International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April 2020

Real-Time Object Detection for Aiding Visually Impaired using Deep Learning

M. Vaidhehi, Varun Seth, Bhavya Singhal

***Abstract*: This research aims to create an assistive device for the people who are suffering from vision loss or impairment. The device is designed for blind people to overcome the daily challenges they face which may be perceived to be trivial to normal people. The device is created by using advance computer science technologies such as deep learning, computer vision and internet of things. The device created would be able to detect and classify daily objects and give a voice feedback to the user who is handicapped with blindness.**

***Keywords*: visually impaired, object detection, deep learning, computer vision, internet of things.**

I. INTRODUCTION

In a cumulative study till 2015, there were around 940 million people around the world who were suffering from a certain level of vision loss. Out of these 940 million people, 240 million had very low vision and 39 million were completely blind.

Visually impaired people face complex issues while performing tasks that may seem very trivial to normal people, such as looking for something, perceiving and identifying objects in the surrounding, navigating through a route indoors and outdoors as well. In general, avoiding obstacles and perceiving the elements in the surrounding is a challenge for them.

This project aims to create an assistive device using concepts of deep learning, computer vision and Internet of Things to help blind people overcome hurdles in their way. The device should be capable of classifying different objects and identify them in real-time and notify the object class to the user through voice feedback.

The opinion of the visually impaired personal is significantly important as they are the ones who will be using the device. Therefore the study follows the opinions of visually impaired. The aim of this project is to create a cost effective solution for the visually impaired people which would make their life much easier and can avoid any unfortunate events which may occur due to their handicap by using advanced computer science techniques like deep learning and image processing to provide the output expected in this study. In the following section, fundamentals will be explained. In section 3, previous studies similar to this project will be given. Implementation of the project is explained in section 4. The conclusion and further future works are discussed in the last section.

Revised Manuscript Received on April 18, 2020.
* Correspondence Author
M. Vaidhehi*, Assistant Professor, Dept. of CSE, SRM Institute of Science and Technology, Chennai, India. Email: m.vaidhehi@gmail.com
Varun Seth, Dept. of CSE, SRM Institute of Science and Technology, Chennai, India. Email: varunseth2303@gmail.com
Bhavya Singhal, Dept. of CSE, SRM Institute of Science and Technology, Chennai, India. Email: bhavyasinghal11@gmail.com

II. FUNDAMENTALS

This project is based on mainly three important computer science techniques, namely Deep Learning, Computer Vision and Internet of Things. They are further elaborated down below.

A. Deep Learning

Deep learning is a sub category of machine learning in Artificial Intelligence which utilizes multiple layers of neural networks to imitate human brain functions in processing data and creating patterns for critical decision making. Deep learning is a function of AI which consists of networks capable of unsupervised learning from dataset which is unstructured or unlabeled.

Deep learning has evolved in the digital era exponentially which has brought upon an explosion of data in all forms and from every region in the entire world. This data, known as the big data, can be obtained from online sources like the social media, e-commerce platforms, internet search engines, and online cinemas among other sources. This enormous amount of data is readily available and can be shared through applications like cloud computing.

However, the dataset used can be unstructured and in such an enormous size that it could take decades for a human to comprehend it and extract relevant information and trends.

Deep learning is an unsupervised learning algorithm which do not require data with desired variables and features. Instead it utilizes an iterative way to get the desired result. Deep learning usually works fine with big data.

Deep learning utilizes neural networks to scan the massive dataset to find patterns and fine correlations automatically. After the model is trained using the dataset, the learned associations are used to interpret new information. An article on Deep learning can be found on [8].

B. Computer Vision

Computer vision is technology which is capable of using machines to process, understand and analyze imagery (both photos and videos).

Although algorithms similar to computer vision have been around since the 1960's, but due to recent advancements in machine learning, as well as leaps forward in data-storage technologies, much improved computing capabilities, and cheap high quality input devices, there is a major improvement in how well the software can explore this kind of data.

Computer vision is a broader term for any computation involving visual content, such as images, videos, icons or pixels in general. But within computer vision, there are a few specific tasks that are the core building blocks.

In object classification, a model is trained on a dataset of specific objects. The trained model classifies the object belonging to one or more of the trained classes.

In object identification, since the model is trained on dataset of an object, it should be able to recognize a specific instance of the object. An article can be found on [9].

