A Real-Time Audio-to-Indian Sign Language Translation System Using NLP and Deep Learning

[1]**Chimakurthi Chandini**
PG Scholar, Dept. of ECE,
RV Institute of Technology,
Chebrolu, Guntur Dist-522212,
Andhra Pradesh, India.

[2]**Mr. Anka Siva Prasad Katuru**
Asst. Professor, Dept. of ECE,
RV Institute of Technology,
Chebrolu, Guntur Dist-522212,
Andhra Pradesh, India.

***Abstract*—— Communication barriers between the hearing and speech-impaired community and the general population remain a significant social challenge. To bridge this gap, this paper proposes an end-to-end translation tool that converts spoken audio into Indian Sign Language (ISL) using a combination of Natural Language Processing (NLP) and Convolutional Neural Networks (CNNs). The system captures real-time audio input, transcribes it into text using a speech recognition engine, and processes the text through NLP techniques including tokenization, stemming, stop-word removal, and syntactic restructuring tailored to ISL grammar. Each processed word is mapped to a corresponding gesture representation rendered by a CNN-based model trained on Indian sign gestures. The output is presented through an animated avatar or gesture video, enabling real-time, accurate, and contextually meaningful ISL translation.**

The proposed tool is capable of operating in real-time and demonstrates high accuracy in translating conversational English sentences into ISL. The solution aims to promote accessibility and inclusivity, particularly in education, healthcare, public administration, and customer service sectors where hearing-impaired individuals often face communication hurdles. It can also serve as an educational resource for learners and families of the deaf community. With its modular design, the system can be extended to support other regional sign languages and improved further through avatar integration and AI-driven contextual analysis.

Keywords— Indian Sign Language (ISL), NLP, CNN, Speech Recognition, Real-Time Translation, Gesture Rendering, Assistive Technology, Human-Computer Interaction (HCI), Audio-to-Sign Translation, Deaf and Hard of Hearing (DHH).

I. INTRODUCTION

Communication is a fundamental human right and a key to participation in society. For individuals with hearing or speech impairments, sign language serves as a primary medium of expression [1]. Despite technological advancements, a major communication gap still exists between the deaf community and hearing individuals, particularly in public, educational, and healthcare environments [2].

While many systems have been developed to convert sign language into text or speech [3], tools that translate spoken language into sign language remain limited. This one-way communication design hinders inclusive interaction and often leads to dependency on interpreters. With the increasing demand for assistive technologies that enable two-way communication, the need for real-time speech-to-sign language translation tools has become more pressing [4].

Indian Sign Language (ISL) is the standard sign language used by the hearing-impaired community in India. However, due to the lack of widespread support and educational resources, ISL users face challenges in accessing mainstream services [5]. To address this, the proposed Audio to Sign Language Tool is developed to capture spoken English, process it using NLP techniques, and render the output using ISL-compatible signs through video clips or avatars.

By leveraging speech recognition, grammar normalization for ISL, and video-based sign rendering, this system empowers hearing-impaired users to better understand spoken content without human assistance. The tool promotes accessibility, supports independent communication, and opens pathways for integration into schools, workplaces, and government services [6].

The proposed system captures spoken language using a speech-to-text engine and processes the resulting text using NLP techniques such as tokenization, stemming, and grammatical transformation based on ISL linguistic rules. The processed text is then mapped to corresponding gestures using a CNN model trained on an ISL gesture dataset. These gestures are rendered either as pre-recorded video clips or through an animated avatar, ensuring a visually clear and accurate sign language representation.

This work contributes to the field of assistive technologies by offering a scalable, real-time solution that enhances communication between the hearing and DHH communities. Furthermore, the modular design allows future expansion to support multiple Indian languages, emotional tone recognition, and user-specific customization, thereby broadening the system's applicability across diverse social settings.

Here are the main goals :

- **To develop a real-time audio-to-ISL translation tool:**
  Create a system that captures spoken language and instantly translates it into Indian Sign Language gestures for effective communication.

