Form 4: Results and conclusion

1. **Team Number:** 20
2. **Project Title:** Multi-Modal Assistive System for people with disabilities
3. **Experiment Environment:** The platform employed for the creation and execution of the code in visual studio code, google colab, jupyter notebook. Flask, HTML, CSS and JavaScript has been used in the creation of the web interface.

Libraries Used: Flask, SQLAlchemy, Tensorflow, speech_recognition, keras,gtts, pillow, pytesseract.

4. Parameters:

1.Sign language translation:

1.The feature extraction in TSM is calculated with the following equation: $Ht=\sum j\sum kWi[j,k]At[a-j,a-k]\backslash$

2.The feature map Z in TSM is calculated with Equation:

:$Z=Ht+Ht+1=\sum j\sum kWA[j,k]At[a-j,a-k]+WB[j,k]At+1[a-j,a-k]$3.The feature map Y in TSM is calculated with Equation

(3): $Y=OutputTSM=\sum l=1c-lZi\&\sum c-lcHt$

2.Visual Question Answering on Images:

1.Learning rate decay: w_i^(t+1) w_i^(t)-α*∇L(w)/∇w_i^(t)

2.Adam optimizer: w_{t+1} = w_t - α * m_t / (√(v_t + ε))

3.OCR- powered image-to-speech and speech-to-text:

1.Kernel Function : f (x) = sgn(X 1 i=1 αiyiK(xi, x) + b)

2.Finding the probability of nearest sample : p(y|q) = P k∈K Wk .1(ky=y)/ P k

5. a) Experiment 1:

I. Visual question and answering

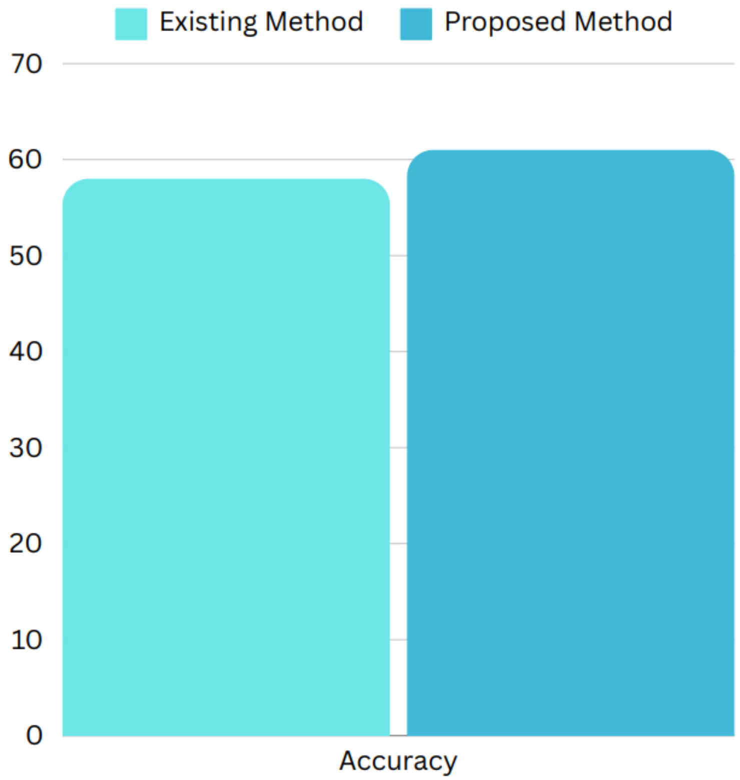
Input: Image, String.

Output: String generated by VQA model

Experiment Analysis:

	Existing Method	**Proposed Method**
Accuracy	58%	61%

Graph:



Findings: The proposed method has highest accuracy with best confidence.