C. Internet of Things

Internet of Things, or IoT is a system of interconnected computing devices, sensors, domestic and industrial appliances or people with a unique identifier (UID). They can communicate and are capable of transferring data over a network without requiring human-to-computer or human-to-human interaction.

A thing in IoT can be any natural or man-made object that can be assigned an Internet Protocol (IP) address and is able to transfer data to other things in the network. Many organizations in the industry are using IoT to work more efficiently, understand customers to deliver better customer services, improve decision making to increase business value.

An IoT system consists of smart devices that use embedded systems, such as SOC, sensors and communicating hardware, to collect data from the environment, send them to other nodes in the ecosystem and act on the data accordingly. All the IoT devices share the sensor data by connecting to a gateway where the data is either sent to a cloud server to be analyzed or analyzed locally. This type system lets the devices do the most of the work without requiring much human interaction.

In this application, we will be using a microcontroller paired with a camera sensor to capture live video feed from the environment and analyze the data for object detection locally on the device itself. More information on the devices are discussed in upcoming sections.

III. RELATED WORK

There is not much product being developed using deep learning and image processing for aiding the visually impaired. Some of the significant related works are mentioned down below in this section.

In [1], an android mobile application is developed for visually impaired, where the live feed from the device's camera was fed as input to a trained machine leaning model. The model uses image processing and object detection to recognize the instances objects which were used to train the model in object classification. The aim of the application is to help visually blind to move easily and alert them if they run into an obstacle.

In [2], similar prototype device is developed which is equipped with binocular vision sensors. These binocular sensors capture images in a fixed frequency amongst which the most informative ones are selected through stereo image quality assessment (SIQA). These captures images are then sent to the cloud for further processing. The detection and automatic result will be provided through the use of Convolutional Neural Network based on big data. Through image analysis, cloud computing will return the requested information to the user, so that the user can make reasonable decisions in further actions.

In [3], a mobile alerting system is developed for visually weakened people which has a current location and construction jobs detection infrastructure which use web facilities on the internet to achieve so. Mobile devices and marked coordinates are determined through GPS feature. The aim of the mobile device is to help move visually impaired people through the distances according to marked coordinates.

In [4], this research aims to detect, estimate distance and relative position of visually blind people to the objects around them, particularly parked motorcycles. The proposed device uses Single Shot Multibox Detector (SSD) for detecting parked motorcycle by testing several different learning algorithms. This uses a pin-hole camera system to estimate distance of the motorcycle from 2 to 5 meters by comparing the actual motorcycle with the image of the motorcycle based on similar triangle principle.

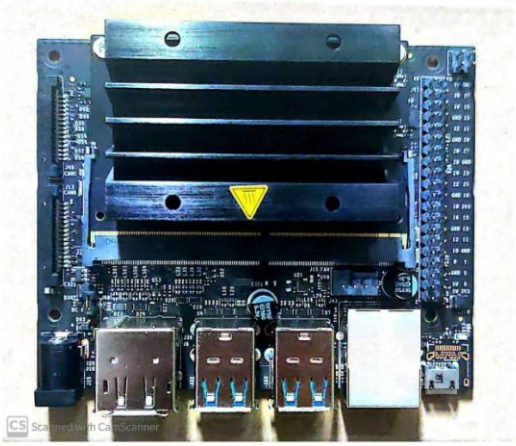
In [5], the research provides its relevant information such as size, and safety for grasping on a flat surface. The pipelines combine a series of the point cloud representation, table plane detection, objects detection and the full model estimation through a robust estimator. In this frame-work recent advantages of deep learning (e.g., RCNN, YOLO) are used that could be an efficient way for the detection task, while the geometry based approaches estimate full 3-D model. This study does not require separating (or segmenting) the interested objects from the background of the surrounding scenes.

IV. EXPERIMENTAL

The assistive device functions on two main pillars – deep learning algorithm used for object classification and detection, and the IoT devices such as a microcontroller capable of running neural networks efficiently and camera sensors to capture the surrounding. The deep learning algorithm will use computer vision which is available as a library in python programming language.

Following are the hardware components required for the device assembly:

- Nvidia Jetson Nano developer kit.
- Raspberry Pi camera version 2.
- USB to 3.5mm audio jack adapter.
- 64GB micro SD card.
- Power source of 5V/2A rating.
- Pair of earphones.



