International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-12, Special Issue-1, March-2024
https://doi.org/10.55524/CSISTW.2024.12.1.7
Computing for Sustainable Innovation: Shaping Tomorrow's World (CSISTW-2024) on 16th-17th February, 2024
Organized by B.N. College of Engineering and Technology, Lucknow, India

Advancement in Sign Language Recognition Technologies: A Comprehensive Review

Vaishnavi Mishra[1], Tatheer Fatima[2], Shreshtha Srivastava[3] and Dr. Komal Asrani [4]

[1,2,3]B. Tech. Scholar, Department of Computer Science and Engineering, Babu Banarasi Das Northern India Institute of Technology, Lucknow, India

[4]Professor, Department of Computer Science and Engineering, Babu Banarasi Das Northern India Institute of Technology, Lucknow, India

Correspondence should be addressed to Vaishnavi Mishra; vaishnavimishra5222@gmail.com

ABSTRACT- This comprehensive review paper delves into the intricate landscape of sign language recognition (SLR) technologies, addressing the pressing need for effective communication solutions for the global population grappling with hearing impairments. With a meticulous examination of the past two decades of SLR research, the paper categorically explores isolated and continuous manual and non-manual SLR, shedding light on diverse modalities and datasets. The review not only emphasizes the significance of sign language as a fully developed language but also underscores the potential of advanced technologies, such as deep learning and unsupervised learning, in revolutionizing SLR. The inclusion of diverse research projects, ranging from real-time Indian Sign Language conversion to text to the development of deep learning models for American Sign Language recognition, adds practical depth to the exploration. The paper's contributions extend to presenting a tabular format for enhanced comprehension, addressing key facets like modality and sign language types, and offering insightful guidance for the trajectory of this burgeoning field. Overall, this review serves as a valuable and timely resource, advocating for continued innovation to foster a more inclusive and interconnected world for individuals with hearing impairments.

KEYWORD- Sign language recognition (SLR), Machine learning methods, Convolutional Neural Networks (CNN), Indian Sign Language (ISL), Hand Gesture Recognition (HGR), Linear Discriminant Analysis (LDA), Deep learning models, American Sign Language (ASL), Neural network-based technique, Real-time conversion, Computer vision

I. INTRODUCTION

This review paper addresses the critical role of sign language in overcoming communication challenges for the 430 million individuals worldwide with hearing impairments. With 70 million of them being deaf, particularly in developing countries, the paper explores how advanced technologies like deep learning and unsupervised learning can revolutionize sign language recognition (SLR). It highlights the need for ongoing research in this field and identifies a gap in existing literature, prompting a comprehensive review of SLR over the past two decades. The paper meticulously examines key facets, including modalities (vision and sensor), sign language types (isolated and continuous, manual and non-manual), datasets, and methodologies. Drawing on significant works from reputable databases, it categorically explores isolated and continuous manual and non-manual SLR. The review also delves into sensing approaches for SLR and presents datasets that cover various sign languages, shedding light on complexities within these datasets. Finally, it dissects the framework of SLR, offering insightful guidance for the future trajectory of this field. In essence, the paper not only enhances understanding of sign language as a vital form of communication but also calls for continued exploration and innovation to create a more inclusive and interconnected world.

II. LITERATURE REVIEW

Sign language recognition (SLR) utilizes various computational algorithms and machine learning methods, including ANN, CNN, SVM, HMM, each with specific strengths and limitations. Vision-based and glove-based approaches, the two primary methods, face challenges related to camera quality, lighting, and distance. Crucial evaluation parameters such as accuracy and recognition speed are pivotal in assessing SLR effectiveness. With over 300 regional sign languages, customization is essential, extending beyond serving the deaf and mute to address communication challenges in noisy public spaces. The review introduces manual and non-manual SLR, including finger spelling, highlighting the need for nuanced approaches in this dynamic and evolving field. The[1] project focuses on developing a system for real-time recognition and conversion of Indian Sign Language into text for individuals with hearing and speech impairments. Challenges in finding suitable datasets led to the creation of a dataset with 600 training images and 150 testing images for each hand sign. The OpenCV library was used for image capture and processing, incorporating filters like grayscale, Gaussian blur, and threshold. The system is designed for single spelling-based signs, representing each alphabet with a distinct hand sign. Notably, it achieved a high accuracy of 97.3% on the testing dataset. The project introduced background noise in the training data to simulate real-world scenarios, training the model to focus solely on hand signs and gestures while disregarding interference. The significance of the project lies in addressing the communication needs of individuals with hearing and speech impairments through real-time sign language to text conversion. The technology utilized includes the OpenCV library, showcasing effectiveness in practical situations.

