Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

AI-Powered Hand Gesture Recognition based on Adaptive Thresholding and Gaussian Blur for Human-Computer Interaction using CNN

Ritesh Kansal, Arsalan Reyaz, Harsh Sharma, Ram Paul, Sanjiv Kumar Tomar
Department of *CSE, ASET, Amity University Uttar Pradesh, Noida, India*

ARTICLE INFO

Received: 24 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

A real-time system to recognize American Sign Language motions has been developed based on Adaptive Thresholding, Gaussian Blur using CNN to help Deaf and Dumb people and others communicate. The main objective of our work is to build a model that can recognize hand gestures from fingerspelling and combine individual gestures to form words. The study classifies American Sign Language fingerspelling movements recorded by a webcam using Convolutional Neural Networks and cutting-edge computer vision algorithms. With the use of highly advanced methods like adaptive thresholding and Gaussian blur using CNN the experimental results improved gesture prediction accuracy to an astounding 98%. This approach achieves remarkable performance by overcoming traditional limitations in gesture detection by integrating two layers of algorithms. Although the study shows the possibility for expanding the system to additional sign languages, the current focus is on American Sign Language.

Keywords: ASL, Hand Gesture Recognition, CNN, Computer Vision, Sign Language Translation

INTRODUCTION

Sign language used in America is the most widely used type. Because they are unable to use spoken languages, Deaf and Dumb (D&M) people can only communicate through sign language. Communication is the process of conveying concepts and messages by voice, gesture, action, and picture. D&M people use a variety of motions with their hands to interact with others. Gestures are nonverbal cues that are only visible to the human eye. The nonverbal communication method used by the deaf and dumb is called sign language. Sign language is a language in which hand shapes, orientation, and other gestures are used to convey meaning instead of sounds.

Letter-by-Letter Spelling	**Vocabulary Through Individual Signs**	**Facial and Bodily Expressions**
This method is used to represent words by signing each letter individually.	Most communication is carried out using specific signs for entire words.	Includes facial expressions, mouth shapes, and body movements.

Fig. 1. Sign language is a visual language and consists of 3 major components

Making sure that everyone communicates verbally becomes a goal, as does closing the verbal communication gap between D&M and non-D&M people. The most natural form of communication for those who are deaf or hard of hearing is sign language translation, which is one of the fastest-growing fields of study. Thanks to hand gesture detection technology, deaf people can now interact with vocal humans without the need for an interpreter. The system is made to convert text and ASL into speech automatically.

The desired gestures to be trained are those depicted in the figure.

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

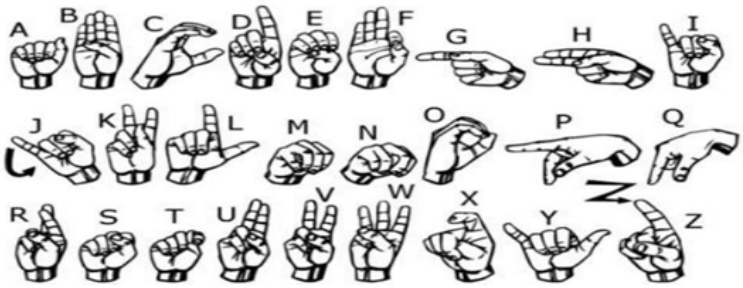

Fig. 2. ASL Sign language and gesture

According to an *evaluation conducted in 2011[7]*, 6.3 crore persons, or 6.3% of India's population, were deaf. Of these, 76–89% of Indians who have hearing loss are illiterate in written, spoken, or sign languages. There are several reasons for this low literacy rate, such as a dearth of resources, interpreters, and courses offered in Indian Sign Language. A speech-to-message software and gesture communication can be useful for a number of reasons.

