DRISHTI
A Vision to Help the Visually Impaired

[1]Sneha Kumari Sinha, [2]Misbah Sirnaik, [3]Priyanka Surnashe, [4]Rachana Dhannawat

[1,2,3]B. Tech Student, [4] Assistant Professor
Usha Mittal Institute of Technology.
SNDT Women's University, Mumbai, India.

Abstract: Visual impairment, also known as vision impairment or vision loss, is a decreased ability to see to a degree that causes problems not fixable by usual means, such as glasses. This project proposes to build a model that is able to help the visually impaired individuals with an Object Detecting Eye Wear and Smart Stick which that can detect obstacle in the range of 20 cm and 30 cm respectively. A Deep Learning Model is proposed which can automatically describe the content and description of a picture using properly formed sentences using VGG-16 and LSTM models. This Image Captioning comes along with a Multi-lingual support structure i.e., English and Hindi for better understanding of the environment description. Lastly, this project proposes an Android Multi-featured Mobile Application will provide various service to the visually impaired with Voicebased Feedback Mechanism to ease their task.

IndexTerms **- Object Detecting Eye Wear, Smart Stick, VGG-16, LSTM, Voice-based Feedback.**

I. INTRODUCTION

Visual impairment is defined as vision loss, it is a type in which some people have partial vision loss or people who are completely blind. Visual impairment is usually caused by an injury, some children have congenital blindness means they have vision loss at birth. This congenital blindness can be caused by things like genetics, infection, disease, etc., This problem of visual impairment can be solved by eyeglasses, contact lenses, eye drops, or other medicines. In some cases, surgery is required.
According to the World Health Organization, globally 2.2 billion people are facing the difficulty of near or distance vision impairment. The majority of people who are facing the difficulty of visual impairment are over the age of 50. Visual impairment can affect people of all ages. Vision loss severely affects the quality of living in the adult population, people with vision loss have lower rates of productivity which causes high depression and anxiety. In the older generation, visual impairment causes social isolation, difficulty in living day to day lifestyle, and a greater possibility of entering into nursing or care homes.
Technology has an outstanding power to improve the lives of visually impaired people by utilizing multiple smart devices such as smartphones, smartwatches, smart glasses, etc., that are available in the market but all of these are used for normal people. The main problem is there is an enormous lack of technology to aid the visually impaired.

In this paper the aim is to design a low-cost smart glass object detecting eyewear using Arduino UNO and an Arduino Pro Mini, a smart stick using Arduino UNO. The stick uses an ultrasonic sensor for obstacle detection. The main purpose of this stick is to detect nearby obstacles and notify the direction of the obstacle to the user. Capturing the images and generating the textual description of that images in which dataset consists of input images and their corresponding output captions. The multipurpose android application which is completely reliant on voice control will be very helpful to the visually impaired.

II. LITERATURE REVIEW

Shubham Melvin Felix, et al. [1] is mainly focussed upon Image Recognition using REST API (used to analyse the image captured), Voice Recognition using Google Cloud Speech API (changes the voice input to text) and Chat Bot using ID3 algorithm and Dialog-Flow platform (used for building and training the chat-bot so that the person can interact with it). But this method requires the person to click images manually, image and textual description are not expressive enough nor does it possess multi language support. [1]
Aviral Chharia, et al. [2] presents an end-to-end real time human-centric model for aiding the visually impaired people. The vision enabled eye wear of the visually impaired person captures the scene as real-time video. The image frame from the video is extracted and sent to a deep recurrent architecture (that uses transfer learning) where a CNN (VGG-16 net) is used to obtain a 4096-dimensional image feature vector. These feature vectors are then fed into an LSTM (which has been trained on Flickr 8K dataset) to generate captions (with max caption words = 30) and the BLEU score is calculated. The generated captions are converted to audio for the visually impaired person to hear and get greater assistance through continuous feedback. But this method faces few limitations, like images which are not accurate (BLEU score > 0.25) are also encountered. Furthermore, it is noticed that in case the scene around the visually impaired person is not changing fast enough like when sitting on a park bench or roadside with people walking down the street, resulting captions repeat frequently. Another challenging field is the evaluation of human emotions during face to face interactions. [2]
According to N. Komal Kumar, [3] deep learning has the way for generation, recognition and detection of image captioning. Image caption generator deals with the given image by capturing the semantic meaning within the image and converting it into tongue . It is tedious work to collaborate both image processing and computer vision. In the paper Regional Object Detector is used to detect, generate and recognize captions using deep learning. This proposed method focuses on further improvement upon the existing image caption generator system. This experiment is conducted on the sparkle 8k using python language to demonstrate the tactic . In their Proposed methodology contains four steps, 1st is object detection and then feature extraction after that creating attributes and the last step is encoding and decoding. The proposed methodology generated the caption in a more descriptive and accurate manner with the help of Flickr 8k dataset which will create 8000 images which are more meaningful than the existing image caption generation generators. In the future by applying a hybrid image caption generator model can produce more accurate captions in the future. [3]

