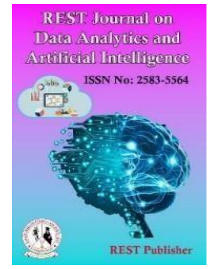
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/61



Sign Language Recognition for Deaf People Using Deep Learning

***Askam Datha Prasad, Bongarala Sowmya, Katukuri Vasu, Manoranjan Dash**
Anurag University, Hyderabad, Telangana, India
**Corresponding Author Email: dathaprasad5303@gmail.com*

Abstract: *Recent deep learning breakthroughs will make sign language easier to decipher, which will be vital to enable deaf individuals to communicate on a daily basis. Deep learning will be utilized in this study to convert sign language gestures into text, making it easier for everyone to know what deaf people are communicating. To enable this to function well, video data of sign language will be gathered using a software named OpenCV and will be thoroughly prepared for analysis. A unique type of neural network named a 3D Convolutional Neural Network (3D-CNN) will be constructed which can identify signs created using a single hand. In this project, various methods will be experimented with to train the neural network, including with optimizers such as Adam and Stochastic Gradient Descent (SGD). These optimizers will assist the network to learn quicker and better from the gathered data. Adam will tune the rate at which the network learns for every section of the data, which will be fantastic for large datasets. SGD will perform well when the data is not very distinguishable or is missing. Once we have done all of that, our system will be capable of translating sign language movements into text with a high degree of accuracy. This will result in it accurately understanding and interpreting signs into words the majority of the time. This technology will be significant because it will facilitate bridging the communication gap between sign language users and all other people. In the future, we aspire to further improve our system so it can assist more individuals in real-time conversations.*

Keyword: *Sign Language Recognition, 3D Convolutional Neural Network (3D-CNN), OpenCV, Real-time Communication, Gesture to Text.*

1. INTRODUCTION

For deaf people interacting with others, sign language is a vital tool of communication. It speaks with hand gestures, facial expressions, and body language. One of the main issues with sign language, nevertheless, is that not everyone speaks it. Deaf persons thus find it difficult to interact with others unfamiliar with sign language. This can be challenging and aggravating in environments including offices, classrooms, or even for something as simple as shopping or asking for help. Fortunately, the advances in technology help to solve this issue. Artificial intelligence (AI) and deep learning are one such development in which sign language motions as words or written text can be detected and understood. This technology teaches computers to "observe" and "transpose" the gestures through convolutional neural networks (CNNs), thus empowering expert models used in image and video recognition tasks.
CNN can be trained to identify video frames of an individual performing sign language gestures, detect the prominent features of the gestures—e.g., hand shape and path—and translate them into comprehensible words or text. This is one way to close the communication gap between sign language users and non-native speakers. With time, these systems are doing ever better. Since researchers are developing better models and expanding data, these systems can identify a wide spectrum of gestures and are more accurate. Additionally, systems for real-time recognition are being developed so that instantaneous translation is enabled in conversation, which has great potential for helping millions of deaf people worldwide.

For example, a deaf person can utilize this technology to communicate in public places like airports, banks, or restaurants without the aid of a human interpreter. This allows deaf people to communicate more freely, become more independent, and enables them to be able to communicate in social, educational, and professional situations more at ease. As these technologies evolve, they have the potential to make a significant improve the deaf community's quality of life, making communication more accessible and inclusive for everyone.

2. LITERATURE REVIEW

This literature review takes into account current IEEE research on sign language recognition, with a focus on the algorithms employed and the limitations. In 2022, a research created an end to-end model based on CNN and LSTM but was not scalable. Another 2022 paper utilized Motion History Images and 3D CNNs for single sign recognition, though its precision was low. In 2023, research proposed a sign language recognition CNN-based system, though its accuracy was compromised by a small training set and are of low precision. These articles demonstrate the potential of CNN based models but highlight challenges such as scalability, accuracy, and data set limitations.