(a)

(b)

Fig. 1.(a). Top view of Nvidia Jetson Nano

Fig .1.(b). Ports of Nvidia Jetson Nano

A. IoT Device/ Microcontroller

For building an assistive device using IoT and deep learning, we need a base device which is capable of running various deep learning algorithms for object classification and detection. Nvidia Jetson Nano is a device which has enough computation power for implementing applications which involve image processing and object detection. That makes it the ideal choice as the base device.

Jetson Nano is small yet powerful device which is capable of implementing small, low-power AI systems and opens a world of innovation in embedded IoT applications.

Jetson Nano is capable of running multiple neural networks for the deep learning algorithms which we will be experimenting on it to find out its performance on different frameworks and datasets.

Jetson Nano has 128-core Maxwell GPU on-board which is a plus for applications which involve image processing. So we should get a decent level of performance considering the form factor of the device. Jetson Nano has support for 2 CSI cameras which enables the use of multiple cameras which can be used to implement a solution to determine the distance between the users and object.

The above topic is further discussed in future works. We are using the Raspberry Pi camera version 2 as it is the most recommended camera by Nvidia.

The above figure Fig. 1. Depicts the top and rear view of the Nvidia Jetson Nano developer kit. Its features and specifications are mentioned in detail down below.

Jetson Nano has the following features;

- Power Jack
- HDMI and display port.
- 4 x USB type A.
- Ethernet port.
- MicroUSB port.
- MicroSD card slot.
- Heatsink.
- 2 x CSI camera connectors.
- M.2 Key E slot.
- Power LED.

The specification of Jetson Nano is as follows;

1) **Graphics Processing Unit (GPU) :** Jetson nano is equipped with a 128 – core Maxwell architecture based GPU which is manufactures by Nvidia itself.
2) **Central Processing Unit (CPU) :** Jetson nano contains a quadcore ARM based CPU clocked at 1.43 GHz.
3) **Memory :** There are multiple configurations available for Nvidia Jetson Nano, the one used in this application contains 4 GB LPDDR4 RAM which is capable of 25.6 GB/s transfer rate.
4) **Storage :** Nvidia Jetson Nano uses a microSD card for storing media files.

B. Camera Sensor

Nvidia recommends for using the Raspberry Pi camera v2 if implementing applications which involve computer vision.

The camera has the following features:

1) The camera uses the Sony IMX219PQ image sensor – high-speed video imaging and high sensitivity.
2) It is equipped with a 8 Megapixel native resolution sensor capable of taking 3280 x 2464 pixel static images.
3) It supports 1080p30, 720p60 and 640x480p90 video recording.
4) It has a fixed focal length, which cannot be used to determine distance (Further discussed later).
5) It has support for Nvidia Jetson Nano and connects via a short ribbon cable.

Fig. 2. Raspberry Pi camera v2.

The base device has support for CSI cameras which are available in market in various variety and configuration. We went with the raspberry pi camera v2 as it was a cost effective solution for a prototype design.

But there are many options available which can be used as the camera sensor for the device such as a smart IP (internet protocol) camera which can be used wirelessly with the device, or a binocular sensors which can be used to calculate the approximate distance of the object from the user. A bi-focal lens camera or even a dual camera setup containing a depth sensor alongside the main camera would also enable the device for calculating distance.

Although these sensor configuration options would open more possibilities of applications, opting for them would increase the cost of the prototype design significantly. The Raspberry Pi camera v2 is perfect for testing the deep learning algorithm on the device. We will calculate the performance and efficiency of the device using this sensor and if we get acceptable performance, we can upgrade the camera setup in future.

C. Dataset and Algorithm

Deep learning is a subset of machine learning which utilizes multi-layered neural networks to recognise complex patterns and gain insights in big data. Deep learning is closely associated with Computer Vision as it enables applications such as image classification and object detection.

There are many deep learning algorithms such as RCNN, faster-RCNN, SSD, YOLO for object detection. In this study, YOLO, which stands for ***'you only look once'***, is used because it provides real-time object detection and gives more accurate results than other algorithms.

Some faster-RCNN algorithms may be superior to YOLO but they require very high computational power. Unlike faster-RCNN, YOLO runs on a single neural network. YOLO is capable of providing real-time object detection in less computation power.

