© May 2024| IJIRT | Volume 10 Issue 12 | ISSN: 2349-6002

Sign Language Conversion to Text and Speech Using Machine Learning

Anitha M[1], Annanya P[1], Chilakala Venkata Shritha Reddy[1], Hargun Kour[1], Mrs. Mangala H S[2]

[1]*Student, Department of CSE, Dayananda Sagar Academy of Technology and Management, Bangalore, India*

[2]*Professor, Department of CSE, Dayananda Sagar Academy of Technology and Management, Bangalore, India*

***Abstract*—The "Sign Language Conversion to text and speech using machine learning" initiative tackles the urgent need for improved communication within this community by introducing an innovative system designed to convert sign language into text and speech. Leveraging state-of-the-art technologies, the system integrates computer vision techniques for real-time sign language gesture detection, employing Convolutional Neural Networks (CNNs) to adapt to diverse signing styles. The multi-stage methodology involves precise gesture recognition, mapping gestures to linguistic components using sequence-to-sequence models, and refining textual output through Natural Language Processing (NLP) techniques. The system further incorporates a sophisticated text-to-speech synthesis module, prioritizing prosody, intonation, and emotion to convey the expressive nature of sign language. The overall methodology ensures not only accuracy and real-time processing but also adaptability to different sign language variants. Promising outcomes are observed, with the system proving effective in overcoming communication barriers, as affirmed by user feedback and evaluations from the Deaf and Hard of Hearing community. The results highlight the system's accuracy, real-time processing capabilities, and its potential to foster inclusive interactions, making it a significant advancement in addressing the communication needs of this community.**

Keywords—Sign Language, Machine Learning, Hand gestures, Hearing-Impaired, Convolutional Neural Networks, User-Centric Experience, Scalability, Sustainability.

I. INTRODUCTION

In the intricate tapestry of human communication, American Sign Language (ASL) emerges as a prominent thread, weaving a unique form of expression for those who navigate the world without the reliance on spoken languages. The silent symphony of ASL becomes

The linguistic bridge for individuals facing the challenges of deafness or muteness, offering them a rich palette of hand gestures and motions to convey thoughts and emotions. The profound necessity for

Sign language arises from the inherent limitations posed by traditional spoken languages for individuals with hearing and speech impairments. As communication unfolds as a dynamic process, the exchange of ideas and messages takes on various forms, encompassing words, signs, actions, and images. For those utilizing sign language, the hands become eloquent messengers, crafting a visual language that transcends the barriers of auditory communication.

Within this realm of nonverbal expression, the term "sign language" encapsulates a diverse array of communication methods specifically tailored for individuals with hearing and speech impairments. Across the global stage, an impressive array of approximately 135 distinct sign languages exists, each with its own nuances and cultural adaptations. Notable among these is American Sign Language (ASL), a linguistic powerhouse that resonates not only within its native context but also echoes across international boundaries, making it one of the most widely utilized sign languages worldwide.

Despite the prevalence of sign languages, individuals facing hearing and speech impairments encounter multifaceted challenges in numerous domains, ranging from conferences and office sessions to educational institutions. In response to these challenges, many resort to text-based communication as a practical means of conveying thoughts. However, as technology evolves at an unprecedented pace, there emerges a growing imperative to facilitate seamless and natural communication for those with hearing and speech impairments.

This ambitious project takes root in recognizing and addressing this imperative, centering its focus on the development of a model capable of identifying hand gestures based on finger spelling. The ultimate goal is to harness the power of technology to empower individuals with hearing and speech impairments, granting them the ability to communicate effortlessly and inclusively. This technological intervention aims to dissolve the barriers that have traditionally hindered their integration into various spheres of society, enabling them to engage with healthcare networks, collaborate with colleagues, and connect with peers, regardless of the interlocutor's proficiency in sign language. Through innovation, this project strives to create a world where communication knows no boundaries, allowing every individual, irrespective of their abilities, to participate fully in the vibrant tapestry of human interaction.