- **Accessible**: Spoken material promotes inclusivity by making it understandable to those who use assistive listening devices or have hearing impairments. By applying discourse analysis to meaning, they can comprehend spoken language and take part in discussions. For certain individuals who are hard of hearing or on the verge of becoming deaf, gesture-based communication is an essential mode of communication. The program makes communication easier for people who are hard of hearing or who are not familiar with gesture-based communication by translating text interpretation into gesture-based communication.
- **Training**: In educational institutions like schools and colleges, the application can be utilized as a teaching tool. Enhance my writing by bridging the knowledge gap between easy and tough learning.
- **Public Speaking and Introductions:** The program can help deafen people making introductions or public speeches by interpreting their pre-planned message into communication through gestures. They may then share their ideas and viewpoints with a larger audience.
- **Interpretation**: The application can help translate verbal communication into gesture-based interpretation when a gesture-based communication translator is unavailable.
- Indian Sign Language has gotten less attention than American Sign Language, despite the latter being the focus of much research, as demonstrated by *Peguda et al. [1]*. The primary goal of the work was to convert each of the six distinct zonal languages' recorded voice into ISL. The model that was suggested took vocal input and produced a series of movements that were shown on a screen. It comprises of text interpretation using LSTM, text planning using sign language, and speech recognition using MFCC combined with GMM based on Wavelet transformation.

A model for translating recorded speech to Indian Sign Language was developed by *Ezhumalai P. et al. [2]*. using Google's audio Recognizer API, the model translated recorded audio into text. This paradigm indicated the proper sign language or gesture for each word in a phrase. This approach was limited to translating from English into Indian Sign Language. The model followed the following flow:

- Text entered, audio files, or spoken words could be utilized.
- For speech recognition, utilize the Google Speech Recognizer API.
- Text preprocessing comprises trimming superfluous words and filler from the text and then doing things like rooting.
- Lastly, web scraping techniques and OpenCV were used to transform text to Sign Language.

Real-time speech to American Sign Language translation was examined by *Khalid El-Darymli et al. [3]* utilizing the proposed voice to Sign Language Interpreter System (SSLIS). Instead of using ASL syntax for translation, this model used Signed English (SE). As was already noted, this model's architecture was:

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

1). Live speech captured with a microphone or other recording device is converted into text by a speech recognition engine. This was done using the Sphinx 3.5 model.

2). The recognized text is added to the ASL database in order to search for matches.

3). In the event of a match, the ASL interpretation will be presented in accordance with English as per the Signed English (SE) handbook, instead of following the ASL sentence structure.

Lalit Goyal et al. [4] state that this method is different from the traditional approach of translating English into Indian Sign Language (ISL) by using artificial animations that spontaneously arise. The system in question consists of six modules:

[1] Text Interpreter: ISL grammar-based sentence structure with no extraneous or empty words

[2] Lemmatization words are converted to SiGML for synthetic animations using the HamNoSys Module. Similar techniques for converting text from English to ISL have been discussed in [5] and [6].

Purush The development of INGIT, a system that converts Hindi handbook strings into Indian Sign Language (ISL) was made possible in large part by *Purushottam Kar et al. [7].* The primary driving force for the creation of this system was the field of railway inquiry. The Hindi alphabet was supplied to the system via FCG. Using the produced module, a thin semantic structure was constructed from the stoner input. Ellipses Resolution removed unnecessary words from the input, which further altered the data. The appropriate ISL-markers were created using the ISL creator based on the kind of building. The semantic framework provided by this system was over sixty-delicious.

A domain-specific system that required input from the English textbook was developed by *Ali et al. [8].* After that, the textbook is transformed into an ISL textbook, which is then further transformed into ISL symbols:

1. These parts are part of the armature portion of the system.
2. The system's architecture consists of the following components:
3. An Input Module for Text Translation
4. A tokenizer that separates sentences into their component words
5. Several railway questions along with an ISL symbol library. Its synonym is used when there isn't any equivalent sign language.
6. Each phrase was matched to its matching symbol using a translator.

The development of a two-phase approach for sign language construction was the aim of *Vij et al. [9].* Preparing Hindi 2 sentences and translating them into ISL grammar comprised the first section. WordNet and Dependency Parser were coupled in the first phase. The relationships between words that represent those heads and the words themselves possess the ability to alter those heads. The grammar acquired in the first phase is translated into Sign language in the second phase using HamNoSys. Subsequently, the produced symbols are converted into XML components using SIGML. The forms made up of these XML components can then be interpreted by a three-dimensional rendering engine.