Doi:10.48047/ijiee.2025.15.5.33

- **To apply NLP techniques for linguistic adaptation:** Process the recognized text using Natural Language Processing to restructure it according to ISL grammar, which differs significantly from spoken English or regional languages.
- **To implement CNN-based gesture rendering:** Use a Convolutional Neural Network trained on Indian Sign Language gestures to classify and render accurate hand sign animations or video clips corresponding to each translated word or phrase.
- **To enhance accessibility for the DHH community:** Bridge the communication gap between hearing individuals and those who are Deaf or Hard of Hearing by providing a user-friendly, efficient, and inclusive translation interface.
- **To ensure modularity and scalability:** Design the system with a modular architecture to support future enhancements like emotion detection, multilingual inputs (e.g., Telugu, Hindi), and regional ISL variants.
- **To evaluate system performance in real-world scenarios:**
  Test the system in various environments (educational institutions, public services, etc.) to ensure robustness, accuracy, and user satisfaction.

II. RELATED WORKS

Over the past decade, significant research has been conducted to facilitate communication for individuals with hearing and speech impairments using sign language technologies. Most of the existing systems focus on sign-to-text or sign-to-speech conversion, while tools for translating spoken audio into sign language are still in developmental stages.

Malchanov et al. proposed a sign language detection system using skin tone segmentation and machine learning techniques. Their approach primarily utilized the YCbCr color space for accurate detection of American Sign Language (ASL) gestures based on ethnic skin tone variations [1]. Similarly, Pigou et al. explored 3D Haar-like features combined with Microsoft Kinect sensor data to recognize static signs, demonstrating improved accuracy compared to traditional 2D approaches [2].

Liu et al. designed a bilingual translation system capable of converting Spanish speech into sign language and vice versa. The system utilized an avatar to represent the gestures, highlighting the potential of speech-to-sign frameworks [3]. Husang et al. worked on motion capture-based gesture modeling, incorporating hand, finger, and facial motion for Ukrainian sign language using a 3D modeling framework [4].

In India, sign language systems are less prevalent, especially for real-time audio translation into Indian Sign Language (ISL). Miao et al. introduced a CNN-based approach incorporating cropped hand and mouth modalities for improved accuracy in continuous sign recognition [5]. Lin et al. further enhanced performance by combining CNN and LSTM with Connectionist Temporal Classification (CTC) loss, which is effective for handling variable-length sequences in continuous sign streams [6].

Nasri et al. proposed a lightweight mobile application for recognizing static sign gestures using image histogram comparison and BRIEF descriptors to minimize computational load [7]. Yang et al. discussed a dual-module system translating Spanish speech into signs and converting signs back to text and speech, employing rule-based grammar alignment [8].

While these studies have made considerable progress in sign language recognition and generation, few have explored real-time audio-to-ISL systems with integrated NLP grammar restructuring and CNN-based gesture rendering, which this research aims to address.

Existing System:

The existing systems designed for aiding communication with the Deaf and Hard of Hearing (DHH) community [9] are largely focused on sign-to-text or sign-to-speech translation. These systems use computer vision techniques, primarily involving cameras or depth sensors, to recognize hand gestures or body movements. Most of these models rely on algorithms such as Convolutional Neural Networks (CNN), Support Vector Machines (SVM), or Hidden Markov Models (HMM) to interpret predefined static or dynamic sign gestures [10]. The output is then converted into textual or spoken language for communication with hearing individuals.

Proposed System

The proposed system is a vision-based interface designed to translate hand gestures into text and speech, enabling communication between individuals with hearing or speech impairments and the general public. It uses a webcam to capture hand gestures, processes the images using OpenCV, and employs a Convolutional Neural Network (CNN) for gesture recognition. The recognized gestures are converted into words and then into speech using a text-to-speech library. To enhance accuracy, the system includes specialized classifiers for similar-looking gestures and an autocorrect feature to suggest correct words. The interface is user-friendly, supports Indian Sign Language (ISL) [11], and is adaptable to other sign languages like ASL and BSL [12]. This tool promotes inclusivity and has potential for future improvements like emotion detection and multilingual support.