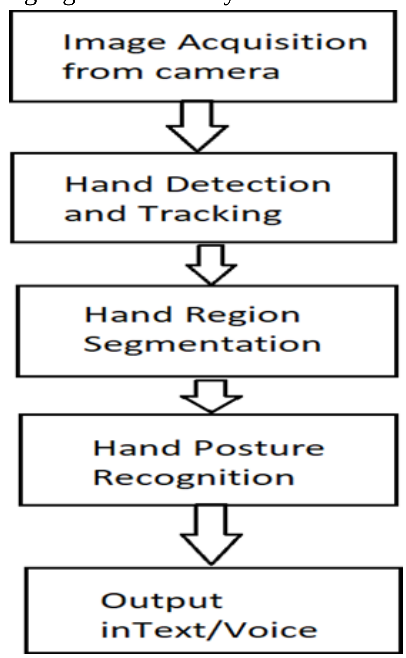
II. LITERATURE REVIEW

The Significant progress has been made in translating sign languages into text and speech, with various systems addressing the associated challenges. The suggestion to capture, recognize, and translate sign language motions using both hardware and algorithmic software programs has been explored. However, challenges persist in accurately translating body language and facial expressions. Technologically, glove-like extensions with translation algorithms have been developed to capture hand movements precisely. Yet, these tools fall short of comprehensively capturing all aspects of sign language, overlooking facial expressions and nonverbal cues. Another approach involves recording video inputs using mobile or specialized devices. For instance, a team in Tamil Nadu created a smartphone application utilizing a camera fixed to a cap to record signers' movements from above.

Sign language is multifaceted, encompassing body language, facial expressions, and hand movements. Current models primarily focus on hand movements, limiting their applicability. Incorporating body language necessitates the shift to video inputs, allowing simultaneous recognition of hand gestures and facial expressions, addressing a key issue in sign language interpretation.

Individualistic variations in signing present additional complications. Differences in behavior, body language, and physical hand structures lead to variations in signs for the same message. To enhance algorithmic performance, a diverse training database capturing variations from different signers is crucial.

Image processing and computer vision algorithms, including skin detection and Canny Edge detection, are applied during processing to input photos or videos. These methods enhance the accuracy of gesture detection algorithms, contributing to the overall efficacy of sign language translation systems.



Various algorithms, such as SURF, SIFT, and Neural Networks, including Convolutional Neural Networks (CNN), have historically been employed based on the model's nature. CNNs, renowned for their effectiveness in image recognition, play a crucial role in sign language translation system development.

The final step in the translation process involves converting the generated text output into speech. Open-source APIs and libraries like Google Text-to-Speech (gTTS) in Python facilitate this conversion, providing flexibility across multiple languages, with some requiring an internet connection for real-time functionality.

The primary motivation behind this research is to develop technology for the translation of Indian Sign Language (ISL), scalable for different sign language dialects. The focus on ISL arises from a lack of comprehensive work in this domain, with the ultimate goal of creating an accessible application for the deaf and mute community in India. This addresses the need for higher translation accuracies and real-time translation capabilities.

In summary, the literature review underscores the evolution of technologies for sign language translation, emphasizing the necessity of a comprehensive approach considering the multifaceted nature of signing and individualistic variations. Integrating advanced algorithms, image processing techniques, and efficient text-to-speech conversion is crucial for developing a robust and accessible system for the deaf and mute community.

III. METHODOLOGY

Embarking on the enchanting odyssey of transmuting the artistry of sign language into the mellifluous cadence of spoken words requires a meticulous methodology, a choreographed dance between technology and empathy.

At the inception of this symphony, computer vision becomes the virtuoso conductor, wielding its algorithms withthe finesse of a maestro interpreting every flourish of expression in sign language. The algorithms, like attentive dancers in a ballet, meticulously decode the intricacies of hand movements, facial expressions, and body language. Their harmonious interplay not only deciphers the linguistic syntax but also captures the emotional nuances embedded in the graceful dance of signing.

As the interpretative ballet unfolds, the natural language processing algorithms gracefully step into the limelight.They function as linguistic poets, translating the silent eloquence of sign language into a textual tapestry, capturing the cultural richness and contextual subtleties. Just as a skilled writer navigates the nuances of a language, these algorithms embark on a journey to preserve the essence of sign language, weaving a narrative that speaks not only of words but of emotions.

