A Real-Time English Audio to Indian Sign Language Converter for Enhanced Communication Accessibility

1st Naga Prasanthi Kundeti
Department of Computer Science and Engineering
Lakireddy Bali Reddy College of Engineering (Autonomous)
Mylavaram, India
knpmadam662@gmail.com

2nd Surya Tejaswini Gonela
Department of Computer Science and Engineering
Lakireddy Bali Reddy College of Engineering (Autonomous)
Mylavaram, India
suryatejaswinig@gmail.com

3rd Venkateswara Reddy Guduru
Department of Computer Science and Engineering
Lakireddy Bali Reddy College of Engineering (Autonomous)
Mylavaram, India
venkateswarareddyguduru51@gmail.co

4th Jahnavi Vuyyuru
Department of Computer Science and Engineering
Lakireddy Bali Reddy College of Engineering (Autonomous)
Mylavaram, India
jahnavivuyyuru123@gmail.com

5th Harsha Vardhan Balusupati
Department of Computer Science and Engineering
Lakireddy Bali Reddy College of Engineering (Autonomous)
Mylavaram, India
Harshaik3001@gmail.com

Abstract— **People who are hard of hearing use Sign Language (SL), sometimes referred to as gesture-based language, to communicate with each other. For those who are not familiar with SL, interpreters are necessary, but they are not always accessible. Consequently, in order to convert the text into SL, an English text-to-Indian Sign Language (ISL) translation system is implemented in this article. The system accepts english text or voice in english language as input and produces a video of Indian Sign Langauge as output. It serves as a tool for human-computer interaction and does away with the requirement for an ISL human interpreter when interacting with the hard of hearing. The system has a large collection of frequently used sentences and English terms. It highlights the value of cultural sensitivity in the development and application of assistive technology in addition to showcasing technological competence.**

Keywords—Real-Time; Indian Sign Language; Speech Recognition; Linguistic Model; Natural Language Preprocessing.

I. INTRODUCTION

In recognition of those lacking the ability to hear, sign language provides a visual and expressive language that helps them express their ideas, feelings, and thoughts. It is an essential form of communication. Sign languages are widely used and have developed naturally in many cultural situations, each having its own vocabulary and grammatical structure. Indian Sign Language (ISL) [1] is one such variation that reflects the diversity and individuality of its linguistic expression and has its roots in the rich cultural tapestry of India [2]. Sign language helps people who are hard of hearing or have speech difficulties communicate with members of the general public who might not be familiar with traditional spoken languages. This introduction aims to shed light on the importance and adaptability of sign languages, with an emphasis on the subtle differences between Indian sign language and other sign languages within the context of Indian culture.

II. RELATED WORK

In recent years, there has been increased interest in and progress in the study of sign language translation and detection. The goal of many research projects and outreach programs has been to close the communication gap that exists for people who have speech and hearing impairments. One noteworthy area of related study is the identification of sign language using machine learning algorithms. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been used in previous studies to efficiently detect and decipher sign language motions. To help the network understand the complex patterns and variances seen in various sign languages, these models frequently depend on large datasets including annotated sign films. Real-time systems for translating sign language have been made possible by the success of these machine learning techniques.

Kavya Dharshini [3] proposed sign languages, relying on manual gestures, are being studied to convert them into text and audio, aiming to bridge communication between people with hearing and speech impairments. Techniques like Media Pipe Holistic, Drawing Landmarks, Open CV, LSTM Neural Network, Google Translator, and GTTS are used. Poonia [4] and team created a simple framework for a continuous neural network-based Indian sign language translator. Neog [5] aims to translate speech/text into Indian Sign Language using Natural Language Processing, a method used by individuals with hearing impairments to communicate and facilitate lip reading. Kumar [6] created an audio-to-text translation system for Indian Sign Language that follows grammatical norms. Word videos are checked, stop words are removed, and stemming and lemmatization are performed. Reddy [7] and the team proposed a system intended to display Indian Sign Language on Android phones and translate it into English and Malayalam, facilitating communication for the deaf and dumb.