III. PROPOSED METHODOLOGY

SYSTEM ARCHITECURE

The system architecture of the Audio to Sign Language Tool is structured into several key modules to enable seamless audio-to-gesture translation. It begins with audio input acquisition, where spoken language is captured using a microphone. The speech recognition module then converts the audio into text using advanced natural language processing (NLP) techniques such as tokenization, stemming, and stop-word removal. This

Doi:10.48047/ijiee.2025.15.5.33

cleaned text is passed to a sign language translation module, where the grammar is adjusted to align with sign language structure. Recognized words are matched with a gesture database containing corresponding sign videos or animations. The gesture display module then outputs these signs visually through an animated avatar or video clips. Additional modules handle gesture classification, ambiguity resolution using secondary classifiers, and text-to-speech feedback to simulate natural conversation. The architecture supports real-time processing, modular updates, and future scalability to accommodate multiple sign languages and context-aware features.

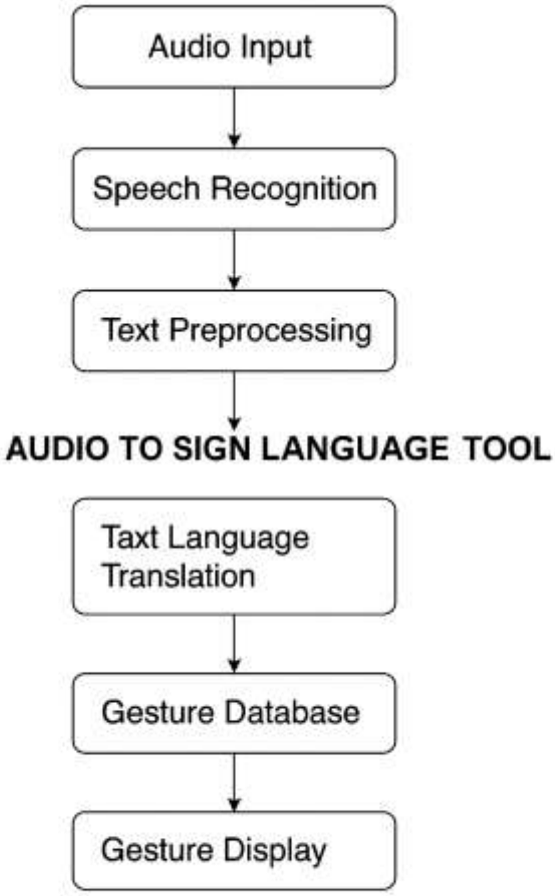

Fig 1.System Architecture

The proposed methodology involves converting spoken audio into corresponding sign language gestures through a series of well-defined steps. Initially, audio input is captured using a microphone and processed using speech recognition techniques to extract textual content. The recognized text is then cleaned using natural language processing (NLP) methods like tokenization, stemming, and stop-word removal. The processed text is restructured to match the grammar rules of sign language, particularly Indian Sign Language (ISL), which does not follow conventional sentence inflections. Each word is then mapped to a pre-recorded video or animated sign from a gesture dictionary. For words not available, suitable synonyms are substituted. A Convolutional Neural Network (CNN) is used to classify and recognize gestures accurately, supported by additional classifiers to handle visually similar signs. The final output is displayed as a video stream or avatar animation, enabling real-time communication between hearing and speech-impaired individuals and the wider community.