The current sign language recognition systems are primarily based on technologies such as machine learning, deep learning, and computer vision to revolutionize hand gestures to semantic words or text. They try to bridge communication gap between the deaf and those who are not sign literate language. Sign language interpretation traditionally has been accomplished manually by human interpreters, which is a good but restricted solution. Human interpreters are not always available, and it is costly to depend on them, especially on a daily-scenarios. As a reaction to this, camera-based sign language recognition systems have emerged. These models track hand motion from video, and machine learning models are taught to identify these gestures using current datasets. One of the most popular models used in sign language recognition is the Convolutional Neural Network (CNN), which is best suited to process image and video data. CNNs are capable of identifying motion and hand shapes from images in order to classify them to their respective signs or gestures. These further extend the existing capabilities by bringing in additional technology like depth sensors, which provide more specific information about hand movement and position in space.

Kinect-based solutions or those based on Leap Motion technology are some examples. These systems record 3D information, which can greatly enhance the precision of gesture recognition through offering a richer insight into the orientation and motion of the hand.

Although all these advances, most current systems are only capable of recognizing isolated gestures, i.e., able to process single signs (e.g., single letters or words) and struggle to comprehend continuous flows of gestures. Most systems are also language-specific, with particular attention to American Sign Language (ASL), limiting their application by other individuals using other sign languages such as British Sign Language (BSL) or Indian Sign Language (ISL).

Briefly, while existing systems provide a foundation for automated sign language recognition, they are constrained by their capacity to interpret gestures in real-world conditions, and they typically rely on special hardware or controlled environments to function effectively.

TABLE 1. Literature Review

S No	Title Of The Paper	Year of Publication	Algorithms
1	Development of an End-to End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation	2022	CNN, LSTM
2	Using Motion History Images with 3D Convolutional Networks in Isolated Sign Language Recognition	2022	CNN
3	A Sign Language Recognition System for Helping Disabled People	2023	CNN

Proposed Model: The proposed system is to create an innovative sign language recognition system based on machine learning and deep learning methods, i.e., a 3D Convolutional Neural Network (3D-CNN). This is an architecture for processing video and image data in real-time and convert sign language movements into text or speech. This will improve the speech and hearing of the deaf and hard-of-hearing persons by a strong and successful gesture understanding device without the requirement of human interpreters. It will be an easy system, installable with little hardware, and scalable to support many sign languages and gestures. Unlike conventional systems that rely greatly on human or manual translation interpreters, this autocorrecting feature does its best to fill in those gaps with current-day deep learning algorithms that can identify intricate patterns and movements in hand movements.

Through real-time processing and an appropriately structured architecture, the system attempts to achieve high accuracy with the capacity to learn from real-world applications.

The following are the primary modules that the proposed system consists of:

1. Input Video Data

- The system starts with capturing live video feed from a camera, typically a webcam, using OpenCV.
- The camera continues to record the user hand-gesturing in real time. The video stream is the input to the recognition system.
- The input must be capable of having a clear view of the user's hands so that the hand movements are identified by the system correctly.

2. Pre-processing

- Here, the raw video frames are preprocessed so that the data is made more compatible for the deep learning model.
- Frame Extraction: Frames are pulled out from the video stream. Rather than processing the entire video, frames are isolated for analysis.
- Grayscale Conversion: The frames are then converted into grayscale, which reduces the data by eliminating unnecessary color information. This reduces the computational burden without sacrificing the significant features of the gesture.
- Standardizes the input to be properly formatted for subsequent processing in the model.

3. Segmentation

- The system distinguishes the regions of interest (ROI), i.e., the hand movements.
- Hand Gesture Segmentation: Here, the OpenCV algorithms are used to segment and detect the hand from the background. Segmentation of the hand regions enables the model to focus only on the gesture without any interference from the environment.
- This is a critical step because it limits the region to be processed, increasing the accuracy of feature detection and gesture recognition.

4. Feature Extraction

- After segmenting the hand gestures, 3D-CNN (3D Convolutional Neural Network) is used to extract features from hand motion data.
- The 3D-CNN considers the spatial and temporal information of the video frames. It extracts:
o Spatial Information: The hand's orientation, movement, and shape in the frame. o Temporal Information: The sequence of movements among a number of frames.
- This dual process of feature extraction is required for distinguishing between different gestures, especially since sign language includes not just static position but also the dynamic movement.