Cherian [8] built the speech and hearing impairments, sign language is essential, yet communication without an interpreter can be challenging. This is addressed by a method that uses convolutional neural networks to represent characteristics in the transcription of gestures into spoken words. Throat [9] by utilizing American and Indian sign languages and convolution neural networks to recognize sign language and translate it into text or voice, this research seeks to lower language barriers for those with hearing and speech impairments.

Joshi [10] Introduced an ISL Translate, the biggest continuous Indian Sign Language (ISL) translation dataset, this article aims to bridge communication barriers between the hard-of-hearing and other communities. Nandi [11] using convolutional neural networks and data augmentation, a fingerspelling recognition system for the Indian sign language alphabet has been presented to close the communication gap between normal and deaf-dumb people, attaining high training and validation accuracy.

Previous studies have also focused on the cultural sensitivity of sign language detection. Understanding that sign languages vary throughout groups and geographical areas, research has attempted to include cultural quirks in machine learning models and training data. A lot of work has gone into selecting datasets that include a wide variety of sign language expressions [12] unique to particular cultural situations. By doing this, these programs hope to develop accurate and culturally appropriate sign language translation systems, facilitating successful communication within certain linguistic and cultural communities. In brief, a wide range of fields, including machine learning, natural language processing, user interface design, and cultural issues, are involved in the related work in sign language identification. Together, these endeavors have established a strong basis for the present investigation, which integrates knowledge from earlier study to provide an all-encompassing and efficient real-time English Audio to Indian Sign Language Converter.

III. DATA COLLECTION

A rigorous and focused strategy was taken throughout the data gathering phase of this study in order to compile a varied and culturally appropriate dataset. Popular video-sharing websites were the main source of sign language videos, with YouTube being the most popular owing to its vast content bank (as shown in Fig.1). A significant amount of work went into selecting a dataset with an emphasis on Indian Sign Language (ISL), acknowledging the significance of cultural background in developing a useful communication aid for those with speech and hearing impairments. The dataset included a wide variety of sign language expressions, including differences in gestures, facial emotions, and body movements specific to ISL, in
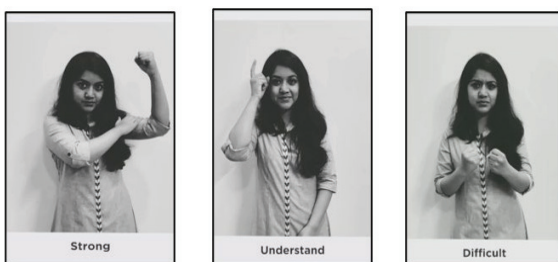
order to assure diversity. A crucial first step in guaranteeing the precision and cultural appropriateness of the created sign language identification system [13] is the meticulous curation of this information.

The collection of a varied and culturally pertinent dataset from YouTube videos required a methodical and comprehensive approach to the data collection procedure for this study.

The study team carefully picked 300 videos that captured a range of signs routinely used in daily life, with an emphasis on the nuances of Indian Sign Language (ISL). The goal of this painstaking curation was to capture the range and depth of ISL expressions, guaranteeing that the final dataset would be complete and representative of the subtle linguistic and cultural aspects of Indian sign language. The larger videos were carefully divided into shorter portions using video editing tool Microsoft Clipchamp, each focusing on a distinct symbol, in order to generate a more detailed knowledge of the numerous indicators (as shown in Fig.2).
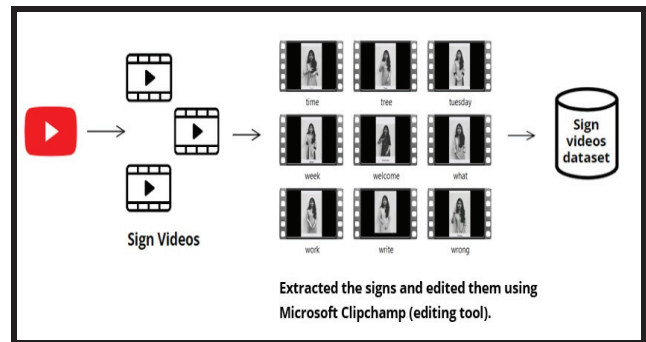

