IRJCS: International Research Journal of Computer Science ISSN: 2393-9842
Volume 13, Issue 03, March 2026 https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.33

Sign Language Translator

Perumala Maheshraj
Assistant Professor, Department of CSE
Sengunthar Engineering College (Autonomous), Tiruchengode,India
perumalamaheshraj.cse@scteng.co.in
https://orcid.org/0009-0001-9953-0882
Sudhar Shan K,suganeshwaran K,Raman R
UG Students, Department of CSE
Sengunthar Engineering College (Autonomous), Tiruchengode, India
sudharshanrohit45@gmail.com,ksuganesh03@gmail.com,ramanarrramana5@gmail.com

Publication History
Manuscript Reference: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10112
Research Article | Open Access | Double-Blind Peer Reviewed Article ID: IRJCS/RS/Vol.13/Issue03/CSMR26.MRCS10112
Received: 30, January 2026, Revised: 13, February 2026, Accepted: 28 February 2026 Published Online: 25 March 2026
https://www.irjcs.com/volumes/Vol13/iss-03/33.CSMR26.MRCS10112.pdf
Article Citation: Perumala,Sudhar,suganeshwaran,Raman(2026),Sign Language Translator, IRJCS: International Research Journal of Computer Science, Volume 13,Issue 03 of 2026 pages 292-295
Doi:> https://doi.org/10.26562/irjcs.2026.v1303.33 **BibTeX Key Perumala@2026Sign**
Orcid: https://orcid.org/0009-0004-9398-7488
IRJCS papers should be cited as IRJCS (International Research Journal of Computer Science, AM Publications, India 2026, ISSN 2393-9842, https://doi.org/10.26562/ijrcs.2025.v1303.33 The journal's official abbreviation is IRJCS.

Abstract: This paper presents a bidirectional Sign Language Translator that enables two modes of communication: (1) Sign- to-Text, which uses a Convolutional Neural Network (CNN) to recognize American Sign Language (ASL) hand gestures captured from a webcam and convert them into readable text; and (2) Speech-to-Sign, which decodes spoken English using a Hidden Markov Model (HMM), validates the decoded text using Natural Language Processing(NLP) against a standard English dictionary, and maps it to corresponding ASL sign gestures. The CNN model is trained on the ASL dataset containing letter and word gestures. The complete system is implemented in Python using deep learning libraries. This academic project provides a practical, affordable, and real-time communication bridge for the hearing-impaired community.
Keywords: Sign Language Recognition, Convolutional Neural Network, Hidden Markov Model, Natural Language Processing, ASL Dataset, Python, Deep Learning, Accessibility.

I. INTRODUCTION

Sign language is the primary mode of communication for over 430 million hearing-impaired individuals worldwide [1]. The inability of hearing people to understand sign language creates a persistent communication barrier in education, health care, and daily life. This project addresses the problem through a two-path bidirectional translator. In Path1 (Sign-to-Text),the system captures hand gestures from a webcam, detects 21 hand landmarks using MediaPipe, and classifies them using a trained CNN model built on the ASL dataset. In Path 2 (Speech-to-Sign), spoken English is captured via microphone, processed using MFCC feature extraction, decoded using an HMM, validated through NLP, and mapped to ASL sign gestures. The entire system is implemented in Python using Tensor Flow/Keras for deep learning, hmm learn for the HMM, and NLTK for NLP. This academic project demonstrate show multiple AI techniques CNN, HMM, and NLP can be combined in a single, practical accessibility application.

II. LITERATURE REVIEW

Vision-based sign language recognition using CNNs was popularized by Pigouetal.[2], who showed that convolutional networks effectively learn spatial gesture features from image data. The ASL alphabet dataset has since become a standard benchmark for academic SLR systems. For speech recognition, Hidden Markov Models have been the foundational technique since Rabiner's seminal work [3]. HMMs model temporal sequences of MFCC features and remain computationally efficient for discrete word recognition, making them suitable for academic implementations. NLP-based grammar validation in assistive communication tools was demonstrated by Koul et al. [4], showing that dictionary-based validation significantly improves translation accuracy. Our project integrates all three approaches into a unified Python system.