The algorithm works like in the following procedure:-

1) YOLO takes real-time feed from an input device and analyse each frame. It divides entire frame into 'S' x 'S' grids. Each grid has a probability of enclosing single or multiple objects. These objects are to be bound under a box within each grid. Therefore, each grid may have B bounding boxes and probabilities of C trained classes in the model.
2) A confidence score is assumed (generally 40% or above) according to which bounding boxes are predicted against probability of C classes of objects. The predictions with invalid confidence score are not projected.
3) Each boundary box prediction includes 5 components: *x, y, w, h, and confidence*. Centre of the box is represented by (x, y) coordinates, it is relative to grid cell location. Bounding box width and height is represented by (w, h). x, y, w, h are normalized to fall between [0, 1]. In OLOv3 here are 80 class probabilities for each cell but only one class probability is predicted per cell. The final prediction is of the form S * S * (B * 5 + C).

The Fig. 3. Illustrates the working of YOLO and the intuition behind it. The latest release YOLOv3 is used in this application with COCO dataset which is preloaded with 80 classes of day to day objects such as a person, car, sofa, cell phone, etc.

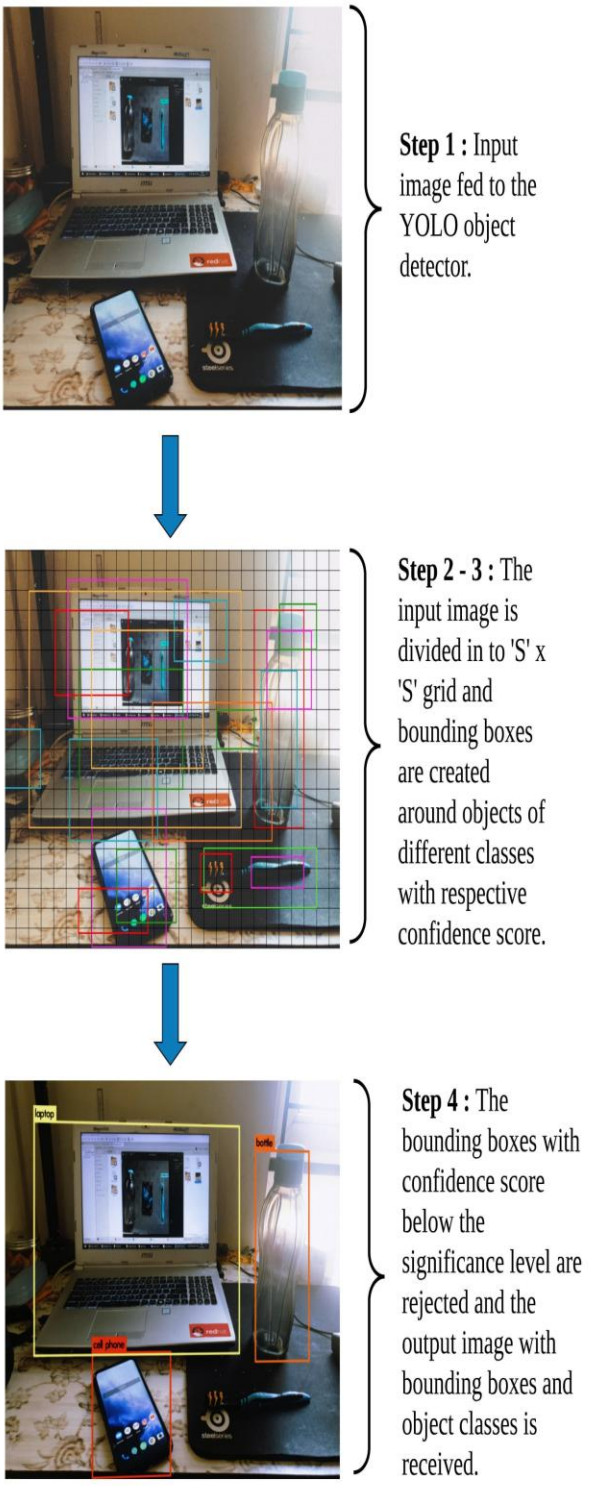


Fig. 3. YOLO Intuition.