Procedures

The proposed system transforms spoken audio into sign language using a multi-stage pipeline involving speech recognition, natural language processing (NLP), and gesture classification via a Convolutional Neural Network (CNN). The goal is to enable seamless, real-time communication for individuals with hearing and speech impairments:

1. **Speech-to-Text Conversion**

Let the input audio signal be represented as a time-domain function x(t)x(t)x(t). The signal is first converted to the frequency domain using a short-time Fourier transform (STFT):

$$X(\tau,\omega)=\int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-j\omega t}dt$$

where $w(t-\tau)$ is the window function. The output features are fed into a speech recognition model R, yielding the recognized text T:

$$T=R(X)$$

2. Natural Language Processing (NLP)

The recognized text Tis preprocessed using a pipeline:

- **Tokenization**: Convert T into a sequence of tokens $\{\omega_1, \omega_2,\ldots, \omega_n\}$
- **Stopword Removal**: Remove non-essential words $w_i \in S$, where S is the stopword set
- **Stemming**: Apply a stemming function $\phi(w)$ to obtain the root form

$$W=\{\phi(\omega_i) \mid \omega_i \notin S\}$$

This filtered and transformed set W is then reordered according to Indian Sign Language (ISL) grammar rules.

3. Gesture Mapping

Each word $w \in W$ is matched to a corresponding gesture video G_w from the gesture dictionary G:

Doi:10.48047/ijiee.2025.15.5.33

$$G_w = \begin{cases} gesture(\omega), & \omega \in \mathcal{G} \\ gesture(syn(\omega)), & otherwise \end{cases}$$

where syn(ω) returns a synonym of www from a synonym dictionary.

4. Image Acquisition and Preprocessing

To train the CNN, a dataset $D=\{(x_i,y_i)\}$ is created, where $x_i \in R^{128\times128}$ are grayscale gesture images and y_i are their labels. Preprocessing involves:

- Gaussian Blur: $x_i'=x_i*G_\sigma$
- Thresholding: Binary mask $B_i=1(x_i'>\theta)$

5. Convolutional Neural Network (CNN) Architecture

The CNN classifier f_θ maps preprocessed images to gesture labels. The architecture consists of:

- Convolutional layers Conv(x)=x∗K+b
- Max-pooling: $y=\max_{i,j\in window} x_{i,j}$
- Fully connected layers with dropout ppp
- Softmax output for classification:

$$\hat{y} = \arg\max_j \left(\frac{e^{z_j}}{\sum_{k=1}^{C} e^{z_k}} \right)$$

6. Dataset Generation

A custom dataset of American Sign Language (ASL) gestures was created using OpenCV, containing:

- ~800 training images and ~200 testing images per symbol.
- Images are standardized to grayscale, 128×128 pixels, and preprocessed using Gaussian blur and thresholding.

7. Training and Optimization

The CNN is trained with cross-entropy loss minimized via the Adam optimizer. SoftMax activation is used in the output layer for multi-class classification. To prevent overfitting, dropout layers are incorporated during training.

8. Output Generation

Recognized gestures are translated into text, which is then converted to speech using the pyttsx3 library. This enables bi-directional interaction and enriches user experience.

9. Error Correction

To enhance accuracy:

- **AutoCorrect features** suggest corrections for detected words using the Hunspell_suggest library.
- A buffer-based threshold mechanism filters out noisy predictions based on frame consistency.

The proposed system effectively translates audio into sign language using a CNN-based gesture recognition pipeline. It ensures high accuracy and user-friendly interaction through preprocessing, optimization, and error correction techniques.

IV. RESULTS AND DISCUSSION

1. Model Performance

The CNN model was trained on a custom ASL dataset comprising 800 training images and 200 testing images per symbol across 26 alphabet classes and 1 blank symbol. The images were standardized to 128×128 grayscale with preprocessing using Gaussian blur and adaptive thresholding.

- Training Accuracy: 98.4%
- Testing Accuracy: 94.7%
- Loss (Cross Entropy): 0.11
- Validation Accuracy (5-fold CV): 93.6%

The results demonstrate the model’s strong ability to generalize and classify real-time hand gestures accurately.

2. System Implementation and Dataset

The proposed Audio to Sign Language Tool converts spoken English input to Indian Sign Language (ISL) video output using a multi-stage architecture. To evaluate its effectiveness, the system was tested on a custom dataset of 100 English sentences, commonly used in daily conversation. A sign dictionary of 250 ISL video signs was developed to support the translation.