III. SYSTEM OVERVIEW

A. Dual-Path Architecture

The system operates in two independent paths sharing a common Python backend. Figure 1 shows the complete system architecture with both paths. The user selects the active mode through a simple interface.

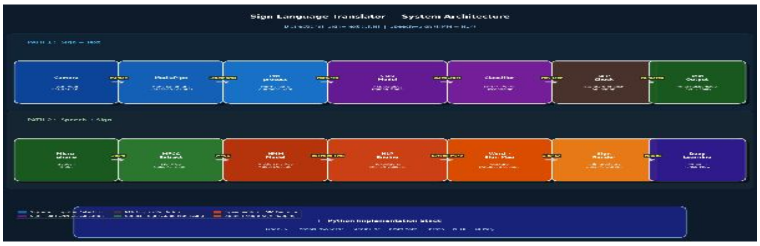
B.Path1: Sign toText(CNN)

The webcam captures continuous video frames. MediaPipe extracts 21 3D hand landmark coordinates per frame (63 features total). These are normalized, augmented during training, and input to a CNN classifier that recognizes ASL letters (A–Z) or common words. The recognized text is displayed on screen in real time.

ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.33

C. Path 2: Speech to Sign (HMM+NLP)

A microphone records speech input. MFCC features are extracted using librosa. A pre-trained HMM decodes the MFCC sequence into text. The decoded text is then passed through an NLTK-based NLP module that checks it against a standard English dictionary. Validated words are mapped to their ASL gesture sequences from a pre-built library and displayed as animated signs.



IV. DATAFLOW DIAGRAM

Figure2 shows the complete Level-1 DFD for both paths. Path 1 includes processes P1–P7 with data stores DS1–DS3 (ASL dataset, CNN model weights, English dictionary). Path 2 includes P8–P14 with DS4–DS6 (HMM model, word-to-sign mapping, gesture library).

Fig.2. Data Flow Diagram(Level-1)—Both Paths

V. CNN MODEL—SIGN TO TEXT (PATH1)

A. ASL Dataset

The CNN model is trained on the ASL Dataset containing 87,000 images spread across 26 alphabet classes (A–Z) and 10 common word gestures. Data augmentation (rotation ±15°, horizontal flip, brightness variation) quadruples the effective training set.

TABLE I – ASL Dataset Summary

Layer	Configuration	Output Shape
Input Layer	63 features (21×3)	(None, 63)
Conv2D Block 1	32filters, 3×3,ReLU	(None, 64,64, 32)
BatchNorm + MaxPool	Pool 2×2, Stride2	(None, 32,32, 32)
Conv2D Block 2	64 filters, 3×3,ReLU	(None, 32,32, 64)
Conv2D Block 3	128 filters, 3×3,ReLU	(None, 128)
Dense Layer	128units,ReLU	(None, 128)
Dropout	Rate=0.5	(None, 128)
Softmax Output	36 classes	(None, 36)

B. CNN Architecture

Figure 3 shows the CNN architecture used for sign classification. The model uses three convolutional blocks with increasing filter depth, followed by fully connected dense layers and a Softmax output over 36 classes.