The symphony crescendos as speech synthesis takes center stage. Here, technology assumes the role of an alchemist, transmuting the written word into audible poetry. State-of-the-art synthesis techniques weave togetherphonetic subtleties, intonations, and rhythms, birthing a synthesized voice that echoes with authenticity. This voice, a digital bard, becomes the vessel through which the silent verses of sign language metamorphose into a melodic language that resonates with the soul.

The methodology, thus, becomes an intricate dance of algorithms and humanity, a waltz that traverses the delicate balance between technological precision and cultural sensitivity. In this choreography, inclusivity is not just a note but the very heartbeat of the composition. The methodology embodies a commitment to breaking barriers, where the convergence of sign language and speech is not merely a technical achievement but a celebration of diversity, an ode to the universal desire for connection and understanding. It is a symphony that harmonizes the silent elegance of signing with the spoken poetry of words, forging a bridge across the realms of sound and silence.

IV. ALGORITHM

The algorithm for the sign language conversion to text and speech system is implemented in Python, chosen for its high-level, interpreted nature and robust support for artificial intelligence applications. Python's openness and extensive library ecosystem make it an ideal platform for advanced development and implementation.

To commence the implementation, a comprehensive dataset was compiled from a pre-existing open-source GitHub repository, focusing on stationary signs in Indian Sign Language (ISL) by a distinct user. Images were captured for each sign from various users with varying degrees of clarity, creating a diverse dataset. This dataset was then partitioned into training and test images to ensure robust model training.

For enhanced flexibility and customization, the implementation utilizes the Keras Deep Learning framework on top of TensorFlow. Keras offers pre-trained networks with precise recognition and categorization of objects, making it suitable for image-related tasks. Leveraging the Convolutional Neural Network (CNN) model within Keras, a specific model adapted to sign language motions is crafted, capable of classifying inputs into appropriate categories.

To achieve high accuracy in gesture recognition, the model undergoes multiple iterations on the training dataset, the number of iterations defined by the chosen number of Epoch cycles. Keras, as a machine learning algorithm, adapts to the vast and varied training dataset, enhancing the model's ability to produce accurate results over successive runs.

The ultimate goal of the algorithm is to generate a text output for the input sign. Once the CNN model processes the input, the resulting text undergoes translation through a text-to-speech translator. This translator converts the text output into speech, serving as the final outcome of the system.

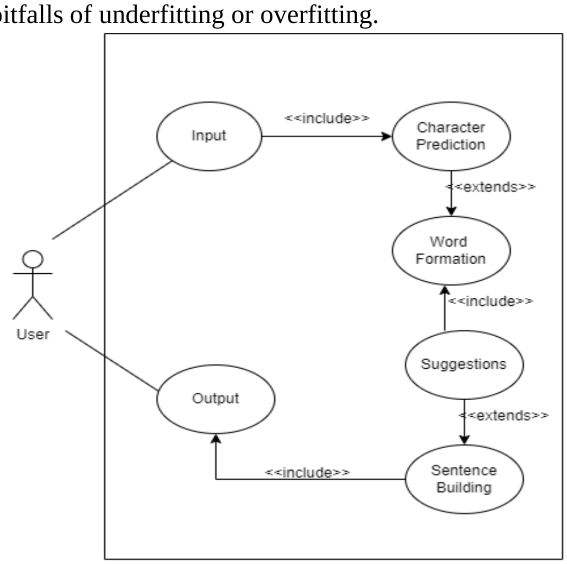
The generated output can be dynamically played or stored as an audio file, providing flexibility in accessing and utilizing the converted sign language to speech. In summary, the algorithm seamlessly integrates data preprocessing, deep learning with CNNs, and text-to-speech translation to achieve the conversion of sign language to speech.

V. IMPLEMENTATION

In the first stage of execution, we focused on building and forecasting models with a collection of 1 motionthat are stationary. Next, a series of tests were run on the algorithm to ascertain how it was true. In order to conduct the experiment, different input data combinations and the number of neural network iterations were tested.