Fig. 2. Sign Data Set Preparation from YouTube videos

In order to extract individual signals and conduct a more in-depth study of the gestures, facial emotions, and body movements connected to each language piece, segmentation was very important. The carefully selected collection contains a variety of signs, such as salutations like "hello" and "hi," sentiments like "love," and basic language elements like the alphabet [41].

By using a comprehensive approach, the dataset was guaranteed to include a wide range of expressions, which aided in the creation of a more reliable and adaptable English Audio to Indian Sign Language Converter.

IV. WORKING METHODOLOGY

The investigation of the initiative's working technique is a multifaceted approach that combines language modeling, state-of-the-art technology, and a methodical user interface design. The speech recognition API in JavaScript is the first step in the basic process that powers the real-time English Audio to Indian Sign Language (ISL) Converter. With the use of this API, the system can record spoken English speech and translate it into text. Then the linguistic model kicks in, emphasizing intelligent spelling correction, lemmatization, and stemming. A polished textual representation of the desired message is the result of this language processing, and it is put into the machine learning model (as shown in Fig.3). Each tokenized and lemmatized word is mapped to a corresponding ISL sign video from an extensive library. ISL has its unique grammatical rules and structure, differing significantly from the grammar of spoken or written

Fig. 1. Data Gathered from Various Video Sources

languages. The system adjusts the order of sign videos and incorporates necessary grammatical signs to ensure the output adheres to ISL syntax and context. The relevant ISL sign videos are concatenated in the adjusted order to form a coherent ISL translation of the input text. The final output is a smoothly concatenated video displaying the ISL translation of the original textual input.

In addition to the technological features, the Django framework-based user interface is essential for enabling smooth communication. Through an intuitive interface, users may enter text, which is then refined by the language model and converted into ISL signs by the machine learning model [14][15][16].
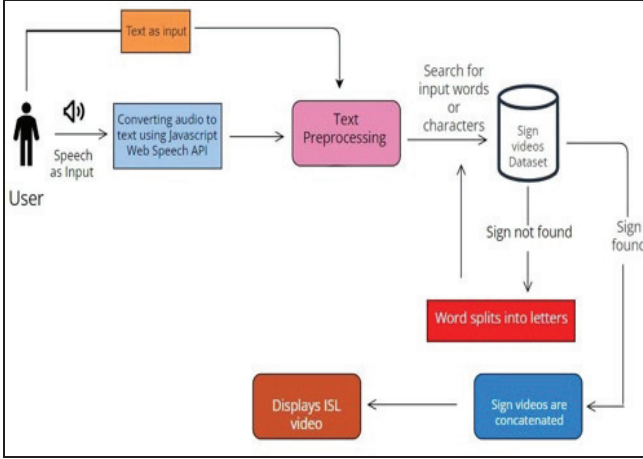

Fig. 3.   Proposed System Model

Additionally, the interface has login and logout capabilities, which improve user experience and make the system available to a wider range of users. This all-encompassing working approach, which encompasses language modeling, machine learning [17][18], speech recognition, and user interface design, comes together to provide a novel and useful tool that supports people with speech and hearing impairments in their communication aspirations.

V. IMPLEMENTATION

The system is engineered to handle two types of inputs – audio and text. It processes these inputs through a series of stages (as shown in Fig.5) to produce a coherent ISL video translation.

A. Input Processing

- *Audio Input:* The system utilizes the JavaScript Web Speech API to capture and transcribe spoken language into text. This feature broadens the system's accessibility, allowing users to input data through speech.