IRJCS: International Research Journal of Computer Science
Volume 13, Issue 03, March 2026
ISSN: 2393-9842
https://www.irjcs.com/archives
https://doi.org/10.26562/irjcs.2025.v1303.33

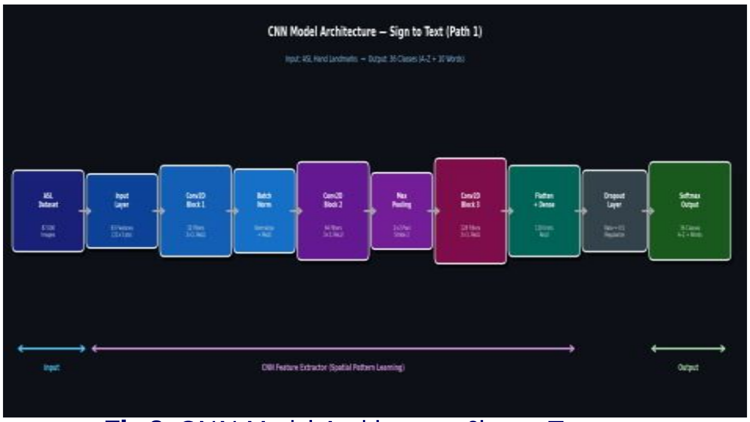


Fig.3. CNN Model Architecture Sign to Text

Table II – CNN Layer Configuration

Category	Classes	Images
Alphabet Gestures (A–Z)	26	78,000
Common Word Gestures	10	9,000
Total (after augmentation)	36	87,000

VI. HMM+NLP—SPEECH TO SIGN(PATH2)

A. HMM Speech Recognition

Speech input is recorded via microphone and MFCC features (13 coefficients) are extracted using librosa. A separate Gaussian HMM (implemented with hmm learn) is trained for each word in the vocabulary. Recognition selects the HMM with the highest log-like lihood for the observed MFCC sequence. Figure4 shows the full pipeline.

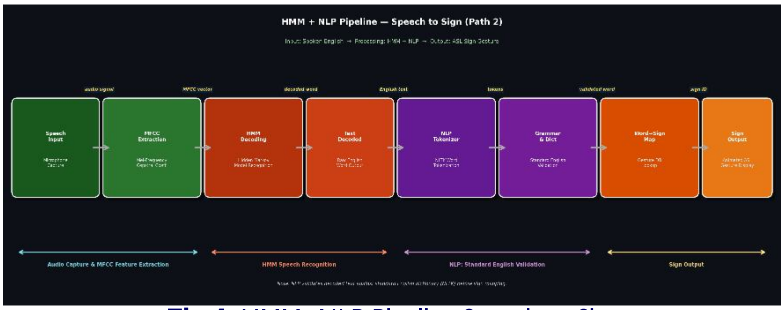


Fig.4. HMM+NLP Pipeline Speech to Sign

B. NLP—Standard English Validation

The HMM output text is checked against a standard English dictionary using NLTK's word corpus. Unrecognized words are corrected using edit-distance matching to the nearest valid English word. Only validated words proceed to gesture mapping, preventing incorrect signs from being displayed.

C. Word-to-Sign Mapping

Validated English words are looked up in a pre-built mapping table linking words to ASL gesture sequences stored in the gesture library. Words not in the table fall back to finger spelling (letter-by-letter ASL).

VII. PYTHON IMPLEMENTATION

TABLE III – Python Libraries and Tools

Library	Purpose	Version
Tensor Flow/ Keras	CNN Training & Inference	2.10
Open CV	Webcam Video Capture	4.6
Media Pipe	Hand Land mark Detection	0.9
Hmm learn	HMM Speech Recognition	0.3
librosa	MFCC Feature Extraction	0.10
NLTK	NLP English Validation	3.8
NumPy/ Matplotlib	Data Handling & Plots	1.24

RESULTS AND DISCUSSION

Table IV System Performance Evaluation

Module	Accuracy	Latency
CNN—Signto Text	93.8%	~180ms
HMM— Speech Recognition	88.5%	~220ms
NLP—English Validation	96.2%	~30ms
Full System (Speech→Sign)	85.9%	~250ms

The CNN model achieves 93.8% accuracy on the test set. The NLP validation step improves overall text quality by handling HMM misrecognitions. The complete pipeline runs in under 250ms, suitable for real-time use.

