Indiana Journal of Multidisciplinary Research
Abbriviate Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.12667831

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-04|Issue-03|2024

A Comprehensive Survey of Gesture-to-Text Conversion Technologies Using Deep Learning Techniques

Adarsha Sagar H V[1], Ramya R[*2], Charane Veeri[3], Naga Harshini V[4]

[1]Assistant Professor, Department of Computer Science & Engineering, R V Institute of Technology and Management, Bengaluru, Karnataka, India.
[2,3,4]Department of Computer Science & Engineering, R V Institute of Technology and Management, Bengaluru, Karnataka, India.

Article History
Received: 20.05.2024
Accepted: 05.06.2024
Published: 30.06.2024

Citation
Sagar, A. H. V., Ramya, R., Veeri, C. & Harshini, N., V. (2024). A Comprehensive Survey of Gesture-to-Text Conversion Technologies Using Deep Learning Techniques. *Indiana Journal of Multidisciplinary Research*, *4*(3), 65-69.

Abstract: Communication is an integral part of our daily lives, facilitating the sharing of experiences and needs while fostering connections with others. It is ingrained in our daily, guaranteeing that activities are completed effectively and elevating professionalism for individuals as well as organizations. While verbal communication predominates, there exists a segment of our society—the impaired community—that faces hurdles in effectively engaging with others. This difficulty extends to accomplishing essential daily tasks. However, members of the impaired community often rely on hand movements and gestures for communication, albeit encountering barriers as a result of lacking understanding from those unfamiliar with sign language. However, as technology develops quickly, chances to close this communication gap between the community of people with disabilities and the general public arise. Advancements in technology hold the potential to increase the accessibility and feasibility of communication for everybody. This survey paper delves into various methodologies employed in recent years for sign language recognition, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and machine learning techniques. It also covers the related restrictions and difficulties in successfully putting these strategies into practice. The research attempts to illuminate the possible paths towards enhancing the impaired community's communication accessibility and promoting more inclusivity in society by means of a thorough analysis of these strategies.

Keywords: Sign Language, Deep Learning, LSTM, Mediapipe

INTRODUCTION

For individuals who happen to be difficult of hearing, sign language is an integral means of communication, facilitating their interaction within their community and bridging the gap with the hearing population. As the prevalence of hearing loss continues to rise globally, with estimates from the World Health Organization projecting a significant increase by 2050, the demand for effective communication tools for the deaf and mute communities becomes increasingly urgent. This survey paper aims to explore and analyze the current landscape of sign language recognition technology, focusing on advancements that aim to enhance communication accessibility for both the deaf and the hearing population. Drawing upon insights from multiple scholarly works, we delve into the intricacies of sign language as a fully developed language with its own grammar, vocabulary, and syntax. From its historical roots dating back to ancient Greece to its widespread presence across different cultures, sign language stands as a testament to human adaptability and creativity in overcoming communication barriers. As we navigate through the advancements and constraints of existing technologies, this survey paper sheds light on the importance of continuous study and advancement in the field of sign language recognition. By addressing the complexities of sign language interpretation and expanding the accessibility of communication tools, we strive towards a more inclusive society where individuals ofall abilities can effectively communicate and connect with one another.

MATERIALS AND METHODS

A system to translate American Sign Language (ASL) into text was proposed [1] in an effort to alleviate communication difficulties for the speech and hearing impaired. Efforts were made to capitalize on colour segmentation and the Viola-Jones algorithm for sign extraction, alongside skin colour segmentation for preprocessing. Promising results were produced, with 72.3% accuracy achieved on signed sentences and 89.5% accuracy on isolated sign words. To improve the flow of information among those who have hearing impairments, a real-time gesture recognition system was developed using supervised and unsupervised machine learning techniques. Recognition rates of up to 98.20% for American Sign Language (ASL) numerals were achieved, demonstrating improved accuracy in identifying sign language gestures through the utilization of hierarchical and multilabel classification algorithms. These high accuracy rates were attained through innovative technologies like depth-based hand sign recognition and sensory gloves, although scalability and

***Corresponding Author:** Ramya R

usability are areas requiring further exploration and improvement. [2]