- *Text Input:* Additionally, the system accepts manually entered text, catering to users who prefer typing or have pre-written content.

B. Text Preprocessing:

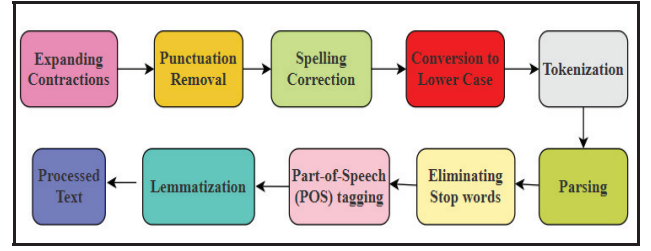Regardless of the source, all textual data undergoes preprocessing which includes steps (as shown in Fig.4)

Fig. 4.   Text Preprocessing Steps

- *Expanding Contractions:* In English, contractions are shortened forms of words or combinations of words created by omitting certain letters and sounds, often replacing them with an apostrophe. Our system identifies these contractions and expands them to their full form.

- *Punctuation Removal:* The system identifies and removes characters that are classified as punctuation, such as commas, periods, question marks, and exclamation marks.

- *Spelling Correction:* The spelling correction feature in our system plays a crucial role in ensuring that both typed and speech-transcribed texts are free of errors. The goal of spelling correction is to identify and correct misspellings in the text.

- *Conversion to Lower Case:* Converting the text's input case to lower case. Even though the keywords "GOOD," "good," and "Good" have the same meaning, if the lower casing is not applied, the model sees them as three different words.

- *Tokenization:* It involves breaking down the input text into smaller units, called tokens, which are typically words, phrases, or other meaningful elements.

- *Parsing:* In the parsing process, the system examines the sentence structure to determine the roles and relationships of individual words and phrases. This includes identifying subjects, verbs, objects, and other grammatical elements. The purpose is to understand the syntax or the arrangement of words and phrases to convey a message. It helps in handling complex sentence structures and ensures that nuances in meaning are correctly conveyed in ISL.

- *Eliminating Stop words:* There are no stop words in the syntax of ISL. In ISL, sentences are constructed entirely of main words. Articles like "the," "a," and "an," as well as connecting words like "am," "are," "is," "be," "to," and "for," are eliminated.

- *Part-of-Speech (POS) tagging*: In POS tagging, the system analyzes the words in a sentence and labels them with their corresponding part-of-speech tags based on their definition and context. This helps in distinguishing between words that can serve multiple grammatical roles.

- *Lemmatization:* In Lemmatization [19], the system processes each word to extract its root form, irrespective of its tense, number, or case. For example, words like "running," "ran," and "runs" are all reduced to their base form, "run." This is crucial

for understanding the true meaning of a sentence, as different grammatical forms of a word should map to the same sign in Indian Sign Language.

C. Token-to-Sign Mapping:

Each tokenized and lemmatized word is mapped to a corresponding ISL sign video from an extensive library. This library contains pre-recorded videos of ISL signs representing a wide range of vocabulary.

D. Grammatical and Contextual Adjustment:

Indian Sign Language (ISL) has its unique grammatical rules and structure, differing significantly from the grammar of spoken or written languages.

ISL has Certain distinct features few of them are:

1. Number representation are done with hand gestures for each hand.
2. In interrogative sentence, all the WH questions are placed in the back of the sentence.
3. The question words like WHAT, WHERE, WHICH, HOW etc. are placed at the end of interrogative sentences.
   i) English: Where is the bank? ISL: BANK WHERE
   ii) English: Who is sick? ISL: SICK WHO