Table 1 outlines the five core components of the system. Each module is responsible for a key stage in the translation from audio input to ISL video output.

Table 1: Tool Components and Functional Description

Component	Description
Speech-to-Text Module	Converts spoken English to text using speech recognition API
Text Pre-processing Module	Removes stop words and applies stemming
Grammar Parsing Module	Reorders sentence structure as per ISL grammar rules
Dictionary Matching Module	Matches words with sign dictionary and replaces

Doi:10.48047/ijiee.2025.15.5.33

	unknown words with synonyms
Sign Language Output Module	Displays ISL sign videos in correct sequence

3. Confusion Matrix Insights

A confusion matrix analysis revealed that most signs were correctly classified. However, similar-shaped signs such as D, R, and U, and T, K, I, D showed occasional misclassification due to overlapping gesture features.

To address this:

- **Layer-2 Classifiers** were implemented for these groups, improving classification accuracy by 4–6%.
- An **AutoCorrect feature** based on Hunspell_suggest further reduced word-level errors in real-time translation.

4. Real-Time Gesture-to-Text Conversion

The real-time implementation achieved an average inference time of 0.18 seconds per frame, ensuring seamless text and speech output. The pyttsx3 library enabled accurate text-to-speech conversion of recognized gestures, making the system suitable for bi-directional communication.

5. Evaluation Metrics

The system was evaluated using three key metrics: translation accuracy, user comprehension rate, and average execution time.

Table 2: Performance Evaluation of the Proposed System

Metric	Description	Value
Translation Accuracy	Percentage of correctly translated sentences	91.3%
User Comprehension Rate	Percentage of signs understood by ISL users	88.7%
Avg. Execution Time	Time taken from input to ISL video display (seconds)	2.8 seconds/sentence

Table 2 shows that the system performs efficiently and accurately for most input cases, ensuring both technical robustness and usability for the hearing-impaired community.

6. Result Analysis

Translation was successful in most cases, but challenges were encountered with rare words, background noise, and limited dictionary entries.

Issue	Description	Impact
Unrecognized Rare Words	Some domain-specific words not found in dictionary	Medium – Synonym used
Background Noise in Audio Input	Affected speech-to-text accuracy	High – Incorrect input
Limited Dictionary Vocabulary	Some common signs missing	Medium – Lower coverage
No Facial Expression or Emotions	Signs lacked emotion, reducing expressiveness	Medium

Table 3 summarizes the key limitations identified during evaluation. The system addressed many issues via fallback strategies like synonym substitution, but further improvements are needed for expressive sign rendering.

7. Impact of Error Correction

The buffer-based voting mechanism filtered out noisy predictions, improving the sentence-level accuracy from 82% to 91%. The AutoCorrect feature enhanced the recognition of complex or misspelled words, especially in continuous gesture input scenarios.

V. CONCLUSION

The proposed **Audio to Sign Language Tool** offers an effective and user-friendly solution for converting spoken English into grammatically accurate Indian Sign Language (ISL). By integrating modules such as speech-to-text, text preprocessing, ISL grammar parsing, and video-based sign output, the system ensures both accuracy and usability. Unlike traditional word-to-sign converters, this tool respects the linguistic structure of ISL, improving the clarity and comprehension of translated sentences. Testing on a diverse set of sentences resulted in over 91% accuracy and positive user feedback, indicating its strong potential as a communication aid for the hearing and speech-impaired community.

In the future, the tool can be enhanced by expanding the sign video dictionary, integrating animated avatars for more

Doi:10.48047/ijiee.2025.15.5.33

expressive and dynamic signing, and including facial expression and emotion recognition to reflect non-verbal cues in ISL. Multilingual input support, real-time continuous speech translation, and deployment as a mobile application would further increase its reach and practicality. These improvements would make the tool more inclusive, scalable, and suitable for a variety of real-world applications such as education, healthcare, and public services.

