Smart AI for Visually Aided People

Ms. Nishigandha Shirsath[1], Ms. Tejal Bhoi[2], Ms. Shweta Gangurde[3], Ms. Sakshi Thete[4], Ms. Prajakta Wankhede[5]

[1] MARATHA VIDYA PRSARAK SAMAJ'S
KARMAVEER BABURAO GANPATRAO THAKARE COLLEGE OF ENGINEERING
Udoji Maratha Boarding Campus, Gangapur Road, Nashik, Maharashtra
nishigandhashirsath01@gmail.com

[2]MARATHA VIDYA PRSARAK SAMAJ'S
KARMAVEER BABURAO GANPATRAO THAKARE COLLEGE OF ENGINEERING
Udoji Maratha Boarding Campus, Gangapur Road, Nashik, Maharashtra
tejalbhoi27@gmail.com

[3]MARATHA VIDYA PRSARAK SAMAJ'S
KARMAVEER BABURAO GANPATRAO THAKARE COLLEGE OF ENGINEERING
Udoji Maratha Boarding Campus, Gangapur Road, Nashik, Maharashtra
gangurdeshweta@kbtcoe.org

[4]MARATHA VIDYA PRSARAK SAMAJ'S
KARMAVEER BABURAO GANPATRAO THAKARE COLLEGE OF ENGINEERING
Udoji Maratha Boarding Campus, Gangapur Road, Nashik, Maharashtra
thetesakshi@kbtcoe.org

[5]MARATHA VIDYA PRSARAK SAMAJ'S
KARMAVEER BABURAO GANPATRAO THAKARE COLLEGE OF ENGINEERING
Udoji Maratha Boarding Campus, Gangapur Road, Nashik, Maharashtra
Prajaktawankhede101@gmail.com

Abstract: The new generation of mobile phones has great tackle capability and briskly processing which is important enough to develop operations which help the stoner to connect and interact with the world at their own comfort zone. This system is an OCR reading system which uses camera operation present in your smart phones combined with OCR(Optical Character Recognition). OCR is a medium which coverts images of compartmented, handwritten, or published textbook into machine decoded textbook. This system will help you to take a picture or overlook the document present with stoner using the phone's camera, the image will be scrutinized and the operation will read the textbook written in English language and convert the affair in speech format. The speech affair is generated using Text To Speech Module. The purpose of delivering the affair in form of voice/ speech is to serve the information that's present on the document to the visually bloodied.

Keywords: Mobile Phone, Optical Character Recognition, Text to Speech, Visually disabled, Automated reading device

1. Introduction

In todays advanced hi- tech world, Due to lack of necessary information in the girding terrain visually disabled people face problems and are at disadvantage since visual information is what they warrant the most. With the help of the advanced technology, the visually aided can be supported. The idea is enforced through Android mobile app that focuses on voice adjunct, image recognition, currency recognition, e-book, converse bot etc. By WHO, 285 million people are prognosticated as visually bloodied worldwide. From that 39 million people are eyeless and 246 million people are having low vision. Reading is one of the main problems faced by eyeless people. With the help of the Braille, utmost of the accoutrements can be read by the blind. However, it's easy for them to read, If a person learns Braille. If they didn't learn the Braille, it's unfit to read. The trouble of Braille is that a mistake in understanding will bring about poring incorrect information and the other debit is that the books, documents etc. Have to be changed into a form of raised blotches for the eyeless to read. In Braille format, the books and papers which are available for the eyeless are fairly less when compared to the vast number of books which are published daily. To pierce the information in a textbook, a person needs a vision. A visually bloodied faces numerous troubles similar as alignment, focus, delicacy, mobility and effectiveness with penetrating published textbook using the being technology. Moment, there are many systems similar as portable bar law compendiums are designed to help bedazzle in order to descry the colorful products that helps the visually bloodied to gain the data about their products through communication and Braille. It's observed that they're still chancing it as a huge problem to move their routine, and it's primary to make introductory strides with the upcoming methods to assist them with living the present independent of their inability.

The app is capable to assist using voice command to recognize objects in the surrounding, do text analysis to recognize the text in the hard copy document. It may be the effective way blind people will interact with other people and may help blind people independent life.