The system adjusts the order of sign videos and incorporates necessary grammatical signs to ensure the output adheres to ISL syntax and context.
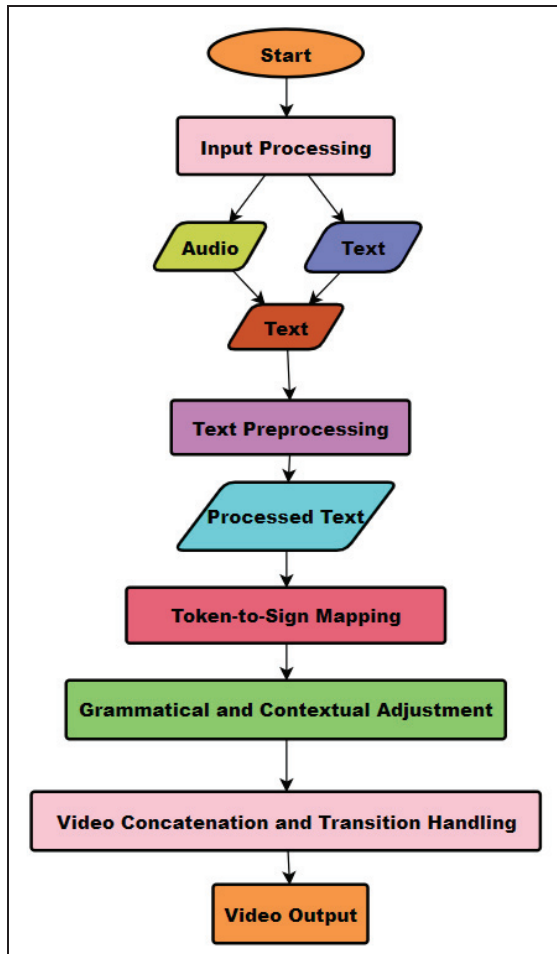
E. Video Concatenation and Transition Handling:

The relevant ISL sign videos are concatenated in the adjusted order to form a coherent ISL translation of the input text. Special attention is given to transitions between signs, ensuring a natural and understandable flow.

F. Video Output:

The final output is a smoothly concatenated video displaying the ISL translation of the original textual input. This output can be presented on various devices, making the system versatile and accessible.

VI. RESULTS

In the thorough evaluation of our audio-to-sign language system, we conducted a series of tests for a comprehensive assessment of its linguistic capabilities. This included correcting 100 misspelled words to test spelling accuracy (as shown in Fig.6), processing 100 sentences with various punctuations to evaluate handling of syntactic nuances (as shown in Fig.7), and lemmatizing 100 words to examine the system's proficiency (as shown in Fig.8) in reducing words to their base forms. Additionally, the system's ability to expand contractions was tested with 100 sentences, crucial for understanding informal language (as shown in Fig.9). A miscellaneous test with a corpus of 500 sentences, combining all these elements, further provided an extensive overview of the system's ability to handle complex linguistic tasks, ensuring its effectiveness in translating spoken language nuances into sign language accurately.

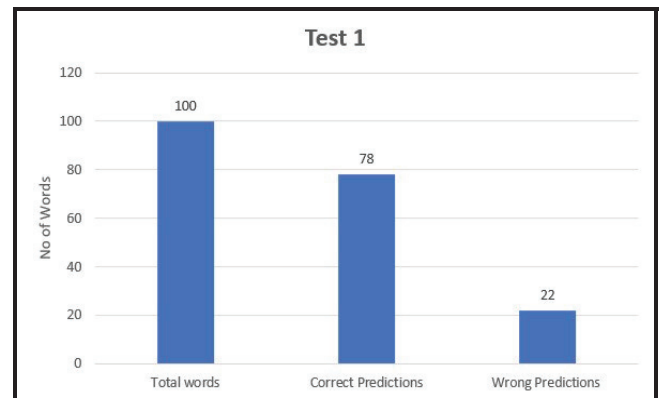
Fig. 5. Data Flow for the Proposed System