Sumitra A. Jakhete, et al. [4] mainly focused on an Android application which recognizes the object around visually impaired people in real time and provides the audio output to guide them as soon as possible. SSD(Single Shot Detector) algorithm is used for object recognition and also detection. This algorithm provides accurate results for real time object detection. Android application uses a android tensorflow APIs and TextToSpeech API to give audio output.This android application significantly helps the visually impaired people to detect and recognize the object in a real time system.But in this android application we can provide a multilingual support for more user friendliness. [4]

Kasthuri R, et al. [5] proposed a concept to help the blind people to navigate the outdoor region using voice control output. When the user speaks something, it is interpreted by Speech Recognition Engine (SRE) which converts that speech into text for direct actions. In this application Selendroid app interface is available which allows users to fetch the latest information from various web servers. This latest information includes news update, weather report, and transport related information. In this application,we can add more number of real time additional features. [5]

Saeed Mian Qaisar, et al. [6] proposed a system which converts effective scenes to text conversion and pronunciation. This method is very useful for people who are visually impaired. This system helps them to understand the written text during their daily lives for example food products, medicines, notices,etc. System captures the image with an integrated camera.Image will be enhanced by the mechanism CLAHE. It will detect a text region using MSER and identify the characters using OCR. After this it will regroup these characters to form a meaningful sentence and word. Detected output finally successfully pronounced by using TTS. A detailed system performance evaluation for standard natural scenes database is a part of future work. [6]

III. PROPOSED TECHNIQUES AND METHODOLOGY

A. OBJECT DETECTION EYE WEAR AND SMART STICK USING ARDUINO:

Visually impaired people face various difficulties detecting obstacles which are present in front of them. This makes it dangerous for them to walk on the street.The Object Detection Smart Eye Wear Figure (1) and Smart Stick Figure (2) comes as a proposed solution which enables them to identify the world around them. In this paper we propose a solution, represented in the smart stick and eyewear with Ultra Sonic Sensor HC-SR04, Arduino Pro Mini and Arduino Uno.

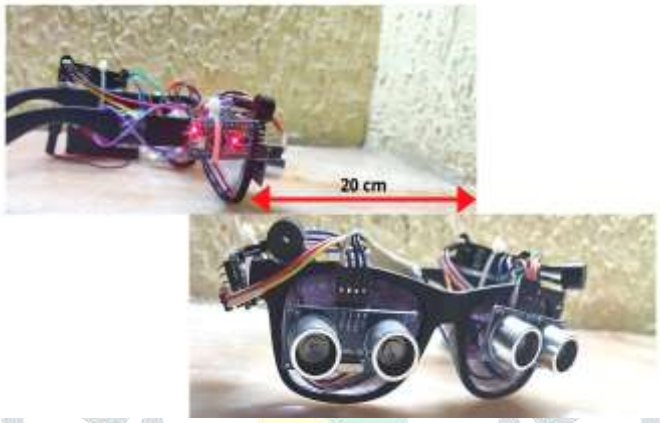

Figure. 1. Smart Eye Wear.

Figure. 2. Smart Stick.

The Smart Stick model has been designed in such a way that it can sense any object that is at the distance of 30 centimeter or less also, the object detection range for the eyewear is set as 20 centimeters. Both the ranges of detection can be changed as per need of the user.

Figure. 3. Requirments Of Object Detection Eye Wear And Smart Stick

Gathering the requirements: Collection of all the components was done in a way that it would be cost-efficient, accurate and user-friendly. Following are the description of the components(i.e., Arduino UNO, Arduino Pro Mini, 5V power supply module, Buzzer, Ultra-Sonic Sensor, etc.) used during the construction of Object detection Eye Wear and Smart Stick which are shown in Figure (3).