2. Related Work

Since 2013 Singh during research finds E-accessibility as an agents for charge to transfer schools teaching practices and assist in learning for disabilities in creating skills that increases education to gain fulfilling employment. The aims to ensure the access for people with disabilities to new information communications technologies and system, including the internet. Method used by Author Singh are some assistive tools and technologies, such as different types of software better for E-accessibilities improve the functionalities of VI people Tools provide an opportunities to communicate, thus helping to gain access to education and skills needed for gainful employment.

During the survey in 2014 Laabidietal research were found to proposed a new approach for the development of accessible E-learning environments consist of three phases Design, Implementation, Validation system helps to generate content the meta-models for accessible the e-learning automatically. Aims to publish basic concepts of e- accessibility, universal design which focus on e-leraning systems. Methods are performed to identify conceptual abstract view which process to allow the translation conceptual model into specific code adaption to the user's individual learning context.

Author Ramos & de Andrade 2014 objective were study of schools for blind and partially sighted students that concentrate on human and material resources. This survey is adopted as a method of data collection founded the research teaching students with visual disabilities.

3. Methodology

3.1. Methodologies

Speech-enabled devices the progress made over the last few decades on Digital Signal Processing, speech processing algorithms enabled the integration of the three main speech technologies compression, synthesis and recognition on everyday devices and products. Embedded speech technologies are offered on both hardware:

1. Talking devices, 2. Speech-activated devices, 3. Spoken dialogue-based devices.

• Camera

The inbuilt camera application in any smart phone is used to capture the image of the document which is to be read. The image will contain textual regions from which the text will be recognized.

OCR module

OCR module that is Optical Character Recognition module is used to preprocess the image and detects and identify the words that in English language. After capturing of the image in standard resolution, the textual regions within the image are localized. The system is only concerned with the textual regions and complex backgrounds are not within the scope of the project. The captured image is processed and the first step in the processing is localization of the textual regions in the image. The identified characters are converted into text that is machine encoded and displayed on the screen with the help of this module.

Text To Speech Module

This module is used to give speech output of the converted text from the image. This module thus helps to hear whatever is printed or written on the scanned document.

User Interface Module

User Interface module will provide features such as a login page, a registration page and also will also display the result of detected text on the screen and give a voice output which is convenient for the end user.

Object Detection Module

It detects all the objects in front of the camera. Firstly, we need to give command as “Detect Object”. Then the camera will capture all the images and will give output with maximum accuracy of object detection.

3.2. Algorithms

Different types of Neural Networks are used for different purposes, for illustration for prognosticating the sequence of words we use intermittent Neural Networks, also for image bracket we use Convolution Neural networks. Convolutional neural networks are distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They've three main types of layers, which are

