

## Voice For B

Shiva Harish Kompelli (21MCME30)

Guided by: Dr. Salman Abdul Moiz

**Abstract**  
This paper presents a method for converting mathematical symbols captured in images into spoken words through database integration. The proposed system employs a multi-stage approach, beginning with optical character recognition (OCR) to extract symbols from images. Following this, a symbol recognition module utilizes machine learning algorithms to identify individual mathematical symbols accurately. A central database is then employed to map recognized symbols to their corresponding spoken representations. This database serves as a key component, containing entries for various mathematical symbols and their phonetic equivalents. The system further integrates a text-to-speech (TTS) engine to synthesize the recognized symbols into spoken words based on the database lookup. Python libraries such as OpenCV, Tesseract-OCR, and TTS engines are utilized for implementation. Challenges such as handling diverse fonts, styles, and languages are addressed to ensure robust symbol recognition. The proposed system offers a valuable tool for individuals with visual impairments or those seeking auditory assistance in comprehending mathematical content presented in images.

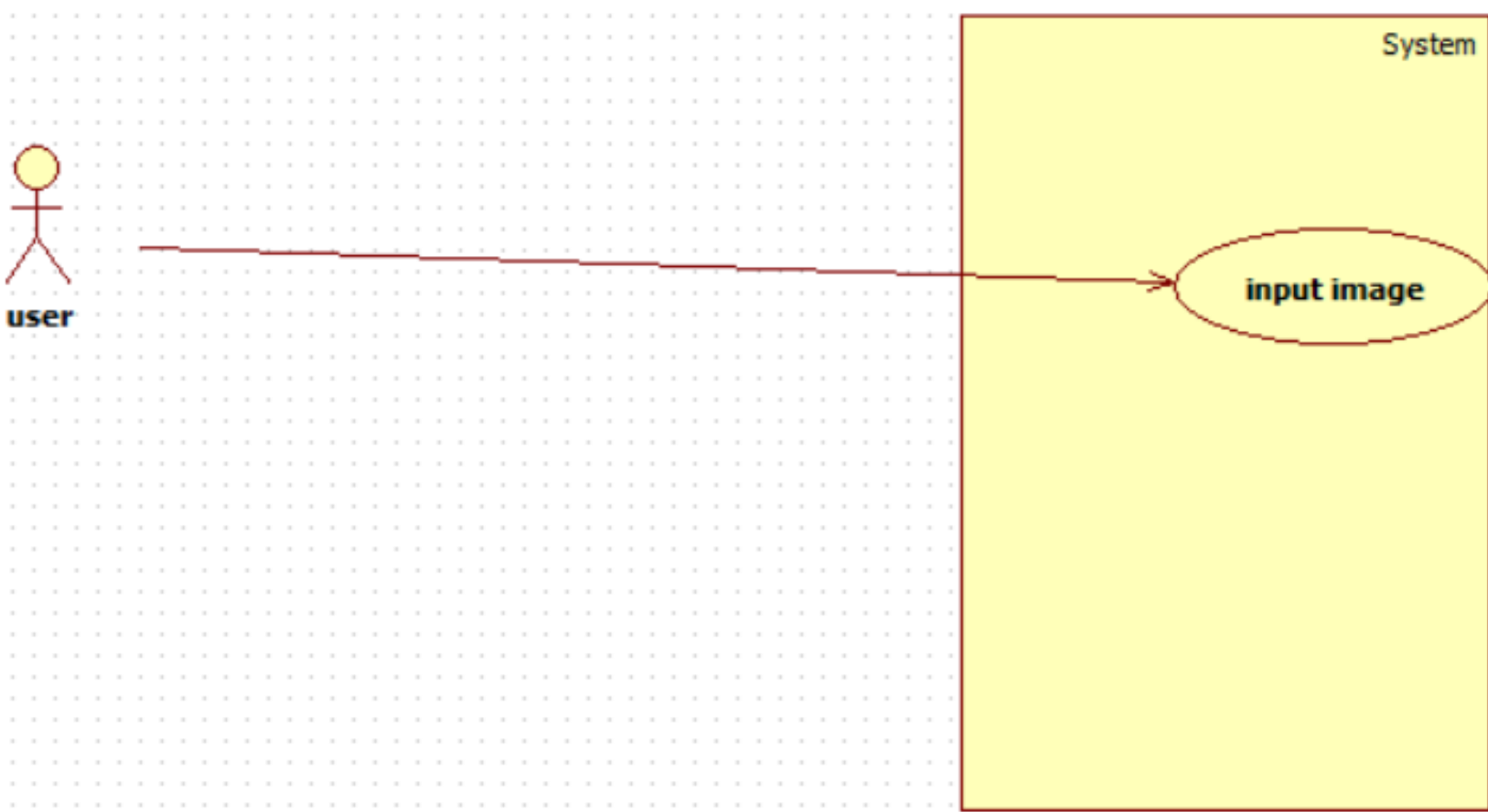
**Problem Statement**  
Symbol recognition in an image and convert it to voice output for visually challenged students.

### System Design-

- Tools used for webpage**
- 1) OCR-Tesseract
  - 2) Python3
  - 3) MYSQL (Database System)
  - 4) HTML
  - 5) CSS
  - 6) Javascript



### Use Case Diagram



**Conclusion:**  
*This design when implemented with the given future enhancements will ease the efforts of blind people in hearing the voice output of the symbols along with that text that present in an image.*

### Future Enhancements

- 1) Provide clear and concise voice guidance throughout the interaction with the site, guiding the user on how to capture images, and interpret results.
- 2) Include audio descriptions of the captured image before and after processing to provide context for blind users. Describe the layout, orientation, and content of the image to help users understand what they are interacting with.
- 3) Enable voice command functionality to allow users to interact with the system using natural language. This can include commands for uploading images, initiating the conversion process and controlling the playback of the spoken output.
- 4) It also extracts the text part along with the symbols part and reads the whole image (upto 98%).
- 5) The user can use his own voice to listen the voice output in his own voice, by taking input of his voice audio file as sample.

### References

- 1. <https://www.simplilearn.com/tutorials/python-tutorial/speech-recognition-in-python>
- 2. [https://commons.wikimedia.org/wiki/File:Tesseract\\_OCR\\_logo\\_\(Google\).png](https://commons.wikimedia.org/wiki/File:Tesseract_OCR_logo_(Google).png)
- 3. <https://www.onlineocr.net/>
- 4. <https://pypi.org/project/pytesseract/>