In *[10], MS Anand et al.* demonstrated a two-way ISL translation. This module also used a submodule to eliminate noise from the voice that was received as input. A system for text sequence decoding and speech recognition then receives the output. Lastly, a text annotation and animation module were used to display the corresponding signs for the spoken input. In the framework created by *Dasgupta and associates [11].* Following its consumption as input, English text was transformed into an ISL structure that complied with ISL grammar rules. The parts of the system were as follows:

1. Text analyzer and syntax parser combined
2. Making use of the LFG f-structure for illustration
3. Use of grammar principles to create suitable ISL sentences

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The input text was parsed using parsers such as Minipar Parser in order to generate a dependency structure. The outcome is an f-structure, which is a structure that can represent grammatical relationships in the input. The grammatical relationship among the input suggests the phrase's subject, object, and tense. The data was represented by an attribute-value pair combination. Every feature was associated with the proper grammatical name. Not only was the F-structured English produced correctly, but it was also reformed to an ISLF-structure. There is not enough correct ISL orthography, making system evaluation challenging.

A publicly accessible dataset containing signs for each letter of the English alphabet was used by *Yogeshwar I. Rokade et al. [12]* to construct their model. A vision-based technique that uses a web camera for the model is generated via hand gesture recognition *[13]*.

A deep neural network-based model was proposed by *Shashi Pal Singh* and colleagues, which can accomplish rule selection, joint translation, reordering, word alignments, and language modeling. Lastly, *[14]* offers a model that offers a method for using MFCC (Mel- Frequency Cepstral Coefficients) in the operation of various algorithms. A voice synthesizer was created utilizing language model creation for speech input, feature extraction, and acoustic model construction, based on research regulated for *TSL (Telugu Sign Language) [15]*. The Sphinx-3 algorithm was used to translate Telugu voice into legible text. The program operates in two phases*: training and decoding [16]*. Speech processing applications include machine translation, segmentation and clustering stem method, stemmer, and *rule-based morphological analyzer [17]*. While little to no work is put into text-to-ISL conversion models, a number *of MFCC models are built for the recognition of speech in Tamil and Telugu [18], [19], and [20]*.

The Kaldi Toolkit was used for speech recognition in a Malayalam investigation, and text notation from *HamNoSys [21]* was then obtained using SiGML. Speech signals in Manipuri, Hindi, and Urdu were recorded and created using the *Goldwave Software Tool [22]*. A two-way method was presented for the goal of *text to gesture (T2G) conversion [23]*. Long-Short Term Memory (LSTM) was used to train 40 samples over 30 epochs for each character in *ISL [24]*. Sentences containing difficult words were parsed using *RTR processing in combination with LSTM [25]*. Not much study has been done on developing a system that can process many languages concurrently to construct Indian Sign Language due to regional linguistic variety and factors like corpus and language variation.

SYSTEM MODEL

A) Feature Extraction and Representation

The representation of a picture as a three-dimensional matrix where each pixel has a height, width, and depth (1 for grayscale pixels and 3 for RGB pixels). CNN also makes use of these pixel values to derive useful properties.

B) Artificial Neural Network

A network of connections between neurons that resembles the structure of the human brain is called an artificial neural network. Each connection allows information to go from one neuron to the next. Inputs are transferred to the second layer of neurons, sometimes referred to as the hidden layers, after being processed by the first layer of neurons. After passing through many tiers of hidden layers, data is sent to the ultimate output layer.