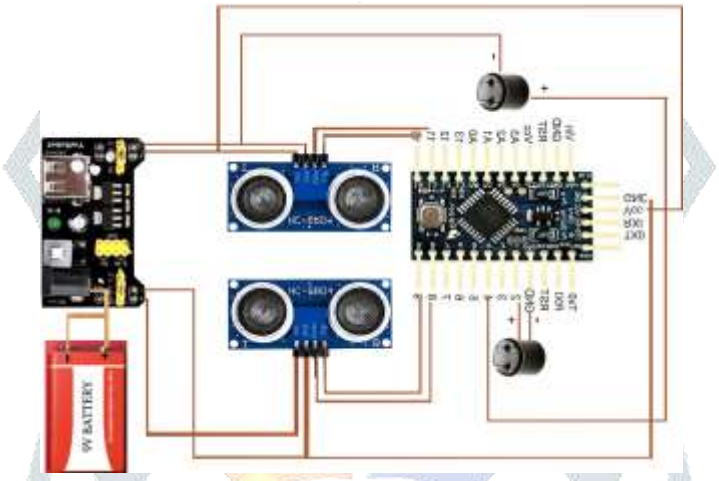

Figure. 4. Circuit of Eye Wear.

The Arduino UNO has a total of 14 digital I/O pins out of which: 6 are analog inputs, 6 can be used as PWM outputs, a power jack, a USB connection, a reset button, a 16 MHz ceramic resonator and ,an ICSP header. The Arduino Pro Mini is a ATmega328 based microcontroller board. There are 14 digital I/O pins out of which 6 can be used as PWM outputs other 6 are analog inputs also an on-board resonator with a reset button, and holes for mounting pin headers. Ultra-Sonic Sensor HC-SR-04 is a 4-pin module,the names of pins being Vcc, Trigger, Echo and Ground respectively; which is used for sensing the obstacles. Other than these 5V power supply module, 9V batteries 10 mm LED, Jumper wires and Soldering Equipment.

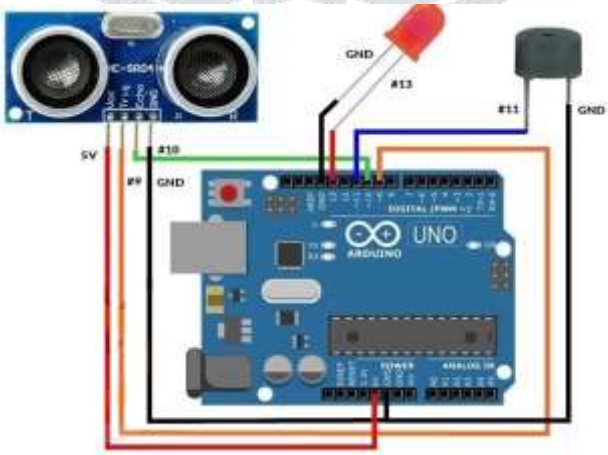

Figure. 5. Circuit of Smart Stick.

After gathering the components, the connections are done by soldering as shown in Figure (4) and Figure (5) and the code was uploaded in Arduino Pro Mini.

B. IMAGE CAPTION GENERATION USING DEEP LEARNING:

The aim here is to build a system that automatically describes the content of the picture using proper English sentences as shown in Figure (6). This task is significantly harder, for example, than the well-studied image classification or the task of object recognition, which has been a main focus for the computer vision community.

Figure. 6. IMAGE CAPTION

Indeed, a well developed description must not only depict the objects contained in an images, but should also define how these objects are relate to each other as well as the attributes and their activities they are involved in. Moreover, the above mentioned knowledge here should be expressed into commonly used natural language like English, which again means that a different language model is needed with visual understanding. In the proposed system, Figure(7) the attempt is to develop a specialized deep learning model for visually impaired people with the feature of multi-lingual support i.e., Hindi and English with an accurate version of image description using attention mechanism.

