial Intelligence has been enhancing the life of the vision impaired. The technology too is vastly growing day by day and improving the other aspects of life those who are in need of someone to take care of them such as disabled, autism, elderly, blind and other such people and Definitely within few decades from now, machines will be able to understand the feeling of the humans and solve problem accordingly. We have made use of these concepts for the people who are impaired visually (special disabled people) so that specially abled humans can too can lead a normal and independent life like other people. Object-detection embedded with facial recognition and emotion intelligence software helps the visually impaired to have a effective understanding about the environment and therefore could react appropriately. Currency recognition helps in easy payment, Text recognition helps in reading and analyzing text. The virtual assistant gives visually impaired people a simple method to access any website. It does away with the need to know difficult keyboard shortcuts or how to use screen readers. Additionally, processing a higher frame resolution and a higher frame rate will be possible if processing performance is enhanced and the latency of the data input from the camera module to the model is decreased, thus allowing more precise acquisition of the environment.

References

[1] Prof. V.R. PAWAR, “Smart Walking Stick”, International Journal of Management Technology And Engineering, vol. 8, no. V, MAY 2018.

[2] M. Naveen Kumar, K, "Smart Blind Stick Using Gsm Technology", Usha International Journal of Engineering Trends and Technology (IJETT), vol. 4, no. 7, July 2013.

[3] K. Hemavani, V. Sai Kumar and S. A. K. Jilani, "Smart Blind Stick", International Journal of Management Technology And Engineering, vol. 8, no. V, May 2018.

[4] K.C. Nalavade, Fatema Bharmal, Trupti Deore and Ajay Patil, "Use of ultrasonic sensors GPS and GSM technology to implement alert and tracking system for Blind Man", International Conference of Advance Research and Innovation (ICARI-2014).

[5] Jismi Johnson, P Nikhal Rajan, Nivya M Thomas, C S Rakendh and Sijo T Varghese, "WALIKG STICK FOR BLIND PEOPLE USING MICROCONTROLLER", International Journal of Engineering Science Invention Research & Development, vol. III, no. IX, March 2017.

[6] N Pratik, Poornesh, Shashikant and Shreedhar Kudva, "SMART VISION STICK", Saritha International Journal of Latest Trends in Engineering and Technology, vol. 9, no. 3, pp. 273-27.

[7] Jismi Johnson, P Nikhal Rajan, Nivya M Thomas, S Rakendh and Sijo, "SMART WALKING STICK FOR BLIND", International Journal of Engineering Science Invention Research & Development, vol. III, no. IX, March 2017.

[8] Mohammed Azher and Therib Al-Awsat, "Smart Blinding Stick with Holes Obstacles and Ponds Detector Based on Microcontroller", Journal of Babylon University/Engineering Sciences, vol. 25, no. 5, pp. 09571, 2017.

[9] Thiyagarajan Manihatty Bojan, Umamaheswaran Raman Kumar and Vishwanathan Manihatty Bojan, "Designing Vehicle Tracking System-An Open Source Approach", International Conference on Vehicular Electronics and Safety (ICVES), December 16-17, 2014.

[10] Harshad Girish Lele, Viten Vilas Lonkar, Varun Vasant Marathe and Mrunmayi Mohan Modak, "Electronic path guidance for visually impaired people", The International Journal Of Engineering And Science (IJES), pp. 09-14, 2013.
[11] Michael Carpenter and Shanker Balasubramaniam, "Interfacing a GPS Receiver with a Microprocessor via UART Application note", November 2009.
[12] N Nandhini, G Vinoth Chakkaravarthy and G. Deepa Priya, "Talking Assistance about Location Finding", International Journal of Innovative Research in Science Engineering and Technology, vol. 3, no. 2, February 2014.
[13] J.B.F. van Erp, J.A. Veltman and H.A.H.C. van Veen, "A Tactile Cockpit Instrument to support Altitude Control", Proceedings Human Factors and Ergonomic Society 47tn annual meeting 2003, 2003.
[14] G. Prasanthi and P. Tejaswidho, "Sensor assisted Stick for the Blind People-Transmission on Engineering and Sciences", Int. Journal, vol. 3, no. 1, pp. 12-16, 2015.
[15] Ulrich and J. Borenstein, "The GuideCane –Applying Mobile Robot Technologies to assist the visually impaired", IEEE Tr. SMC, vol. 31, no. 2, March 2001.
[16] N. Senthil kumar, A. Abinaya, E. Arthi, M. Atchaya, M. Elakkiya, "Smart Eye for the Visually Impaired People", International Research Journal of Engineeringand Technology, Volume: 07 Issue: 06, June 2020.
[17] wwwhttps://www.geeksforgeeks.org/introduction-deep-learning/
[18] Priya U. Thakare, Kote Shubham, Pawale Ankit, Rajguru Ajinkya, Shelke Om, Smart Assistance System for the Visually Impaired, International, Journal of Scientific and Research Publications, Volume 7, Issue 12, December 2017.
[19] Tufel Ali Qureshi, Mahima Rajbhar, Yukta Pisat, Prof. Vijay Bhosale, "Ai based app for blind", International Research Journal of Engineering and Technology (IRJET),India,volume-08, Issue-03, 2021
[20] Mr.Vinayak Iyer, Mr.Kshitij Shah, Mr.Sahil Sheth, Mr.Kailas Devadkar,Virtual assistant for visually impaired, International Journal of engineering research & Technology, 2012.
[21] Ms. Kavya. S ,Ms. Swathi ,Mrs. Mimitha Shetty, Assistance System for Visually Impaired using AI, International Journal of Engineering Research & Technology (IJERT),
[22] Ms.Padma Shneha, Mr.Prathyusha Reddy ,Ms.V.M.Megala,AI for visually impaired, International Journal of engineering research and technology (IJERT),
[23] R. Gnana Praveen, Roy P Paily," Blind Navigation Assistance for Visually Impaired Based on Local Depth Hypothesis from a Single Image", International Conference on Design & Manufacturing, IConDM 2013.