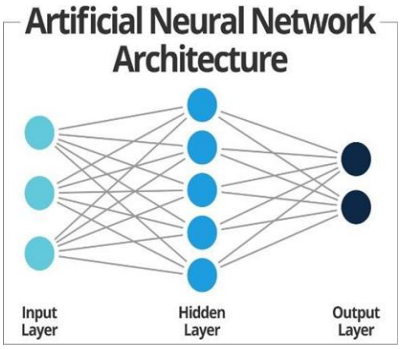

Fig. 3. The ANN(Artificial Neural network) Architecture

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

These require training and have the capacity to learn. Various learning methodologies exist:

- Unsupervised Learning
- Supervised Learning
- Reinforcement Learning

C) Convolutional Neural Network (CNN):

Unlike traditional neural networks, CNN has layers with neurons organized in three dimensions: depth, breadth, and height. The neurons in a layer are not linked to every other neuron in the layer; instead, they are only related to a small section of the layer (window size) that comes before it. Additionally, at the conclusion of the CNN architecture, the entire picture would be converted to the final output layer would have dimensions (number of classes) since there would only be one vector of class scores.

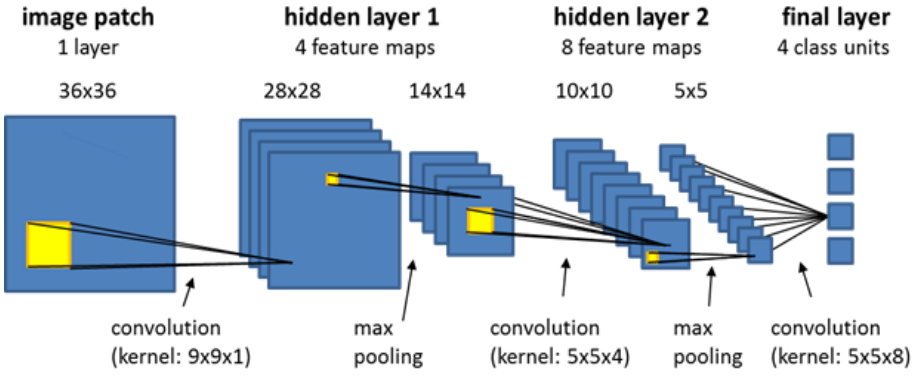

Fig. 4. Convolutional Neural Network

I. Convolution Layer

To reach the depth of the input matrix, we employ a small window size (often 5 by 5) in the convolution layer. The layer consists of learnable filters that are window sized. In every iteration, we move the window by the stride size, which is typically 1, after determining the dot result of the provided values at a certain location. Along the way, a two-dimensional activation matrix will be created, which will display the matrix's reaction at each spatial position. In other words, filters will activate when the network detects particular types of visual characteristics, such an edge with a specific direction or a blotch with a certain color.

II. Pooling Level

Eventually, we use a pooling layer to reduce the size of the learnable parameters and the activation matrix**.**

Maxim Pooling: This method selects only the highest four values using a window dimension (for example, 2*2). We will ultimately have an activation matrix that is half the size of its initial size if we close this window and proceed as described. All values inside a window are used when **average pooling** is used.

D) Tensor-Flow

Tensor-Flow is a complete, open-source machine learning platform. Its extensive and adaptable network of libraries, tools, and community resources makes it easier for developers to create and deploy machine learning programs while also enabling academics to advance the field.

You can choose the right abstraction level with TensorFlow based on your needs. You may quickly start creating and training models, using TensorFlow and machine learning, by utilizing the high-level Keras API. Quick iteration and intuitive debugging are made possible by eager execution, if you want greater flexibility. Without altering the model specification, you may spread training across various hardware configurations for large-scale machine learning applications by using the Distribution Strategy API.

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

E) Keras

Keras, a high-level neural network framework written in Python, is wrapped around TensorFlow. When it's wanted to swiftly create and test a neural network with few lines of code, it's utilized. It provides tools to facilitate working with text and picture data in addition to solutions for commonly used artificial neural network elements, including optimizers, layers, objectives, and activation functions.

METHODOLOGIES USED

We used state-of-the-art approach and modern technologies to fully explore our study issues. We made sure that the results were solid and trustworthy by using state-of-the-art techniques for data collecting and analysis. This strategy demonstrated our dedication to quality and innovation by enabling us to overcome conventional constraints and obtain deeper insights. Our conclusions are more credible because we also carried out thorough testing to confirm our procedures. We were able to successfully handle difficult problems because to our all-encompassing approach.