To assess the correctness of the model, multiple instances of the input dataset were constructed and subjected to the procedure for varied numbers of Epoch cycles. A dataset including a collection of gestures was used to train the model for 25, 50, 75, and 100 iterations. With every cycle, there was a continuous decrease in the Log-loss function and an increase in accuracy. Notably, the stability of the model increased wit.

For the training dataset, the achieved model accuracy varied from 85–95%, while for the validation dataset,it was 75–85%. The accuracy trends across Epoch cycles for both training and validation datasets indicated that the model was appropriately trained, avoiding the pitfalls of underfitting or overfitting.



Graphical representations of the loss function and accuracies are presented in the accompanying images. Generally, the model demonstrated stability after approximately 50 Epoch cycles, a consistent observation across various input datasets.

The model was used to recognize test indications after training. This required calculating the intervals for every sign the model was trained on and choosing the highest interval to create a confidence interval for a test image. The model then started text-to-speech translation and named the sign. Nevertheless, there were times when the model had trouble telling one indicator from another. To overcome this constraint, further iterations and the algorithm's integration of deep learning models are needed.

In conclusion, the implementation phase validated the model's training efficacy, stability, and capability to translate sign language to text and speech. The confidence interval approach for identifying signs showcased the adaptability of the model. Ongoing improvements will focus on refining the model's ability to distinguish between similar signs, emphasizing the dynamic nature of the algorithm for real-world applications.

VI. CONCLUSION

In a country like India, where instances of bias against the speech and hearing-impaired communities persist, the development of a tool facilitating communication between these communities and the speaking population becomes a significant step toward inclusivity. The project, "Sign Language Conversion to Text and Speech," serves as a valuable interface to bridge the communication gap, thereby reducing barriers faced by the hearing-impaired.

A key strength of this model lies in its real-time functionality, positioning it as an accessible resource for the masses. This model's reach is further increased by turning it into a mobile application, which guarantees efficient dissemination among the intended audiences.

However, it is crucial to acknowledge certain limitations inherent in the proposed system. Like any translation system, achieving a hundred percent accuracy in sign language translation ischallenging. The real-time nature of this system amplifies the difficulty in identifying and rectifying translation inaccuracies, as users engaged in conversations may not readily flag errors. Despite this challenge, the integration of in-built machine learning algorithms provides an avenuefor users to flag errors, contributing to continuous improvement.

Additionally, the algorithm's reliance on recognizing clean and distinctly made hand gestures limits its ability to contextualize signs based on mood or related factors. The absence of context recognition is a notable consideration for future enhancements.

Furthermore, the installation of several libraries necessitates an active internet connection, suggesting that internet access will be necessary for the final system to function properly.

This restriction needs to be considered in situations where internet connectivity might be restricted.

In conclusion, while the "Sign Language Conversion to Text and Speech" system presents a significant advancement in fostering communication inclusivity, continuous refinement is imperative to address inherent challenges. The prospect of real-time error identification and the need for context recognition represent areas for future research and development to further enhance the effectiveness of the system.

VII. FUTURE SCOPE

The envisioned system, presently focused on translating a specific sign language variant into spoken English, serves as the foundation for future developments with a potential for diversity and expansiveness. A significant avenue for enhancement involves extending the system's capabilities to encompass a wide array of sign language dialects. While the current model supports a specific variety, like Indian Sign Language, the goal is to develop a more adaptable and global approach. This expansion would enable the system to translate between multiple sign language dialects—such as American, British, or Indian sign language—and any spoken language chosen by the user. This has the potential to facilitate cross-cultural communication, allowing individuals signing in one language, such as Pakistani Sign Language, to seamlessly communicate with someone proficient in an entirely different spoken language like Italian.

Another promising dimension for future development is the incorporation of sentiment analysis into the system. Given that sign language intricately involves facial expressions and body language, integrating sentiment analysis alongside recognized gestures could provide a deeper understanding of the signer's intent or emotional state. This additional layer of information has the potential to significantly enhance the system's accuracy. Certain signs are contextually influenced by emotions, and capturing this nuanced interplay could lead to a more refined and precise translation.