In the paper[2] by Mahesh Kumar N B, the focus is on converting sign language into text, particularly emphasizing Hand Gesture Recognition in Human-Computer Interaction. The methodology involves the integration of Euclidean Distance, Eigen Values, and Eigen Vectors, with a specific emphasis on Generalized Fisher's Linear Discriminant (FLD) or Linear Discriminant Analysis (LDA). LDA, commonly used in statistics, pattern recognition, and machine learning, creates a linear combination of features for object or event classification. The LDA approach includes computing mean vectors, scatter matrices, deriving eigenvectors and eigenvalues, and using them for dimensionality reduction and classification through the equation $Y = X \times W$, where X represents samples, and W is a matrix of eigenvectors. In the context of sign language recognition, the system block diagram includes data acquisition, pre-processing, feature extraction, and sign recognition. Existing research primarily focuses on static Indian Sign Language (ISL) signs captured in controlled conditions, utilizing the LDA algorithm for sign recognition to achieve dimensionality reduction, noise reduction, and enhanced accuracy. Future enhancements may involve recognizing numerical values represented in words, incorporating image processing concepts, and fundamental properties. The successful application of LDA algorithms demonstrates efficient gesture recognition, aligning with the overarching project goal of fostering inclusivity by integrating hearing-impaired individuals into daily life.

Shubham Thakar and team introduce a groundbreaking real-time sign language-to-text conversion model in [3] that leverages deep learning, specifically incorporating transfer learning and Convolutional Neural Networks (CNN). The model, featuring three convolutional layers and two dense layers, is designed with ReLU activation and softmax for classification. Implemented as an application, the model captures user actions through a camera, converts data to base64, and transmits it to a Django backend. The backend processes the data, utilizing transfer learning based on VGG16, achieving an impressive 98.7% accuracy on the ASL dataset with 87,000 images and 29 labels. The model's robustness is confirmed through cross-validation, showcasing a 4% improvement over existing CNN models and opening avenues for broader applications in gesture-based communication systems. The study also suggests potential enhancements, including diversifying the model across various sign language datasets and addressing image smoothness concerns through augmentation techniques.

Paper [4] centers on the development of a deep learning-based hand gesture recognition system for American Sign Language (ASL), aiming to facilitate efficient communication for the deaf and mute community. The authors leverage convolutional neural networks (CNNs) to recognize numerical signs (0 to 9) and prioritize a system deployable on mobile applications or embedded single-board computers. To enhance CNN efficiency, the authors utilize image processing techniques, converting RGB data to grayscale for reduced storage requirements and training time. The comparison of various neural network models results in a final 10-layer architecture with a training accuracy of 91.37% and testing accuracy of 87.5%. The paper emphasizes the challenges of conventional algorithms in recognizing ASL expressions and underscores the effectiveness of CNNs in addressing this issue. The proposed system targets real-time and portable sign language recognition, particularly on lower-end embedded systems and smartphones. Furthermore, the paper references related works in sign language recognition, emphasizing their collective significance in fostering effective communication between deaf individuals and the broader public.

The [5] paper introduces a real-time neural network-based technique for finger-spelling-based American Sign Language (ASL), targeting communication aids for the Deaf and Mute community. The model achieves a significant 96% accuracy for each ASL alphabet, using a dataset of 800 training and 200 testing images per sign, processed with OpenCV. The Convolutional Neural Network (CNN) incorporates two algorithm layers, resulting in an outstanding 98.0% accuracy, surpassing existing ASL research. Noteworthy is the use of a standard laptop camera for accessibility and cost-effectiveness. Preprocessing involves RGB to grayscale conversion, Gaussian blur, and adaptive thresholding. The study underscores the importance of easily available resources for a broader audience, supported by confusion matrices illustrating the model's effectiveness in gesture recognition. In summary, the vision-based ASL recognition system offers a practical, accurate, and accessible solution for Deaf and Mute individuals, with potential for further enhancements.