1) Dataset Generation

The foundation of the system is a vision. Since all signs are made by hand, the issue of creating false bias for commercial purposes is resolved. We looked for already produced datasets for the design, but we were unable to locate any in the format of unedited photos that meet our requirements. The only datasets that we could locate were RGB value datasets. Our decision to create our own dataset was motivated by this. The following are the steps we took to create our dataset. Our dataset was created using the Open Computer Vision (OpenCV) framework. For training reasons, we first took around 800 pictures of each sign in American Sign Language (ASL), and for testing, we took about 200 pictures of each symbol.

For training reasons, we first took around 800 pictures of each ASL sign, and for testing, we took about 200 pictures of each symbol. Initially, we record every frame that the webcam on our computer displays. In each picture, we designate an area of interest, which is denoted by a blue-bounded border, as seen in the image below

Fig. 5. Creating dataset

The Gaussian Blur Filter is then applied to the image, allowing us to extract different characteristics from it. After adding Gaussian Blur, the image has the following appearance:

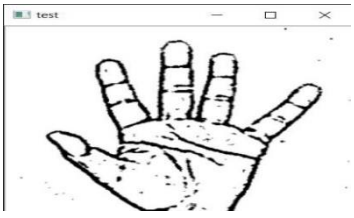

Fig. 6. Image after applying Gaussian Blur Filter

2) Gesture Classification

Our method predicts the user's final symbol using two layers of algorithms

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

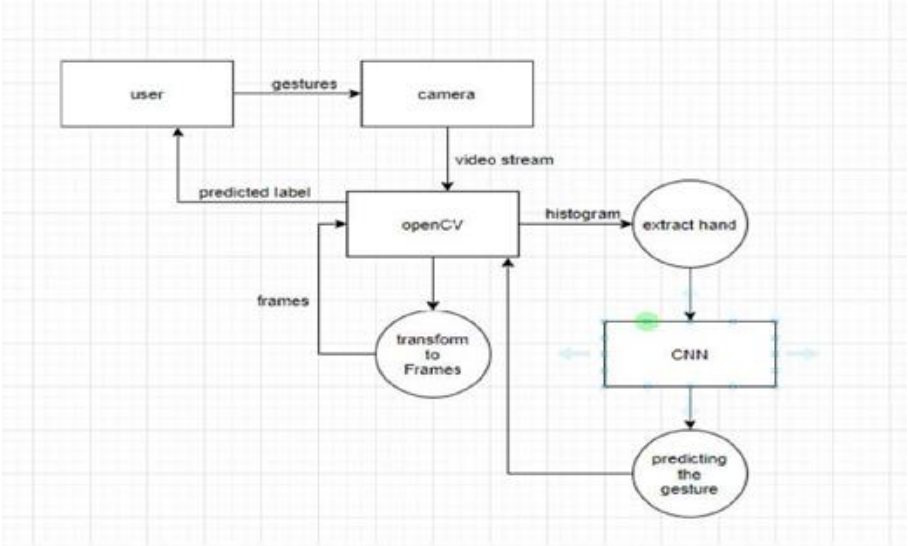

Fig. 7. Flow Diagram

Our approach uses two tiers of algorithms to predict the user's final symbol:

A. Level I

After point birth, frame obtained using OpenCV is subjected to a Gaussian Blur Slurry and threshold in order to obtain the reused picture. The CNN model is encouraged to make predictions using this repurposed image. However, a letter that is discovered for more than fifty frames is published and utilized to create the term. To show the space between words, use the blank symbol.

B. Level II

Vibrant collections of symbols displaying corresponding findings are connected. To distinguish between these sets, classifiers trained for these sets are used.

Layer 1

We utilize a two-layered algorithm approach to enhance symbol detection accuracy by verifying and predicting symbols that closely resemble each other.