A system was developed as described in this paper [3], aiming to utilize Convolutional Neural Networks(CNN) and Faster R-CNN to extract patterns in the feature vectors of each sign of twenty-six alphabets and recognize those signs based on the results. The glove detection method involves sensors within the glove collecting information to detect signs, using pressure sensors to gauge pressure between fingers and process the data for gesture detection. The model was optimized for deployment on Raspberry Pi, necessitating a small model size to decrease prediction time and enable real-time functionality. An accuracy of 93.39% was achieved. In another paper [4], a deep learning model was presented, achieving high accuracy through Transfer Learning and comparing favourably with existing CNN models. The ASL dataset, comprising over 87,000 images, was used to train and test the video. To expedite training while preserving image details, images were downsized to 64 by 64 pixels. The model, leveraging transfer learning with VGG16 and Imagenet weights, achieved an accuracy of 98.7%, representing an improvement of over 4%. The complete model was transformed into an application capturing user actions via camera and transmitting data to the REST API in base64 format. Efficient preprocessing methods for datasets were explored in this paper [5], outlining five stages: grayscale conversion,Gaussian blur for noise reduction, adaptive thresholding for image segmentation based on intensity levels, binarization to transform grayscale images to black and white, and resizing to balance efficiency with information preservation. An innovative approach leveraging Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks was presented in this paper [6], designed to capture semantic dependencies more effectively. The proposed model, comprising an LSTM and a single-layer GRU, demonstrated remarkable performance with an accuracy of approximately 9711 distinct markers. Dropout regularization was applied to mitigate overfitting and enhance generalization, augmenting the model's robustness. The final output was derived using the 'SoftMax' function.

Furthermore, several inherent limitations prevalent in Sign Language Recognition (SLR) systems were discussed. Primary challenges include the limited availability of large-scale datasets for sign language and the need to develop more robust models capable of accommodating variations in sign language gestures across different users, encompassing fluctuations in hand shape, speed, and direction. Addressing these limitations is crucial for advancing the efficacy and applicability of SLR systems in real-world scenarios. The IISL2020 dataset comprises 11 words, with approximately 1,100 video samples per word from 16 study participants (male and female). This dataset was built under natural conditions, without controlled lighting, orientation, additional background settings, gloves, etc.

RESULTS

Software-based sign language recognition systems are preferred over their hardware counterparts due to their superior flexibility, cost-effectiveness, accessibility, scalability, integration capabilities, portability,and user experience. The benefits offered by software-based systems include easy updates, customization, and installation on various devices, making them widely accessible and flexible enoughto user needs. Furthermore, they seamlessly incorporate more software programs, providing a more intuitive and versatile communication experience. Efficient preprocessing capabilities are provided by Mediapipe for real-time sign language recognition systems, ensuring rapid data processing without sacrificing accuracy. With optimized algorithms and hardware acceleration, fast and accurate hand tracking, gesture recognition, and pose estimation are enabled. Its user-friendly framework facilitates easy integration of pre-built components, while also allowing customization to adapt to specific needs or dialects. Moreover, Mediapipe's scalability enables deployment across a wide range of hardware platforms, making sign language recognition accessible to diverse users. Overall, its speed, accuracy, ease of integration, customization options, and scalability [7] collectively contribute to seamless and efficient communication between signers and non-signers in real-time scenarios. Sign language recognition benefits from LSTM networks due to their capacity to retain information from previous frames, [8] addressing the challenges of capturing long-term dependencies. Unlike traditional RNNs, LSTMs mitigate the vanishing gradient problem through gated memory cells, enabling selective retention of crucial information over extended sequences. This ability allows LSTMs to comprehend the context of each sign within a sequence, enhancing accuracy by considering the relationship between signs. Ultimately, LSTM's memory retention capability makes them adept at modelling the temporaldynamics of sign language gestures, contributing significantly to their effectiveness in recognition systems. In the development of a real-time sign language recognition system with Mediapipe and LSTM, Mediapipe plays a crucial role in preprocessing. Initially, hand movements are efficiently tracked, providing accurate detection. Relevant spatial-temporal features are then extracted, serving as input for the LSTM model. Additionally, preprocessing tasks like normalization and scaling are handledby Mediapipe to enhance data suitability. Seamlessly integrated with LSTM, it enables learning of temporal patterns for accurate gesture recognition.

Mediapipe ensures real-time processing, allowing instant interpretation of sign language gestures. Together, Mediapipe and LSTM form a powerful framework for real-time sign language recognition,with Mediapipe handling frontend preprocessing and LSTM

***Corresponding Author:** Ramya R

managing sequence modelling and classification. One significant disadvantage of using CNNs for preprocessing in sign language recognition systems is their computational complexity, leading to longer processing times and increasedresource requirements [9]. This complexity can be overcome by leveraging Mediapipe, which offers efficient and optimized algorithms specifically designed for real-time. multimedia processing tasks.

Unlike CNNs,[10] lightweight algorithms tailored for tasks like hand tracking and pose estimation are employed by Mediapipe, ensuring fast and responsive processing even on resource-constrained devices.This makes Mediapipe a practical and efficient solution for real-time sign language recognition, overcoming the computational challenges associated with CNNs.