• Convolutional layer
• Pooling layer
• Completely- connected(FC) layer

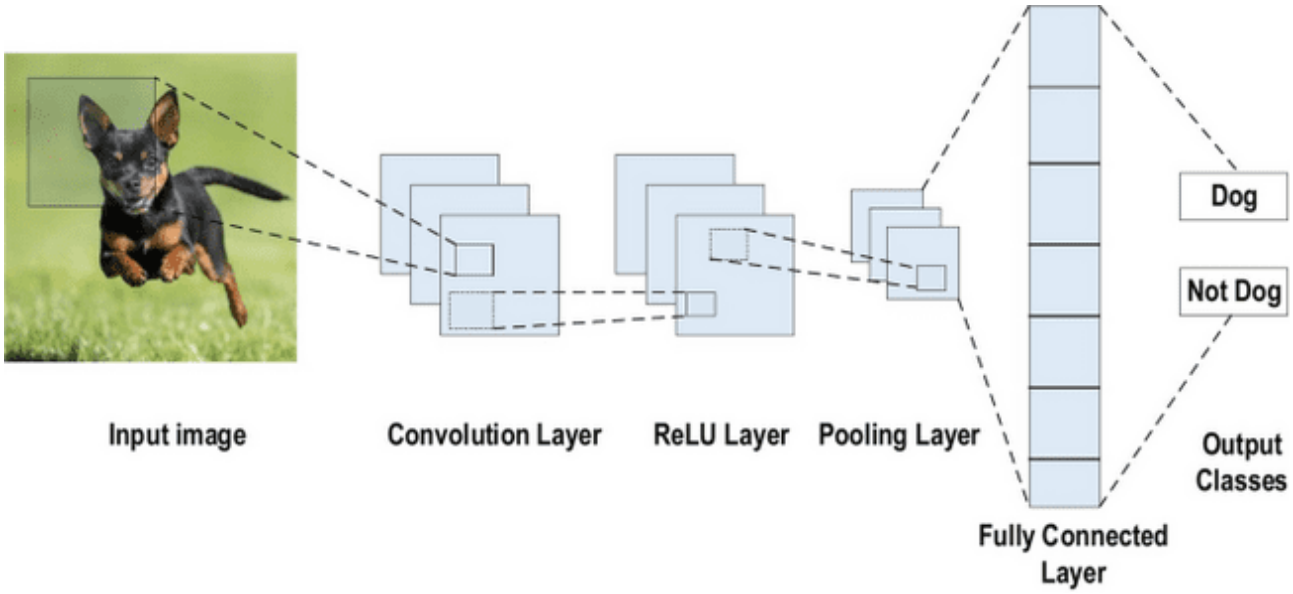


Fig: Implementation of CNN Architecture

1. Convolutional Layer
The convolutional Layer is the core structure block of a CNN, and it's where the maturity of calculation occurs. It requires many factors, which are input data, sludge, and a point chart. Let's assume that the input will be a color image, which is made up of a matrix of pixels in 3D. This means that the input will have three confines - a height, range, and depth which correspond to RGB in an image. We also have a point sensor, also known as a kernel or a sludge, which will move across the open fields of the image, checking if the point is present. This process is known as a complication.

2. Pooling Layer
Pooling layers, also known as down slice, conducts dimensionality reduction, reducing the number of parameters in the input. Analogous to the convolutional layer, the pooling operation sweeps sludge across the entire input, but the difference is that this sludge doesn't have any weights rather; the kernel applies an aggregation function to the values within the open field, colonizing the affair array. There are two main types of pooling

a) Max pooling as the sludge moves across the input, it selects the pixel with the maximum value to shoot to the affair array. As an away, this approach tends to be used more frequently compared to average pooling.

b) Average pooling as the sludge moves across the input, it calculates the average value within the open field to shoot to the affair array. While a lot of information is lost in the pooling layer, it also has a number of benefits to the CNN. They help to reduce complexity, ameliorate effectiveness, and limit threat of over fitting completely

3. Connected Layer
The name of the full- connected layer aptly describes itself. As mentioned before, the pixel values of the input image aren't directly connected to the affair layer in incompletely connected layers. still, in the completely-

system using OCR is an artificial intelligence reading system developed using a smart phones camera combined with connected layer, each knot in the affair layer connects directly to a knot in the former layer.

This layer performs the task of bracket grounded on the features uprooted through the former layers and their different pollutants. While convolutional and pooling layers tend to use ReLu functions, FC layers generally work a SoftMax activation function to classify inputs meetly, producing a probability from 0 to 1.

4. System Architecture

A system armature is the abstract model that defines the structure, gets, and further views of a system. (1) An armature description is a formal description and representation of a system, organized in a way that supports logic about the structures and actions of the system. A system armature can correspond of system factors and the sub-systems developed, that will work together to apply the overall system. There have been sweats to formalize languages to describe system armature, inclusively these are called armature description languages (ADLs) colorful associations can define systems armature in different ways, including the abecedarian association of a system, embodied in its factors, their connections to each other and to the terrain, and the principles governing its design and elaboration.

A representation of a system, including a mapping of functionality onto tackle and software factors, a mapping of the software armature onto the tackle armature, and mortal commerce with these factors. An allocated arrangement of physical rudiments which provides the design result for a consumer product or life- cycle process intended to satisfy the conditions of the functional armature and the conditions birth. An armature consists of the most important, pervasive, top- position, strategic inventions, opinions, and their associated accounts about the overall structure (i.e., essential rudiments and their connections) and associated characteristics and gets. A description of the design and contents of a computer system. However, it may include information similar as a detailed force of current tackle, software and networking capabilities; a description of long- range plans and precedence's for unborn purchases, if proved. A formal description of a system, or a detailed plan of the system at element position to guide its perpetration. The compound of the design infrastructures for products and their life- cycle processes.