CNN Model:

1. **First Convolution Level:** The first convolutional layer processes the input image, which has a resolution of 128x128 pixels, by 32 sorting weights (3 × 3 pixels each) are used. This results in a 126 out of by 126 pixel picture for each filter weight.
2. **1st Pooling Level**: Max pooling of 2x2 down samples the images, retaining the maximum value within each 2 by 2 square array. Consequently, the picture is downgraded to a resolution of 63 by 63 pixels.
3. **Second Convolution Level**: The second convolutional level receives its input as 63x63 pixels, which are the output of the first pooling layer. This produces an image of 60x60 pixels after 32 filter loads (3x3 pixels each) are applied.
4. **Second Pooling Level:** The processed images undergo further downsampling via max pooling of 2x2, reducing them to a pixel resolution of 30 by 30.
5. **First Densely Connected Level:** Next, a fully linked level with 128 neurons receives the pictures. This layer receives the second convolutional level's output, which is transformed into an array of (30 by 30 by 32) = 28,800 values. To avoid overfitting, a dropout layer with a dropout rate of 0.50 is used.

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

6. **2nd Densely Connected Level**: In the first highly connected level, the output is received as input by a totally linked level with 96 neurons.
7. **Last Level:** The output from the second densely linked level is sent to the final level, which contains as many neurons as the number of classes being detected (including letters and an empty sign).

Activation Function:

Every layer uses a Rectified Linear Unit (ReLU), which includes fully linked and convolutional neurons. ReLU adds nonlinearity to the algorithm by computing learning complicated characteristics by calculating max(x,0) for every data pixel. By reducing computation times, it expedites training and resolves the disappearing gradient problem.

Pooling Layer:

With a pool size of (2, 2), max pooling is applied to the input picture in conjunction with the ReLU activation function. As a result, overfitting is lessened, and computing costs are decreased by decreasing the number of parameters.

Dropout Layers:

Overfitting arises when the network's weights excessively adapt to training examples, leading to poor generalization on new data. Dropout layers mitigate this issue by randomly deactivating a subset of activations within the layer, ensuring accurate classifications or outputs for specific examples, even when some activations are dropped out.

Optimizer:

We've utilized the Adam optimizer to update the model according to the loss function output. Adam optimizer merges the advantages of two stochastic gradient descent algorithm extensions: adaptive gradient algorithm (ADA GRAD) and root mean square propagation (RMSProp).

Layer 2

We utilize a two-layered algorithm approach to enhance symbol detection accuracy by verifying and predicting symbols that closely resemble each other. Our testing revealed certain symbols that were not displaying accurately and were being confused with other symbols

To handle the aforementioned situations, we consequently developed three unique classifiers for these sets**:**

- {D, L, U}
- {T, A, D, I}
- {S, M, K}

3) Formation of Finger Spelling Sentence Formation

1. A detected letter is printed and added to the current string when it crosses a certain threshold and no further letters are found within a predetermined radius. The difference threshold is set to 20 and the threshold to 50 by the code.
2. If no suitable letter is found, the current dictionary, which holds the count of detections for the present symbol, is cleared to avoid inaccurate predictions.
3. No spaces are identified if the number of detected blanks above a predetermined threshold value and the buffer is currently empty.
4. If not, the current buffer is appended to the phrase below and a space is displayed to denote the end of a word.