V. IMPLEMENTATION

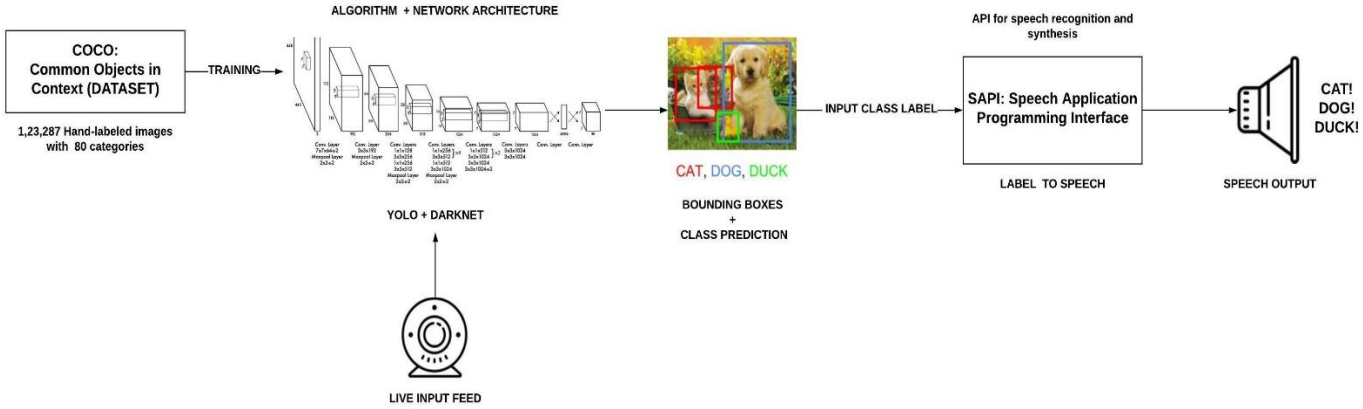


Fig. 4. Object classification and real-time detection process with voice feedback.

After setting up Jetson Nano, deep learning framework is installed on the device. The deep learning algorithms which will be used to test the device will work on their respective framework. We have intensively tested the device with a popular algorithm known as 'YOLO', as it only requires a single neural network to work. This means if we use YOLO, we should get decent performance from the device considering there is no cooling solution installed on the device such as a fan. YOLO works on a framework known as 'Darknet'. The implementation of the device is as follows.

A. Setting up Jetson Nano

After taking Jetson Nano out of its box, we need to install the Linux based Operating System provided by Nvidia. A microSD card (recommended 32 GB), power source and internet connection (optional) is required for installing the OS. Refer to the Nvidia Forum [10] for this step.

B. Installing darknet framework

Darknet is easy to install and provides optional dependencies – OpenCV and CUDA, in case the application requires wider range of supported image type and GPU support.

[6] contains the necessary files for installing darknet. After cloning, darknet directory can be accessed by the root user and the framework can be configured to use YOLO as per the application.

C. Acquiring datasets

For training our model we need datasets which contain labelled images of daily objects. Many datasets can be found on the internet or the dataset can be made manually by creating an image classifier.

The COCO dataset [11], is a clean and preprocessed dataset. It should train the model so that it gives us the most accurate results. We have acquired various datasets in this project and tested the different versions of YOLO against each dataset to get the performance of the device and accuracy of the model in terms of FPS and mAP values respectively. The link for acquiring datasets will mentioned.

Table. 1: Detection framework on COCO dataset.

Algorithm	Train	FPS	mAP
YOLOv3-416	COCO trainval	1.9	55.3
YOLOv2-416	COCO trainval	2.9	48.1
\YOLOv2-tiny	VOC 2007+2012	6.2	57.1
YOLOv3-tiny	**COCO trainval**	**12.8**	**33.1**

VI. RESULT DISCUSSION

The dataset (COCO trainval) which is used with YOLOv3-tiny algorithm on the Nvidia Jetson Nano has worked efficiently. We were able to obtain the best FPS value with the mentioned algorithm with minor tradeoff with accuracy. Even though the accuracy (mAP) is not better than other YOLO algorithms, it is well above the acceptable level and we are getting much better performance on the device. The algorithm and dataset used is mentioned in the table below.

Table. 2: The algorithm used.