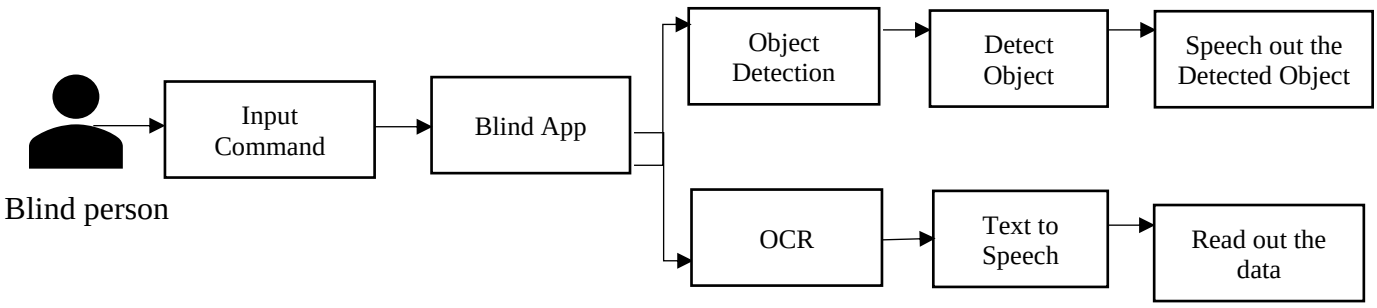


Fig: Proposed Architecture System

5. Implemented System

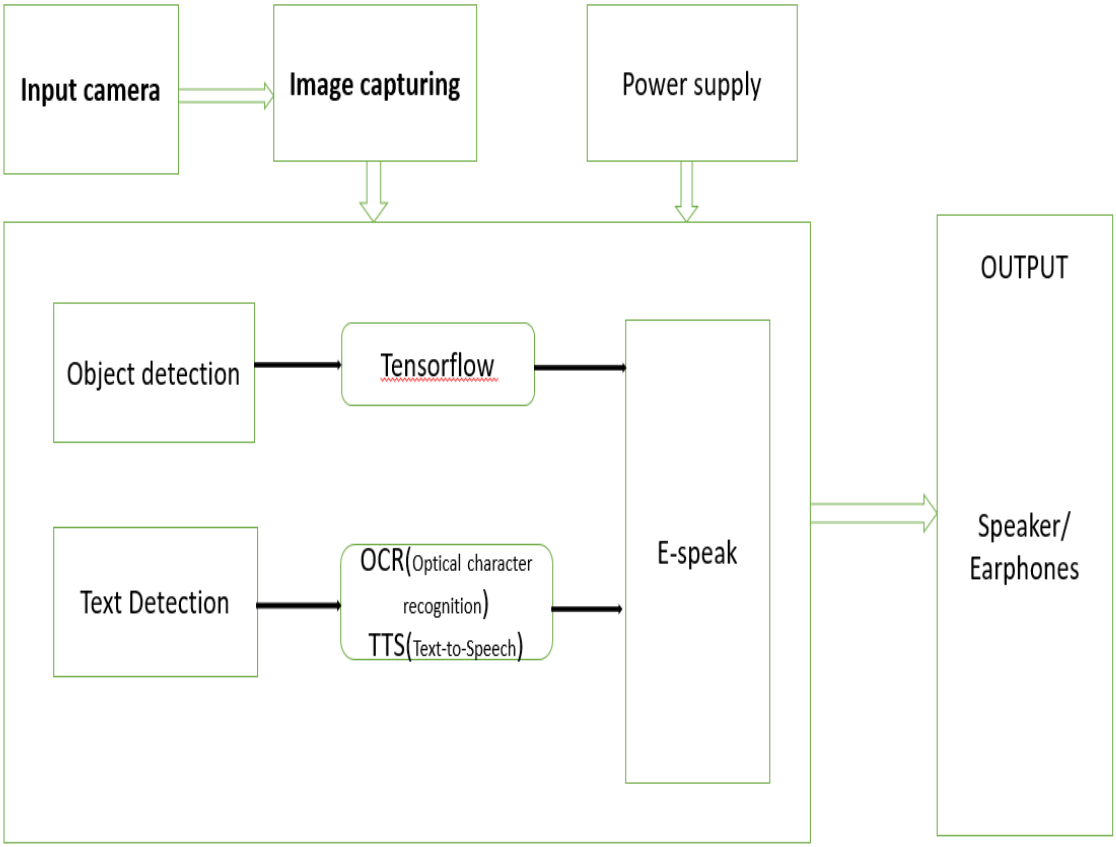
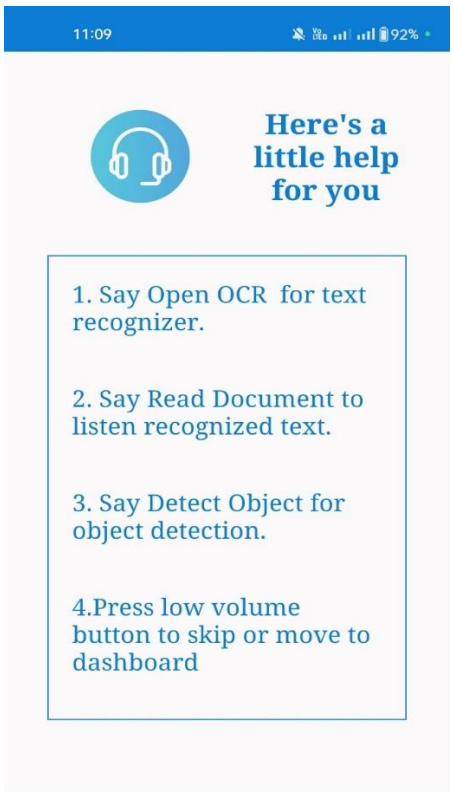