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

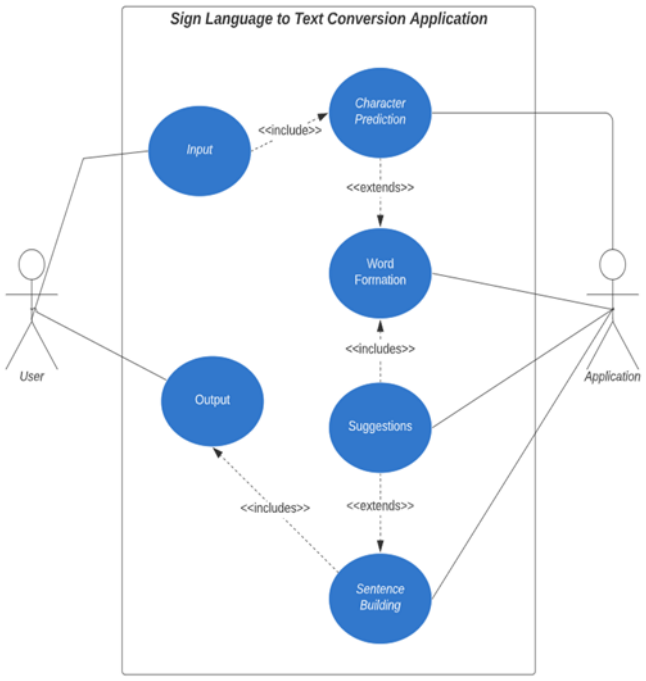

Fig. 8. Uses Case Diagram

4) The Autocorrect Feature

Hunspell, a Python package, is used to offer correction recommendations for each wrong word entered. The user can choose a word to add to the current phrase by selecting from a list of terms that fit the current word. This feature helps anticipate difficult words and lessen spelling mistakes. Additionally, Hunspell leverages extensive dictionaries along with custom word lists to recognize both common and domain-specific vocabulary. It ranks suggestions by frequency and context, ensuring the most relevant corrections are easily accessible. Furthermore, its real-time integration enables seamless text processing, enhancing user experience during input and reducing errors significantly.

5) Training and Testing

To get rid of extraneous noise, we use Gaussian blur after converting our RGB input photos to grayscale. The hand is then separated from the backdrop using an adaptive threshold, and the photos resolution is 128 × 128 for consistency.

The pre-processed images are fed into our model for training and testing. The model's prediction layer uses the SoftMax function to normalize output probabilities between 0 and 1, ensuring a sum of 1 across all classes for meaningful interpretation.

Initially, the output of the prediction layer may deviate from true values. To improve accuracy, we employ supervised learning with labeled data. Cross-entropy, a common performance metric for classification, is minimized through weight adjustments in our neural network. TensorFlow offers built-in functionality for calculating cross-entropy, simplifying its implementation within our network architecture.

RESULTS AND DISCUSSIONS

Our model's accuracy of **95.8%** is achieved with only layer 1 of our algorithm; when layers 1 and 2 are combined, it increases to **98.0%,** which is higher than the majority of research articles now accessible on American sign language.

Most research publications concentrate on hand detection with sensors similar to Kinect.

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Using convolutional neural networks and Kinect, they create a Flemish sign language recognition system in [7] with a 2.5% mistake rate. In [8], a thirty-word vocabulary and a hidden Markov model classifier are used to build a recognition model with a **10.90% e**rror rate. For 41 static Japanese sign language motions, they achieve an average accuracy of **86%** in [9]. Using the depth sensors map, an accuracy of **83.58%** and **85.49%** for new signers and **99.99%** for observed signers was obtained [10].

Their appreciation program included CNN as well. Notably, unlike most of the models shown above, our model does not employ a background subtraction method. Our work's accuracy may thus differ when we attempt to incorporate background subtraction. Our primary goal was to design a model that could be completed with easily obtained materials, even if the majority of the previously stated models make use of Kinect devices. One major benefit of our concept is that it makes use of the laptop's built-in webcam rather than an expensive sensor like Kinect, which is out of reach for most of the audience and also costlyll paragraphs must be invented. Below are the confusion matrices for our results