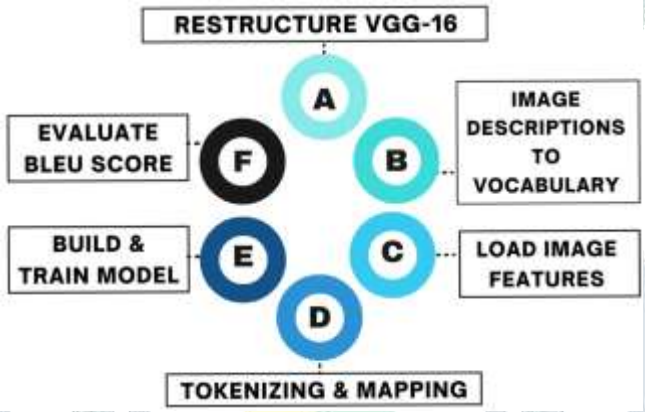

Figure. 7. Image Captioning Model Flow

Image Captioning is the process of generating captions in the form of textual description of a given image.

A) Restructure VGG16 Model:

VGG16 Model is a pre-trained model used to extract features of images using a dataset named Flickr 8K. VGG16 is a convolutional neural network model shown in Figure (8) from Automatic localization of casting defects with convolutional neural networks Conference paper. It achieves 92.7% accuracy.

- INPUT- 8000 Unique Images with their Descriptions.
- OUTPUT- Internal Representation of the image.
- DATASET- Flickr8K Dataset.

B) Convert image description to vocabulary:

The dataset of description contains multiple descriptions for every photo and the text of the descriptions requires some small amount of cleaning. The steps for converting image description to vocabulary are as follows:

1) Reading the image as shown in Figure(9) that is present in the dataset.
2) Extract description for image taken into consideration (refer Figure 10)
3) Cleaning description (getting rid of punctuations, tokens with numbers, etc.) as shown in Figure(11).
   - Converting every word to lowercase.
   - Removing every punctuation.
   - Cleaning all words which are 1 character or small in length (e.g. ‘s’).
   - Removing the words with numbers present in them.
4) Convert loaded description into vocabulary of words.
5) Save description into a file.

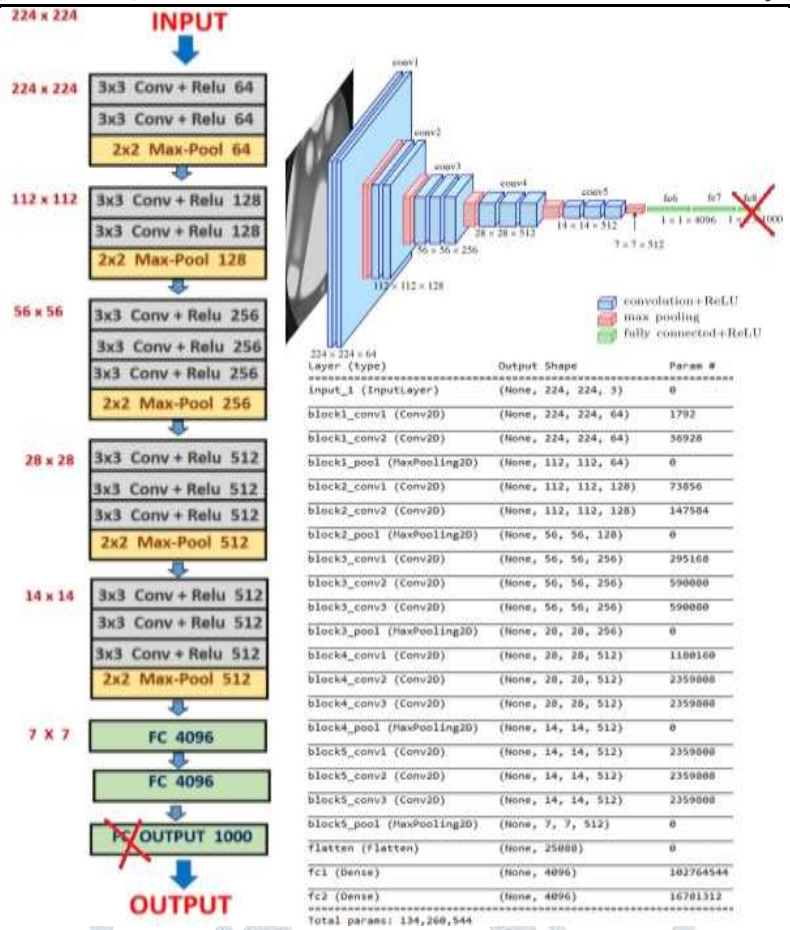

Figure. 8. Restructure VGG-16 Model

Figure. 9. An image in Flickr8K dataset taken as an instance

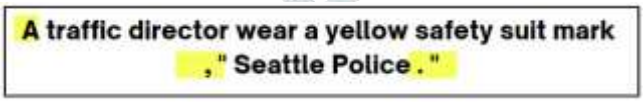

Fig. 10. Extracted Description of the given image

Figure. 11. Preprocessed Description of the Image

C) Loading image features:

Here the main purpose of the process is to train the dataset on every image and its corresponding description provided in the training dataset. The steps for the same are as follows:

1. Loading the prepared image and text data to fit the working model.
2. Train the data on all of the photos and captions in the training dataset.
3. The training and development of dataset have been predefined in the Flickr 8k.trainImages.txt and Flickr 8k.devImages.txt files respectively, that both contain lists of photo file names.

4. Extracting the predefined set of training and developing identifiers and use these identifiers to filter images and text descriptions for each set.

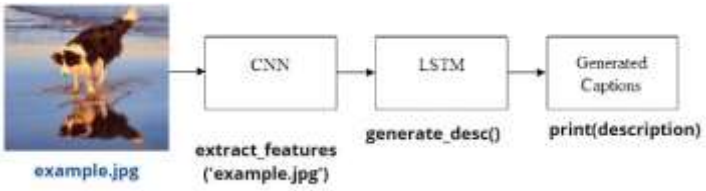

Figure. 12. Flow for Image Caption Generation

D) Tokenise descriptions and map it to numerical values:

The description text data is to be encoded into numerical values before it can be used in the model as an input or for comparison for the model's predictions. The steps for the same are as follows:

1. Converting the dictionary of clean data text into a list of descriptions.
2. Fit a tokenizer for the given captions
3. Obtain vocabulary size of the image
4. Calculating length of description with max words
5. Creating series of pictures, input descriptions and output statements for an image.
6. Define the captioning model of the input image.

E) Build and train image caption model:

1. Firstly, merging the VGG16 and LSTM Models and to train the model – GPU based machines on cloud. This is described in Figure(12)
2. Train model progressively.

F) Evaluate model using BLEU score and deploy:

In our image caption model, we have generated the captions programs as shown in Figure (13) and compared the description with the actual caption in order to come up with a BLEU score.

On merging VGG16 and LSTM deep learning techniques, we obtained an accurate bleu score. This entire flow is shown in Figure (14).

Figure. 13. Image Caption Generation using LSTM and VGG16

C. DRISHTI ANDROID APPLICATION:

The perception that people who have visual impairment can't use devices like smartphones, smartwatches and computers is now turning wrong with advancement in technology.

There are several applications now available to make work easy for them. But still using different applications becomes a confusing task for the visually impaired.

Hence, Drishti Android app is a service that helps blind and vision-impaired users interact with their devices in such a way that they don't have to browse their way to multiple applications. Drishti Android Application is a multi-feature application which proposes an android application, designed specifically for visually impaired individuals. It has spoken and audible feedback to your smart phone device. With the help of just this one single application people will be able to perform various functionalities.

As the application will be opened the user will be able to access the dashboard as shown in Figure (15). The main functionality here is that it can be accessed through voice- based input mechanism using a single click. The application is integrated with voice-based input /output mechanism. In the application the About section Figure (16) helps in understanding the features and functionalities.

Figure. 15. Dashboard of Drishti Application

Figure. 16. About Section of Drishti

A Voice-based email system that will help visually impaired people to access email in an easier manner can be seen in Figure(17). Together with providing usage of mail services simplicity and with efficiency.