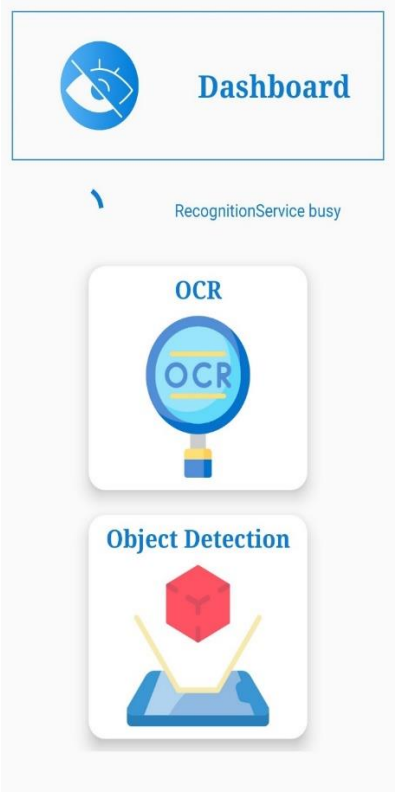
Fig: Implementation of App

1. The input is taken in the form of an image captured from the web camera.
2. The image is then processed either for the purpose of text reading or for object detection based on user choice.
3. The text reading is supported by software named OCR.
4. The read text is changed into an audio output using the TTS Synthesis.
5. For object detection, the input image is broken down into several components. Then it draws boxes bound together around the segments, spanning the entire input image.
6. Each broken-down component follows the process of feature extraction as the model begins to intensify. If there are any visual features in any bounding boxes, the model predicts the objects in that bounded box.
7. Finally, the output is converted into an audio signal.