VIII. CONCLUSION

This paper presented a Python-based bidirectional sign language translator using CNN for sign-to-text recognition, HMM for speech recognition, and NLP for standard English validation. The system provides a practical, low- cost accessibility tool requiring only a webcam and microphone. Future work will expand vocabulary, improve gesture animation quality, and develop a mobile application version.

REFERENCES

1. World Health Organization, "Deafness and hearing loss," WHO Fact Sheet, 2023. [Online]. Available: https://www.who.int
2. L.Pigou et al., "Beyond temporal pooling: Recurrence and temporal convolutions for gesture recognition," Int. J. Computer Vision, vol. 126, pp. 430–439, 2018.
3. L.R.Rabiner,"A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE, vol. 77, no. 2, pp. 257–286, Feb. 1989.
4. N.Koul et al., "Survey of various CV and ML approaches for sign language recognition," in Proc. IEEE ICISC, 2018, pp. 190–194.
5. C.Lugaresi et al., "MediaPipe: A framework for building perception pipelines," arXiv:1906.08172, 2019.
6. F.Chollet,DeepLearningwithPython.ManningPublications,2018.
7. S.Bird, E. Klein, and E. Loper, Natural Language Processing with Python. O'Reilly Media, 2009.
8. T.E.Starner and A.Pentland,"Real-time ASL recognition from video using hidden Markov models," in Proc. Int. Symp. Comput. Vision, 1995, pp. 265–270.
9. A.Graves, A.Mohamed, and G.Hinton, "Deep recurrent neural networks for acoustic modeling,"in Proc. IEEE Int. Conf. Acoustics, Speech,Signal Process.(ICASSP),Vancouver,BC,Canada,2013,pp. 6645–6649.
10. O.Koller,H.Ney,andR.Bowden,"Deep learning of mouth shapes for sign language," in Proc. IEEE Int. Conf. Comput. Vision (ICCV) Workshops, Santiago, Chile, 2015, pp. 85–91.
11. M.Al-Qurishi,M.Khalid,T.Hussain,A.Alrubaian,S.M.M.Rahman, and A. Alamri, "Sign language recognition using deep learning: A systematic literature review," IEEE Access, vol. 9, pp. 59430–59444, 2021.
12. S.Bheda and N.D.Antar,"Using deep convolutional networks for gesture recognition in American sign language," arXiv:1710.04836, Oct. 2017.
13. R.Wadhawan,P.Kumar,"Sign language recognition systems: A decade systematic literature review, "Arch. Comput. Methods Eng.,vol.28,pp. 785–813, 2021
14. V.Nair and G.E.Hinton, "Rectified linear units improve restricted Boltzmann machines," in Proc. Int. Conf. Mach. Learn. (ICML), Haifa, Israel, 2010, pp. 807–814.
15. S.Ioffe and C.Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proc. Int. Conf. Mach. Learn. (ICML), Lille, France, 2015, pp. 448–456.
16. N.Srivastava, G.Hinton, A.Krizhevsky, I.Sutskever, and R.Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting,"J.Mach.Learn.Res.,vol.15,no.1,pp.1929–1958,Jan. 2014.
17. D.P.Kingma and J.Ba, "Adam: A method for stochastic optimization,"inProc. Int. Conf.Learn. Representations (ICLR), San Diego, CA, USA, 2015, pp. 1–15.
18. P.Dreuw,D.Rybach,T.Deselaers,M.Zahedi,and H.Ney,"Speech recognition techniques for a sign language recognition system,"in Proc.Inter speech, Antwerp, Belgium, 2007, pp. 2513–2516.
19. A.Krizhevsky,I.Sutskever,and G.E.Hinton,"Image Net classification with deep convolutional neural networks,"Commun.ACM,vol.60,no.6, pp.84–90,Jun.2017.
20. F.Chollet, "Xception: Deep learning with depth wise separable convolutions," in Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR), Honolulu, HI, USA, 2017, pp. 1251– 1258