These envisioned expansions align with the innate richness of sign language, transforming the system into a more context-aware and empathetic communication tool. By considering not only the gestures but also the emotional context and linguistic diversity, the future development of this project aims to break down communication barriers. The goal is to foster smoother and more authentic conversations in real-world scenarios, making the technology adaptable and inclusive.

In conclusion, the future scope of the "Sign Language Conversion to Text and Speech" project extends beyond its current capabilities. The envisioned enhancements aim to create a more universal, emotionally aware, and linguistically inclusive communication tool, contributing to a more connected and understanding global community.

ACKNOWLEDGEMENT

We express our sincere gratitude to all those who contributed to the successful realization of the project on "Sign Language Conversion to Text and Speech." The dedicated research and development team played a crucial role in designing and implementing the innovative system. We are thankful to the contributors whoprovided access to diverse sign language datasets, laying the groundwork for robust model training. Appreciation goes to the advisors and mentors for their invaluable guidance throughout the project's development. Special thanks to the open-source community for their contributions to the project's success. Additionally, the financial support from our funding partners significantly contributed to the project's execution. This collaborative effort has resulted in a unique solution for sign language conversion, and we extend our gratitude to all individuals and entities involved in this endeavor.

REFERENCE

[1] S. Ma, H. Jiang, M. Han, J. Xie, and C. Li, "Research on Automatic Parking Systems Based on Parking Scene Recognition", IEEE Access, vol.5, pp.2264422650, Oct.18,2017. DOI:10.1109/ACCESS.2017.276020.

[2] J. Kaur, "Implementation of Smart Parking using Artificial Intelligence," Int. J. Sci. Dev. Res., vol. 4, no.

8, pp. 284, Aug. 2019, ISSN: 2455-2631,
[3] R. K. Gupta and G. Rani, "Machine Learning and IoT based Real Time Parking System: Challenges and Implementation," in Proceedings of the International Conference on Innovative Computing & Communications (ICICC) 2020, pp. 1-5, Apr. 1, 2020.
[4] N. Mago, M. Mittal, U. Bhimavarapu, and G. Battineni, "Optimized outdoor parking system for smart cities using advanced saliency detection method and hybrid features extraction model," Sensors and Materials, vol. 34, no. 5, pp. 401-414, May 5, 2022. DOI: 10.1080/16583655.2022.2068325.
[5] A. Pawar, A. Pawar, A. Pawar, G. Pawar, and A. Chaudhari, "An Elaborative Study of Smart Parking Systems," International Journal of Engineering Research & Technology, vol. 10, no. 10, pp. 10895-10900, Oct. 18, 2021. DOI: 10.17577/IJERTV10IS100056.
[6] A. Fahim, M. Hasan, and M. A. Chowdhury, "Smart parking systems: comprehensive review based on various aspects," Heliyon, vol. 7, no. 5, e07050, May 15, 2021. DOI: 10.1016/j.heliyon.2021.e07050.
[7] A. H. Rangkuti, A. Enrico, A. C. Chen, L. Leonardo, and S. Wisely, "Comparison of Car Parking Space Using Pre-Trained Models and Computer Vision Technique," Int. J. Intell. Syst. Appl. Eng. (IJISAE), vol. 11, no. 3, pp. 184-192, 2023.
[8] I. H. Jung, J. M. Lee, and K. Hwang, "Smart Parking Management System Using AI," Webology, vol. 19, no. 1, pp. 4629, Jan. 2022. DOI: 10.14704/WEB/V19I1/WEB19307.
[9] R. Astya, S. Jain, S. Sachan, and A. Aggarwal, "Smart Car Parking System," Int. J. Mech. Eng., vol. 7, no. 6, pp. 453-458, Jun. 2022, ISSN: 0974-5823.
[10] H. B. Razavi and A. Sherafati, "Effective Parameters to Design an Automatic Parking System," Universal Journal of Computer Sciences and Communications, vol. 1, pp. 550, Dec. 13, 2022.