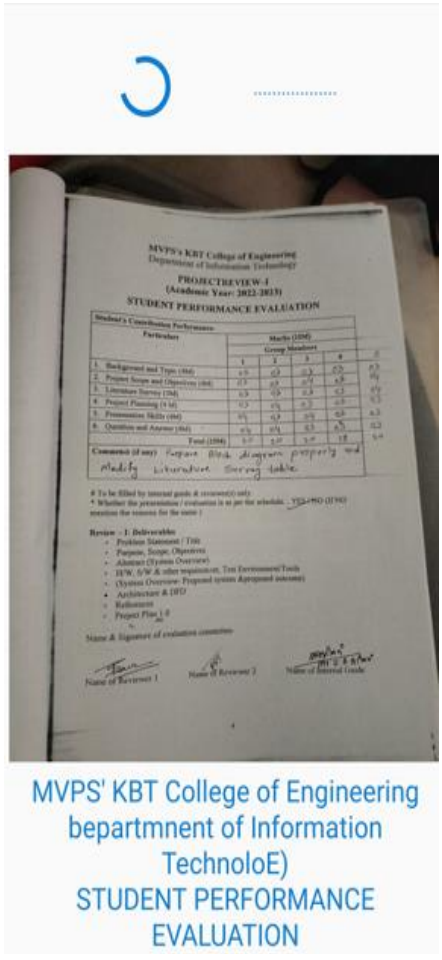
1. Open the app



2. Command Open OCR



3.Scan the document



4.For object detection give Detect Object command

5. Blind Hand

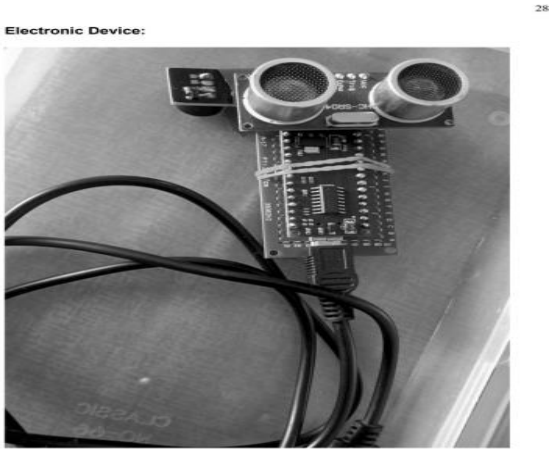


Fig: Screenshots of Implemented System

6. Conclusion

A visually challenged people need an adjunct to carry on work on diurnal base. We've bandied the challenges faced by eyeless people and tried to give a satisfactory result to them for working everyday life.

Our operation will be suitable to fetch the textbook captured by a mobile phone camera, display the restatement affect back onto the screen of the mobile phone, and produce the speech of the restated textbook. It will detect the object in front of the blind person with maximum efficiency. It will help the blind person in his/her daily life to overcome the difficulties.

7. References

[1] Christopher G Relf, "Image Acquisition and Processing with Lab VIEW", CRC Press, 2004

[2] Raviraj S Patkar , S. Pratap Singh , Ms. Swati V. Birje, "Marker Based Augmented Reality Using Android OS", International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 3, Issue 5, May 2013.

[3] World Health Organization, “Visual Impairment and Blindness,” WHO Factsheet no. FS282 , Dec. 2014.

[4] Shagufta Md.Rafique Bagwan, Prof. L.J.Sankpal,” VisualPal: A Mobile App for Oject Recognition for the Visually Impaired”, IEEE International Conference on Computer, Communication and Control (IC4-2015).

[5] Jisha Gopinath, Aravind S, Pooja Chandran, Saranya S S, "Text to Speech Conversion System using OCR", International Journal of Emerging Technology and Advanced Engineering , Volume 5, Issue 1, January 2015.

[6] Aaron James S, Sanjana S, Monisha M, "OCR based automatic book reader for the visually impaired using Raspberry PI", Vol. 4, Issue 7, January 2016.

[7] Shahed Anzarus Sabab, Md. Hamjajul Ashmafee, “Blind Reader: An Intelligent Assistant for Blind”, 19th International Conference on Computer and Information Technology, December 18- 20, 2016, North South University, Dhaka, Bangladesh.

[8] Ashwani Kumar, Ankush Chourasia, "Blind Navigation System Using Artificial Intelligence", March 2018

[9] Liang – Bi Chen, Ming-Che Chen, "An implementation of an intelligent assistace system for visually impaired/blind people, "IEEE, 2018

[10] N. Senthil kumar, A. Abinaya, E. Arthi, M. Atchaya, M. Elakkiya, "SMART EYE FOR VISUALLY IMPAIRED PEOPLE", International Research Journal of Engineering and Technology, Volume: 07 Issue: 06, June 2020.

Author Profile

Tejal Bhoi -- Currently pursuing undergraduate graduation in information technology from MVP'S KBT Collage of Engineering, Nashik, Maharashtra, India

Shweta Gangurde -- Currently pursuing undergraduate graduation in information technology from MVP'S KBT Collage of Engineering, Nashik, Maharashtra, India

Sakshi Thete -- Currently pursuing undergraduate graduation in information technology from MVP'S KBT Collage of Engineering, Nashik, Maharashtra, India

Nishigandha Shirsath -- Currently pursuing undergraduate graduation in information technology from MVP'S KBT Collage of Engineering, Nashik, Maharashtra, India

Prajakta Wankhede -- Currently pursuing undergraduate graduation in information technology from MVP'S KBT Collage of Engineering, Nashik, Maharashtra, India