5. Model Training

- The attributes are utilized to train the 3D-CNN model during the training process.
- The model is trained on a huge set of labeled video sequences, where a sequence is associated with a specific sign language gesture.
- Adam Optimizer: The model is optimized using the Adam optimizer, a robust and effective optimization algorithm commonly used in deep learning. It adapts the learning rate dynamically and is thus especially suitable for training on different datasets.
- In this stage, the model learns the pattern and structure of various gestures in the dataset so that it can make correct predictions when run for real-time execution.

6. Prediction

- The model is employed in real-time identification once it has been trained.
- The pre-processed video frames (of the ongoing input) are input into the trained 3D-CNN model to estimate the corresponding gesture.
- The model compares the input gesture to its learned patterns and gives the anticipated sign.

7. Post-Processing

- Once the gesture is expected, post-processing techniques are applied by the system to fine tune the output:

o Minor Error Corrections: Minor errors may be caused by noise, variability of hand motion. These errors are corrected by post-processing.
o Inter-frame consistency: The system will be able to verify across several frames whether or not the gesture is consistent. If one of the frames includes ambiguity, the system can rely on context information in nearby frames to make an improved estimate.
o Multiple Predictions: If the gesture is complex, the system can make multiple predictions across multiple frames to improve accuracy.

8. Output: Gesture Display

• Finally, the recognized gesture is translated into words or text.
• The system shows the identified sign, which may be a word, command, or sentence, depending on the input gesture.
• The identified sign is made clear to non-sign language users, and it is simple for sign language users to interact with other people.

Challenges and Solutions:

1. **Single Gesture Recognition**: In this test case, the system is tasked with recognizing a single, static gesture, such as "hello," "yes," "no," "please," or "thank you." This evaluates the model's ability to correctly classify individual gestures when performed in isolation. The goal is to ensure that each gesture is correctly identified, without confusion with other gestures. This forms the basis of gesture recognition accuracy and is fundamental to system performance.
2. **Multiple Gestures in Sequence**: In this scenario, the system is tested with a series of different gestures performed one after another in real-time. The challenge here is to see if the model can smoothly transition between gestures without misclassifying them. This scenario is vital for determining how the system handles dynamic changes in hand movements, simulating real-world conversations in sign language.
3. **Real-Time Gesture Recognition**: The system's performance is assessed using a live webcam feed to simulate real-world conditions where gestures are detected and recognized in real-time. This test case is critical because the system must not only recognize gestures correctly but also do so with minimal delay to ensure smooth communication. The success of this test case ensures that the system can be deployed in interactive environments, where users can sign gestures live, and receive immediate feedback.
4. **Variable Lighting and Backgrounds**: The system is tested under different lighting conditions and with varying backgrounds to assess its robustness. These tests simulate real world environments where lighting may not be optimal, or the background may be cluttered or noisy. This is a crucial test case because variations in lighting and background could lead to performance degradation in a real-world setting. The system should be able to maintain high accuracy across a range of different environments, ensuring it is usable in diverse conditions, such as indoors, outdoors, and in low light.

3. CONCLUSION

The deep learning-based sign language recognition system demonstrates considerable scope for closing the communication gap in the case of hearing-impaired individuals. Communication barriers typically hinder interactions between non-hearing and hearing individuals, and such a system offers a good solution by detecting hand movements and translating them into readable text in real-time. The use of 3D Convolutional Neural Networks (3D-CNN) is of great significance in the success of the model in successfully extracting both spatial and temporal features from videos. As opposed to classic image-based models, 3D-CNN can learn movements over time, which is particularly well-suited for detecting dynamic gestures which form sign language. The model was also trained on a well-built dataset, with each movement split into a series of frames, to allow accurate identification of intricate hand movements. OpenCV library for computer vision was utilized for video data preprocessing. It provided unintrusive hand movement capture from video frames and was dedicated only on topics of interest—hands—ensuring the system learned the appropriate used patterns effectively. The gesture information from videos was divided into 80% training data and 20% test data, amounting to a highly precise model with a great recognition accuracy of 99.58%. The ability to recognize nine different gestures, including "Yes," "No," "Hello," "Thank You," "Drink," and others, is indicative of the system's flexibility. The real-time aspect of the recognition process so that the model can be applied to actual cases with little or no lag. Even under slightly changed illumination and background conditions, the system possesses high performance, to ensure that it is robust enough for applied usage.