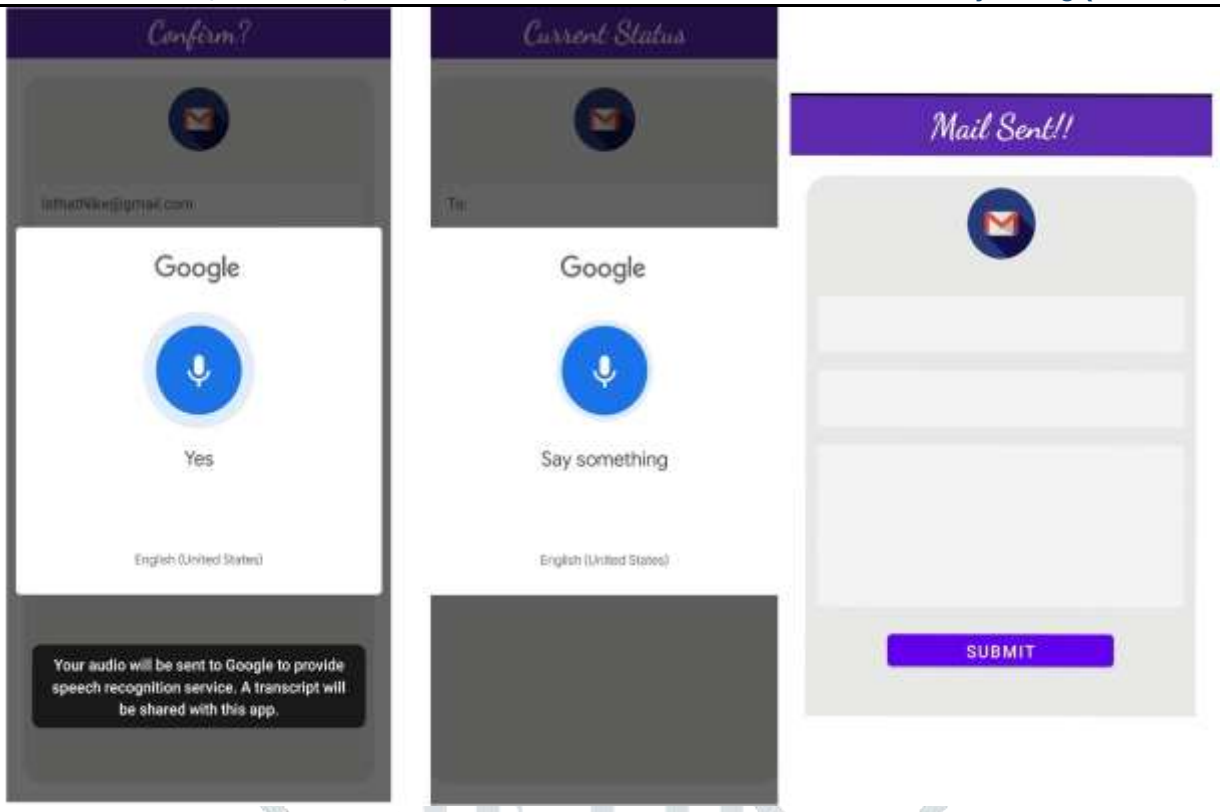

Figure. 17. Voice-based Email Feature

A Voice based calling system that will facilitate visually impaired people to access calling feature without facing any problem by their voice can be seen in Figure (18).

