Journal of Science Engineering Technology and Management Science
Volume 02, Issue 07, July 2025 *ISSN: 3049-0952*
www.jsetms.com *DOI:10.63590/jsetms.2025.v02.i07.504-510*

A CNN-BASED APPROACH TO MULTI-CULTURAL SIGN LANGUAGE DETECTION AND INTERPRETATION

K.KRISHNA[1], YEGETELA RAM KUMAR[2], SHIVARAO YANNAM[3], BANDARI RAJESH[4], BUJPA SHIVAKARTHIK[5], AJMEERA RAMAKRISHNA[6]

[1-6]Assistant professor, Department of Computer Science and Engineering, MallaReddy College of Engineering, Maisammaguda, Secunderabad Telangana,India-500100.
Email{[1]krishnak2k13@gmail.com,[2]yegetela.ramkumar@gmail.com,[3]shivarao.yannam@gmail.com,[4]bandari.rajesh9@gmail.com,[5]b.shivakarthik@gmail.com,[6]ramakrishnasit@gmail.com}

To Cite this Article
K.Krishna, Yegetela Ram Kumar, Shivarao Yannam, Bandari Rajesh,Bujpa Shivakarthik,Ajmeera Ramakrishna, "A CNN-Based Approach To Multi-Cultural Sign Language Detection And Interpretation", Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 07, July 2025,pp: 504-510, DOI: http://doi.org/10.63590/jsetms.2025.v02.i07.pp504-510

Submitted: 14-05-2025 *Accepted: 21-06-2025* *Published: 30-06-2025*

ABSTRACT

Sign language is the main form of communication for the community of people who are speech- and hearing-impaired. The general public finds it very difficult to fully understand or interpret sign language. The creation of a sign language recognition system is necessary to overcome this communication obstacle. Since wearable sensors are the foundation of the majority of sign language identification systems now in use, most people cannot afford the recognition system. Furthermore, not all of the spatial and temporal information needed for precise recognition is taken into account by the vision-based sign recognition frameworks that are now in use. This paper proposes a novel vision-based hybrid deep neural net methodology for custom sign gestures and American Sign Language (ASL) recognition. The objective of the suggested framework is to provide a unified system for monitoring and obtaining multi-semantic attributes, like non-manual components. include co-articulations by hand. Besides, a Hybrid Deep Neural Network (HDNN) with atrous convolutions is used for spatial feature extraction from the sign gestures. HDNN based on attention is utilized for the extraction of sequential and temporal features. Moreover, altered auto encoders are used for the extraction of the unique abstract features. The hybrid attention module is utilized in the discriminative feature extraction process to distinguish desired transition gestures from sign gestures. Experiments with the innovative multi-signer ASL and custom sign language dataset have been conducted to test the proposed model. The hybrid neural net architecture that is being presented, which uses HDNN in particular, performs better than existing state-of-the-art frameworks for sign language recognition. Moreover, Flask web is used to integrate a detection module that enables manual picture and sign input for real-time acknowledgement.

INTRODUCTION

Combining different vocal and instrumental sounds to The primary form of communication for people who are hard of hearing or have speech impairments is sign language, or SL. Sign Language appears to have an underlying structure and set of grammatical rules, much like any other language, that enable users to express themselves and communicate effectively. Furthermore, non-manual articulations like eye gaze, facial expressions, lip movements, etc., as well as manual components like hand motion and hand position are typically used to communicate SL. The components of the multi-semantic feature are made up of both manual and non-manual elements. The hearing community needs to put in a lot of work to master sign language, which necessitates creating a system for sign language recognition (SLR). The following sign language datasets have recently been developed: Indian Sign Language (ISL), American, Arabic, German, Chinese, Turkish, Bhutanese, Russian. SLR has been thoroughly studied to assist hearing persons in understanding sign language and improve the quality of life for the community of people who are hard of hearing and speech impaired. The SLR frameworks are designed to identify and detect sign language that a sign interpreter performs using a visual medium. Signer-dependent variations, local and global elements, feature extraction from heterogeneous backdrops, multi-modality, occlusion, and movement epenthesis are among the many issues that arise when building an SLR. Despite a great deal of research being done on SLR, most problems are still unsolved. The majority of SLRs that have been created use numerous depth sensor cameras, color-coded gloves, or wearable sensors to collect data, which makes it extremely difficult for signers to communicate their signs in natural settings.