In [6], the paper addresses the challenge of Arabic Sign Language (ArSL) recognition by developing a lightweight deep learning model tailored for mobile applications. Leveraging EfficientNet, the study aims to balance classification performance and computational cost. The real dataset, encompassing diverse visual properties for thirty Arabic alphabets, involves over 5000 images collected by various volunteers. The proposed system comprises two phases: feature extraction and classification, utilizing EfficientNet as the backbone structure, with modifications to channels, depth, and resolution. Experimental scenarios explore segmentation impact, different loss functions, and various versions of EfficientNet Lite models. The results highlight that the EfficientNet Lite 0 architecture, with segmented data and label smooth loss function, achieves an impressive accuracy of 94.30%. Recognizing communication challenges faced by the hearing-impaired community, the study underscores the need for efficient, real-time ArSL recognition systems. The proposed lightweight model holds promise for accurate ArSL gesture interpretations, with potential applications in improving communication for the hearing-impaired. The paper suggests future exploration of Transformer-based architectures and expanding research to recognize additional aspects of sign language.

Paper [7] introduces an innovative approach to converting sign language into text, with the primary goal of improving communication for deaf and mute individuals. The system utilizes computer vision and deep learning techniques, employing key point detection through MediaPipe, data pre-processing, labeling, feature generation, and LSTM neural network training. A diverse dataset of Indian Sign Language gestures is continuously updated to ensure the representation of a wide range of hand gestures. The system focuses on real-time tracking of hand gestures for accurate recognition. Pre-processing is crucial, involving resizing, normalization, and transformation of hand gesture images to ensure consistent input for the model. Proper labeling of hand gestures based on Indian Sign Language enhances the accuracy of the system. The training phase involves feeding pre-processed images to an LSTM model, designed with multiple layers to capture complex patterns in sign language gestures. The proposed system serves as a valuable tool for real-time recognition and translation of hand gestures into meaningful text, addressing the challenges faced by individuals with hearing impairments and offering significant potential for improved communication. Ongoing improvements and experiments with complete language datasets are part of the system's future development.

The study in paper [8] focuses on the conversion of American Sign Language (ASL) signed hand gestures into text and speech using machine learning techniques, particularly unsupervised feature learning. Communication barriers for the hearing impaired are addressed, and a teaching aid for sign language is provided. The methodology involves image segmentation, feature detection, and extraction using FAST and SURF algorithms, as well as supervised and unsupervised classification with K-Nearest Neighbour (KNN) algorithms. The classified features are then converted into text and speech using Text-to-Speech (TTS) synthesis. Key steps include image capturing with a KINECT sensor, grayscale conversion, thresholding, and resizing for binary image creation. Feature extraction is performed using FAST and SURF algorithms, and unsupervised classification relies on the KNN algorithm. The system achieves 92% accuracy in supervised feature learning and 78% accuracy in unsupervised feature learning. The dataset comprises around 500 data samples, with each sign occurring five to ten times, collected using the Kinect sensor. The dataset includes 49 different ASL signs used for training. The study concludes that the proposed system effectively converts ASL signed hand gestures into text and speech, offering potential benefits for communication and learning for the hearing impaired.

This [9] paper addresses the machine learning challenge of converting sign language into natural language, emphasizing the linguistic structure and context of sentences. The proposed deep learning approach utilizes bidirectional translation using GRU and LSTM, incorporating Bahdanau and Luong's attention mechanisms. Experimentation on two sign language corpora, ASLG-PC12, and Phoenix-2014T, involving 16 models, demonstrates the superior performance of the proposed approach. Notably, the GRU model with Bahdanau attention achieves outstanding results, with a ROUGE score of 94.37% and BLEU-4 score of 83.98% in translating from text to gloss. Similarly, in gloss-to-text translation, the GRU model with Bahdanau attention excels with a ROUGE score of 87.31% and a BLEU-4 score of 66.59%. On the Phoenix-2014T corpus, the GRU model with Bahdanau attention leads in text-to-gloss translation with a ROUGE score of 42.96%, while the GRU model with Luong attention excels in gloss-to-text translation with a BLEU-4 score of 19.56% and a ROUGE score of 45.69%. Overall, the proposed models outperform related work, showcasing the potential of deep learning for sign language translation. The paper suggests future exploration of pose estimation to enhance sign language translations further.