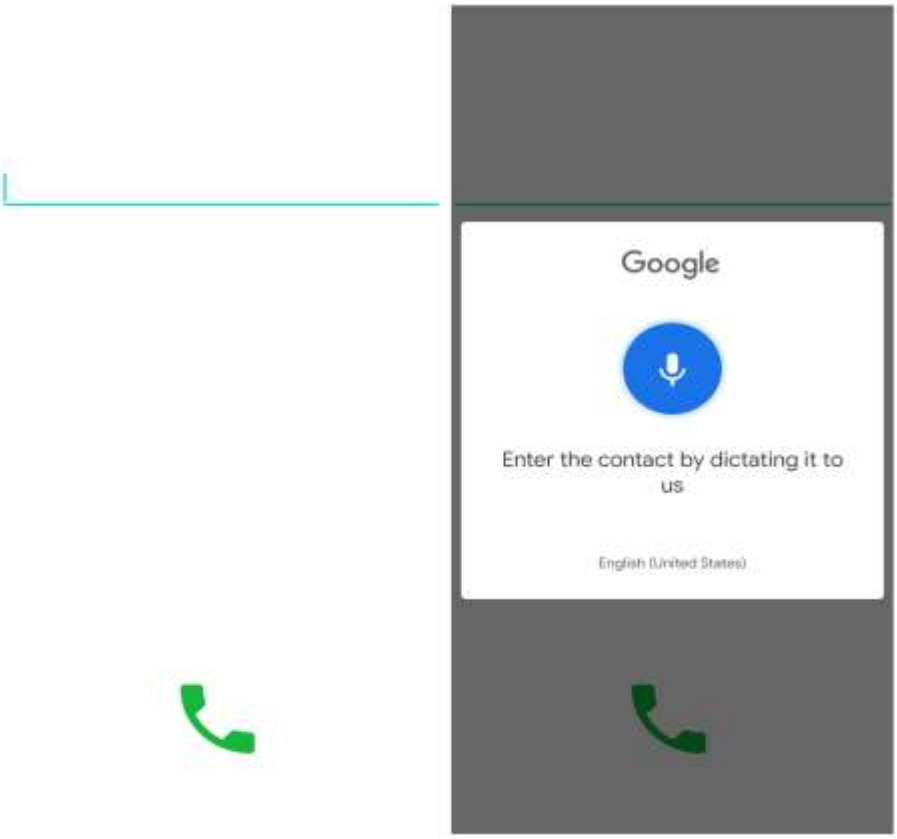

Figure. 18. Voice-based Calling Feature

Voice feedback of Current Time and Battery Level feature for the visually impaired people. As they cannot see the time or the battery percentage this feature will help them know it by a single action.

IV. RESULTS AND DISCUSSION

In this paper, we constructed and programmed working model of the Object Detecting Eye Wear and Smart Stick with Arduino UNO, Arduino Pro Mini, Ultra-Sonic Sensors, Buzzer, etc. which helps the visually impaired learn more about the surrounding and obstacles in path.

Additionally, we built an Image Caption Generation model for Assistive Vision by means of VGG-16 and LSTM on Flickr8k Dataset as shown in Figure (13). Image captioning is a multi-model technique has achieved the BLEU score which shows in Table1.

Table 1: Final BLEU Score Obtained

BLEU SCORE	STANDARD BLEU SCORE RANGE	OUTPUT BLEU SCORE
BLEU 1	0.401 to 0.578	0.541
BLEU 2	0.176 to 0.390	0.301
BLEU 3	0.099 to 0.260	0.209
BLEU 4	0.059 to 0.170	0.100

Also, we successfully developed a working Multi-functionality Android Application- Drishti which will help the visually impaired individual with its Voice-based Feedback Mechanism consisting of features like smart calling, voice based email, allowing user to identify the battery level and current timing, etc.

IV. CONCLUSION

One of biggest challenge for a visually impaired person, especially the one with complete loss of vision is to navigate around places safely without getting hit by an obstacle in path. Our object detecting Eye Wear and Smart Stick help overcome this problem with the help of Arduino and Sensors. This project serves as a helping hand by proper captioning of the given image. Lastly our Drishti Android Application provides users an environment where they can use multiple functionalities with a simple voice based system.

.

REFERENCES

[1] Shubham Melvin Felix, Sumer Kumar, A. Veeramuthu, “A Smart Personal AI Assistant for Visually Impaired People”, 2018 2nd International Conference on Trends in Electronic and Informatics(ICOEI), Tirunelveli, India. May 11-12, 2018,pp. 1245-1250.

[2] Aviral Chharia, Rahul Upadhyay, “Deep Recurrent Architecture based Scene Description Generator for Visually Impaired”, 2020 12th International Congress on Ultra Modern Telecommunications and Control System and Workshops(ICUMT), Brno, Czech Republic. 5-7 Oct. 2020, pp. 136-141.

[3] N. Komal Kumar, D. Veghneshwari, A. Mohan, K. Laxman, J. Yuvraj “Detection And Recognition Of Objects In Image Caption Generator System: A Deep Learning Approach”, 2019 5th International Conference on Advanced Computing Communication Systems (ICACCS), pp.107109.

[4] Sumitra A. Jakhete, Avanti Dorle, Piyush Pimplikar, “Object Recognition App for Visually Impaired”, 2019 IEEE Pune Section International Conference (PuneCon) MIT World Peace University, Pune, India. Dec 18-20, 2019, pp.1-4essing, Information Computing and Communication (ICGTSPICC), Jalgaon, 2016, pp.1-4.

[5] Kasthuri R, Nivetha B, Shabana S, “Smart Device for Visually Impaired People”, 2017 Third International Conference on Science Technology Engineering Management (ICONSTEM), 2017, Chennai, India, pp. 5459.

[6] Saeed Mian Qaisar, Raviha Khan, Noofa Hammad, “Scene to Text Conversion and Pronunciation for Visually Imapaired People”, 2019 Advances in Science and Engineering Technology International Conference(ASET), Dubai, United Arab Emirates, 2019, pp. 1-4.