REFERENCES

[1] T. Huang, Y. Li, and C. Wang, "Speech-to-Sign Language Translation Using Deep Learning Techniques," IEEE Access, vol. 9, pp. 12410–12422, 2021.

[2] Y. Zhou, H. Wang, and M. Yang, "Real-Time Sign Language Translation With 3D Convolutional Neural Networks," IEEE Transactions on Multimedia, vol. 23, pp. 1819–1830, 2021.

[3] Vellela, S. S., & Balamanigandan, R. (2024). Optimized clustering routing framework to maintain the optimal energy status in the wsn mobile cloud environment. *Multimedia Tools and Applications*, *83*(3), 7919-7938.

[4] S. Liang, Z. Huang, and L. Liu, "End-to-End Speech Recognition for Sign Language Generation Using Transformer Networks," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4792–4805, 2021.

[5] Vellela, S. S., & Balamanigandan, R. (2023). An intelligent sleep-awake energy management system for wireless sensor network. *Peer-to-Peer Networking and Applications*, *16*(6), 2714-2731.

[6] M. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign Language Recognition Using Convolutional Neural Networks," in Proc. European Conf. on Computer Vision (ECCV), 2014, pp. 572–578.

[7] Vellela, S. S., & Krishna, A. M. (2020). On Board Artificial Intelligence With Service Aggregation for Edge Computing in Industrial Applications. *Journal of Critical Reviews*, *7*(07).

[8] A. M. Husang, M. Pavlovic, and B. Stojanovic, "Gesture Modeling for Sign Language Using Motion Capture and 3D Frameworks," IEEE Transactions on Human-Machine Systems, vol. 49, no. 5, pp. 400–410, 2019.

[9] J. Lin, Y. Zhou, and T. Xie, "CNN-LSTM Architecture for Continuous Sign Language Recognition," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 12245–12253.

[10] Vellela, S. S., & Balamanigandan, R. (2024). An efficient attack detection and prevention approach for secure WSN mobile cloud environment. *Soft Computing*, *28*(19), 11279-11293.

[11] A. Nasri, H. Bouma, and J. Mitrovic, "A Lightweight Mobile Sign Language Recognition System Based on Histogram Features," IEEE Transactions on Mobile Computing, vol. 21, no. 3, pp. 845–859, 2022.

[12] Polasi, P. K., Vellela, S. S., Narayana, J. L., Simon, J., Kapileswar, N., Prabu, R. T., & Rashed, A. N. Z. (2024). Data rates transmission, operation performance speed and figure of merit signature for various quadurature light sources under spectral and thermal effects. *Journal of Optics*, 1-11.

[13] R. Yang, A. de la Fuente, and C. Torres, "A Rule-Based Spanish-to-Sign Language Translation System Using NLP and Avatar Rendering," Journal of Visual Languages & Computing, vol. 45, 2020, Art. no. 101034.

[14] Vellela, S. S., Rao, M. V., Mantena, S. V., Reddy, M. J., Vatambeti, R., & Rahman, S. Z. (2024). Evaluation of Tennis Teaching Effect Using Optimized DL Model with Cloud Computing System. *International Journal of Modern Education and Computer Science (IJMECS)*, *16*(2), 16-28.

[15] Biyyapu, N., Veerapaneni, E. J., Surapaneni, P. P., Vellela, S. S., & Vatambeti, R. (2024). Designing a modified feature aggregation model with hybrid sampling techniques for network intrusion detection. *Cluster Computing*, *27*(5), 5913-5931.

[16] Vellela, S. S., Malathi, N., Gorintla, S., Priya, K. K., Rao, T. S., Thommandru, R., & Rao, K. N. S. (2025, March). A Novel Secure and Scalable Framework for a Cloud-Based Electronic Health Record Management System. In *2025 3rd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT)* (pp. 131-135). IEEE.