II. **Text/Speech to Sign:**

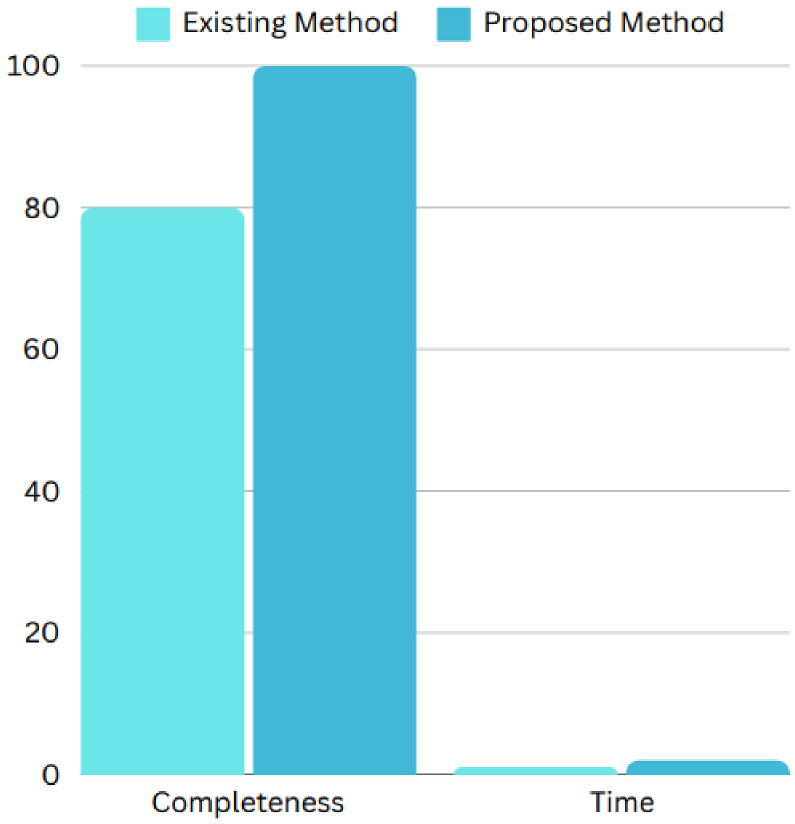
Input: String.

Output: A video consisting sign language

Experiment Analysis:

	Existing Method	**Proposed Method**
Completeness	80%	100%
Time	2 sec	1 sec

Graph:



Findings: The proposed method is both completeness and time efficient compared to existing system

III. Image to Speech:

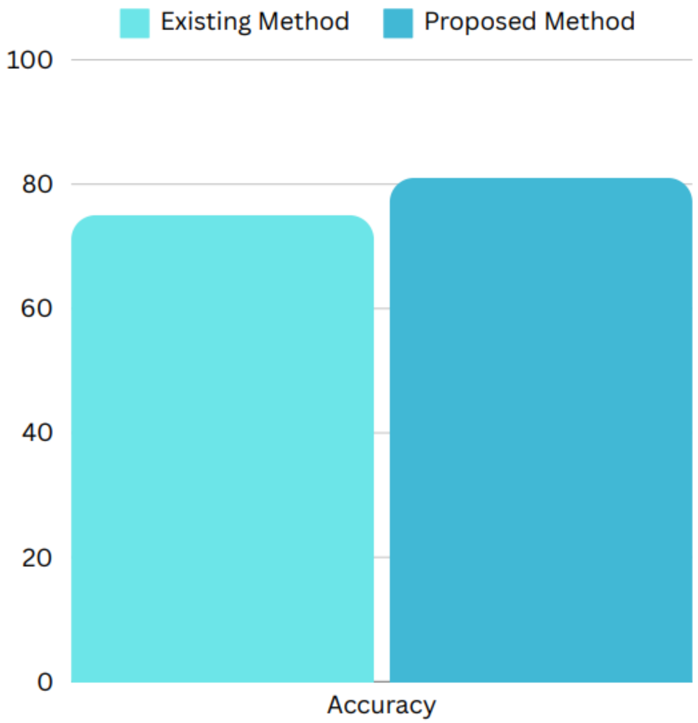
Input: Image consisting of written or typed text.

Output: Audio consisting of recognized text from image.

Experiment Analysis:

	Existing Method	**Proposed Method**
Accuracy	75%	80%

Graph:



Findings: The proposed method accuracy is increased as we used Pre-processing the Input Image and combination of pytesseract with deep learning for text recognition.

4. Parameter comparison table

Parameter	Previous methods	Proposed method
Accuracy (VQA)	In the previously used method, the accuracy is low	The accuracy is increased as we used transformers with customized learning rate and regularization
Accuracy (image to speech)	In the previously used method, the accuracy is low	The accuracy is increased as we used Pre-processing the Input Image combination of pytesseract with deep learning for text recognition
Time	In the previously used method, the generation time of video is low	The generation time of video is fast as we load the videos from the existence
Completeness	In the previously used method, for some sentences they could not generate the complete sign representation	Our method can completely generate any sign video for any sentence as we generate video by combining the individual letters or words of the sign representations

5. Final Conclusion Statements:

In summary, the project integrates visual question answering, sign language recognition, and text/image-to-speech conversion to enhance accessibility for individuals with visual and auditory impairments. Compared with existing systems, notable improvements in accuracy have been achieved across individual modules. This advancement holds promise for addressing communication barriers and improving the overall user experience for the blind and deaf community. Moving forward, continued refinement and optimization of these technologies are essential to furthering accessibility standards and empowering individuals with disabilities to participate more fully in society.

Signature Supervisor
G. Kiran Kumar
(Asst Professor)