OBJECTIVE

The primary goal of this project is to create a Sign Language Recognition (SLR) system that can accurately and economically recognize American Sign Language (ASL) as well as custom sign gestures. By maximizing feature extraction with HDNN with atrous convolutions, attention-based methods, and enhanced auto encoders, we hope to increase recognition accuracy. Furthermore, real-time identification will be made possible via the project's Flask web-based detection module, opening up manual input for more generally accessible and practical applications.

PROBLEM STATEMENT

In order to precisely translate American Sign Language (ASL) and unique sign motions, the project intends to create a revolutionary vision-based sign language recognition framework utilizing a hybrid deep neural network (HDNN) technique. The system will efficiently handle spatial and temporal complications by tracking and extracting multi-semantic features, such as non-manual components and manual co-articulations, by utilizing cutting-edge deep learning algorithms. HDNN with atrous convolutions will be used to extract spatial features, while attention-based HDNN will handle the extraction of temporal and sequential features, allowing sign gestures to be dynamically captured. To improve discriminative capabilities, modified auto encoders will be used for abstract feature extraction. In order to differentiate between undesired transition motions and sign gestures, the system will also include a hybrid attention module, which will increase recognition accuracy.

EXISTING SYSTEM

- The gesture recognition component of the current sign language recognition system (SLRS) helps signers and non-signers communicate more easily. Using a deep learning strategy, the system applies computer-vision methods. Of special note is the Indian Sign Language (ISL) dataset that was generated from 65 users in an unmonitored setting.
- Three extra copies of each training image with different affine transformations are generated using data augmentation, which also increases intra-class variance to improve generalization.
- With an accuracy of 92.43% on the self-collected ISL dataset, the model uses Convolution Neural Networks (CNN) for feature extraction and ISL gesture classification. Recall, f-score, precision, and system processing time are all included in the evaluation procedure.

Disadvantage of Existing System

- When working with dynamic components of sign language, where a detailed understanding of temporal patterns is required, this could provide a problem.
- Lack of labelled training data could lead to reduced performance as it could be challenging to generalize.
- They might not be as effective at handling sequential data.

PROPOSED SYSTEM

- The proposed system addresses the challenges faced by the community of individuals with speech and hearing impairments by introducing a novel approach to sign language recognition. Our strategy is to provide an affordable and easily accessible way for the general population to learn and interpret sign language.
- Our approach leverages vision-based technology and enhances both spatial and temporal information to accomplish accurate recognition, in contrast to existing systems that rely on wearable sensors. The system also features a user-friendly Flask web interface for real-time sign language recognition and manual photo entry.

Advantages of Proposed System

- Consciousness of intricate patterns and connections among several sign movements. This flexibility contributes to improved accuracy, especially in capturing the vibrant motions and co-articulations that define sign language.
- The system's applicability can be strengthened and expanded by recognizing indicators provided by a variety of individuals and in a variety of settings.

RELATED WORKS

Prior studies largely concentrated on wearable sensors or disregarded important temporal and spatial information needed for the recognition of sign language. Although there are some vision-based methods, they frequently have trouble capturing all of the semantic aspects of sign motions. While deep learning techniques like CNNs and RNNs have been studied recently, a thorough framework for multi-semantic property extraction is still lacking. Validated on a novel multi-signer ASL and custom sign language dataset, our proposed methodology combines HDNNs with atrous convolutions, attention mechanisms, and modified auto encoders to overcome these drawbacks and achieve superior recognition performance, with real-time recognition enabled through a Flask web-based detection module.

METHODLOGY OF PROJECT

In order to recognize American Sign Language (ASL) and unique sign movements, the research suggests using a revolutionary vision-based hybrid deep neural network (HDNN) technology. The system attempts to precisely capture the multi-semantic features of sign language by utilizing HDNN with atrous convolutions for spatial feature extraction and attention-based HDNN for temporal and sequential feature extraction. Furthermore, abstract feature extraction is performed using modified auto encoders, which improves discriminative power. Supported by the integration of a Flask web detection module for real-time recognition, experimental validation on a multi-signer ASL and custom sign language dataset indicates the superiority of the proposed framework over current state-of-the-art approaches.