Fig. 6. Spelling Correction Test

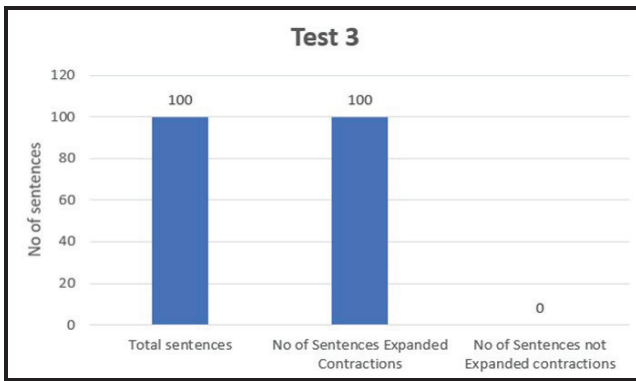Fig. 7. Sentence Punctuation Test
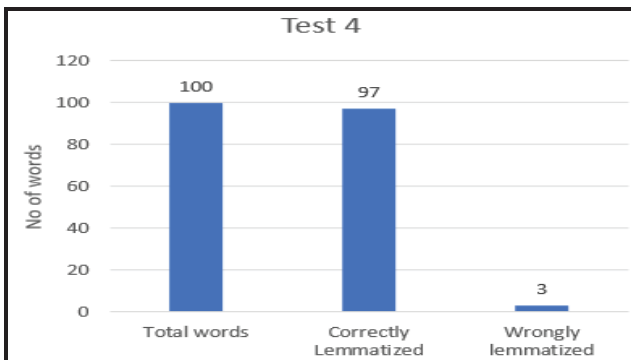
Fig. 8. Contraction Expansion Test

Fig. 9. Lemmatization Test

A corpus of 500 frequently used sentences was used to test the suggested approach. The score obtained from the BiLingual Evaluation Understudy (BLEU) is used to evaluate the proposed system by comparing its output to the reference output, which is produced manually by consulting the ISL regulations. The score for the BLEU is calculated based on how closely the system output resembles the manually generated output of the reference output. Every time, the BLEU score falls among 0 and 1. It shows how similar the translated and intended output are to one other; a value close to 1 indicates that both outputs are equivalent, and a value close to 0 indicates that they are not. The score for the BLEU is now determined for every translated sentence by comparing it to the reference phrases, which are manually constructed using the ISL criteria.

When computing the BLEU score, the total number of position-independent fits is counted by comparing the n-grams of the two sentences. Thereafter, an average of these scores is calculated to estimate the overall translation quality of the system. Using a collection of 500 sentences, the suggested approach obtained the BLEU score of 0.96.

With important features like login and logout choices (as shown in Fig.10), the English Audio to Indian Sign Language Converter's results page (as shown in Fig.11) provides a dynamic and user-centric environment. Additionally, the results page's login option is essential for enhancing the efficacy and correctness of the translation (as shown in Figure 12). Over time, this individualized data helps to improve the language model, leading to more precise and culturally appropriate interpretations of Indian Sign Language (ISL) signs (as shown in Figure 13).

Fig. 10. User Interface

Fig. 11. Text input for converting English Audio to Indian Sign Language

Fig. 12. Text preprocessing on the input text

Fig. 13. Indian Sign Language converter based on text given by user

VII. CONCLUSION

An innovative audio-to-sign language translation system is implemented, integrating human facial expressions to enhance expressiveness and accessibility for the deaf and hard-of-hearing community. Rigorous testing is done on the

5

system with misspelled words, sentences with varying punctuation, contractions and lemmatization challenges. The system demonstrated considerable proficiency. However, it's important to acknowledge its current limitation due to a constrained sign language vocabulary. This restriction highlights the need for future research to expand the system's lexicon, thereby broadening its utility and inclusivity. Overall, this project lays a foundational step towards bridging communication gaps, emphasizing the potential for advanced technologies to create a more inclusive environment for sign language users.

REFERENCES

[1] Intwala, Nishi, Arkav Banerjee, and Nikhil Gala. "Indian sign language converter using convolutional neural networks." 2019 IEEE 5th international conference for convergence in technology (I2CT). IEEE, 2019.