Algorithm	Dataset	FPS	mAP
YOLOv3-tiny	COCO trainval	12.8	33.1

The efficiency and the performance of the device can be increased if we use a fan as a cooling solution for our device as thermal throttling can lead to performance loss. But it will become difficult to carry the device around. To solve this problem, a wireless sensor approach can be implemented with an IP camera (Further discussed in section VII).

VII. CONCLUSION AND FUTURE WORKS

This application aims to make the daily lives of visually impaired people more comfortable by creating a device with which they can interact with the environment. The prototype device built is semi-portable and the camera of the device acts as a pseudo eye for the visually impaired people. This work helped to get one step closer to creating a smart robotic eye which can be used in multiple applications.

We have used a ribbon-wired camera as our sensor for the prototype device. It is possible to use an IP camera or a smart binocular sensor for creating a wireless device sensor. This smart camera would send the live feed to the main device where it can be processed. The device would be kept stationary in close proximity and it won't be necessary to always carry it around. The sensor would communicate with the device wirelessly and would also be capable of calculating distance of the object from the user if a depth sensor is integrated alongside the main camera module.

The device application is majorly designed to help visually blind in transportation. It is planned to add GPS feature to the device which would take coordinates of current and destination location to determine the shortest path. The device would then navigate the blind user to the destination while constantly detecting and alerting about the surrounding so that there are lesser threats to the user.

REFERENCES

1. S. TOSUN and E. KARAARSLAN, "Real-Time Object Detection Application for Visually Impaired People: Third Eye," 2018 International Conference on Artificial Intelligence and Data Processing (IDAP), Malatya, Turkey, 2018, pp. 1-6. doi: 10.1109/IDAP.2018.8620773
2. B. Jiang , J. Yang, Z. Lv and H. Song, "Wearable Vision Assistance System Based on Binocular Sensors for Visually Impaired Users," in IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1375-1383, April 2019. doi: 10.1109/JIOT.2018.2842229
3. ÜNAL, E., & YÜCE, H. (2017). Development Of Mobile Warning System For The Visually Impaired People. Marmara Fen Bilimleri Journal, 3: 102-110, DOI: 10.7240/marufbd.298380
4. Indrabayu, N. L. Jamaluddin and I. S. Areni, "Detection and Distance Estimation against Motorcycles as Navigation Aids for Visually-impaired People," 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 2019, pp. 224-228. doi: 10.1109/ICTS.2019.8850936
5. V. Le, H. Vu and T. T. Nguyen, "A Frame-work assisting the Visually Impaired People: Common Object Detection and Pose Estimation in Surrounding Environment," 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, 2018, pp. 216-221. doi: 10.1109/NICS.2018.8606899
6. Joseph Redmon, Ali Farhadi - "Yolov3: An incremental improvement", arVix preprint arXiv:1804.02767, (Submitted on 8 Apr 2018)
7. 2018 J Redmon, A Farhadi - arXiv preprint arXiv:1804.02767, 2019
8. Marshall Hargrave, "Deep Learning" 2019, [Online]. Available: https://www.investopedia.com/terms/d/deep-learning.asp, (accessed on February 9, 2020)
9. Algorithmia, "Introduction to computer vision: what it is and how it works" 2018, [Online]. Available: https://algorithmia.com/blog/introduction-to-computer-vision, (accessed on February 9, 2020)
10. Nvidia, "Getting Started with Jetson Nano Developer Kit", [Online]. Available: https://developer.nvidia.com/embedded/learn/get-started-jetson-nano-devkit, (accessed on February 9, 2020)
11. "COCO – Common Objects in Context". [Online]. Available: http://cocodataset.org/#home, (accessed on February 9, 2020)

AUTHORS PROFILE

Ms. M. Vaidhehi has completed her Masters of engineering in computer science engineering. Currently she is working as an assistant professor in SRMIST Kattankulathur and has gained a teaching experience of 7 years. Her field of research is in machine learning, deep learning and artificial intelligence.

Varun Seth is a final year student at SRMIST Kattankulathur. He is pursuing the degree of B.Tech in the department of computer science engineering. He is a skilled programmer and passionate about technologies like machine learning and artificial intelligence.

Bhavya Singhal is a final year student at SRMIST Kattankulathur. . He is pursuing the degree of B.Tech in the department of computer science engineering. His research areas include Computer Vision and IoT. He has expertise in Linux system administration.