MODULES DESCSRIPTION:

1) Dataset

A customized dataset is meant to be used with the system's Multi-Cultural Sign Language Detection and Recognition capabilities. Together with signs that match the English alphabet from A to Z, the collection also includes extra signs like 'del' (delete), 'nothing', and'space'. The terms "work," "today," "time," "thank you," "rain," "practice," "love," "internet," "home," "help," "hello," "happy," "game," "friends," "food," "family," "delete," "chocolate," "brave," and "beautiful" are also marked with indicators. There are a total of 29 classes for ASL and 20 classes for the Custom dataset.

2) Importing Required Libraries

The programming language of choice is Python, and the libraries needed for model building and training are imported. These libraries include standard libraries like matplotlib, pandas, numpy, and tensorflow; additionally, Keras is used for model building, scikit-learn is used for data partitioning, and PIL is used for image processing.

3) Finding and Preparing Images

Photos and captions are extracted from the assigned folder (train_folder). The images are subjected to preprocessing operations such as numpy array conversion, scaling to (64, 64), and edge detection.

4) Dividing the Dataset

10% of the dataset is meant for testing, while 90% is meant for training**.**

5) Constructing the HDNN Model

A Hierarchical Deep Neural Network (HDNN) model is constructed with Keras. Dense layers, recurrent layers (LSTM), flatten layers, and convolutional layers (Conv2D) with max-pooling make up the model architecture. 29 nodes are produced by the final dense layer, one for each of the 29 classes. A different HDNN from the last dense layer creates 20 nodes that match the 20 classes by using the softmax activation function for multiclass classification.

6) Educating the Model

Using the Adam optimizer and categorical cross-entropy loss, the model is built. On the prepared dataset, it is then trained using the fit function. In addition to tracking the training process, accuracy and loss graphs are displayed.

7) Accuracy on Test Set

The accuracy of the model is assessed on the test set following training, and it achieves a 99.7% accuracy rate.

8) Preserving the Trained Model

To store the trained HDNN model in an HDF5 file named "HDNNsign.h5," using the tensorflow Keras model.save() method. This stage must be finished before deploying the model in production-ready environments. Pickle library is also touted as a tool for storing the model as a.H5 file, although the provided code does not have the actual implementation.

ALGORITHM USED IN PROJECT

- Hybrid Deep Neural Network (HDNN)-based Method
- The proposed method enables strong sign language identification by utilizing a novel vision-based hybrid deep neural network. This system can recognize both custom sign motions and American Sign Language (ASL). An enhanced method for extracting spatial information and providing a comprehensive understanding of the sign gestures is a Hybrid Deep Neural Network (HDNN) equipped with atrous convolutions. Unlike traditional LSTM, our approach leverages attention-based HDNN to effectively extract temporal and sequential characteristics. Distinctive abstract features are extracted using modified autoencoders. The hybrid attention

module is used to extract discriminative features and separate unwanted transition movements from sign motions.

- ➢ The testing on a multi-signer ASL and custom sign language dataset shows that our proposed technique performs better than previous state-of-the-art frameworks. The system's usability and accessibility are enhanced by the use of Flask web, which enables real-time sign language identification with human input.

SYSTEM ARCHITECTURE

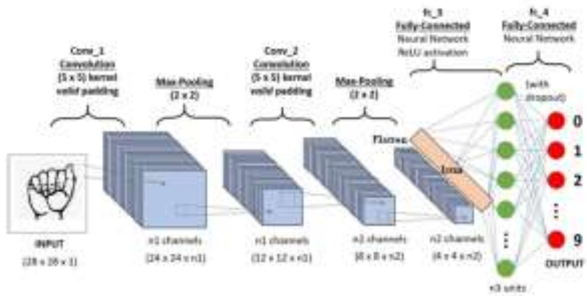
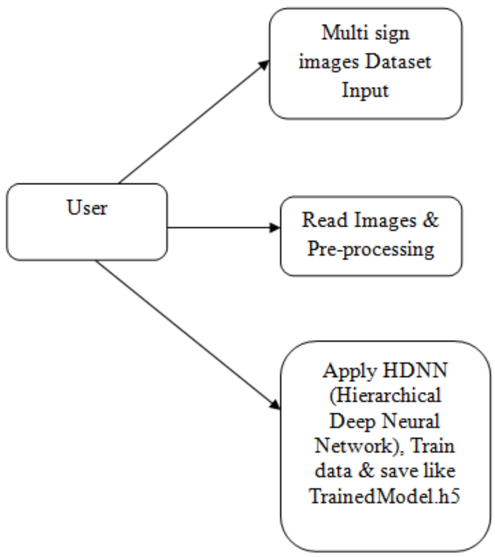