This technology is one step further towards greater accessibility for the deaf community. Converting gestures into text, the system provides a more seamless interaction between individuals with hearing handicaps and the rest of society, and promoting inclusive communication in work and personal life. With future enhancements, such as the addition of more gesture classes for wider sign language coverage, to support multiple sign languages (e.g., ASL, BSL), tuning the model for deployment in mobile devices or embedded systems, this system has a great impact in everyday interactions. It can be applied in in schools, hospitals, workplaces, and public areas, providing make communication more accessible and inclusive. Briefly, the project's accuracy of 99.58% validates the effectiveness and the accuracy of the model.

REFERENCES

[1]. K. K. Podder, M. E. H. Chowdhury, A. M. Tahir, Z. B. Mahbub, A.Khandakar, M.S. Hossain and M. A. Kadir, "Bangla Sign Language (BdSL) Alphabets and Numerals Classification Using a Deep Learning Model," Sensors, vol. 22, no. 2, pp. 574(1-18), 2022.
[2]. F. Tanrisever and K. A. W. Voorbraak, "Crowdfunding for Financing Wearable Technologies," 2016 49th Hawaii Int. Conf. System Sciences (HICSS), pp. 1800 1807, 2016. 3.
[3]. A. Mannan, A. Abbasi, A. R. Javed, A. Ahsan, T. R. Gadekallu and Q. Xin, "Hypertuned Deep Convolutional Neural Network for Sign Language Recognition," Computational
[4]. Intelligence and Neuroscience, vol. 2022, Article ID 1450822, pp. 1-10, 2022. 4. S. C. Das, M. B. Alam, M. S. A. Moon and M. S. Mia, "An Application Programming Interface to Recognize Emotion using Speech Features," 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), pp.16, 2022.
[5]. A.Mavi, "27 Class Sign Language Dataset,â Kaggle, [Online]. Available:https://www.kaggle.com/datasets/ardamavi/27-class-sign-language dataset. [Accessed: 26-Oct-2023].
[6]. Q. M. Areeb, Maryam, M. Nadeem, R. Alroobaea and F. Anwer, "Helping Hearing Impaired in Emergency Situations: A Deep Learning Based Approach," IEEE Access, vol. 10, pp. 8502- 8517, 2022.
[7]. A.Halder and A. Tayade, "Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning," Int. J. Research Publication and Reviews, vol. 2, no. 5, pp. 9-17, 2021.
[8]. N. Mohamed, M. B. Mustafa and N. Jomhari, "A Review of the Hand Gesture Recognition System: Current Progress and Future Directions," IEEE Access, vol. 9, pp. 157422-157436, 2021.
[9]. Y. Obi, K. S. Claudio, V. M. Budiman, S. Achmad and A. Kurniawan, "Sign language recognition system for communicating to people with disabilities," Procedia Computer Science, vol. 216, pp. 13-20, 2023.
[10]. V. Adithya and R. Rajesh, "Hand gestures for emergency situations: A video dataset based on words from Indian sign language," Data in Brief, vol. 31, pp. 1-7, 2020.
[11]. H. R. V. Joze and O. Koller, MS-ASL: A large-scale data set and benchmark for understanding American sign language, 2018, arXiv:1812.01053.
[12]. J.Huang,W.Zhou,H.Li,andW.Li, Attention-based3D-CNNsforlarge vocabulary sign language recognition, IEEE Trans. Circuits Syst. Video Technol., vol. 29, no. 9, pp. 28222832, Sep. 2019.
[13]. D. Li, C. R. Opazo, X. Yu, and H. Li, Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison, in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Mar.2020, pp. 14591469.
[14]. O. M. Sincan and H. Y. Keles, AUTSL: A large scale multi-modal Turkish sign language dataset and baseline methods, IEEE Access, vol. 8, pp. 181340181355, 2020. 15. O.Özdemir,A.A.Kndro lu,N.C.Camgöz,andL.Akarun, Bosphorus sign22k sign language recognition dataset, in Proc. LREC 9th Workshop Represent. Process. Sign Languages, Sign Lang. Resour. Service Lang, Community, Technol. Challenges Appl. Perspect., 2020, pp. 181188.