[2] Weisgrau, Maxine, Abraham Rosman, and Paula G. Rubel. The tapestry of culture: An introduction to cultural anthropology. Rowman & Littlefield, 2023.

[3] Kaviyadharshini, K., et al. "Conversion of Sign Language to Text and Audio Using Deep Learning Techniques." Advances in Information Communication Technology and Computing: Proceedings of AICTC 2022. Singapore: Springer Nature Singapore, 2023. 285-293.

[4] Poonia, Ramesh Chandra. "LiST: A Lightweight Framework for Continuous Indian Sign Language Translation." Information 14.2 (2023): 79.

[5] Neog, Anannya Priyadarshini, Arunabh Kalita, and Ms Nithyakani Pandiyarajan. "SPEECH/TEXT TO INDIAN SIGN LANGUAGE USING NATURAL LANGUAGE PROCESSING." (2023).

[6] Kumar, Chevella Anil, and Kancharla Anitha Sheela. "Development of a Speech to Indian Sign Language Translator." International Conference on Smart Computing and Communication. Singapore: Springer Nature Singapore, 2023.

[7] Reddy, Bandi Rupendra, et al. "Indian Sign Language Generation from Live Audio or Text for Tamil." 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2023.

[8] Cherian, Dona Mary, and Jincy J. Fernandez. "Real-time sign language recognition and speech conversion using VGG16." International Journal of Computational Vision and Robotics 13.2 (2023): 174-185.

[9] Thorat, Sumitra Sambhaji. Sign Language Recognition and Text-to-Speech Translation. Diss. Dublin, National College of Ireland, 2023.

[10] Joshi, Abhinav, Susmit Agrawal, and Ashutosh Modi. "ISLTranslate: Dataset for Translating Indian Sign Language." arXiv preprint arXiv:2307.05440 (2023).

[11] Nandi, Utpal, et al. "Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling." Multimedia Tools and Applications 82.7 (2023): 9627-9648.

[12] Sonekar, S. V., et al. "Real-Time Sign Language Identification using KNN: A Machine Learning Approach." 2023 11th International Conference on Emerging Trends in Engineering & Technology-Signal and Information Processing (ICETET-SIP). IEEE, 2023.

[13] Rahaman, Muhammad Aminur, et al. "Computer vision-based six layered ConvNeural network to recognize sign language for both numeral and alphabet signs." Biomimetic Intelligence and Robotics (2023): 100141.

[14] Yaseen, Zaher Mundher. "An insight into machine learning models era in simulating soil, water bodies and adsorption heavy metals: Review, challenges and solutions." Chemosphere 277 (2021): 130126.

[15] S.J. Basha, D. Veeraiah, G. Pavani, S.T. Afreen, P. Rajesh and M.S. Sasank, " A Novel Approach for Optical Character Recognition of Handwritten Telugu Alphabets using Convolutional Neural Networks", Second International Conference on Electronics and Sustainable Communication Systems(ICESC), Coimbatore, India,2021, pp.1494-1500.

[16] B Swathi, K Shalini, KN Prasanthi, " A Review on Steganography using Images", Asian Journal of Computer Science and Information Technology, Vol.2, Issue 8,2012.

[17] K. Kundeti, N.P., Naga Sai Damodhar, S., Gayathri, K., Bala Vamsi, M., Rahul Pramod, "Image caption to voice generation using deep learning", Journal of Advanced research in dynamical & control systems, vol.12(2),pp.897-904,2020.

[18] Hindman, Matthew. "Building better models: Prediction, replication, and machine learning in the social sciences." The Annals of the American Academy of Political and Social Science 659.1 (2015): 48-62.

[19] Hafeez, Rabab, et al. "Contextual Urdu Lemmatization Using Recurrent Neural Network Models." Mathematics 11.2 (2023): 435.