Fig: 7 SYSTEM ARCHITECTURE OF PROJECT

DATA FLOW DIAGRAM



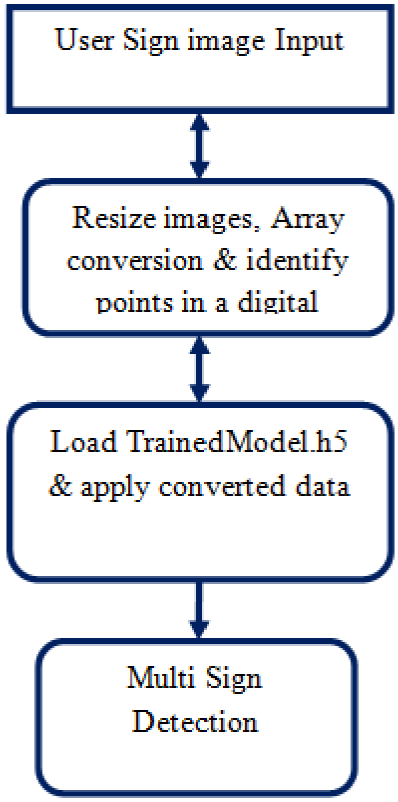


Fig: 8 Flow Diagramsof Modules

RESULTS

FUTURE ENHANCEMENT

As a part of future work, we would like to extend our study toward continuous sign sentence recognition. We would also like to design a framework for handling the segmentation ambiguities and moment epenthesis in continuous sign language recognition. The isolated word gesture recognition framework could also be integrated to enhance sign spotting from continuous sign video stream for recognition sign sentences from continuous sign gestures. We also intend to increase the dataset and publicly publish it for further research.

CONCLUSION

By building an HDNN architecture, a novel Sign Language Recognition system has been constructed. The goal of the suggested HDNN architecture is to extract non-manual and semantic manual co-articulations, which are essential elements required for sign recognition. Accurate recognition also takes into account temporal, sequential, and spatial information. Additionally, to distinguish between real and useless gestures, abstract and discriminative feature extraction is also done. Afterwards, the real gestures are employed to identify the Sign gesture representations, which minimizes the computational overhead. The ASL Dataset and newly developed Custom data have been used to test the proposed HDNN architecture. The output shows a proficient and effective performance.

REFERENCES:

[1]. Y. Saleh and G. F. Issa, ‘‘Arabic sign language recognition through deep neural networks fine-tuning,’’ Int. J. Online Biomed. Eng., vol. 16, no. 5, pp. 71–83, 2020.

[2]. X. Jiang, M. Lu, and S.-H. Wang, ‘‘An eight-layer convolutional neural network with stochastic pooling, batch normalization and dropout for fingerspelling recognition of Chinese sign language,’’ Multimedia Tools Appl., vol. 79, nos. 21–22, pp. 15697–15715, Jun. 2020.

[3]. O. Sevli and N. Kemaloğlu, ‘‘Turkish sign language digits classification with CNN using different optimizers,’’ Int. Adv. Researches Eng. J., vol. 4, no. 3, pp. 200–207, Dec. 2020.

[4]. K. Wangchuk, K. Wangchuk, and P. Riyamongkol, ‘‘Bhutanese sign language hand-shaped alphabets and digits detection and recognition,’’ Ph.D. dissertation, Dept. Comput. Eng., Naresuan Univ., Phitsanulok, Thailand, 2020.

[5]. R. Elakkiya and E. Rajalakshmi. ISLAN. Mendeley Data. Accessed: Jan. 8, 2021. [Online]. Available: https://data.mendeley.com/datasets/rc349 j45m5/1

[6]. W. Sandler, ‘‘The phonological organization of sign languages,’’ Lang. Linguistics Compass, vol. 6, no. 3, pp. 162–182, Mar. 2012.