III. COMPARISON TABLE

A. Accuracy and Methods:

Sign language recognition (SLR) employs diverse computational techniques like CNNs and LDA to achieve accuracies ranging from 87.5% to 98.7%. These systems facilitate real-time conversion of sign language into text and speech, enhancing communication accessibility for the hearing impaired. Challenges like dataset availability and lighting conditions persist, but ongoing research aims to refine SLR systems for broader applications. Through innovations in deep learning and transfer learning, SLR continues to bridge communication gaps and promote inclusivity in diverse contexts.

Ref.	Method/Algorithms	Accuracy
[1]	HMM, CNN& KNNDW	95.7%
[2]	LDA	unavailable
[3]	CNN (VGG16)	98.7%
[4]	CNN	87.5%
[5]	CNN& OpenCV	98%
[6]	Efficient Net &CNN	94.30%
[7]	LSTM & RNN	96.66%
[8]	FAST & SURF With KNN	78%(unsupervised) 92%(supervised)
[9]	LSTM & GRU	66.59% to 94.37%.

B. Advantages and Disadvantages:

Sign language recognition (SLR) is a dynamic field employing various computational algorithms and machine learning methods such as ANN, CNN, SVM, and HMM, each with specific strengths and limitations. Vision-based and glove-based approaches are the primary methods, facing challenges related to camera quality, lighting, and distance. Crucial evaluation parameters include accuracy and recognition speed, pivotal for assessing SLR effectiveness. With over 300 regional sign languages, customization is essential, extending beyond serving the deaf and mute to address communication challenges in noisy public spaces.

Ref.	Method/ Algorithms	Advantages	Disadvantages	Year
[1]	HMM, CNN& KNNDW	High precision (95.7%) in real-time conversion of Indian Sign Language to text facilitates communication for deaf and mute individuals.	Background noise in training data may affect model generalization to real-world scenarios, potentially impacting accuracy (97.3%).	2021
[2]	LDA (Linear Discriminant Analysis)	LDA algorithm reduces noise and enhances accuracy in recognizing Indian Sign Language gestures.	Limited applicability to dynamic gestures and real-world scenarios due to focus on static signs in controlled conditions.	2018
[3]	CNN (VGG16)	The deep learning model achieves high accuracy (98.7%) through transfer learning with VGG16 architecture, improving upon existing CNN models by 5%.	Dependency on specific image size (64x64 pixels) may limit versatility in handling varied input resolutions.	2022
[4]	CNN	Achieves high recognition accuracy (87.5%) for ASL digits with optimized model complexity, suitable for mobile and embedded systems.	Transfer learning with ImageNet weights is ineffective, requiring manual hyperparameter tuning for satisfactory results.	----
[5]	CNN& OpenCV	High accuracy (98.0%) in real-time American Sign Language recognition aids communication for Deaf and Mute individuals.	Requires ideal conditions for accurate recognition, limiting usability in varied real-world environments.	2021
[6]	Efficient Net &CNN	Lightweight EfficientNet models achieve 94% accuracy in Arabic Sign Language recognition, aiding communication for the hearing-impaired.	Reliance on preprocessing and data augmentation may limit robustness to real-world variations, affecting recognition accuracy.	2022
[7]	LSTM & RNN	Efficiently converts sign language gestures to text, improving communication for deaf and mute individuals.	Limited generalization to diverse sign languages and gestures, reliant on specific algorithms and models.	2023s
[8]	FAST & SURF With KNN	Achieves 78% accuracy in converting ASL hand gestures to text and speech, facilitating communication for the hearing impaired.	Reliance on segmentation and feature detection techniques may limit robustness to real-world variations, potentially affecting accuracy.	2018
[9]	LSTM & GRU	Achieves high translation accuracy for sign language to text and vice versa, surpassing previous methods.	Limited generalization to other sign languages and linguistic structures due to reliance on specific deep learning architectures and attention mechanisms.	2021