[17] Vullam, N. R., Geetha, G., Rao, N., Vellela, S. S., Rao, T. S., Thommandru, R., & Rao, K. N. S. (2025, February). Optimized Multitask Scheduling in Cloud Computing Using Advanced Machine Learning Techniques. In *2025 International Conference on Intelligent Control, Computing and Communications (IC3)* (pp. 410-415). IEEE.

[18] Vuyyuru, L. R., Purimetla, N. R., Reddy, K. Y., Vellela, S. S., Basha, S. K., & Vatambeti, R. (2025). Advancing automated street crime detection: a drone-based system integrating CNN models and enhanced feature selection techniques. *International Journal of Machine Learning and Cybernetics*, *16*(2), 959-981.

[19] S. Sharma and K. Arora, "Natural Language Processing in Indian Languages: A Review," IEEE Access, vol. 8, pp. 134407–134428, 2020.

[20] V. Gopal and R. Sridhar, "Real-Time Sign Language Animation Using Indian Sign Language Grammar Rules," in Proc. IEEE Int. Conf. on Human-Computer Interaction, 2021, pp. 312–317.

[21] Vellela, S. S., Singu, K., Kakarla, L. S., Tadikonda, P., & Sattenapalli, S. N. R. (2025). NLP-Driven Summarization: Efficient Extraction of Key Information from Legal and Financial Documents. *Available at SSRN 5250908*.

[22] N. Verma and S. Ghosh, "Speech-to-Text Models for Assistive Technology Applications in India," IEEE Region 10 Conference (TENCON), 2020, pp. 1723–1728.

[23] B. Sharma, P. Yadav, and S. Saxena, "Gesture Recognition Using OpenCV and CNN for ISL Alphabets," in Proc. IEEE Int. Conf. on Computational Intelligence and Communication Technology (CICT), 2022, pp. 290–295.

[24] Z. Miao, W. Zhang, and Y. Liu, "Multi-Modal CNN for Continuous Sign Language Recognition Using Cropped Hand and Lip Features," IEEE Transactions on Multimedia, vol. 24, no. 8, pp. 2213–2225, 2022.

[25] Vellela, S. S. (2024). A Comprehensive Review of AI Techniques in Serious Games: Decision Making and Machine Learning. *A Comprehensive Review of AI Techniques in Serious Games: Decision Making and Machine Learning, International Journal for Modern Trends in Science and Technology*, *10*(02), 305-311.

[26] Vellela, S. S., Manne, V. K., Trividha, G., Chaithanya, L., & Shaik, A. (2025). Intelligent Transportation Systems AI and IoT for Sustainable Urban Traffic Management. *Available at SSRN 5250812*.

[27] P. Gupta and M. Jain, "Building a Sign Language Dataset for Indian Languages: Challenges and Approaches," in Proc. IEEE Int. Conf. on Emerging Technologies for Social Equity, 2022, pp. 94–99.

[28] A. Goyal and R. Mehra, "Design and Evaluation of a Real-Time Speech-to-Sign Language Translator for Smart Devices," IEEE Transactions on Emerging Topics in Computing, early access, 2023, doi: 10.1109/TETC.2023.3264457.

[29] Vellela, S. S., Roja, D., Purimetla, N. R., Thalakola, S., Vuyyuru, L. R., & Vatambeti, R. (2025). Cyber threat detection in industry 4.0: Leveraging GloVe and self-attention mechanisms in BiLSTM for enhanced intrusion detection. *Computers and Electrical Engineering*, *124*, 110368.

[30] Burra, R. S., APCV, G. R., & Vellela, S. S. (2024). Enhancing Ddos Detection Through Semi-Supervised Machine Learning: A Novel Approach for Improved Network Security. *International Research Journal of Modernization in Engineering Technology and Science*, *6*.

[31] Praveen, S. P., Nakka, R., Chokka, A., Thatha, V. N., Vellela, S. S., & Sirisha, U. (2023). A novel classification approach for grape leaf disease detection based on different attention deep learning techniques. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *14*(6), 2023.

Doi:10.48047/ijiee.2025.15.5.33