[7]. R. Elakkiya, ‘‘Retraction note to: Machine learning based sign language recognition: A review and its research frontier,’’ J. Ambient Intell. Human-ized Comput., vol. 12, no. 7, pp. 7205–7224, Jul. 2022.

[8]. S. Diwakar and A. Basu, ‘‘A multilingual multimedia Indian sign language dictionary tool,’’ in Proc. IJCNLP, 2008, p. 57.

[9]. S. K. Liddell and R. E. Johnson, ‘‘American sign language: The phonological base,’’ Sign Lang. Stud., vol. 1064, no. 1, pp. 195–277, 1989.

[10]. P. Eccarius and D. Brentari, ‘‘Symmetry and dominance: A cross-linguistic study of signs and classifier constructions,’’ Lingua, vol. 117, no. 7, pp. 1169–1201, Jul. 2007.

[11]. Z. Ren, J. Yuan, J. Meng, and Z. Zhang, ‘‘Robust part-based hand gesture recognition using Kinect sensor,’’ IEEE Trans. Multimedia, vol. 15, no. 5, pp. 1110–1120, Aug. 2013.

[12]. C. Wang, Z. Liu, and S.-C. Chan, ‘‘Superpixel-based hand gesture recog-nition with Kinect depth camera,’’ IEEE Trans. Multimedia, vol. 17, no. 1, pp. 29–39, Jan. 2015.

[13]. Z. Liu, X. Chai, Z. Liu, and X. Chen, ‘‘Continuous gesture recognition with hand-oriented spatiotemporal feature,’’ in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), Oct. 2017, pp. 3056–3064.

[14]. P. Molchanov, X. Yang, S. Gupta, K. Kim, S. Tyree, and J. Kautz, ‘‘Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 4207–4215.

[15]. R. Agarwal, ‘‘Bayesian K-nearest neighbour based redundancy removal and hand gesture recognition in isolated Indian sign language without materials support,’’ IOP Conf. Ser., Mater. Sci. Eng., vol. 1116, no. 1, 2021, Art. no. 012126.

[16]. O. Aran, ‘‘SignTutor: An interactive system for sign language tutoring,’’ IEEE MultiMedia vol. 16, no. 1, pp. 81–93, Mar. 2009.

[17]. I. N. Sandjaja and N. Marcos, ‘‘Sign language number recognition,’’ in Proc. 5th Int. Joint Conf. INC, IMS IDC, 2009, pp. 1503–1508.

[18]. Q. Yang, ‘‘Chinese sign language recognition based on video sequence appearance modeling,’’ in Proc. 5th IEEE Conf. Ind. Electron. Appl., Jun. 2010, pp. 1537–1542.

[19]. S. Hore, ‘‘Indian sign language recognition using optimized neural net-works,’’ in Information Technology and Intelligent Transportation Systems. Cham, Switzerland: Springer, 2017, pp. 553–563.

[20]. P. Malhotra, ‘‘Indian sign language recognition system using openpose,’’ I-Manager’s J. Comput. Sci., vol. 7, no. 2, p. 43, 2019.

[21]. S. Reshna and M. Jayaraju, ‘‘Spotting and recognition of hand gesture for Indian sign language recognition system with skin segmentation and SVM,’’ in Proc. Int. Conf. Wireless Commun., Signal Process. Netw.(WiSPNET), Mar. 2017, pp. 386–390.

[22]. J. Raheja, A. Mishra, and A. Chaudhary, ‘‘Indian sign language recognition using SVM,’’ Pattern Recognit. Image Anal., vol. 26, no. 2, pp. 434–441, 2016.

[23]. K. Shenoy, T. Dastane, V. Rao, and D. Vyavaharkar, ‘‘Real-time Indian sign language (ISL) recognition,’’ in Proc. 9th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2018, pp. 1–9.

[24]. A. Tyagi and S. Bansal, ‘‘Feature extraction technique for vision-based Indian sign language recognition system: A review,’’ in Computational Methods and Data Engineering. Singapore: Springer, 2021, pp. 39–53.