IV. CONCLUSION

In conclusion, Sign Language Recognition (SLR) has progressed significantly with diverse computational algorithms (ANN, CNN, SVM, HMM) tailored to address specific challenges. Vision-based and glove-based methodologies, despite camera and lighting challenges, reflect nuanced approaches for effective recognition. Projects focusing on ISL, ASL, and ArSL highlight global efforts to enhance communication for individuals with hearing and speech impairments. Technologies from OpenCV to CNNs and EfficientNet demonstrate practical solutions for real-time conversion, finger spelling, and background noise. Consistently high accuracies showcase SLR's positive impact. Accessibility considerations for devices like smartphones contribute to its broader societal applications. While benefiting the deaf and mute community, SLR's role in noisy public spaces underscores its broader impact. Ongoing innovations in attention mechanisms and natural language processing signify dedication to advancing SLR capabilities. In essence, SLR emerges as a crucial tool for enhancing communication, contributing to inclusivity and accessibility goals.

V. FUTURE SCOPE

Future advancements in sign language recognition (SLR) will refine algorithms and machine learning methods for improved accuracy. Customization for regional sign languages will address communication challenges in noisy settings. Real-time systems converting sign language to text will be developed, integrating deep learning techniques. Enhanced datasets and lightweight mobile models will improve accessibility and communication for the hearing impaired.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Professor Dr. Komal Asrani for her invaluable guidance and support throughout the research process. Additionally, we extend our appreciation to the anonymous reviewer for their insightful feedback and constructive criticism, which significantly contributed to the improvement of this paper.

CONFLICTS OF INTRESTS

The authors declare that they have no conflicts of interest.

REFERENCES

[1]. “Indian sign language to text conversion in real-time using machine learning “by Yash Narang, Aditya Sharma Computer Science & Engineering, SRM Institute of Science & Technology.

[2]. “Conversion of Sign Language into Text “ by Mahesh Kumar N B1 1Assistant Professor (Senior Grade), Bannari Amman Institute of Technology, Sathyamangalam, Erode, India.

[3]. “ Sign Language to Text Conversion in Real-Time using Transfer Learning” by Shubham Thakar, Samveg Shah, Bhavya Shah and Anant V. Nimkar Department of Computer Engineering Sardar Patel Institute of Technology.

[4]. “Convolutional Neural Network Hand Gesture Recognition for American Sign Language” by Shruti Chavan, Xinrui Yu and Jafar Saniie Embedded Computing and Signal Processing Research Laboratory Department of Electrical and Computer Engineering Illinois Institute of Technology, Chicago, IL, U.S.A.

[5]. “Sign To Text Conversion- Helping Aid” by Vatsal Patel, Maahi Patel Devang Patel Institute of Advance Technology and Research, Charusat University.

[6]. “Arabic Sign Language Recognition using Lightweight CNN-based Architecture” by Batool Yahya AlKhuraym, Mohamed Maher Ben Ismail, Ouiem Bchir College of Computer and Information Sciences King Saud University, Riyadh, KSA.

[7]. “Conversion of Sign Language to Text” by Akash Kamble, Jitendra Musale, Rahul Chalavade, Rahul Dalvi, Shrikar ShriyalUG Student, Professor, Department of Computer Engineering, Savitribai Phule Pune University, India.

[8]. “Conversion of Sign Language To Text And Speech Using Machine Learning Techniques” by Victoria Adebimpe Akano, Crawford University and Adejoke O Olamiti, University of Ibadan. JOURNAL OF RESEARCH AND REVIEW IN SCIENCE, December 2018.

[9]. “Sign Language Gloss Translation using Deep Learning Models “ by Mohamed Amin Hesahm Hefny, Ammar Mohammed Department of Computer Science FGSSR, Cairo University, Egypt.