DISCUSSION

Our proposed sign language-to-text conversion system integrates several advanced techniques to enable the real-time translation of sign language gestures into textual output. Leveraging the robustness of the MediaPipe framework, feature extraction is performed, eliminating discrimination based on skin colour,gender, and other factors. This framework ensures reliable extraction of key features, including the left hand, right hand, pose, and facial expressions. By representing these features as NumPy arrays internally, the model size is effectively reduced, optimizing computational efficiency. The extracted features are then fed into a Long short-term memory (LSTM) network for training. LSTM networks have demonstrated superior performance in sequential data modelling tasks, making them well-suited for capturing the temporal dynamics [11] inherent in sign language gestures. Unlike Convolutional Neural Networks (CNNs) or traditional machine learning approaches, LSTM networks excel at handling time-series data and preserving long-term dependencies, thereby enhancing the accuracy and robustness of our translation system. Once the LSTM model is trained on the extracted features, our system utilizes the webcam of a laptop to achieve real-time translation of sign language. By leveraging the capabilities of the webcam, our system enables users to interact with the interface effortlessly, providing instantaneous textual output corresponding to the input sign language gestures. This real-time translation capability enhances accessibility and inclusivity, empowering individuals with hearing impairments to communicate effectively in various settings. In the context of sign language recognition,the ability to retain contextual information is paramount for accurate interpretation. Unlike static images, sign language gestures and actions convey meaning through sequences of movements and hand configurations [12]. LSTM networks, with their inherent capability to retain a memory of past inputs, prove to be highly effective in capturing the temporal dynamics and contextual information present in sign language gestures. Traditional machine learning approaches or CNNs, while proficient at recognizing static patterns, may struggle to interpret the sequential nature of sign language gestures. This limitation becomes evident when attempting to classify complex gestures or sequences of gestures, where the temporal order of movements is crucial for accurate interpretation. The table summarizes the performance of various reference methods in terms of accuracy, providing insights into their efficacy for classification tasks. From traditional methods like SVM to sophisticated deep learning architecturessuch as DNNs, DBNs, CNNs, and LSTMs, the results demonstrate a progressive improvement in accuracy. Notably, hybrid models like HMM-BLSTMNN showcase effective integration of probabilisticand deep learning approaches, while dimensionality reduction techniques like RPCA exhibit remarkableaccuracy gains. Overall, the findings underscore the versatility and robustness of deep learning methodologies across a range of applications, reaffirming their status as state-of-the-art solutions in machine learning tasks.

TABLE I
COMPARISON OF METHODS WITH ACCURACY

S.No	Reference	Method	Accuracy
1	13	SVM	85
2	13	DNN	78
3	14	RPCA	92
4	15	GMM-HMM	80
5	16	HMM-BLSTMNN	88
6	17	DBN	95
7	18	CNN	83
8	15	3D CNN	90
9	19	LSTM	79

CONCLUSION

In conclusion, the transformative potential of sign language recognition technology in fostering inclusive communication environments for individuals with hearing impairments is underscored by the advancements showcased in this survey paper. Sign language, serving as a crucial bridge between the deaf and mute communities and the broader society, still faces the barrier of widespread knowledge of the language, hindering effective communication. The barriers are being addressed and lowered, and individuals are being empowered to communicate more seamlessly across linguistic boundaries through the projects and methodologies outlined in this survey. From proof-of-concept endeavors to sophisticated deep learning models, a diverse range of approaches to recognize and interpret sign language has been explored by researchers, ultimately enabling ordinary individuals to engage meaningfully with the deaf and mute population. Furthermore, the potential applications of sign language recognition technology extend far beyond individual interactions. Integrating these systems into educational institutions, healthcare settings, airports, and legal

***Corresponding Author:** Ramya R

proceedings has the potential to revolutionize accessibility and inclusivity for individuals with hearing impairments in diverse contexts. Looking ahead, continued research and development efforts are essential to further refine sign language recognition technology and address remaining challenges. By expanding datasets, refining algorithms, and exploring innovative methodologies, even greater accuracy and usability can be achieved in sign language recognition systems.

ACKNOWLEDGEMENTS

The successful presentation of our paper would be incomplete without the mention of the people who made it possible and whose constant guidance crowned my effort with success. We would like to extend my gratitude to the RV Institute of Technology and Management, Bengaluru, and Dr. Jayapal R, Principal, RV Institute of Technology and Management, Bengaluru for providing all the facilities to carry out this research article. We thank Dr. Malini M Patil, Professor and Head, Department of Computer Science and Engineering, RV Institute of Technology and Management, Bengaluru, for her initiative and encouragement. We would also like to thank our guide, Dr Adarsha Sagar H V,, Assistant Professor, Department of Computer Science and Engineering, RV Institute of Technology and Management, Bengaluru, for his constant guidance and input. We would like to thank all the Teaching Staff and Non-Teaching Staff of the college for their cooperation. Finally, we extend my heartfelt gratitude to our families for their encouragement and support without which we would not have come so far. Moreover, we thank all our friends for their invaluablesupport and cooperation.

REFERENCES

1. Deshpande, A. Shriwas, V. Deshmukh and S. Kale. (2023). Sign Language Recognition System using CNN. *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, 906-911.
2. S. Anthoniraj, V. Ganashree, B. J. R. Umdor, G. D. Sai and B. R. Navya. (2021). Sign Language Interpreter Using Machine Learning. *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, 1-6.
3. Guda, Harsha Vardhan and Guntur, Srivenkat and M, Gowri Pratyusha and Gupta, Kunal and Volam, Priyanka and P V, Sudeep. (2020). Hardware Implementation of Sign Language to Text Converter Using Deep Neural Networks. *Proceedings of the International Conference on Advances in Electronics, Electrical & Computational Intelligence (ICAEEC)*, 67-74.
4. S. Thakar, S. Shah, B. Shah and A. V. Nimkar. (2022). Sign Language to Text Conversion in Real-Time using Transfer Learning. 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), 1-5.
5. Bhat, V. Yadav, V. Dargan and Yash. (2022) Sign Language to Text Conversion using Deep Learning. *2022 3rd International Conference for Emerging Technology (INCET)*,1-7.
6. Sheth, Pranav & Rajora, Sanju. (2023). Sign Language Recognition Application Using LSTM and GRU (RNN). *International Research Journal of Engineering and Technology (IRJET) Volume: 11 Issue: 03*, 1319-1327.
7. Chandwani, Laveen & Khilari, Jaydeep & Gurjar, Kunal & Maragale, Pravin & Sonare, Ashwin& Kakade, Suhas & Bhatt, Abhishek & Kulkarni, Rohan. (2023). Gesture-based Sign Language Recognition system using Mediapipe. *International Conference on Recent Advances in Science and Engineering.* 1-5.
8. Mittal, Anshul & Kumar, Pradeep & Roy, Partha & Balasubramanian, Raman & Chaudhuri, Bidyut. (2019). A Modified LSTM Model for Continuous Sign Language Recognition Using Leap Motion. *IEEE Sensors Journal*, 7056-7063.
9. Suharjito, Suharjito & Anderson, Ricky & Wiryana, Fanny & Ariesta, Meita & Kusuma Negara, I Gede Putra. (2017). Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output. *Procedia Computer Science. 116.* 441-448.
10. Attia NF, Ahmed MTFS, Alshewimy MAM. (2023) Efficient deep learning models based on tension techniques for sign language recognition. *Elsevier Volume 20.* 1-6.
11. Vyavahare, P., Dhawale, S., Takale, P., Koli, V., Kanawade, B., & Khonde, S. (2023). Detectionand Interpretation of Indian Sign Language Using LSTM Networks. *J. Intell Syst. Control, 2(3)*
12. Sahoo, Ashok & Mishra, Gouri & Ravulakollu, Kiran. (2014). Sign language recognition: State of the art. *ARPN Journal of Engineering and Applied Sciences. 9.* 116-134.
13. T.-W. Chong and B.-G. Lee, (2018) American sign language recognition using leap motion controller with a machine learning approach, *Sensors, vol. 18, no. 10*, 3554-3564.
14. S.M. Kamal, Y.Chen, S.Li, X.Shi, and J.Zheng. (2019) Technical approaches to Chinese sign language processing: A review, *IEEE Access, vol. 7*, 9692696935.
15. J. Huang, W. Zhou, H. Li, and W. Li. (2015) Sign language recognition using 3D convolutional neural networks, *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 16.
16. J. Joy, K. Balakrishnan, and M. Sreeraj. (2019) SignQuiz: A quiz-based tool for learning fingerspelled signs in Indian sign language using ASLR. *IEEE Access, vol. 7*, 2836328371.
17. J. Huang, W. Zhou, H. Li, and W. Li. (2015). Sign language recognition using real-sense, in Proc. *IEEE China Summit Int. Conf. Signal Inf. Process. (ChinaSIP)*, 166170.\

***Corresponding Author:** Ramya R

18. F. Yasir, P. W. C. Prasad, A. Alsadoon, A. Elchouemi, and S. Sreedharan, (2017)Bangla sign language recognition using convolutional neural network, *Proc. Int. Conf. Intell. Comput., Instrum. Control Technol. (ICICICT)*, 4953.

19. D. M. Adimas, E. Rakun, and D. Hardianto. (2019). Recognizing Indonesian sign language gestures using features generated by elliptical model tracking and angular projection, *in Proc. 2ndInt. Conf. Intell. Auto. Syst. (ICoIAS)*, 2531.

***Corresponding Author:** Ramya R