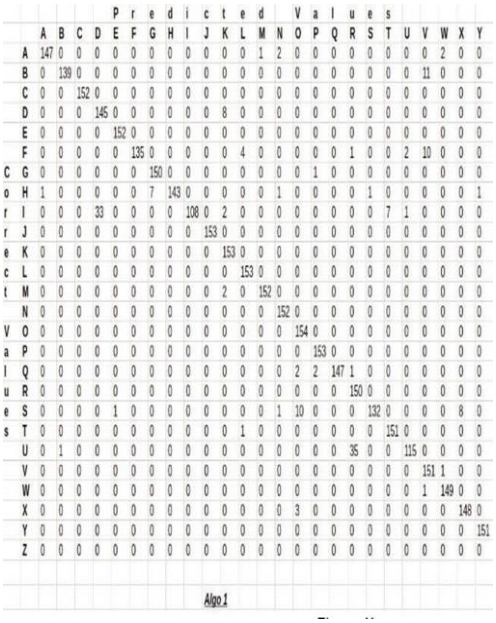

Fig 9. Result for Algorithm 1

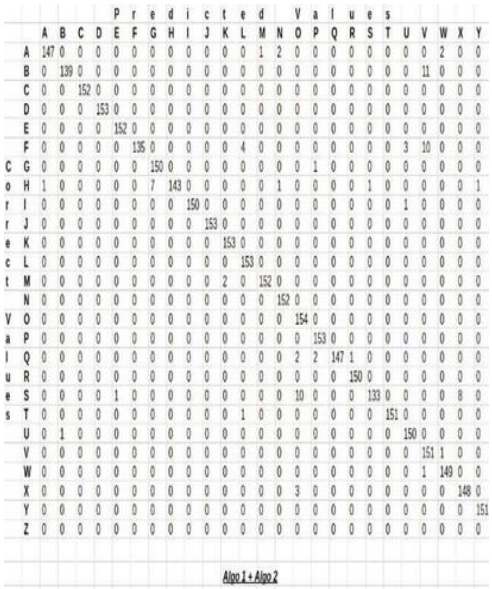

Fig 10. Result for Algorithm 1 + Algorithm2.

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

During the production, many challenges had to be solved. The data set presented the first challenge for us. As it is much more practical to work with only square images, we planned to use raw photos as CNN in Keras, but only in square format**.**

Since we couldn't find any pre-existing data sets that suited our demands, we generated our own. Selecting which filter to apply to our images to extract the necessary features so that we could utilize the picture as an input for the CNN model presented our second obstacle.

We finally chose to employ the Gaussian Blur Filter after experimenting with a few other filters, including binary threshold, canny edge detection, and Gaussian blur.

The precision of the model that we had trained earlier in the process presented us with more issues. Ultimately, this problem was fixed by increasing the input image's size and improving the data gathering.

CONCLUSIONS

The creation of a useful real-time vision-based system enabling D&M patients to recognize American Sign Language utilizing the ASL alphabets is described in this article. We were able to achieve 98.0% final accuracy with our dataset. By using two layers of algorithms to evaluate and forecast more comparable symbols, we were able to improve the accuracy of our prediction. Almost any sign may be recognized as long as it is displayed correctly, there is no background noise, and there is sufficient lighting.

This research demonstrates the potential of deep learning and computer vision in facilitating seamless communication for individuals with hearing and speech impairments. By leveraging Convolutional Neural Networks (CNN) along with adaptive thresholding and Gaussian blur, our model efficiently recognizes ASL signs with high precision. The robustness of this system showcases the advancements in artificial intelligence that can contribute to social welfare by reducing the communication gap between D&M individuals and the broader society.

One of the significant strengths of this work is its reliance on readily available resources, such as webcams, rather than expensive motion-sensing hardware like Kinect. This makes the system highly scalable and accessible for real-world applications. The ability to recognize hand gestures with a simple webcam ensures that the model can be deployed in educational institutions, workplaces, and public spaces, allowing individuals with hearing and speech impairments to communicate more effectively without the constant need for interpreters.

Additionally, the methodology adopted in this research can serve as a foundation for developing more advanced sign language recognition systems that incorporate real-time sentence formation and contextual understanding. The current model is limited to recognizing ASL alphabets through fingerspelling, but future enhancements can integrate full-word gestures and grammar structures, making the system more practical for everyday use. The application of Natural Language Processing (NLP) techniques could further enhance contextual accuracy, enabling a fluid translation of sign language into spoken or written text.

Moreover, as technology progresses, incorporating real-time background subtraction, low-light recognition capabilities, and multimodal inputs (such as voice commands or facial expressions) could significantly improve the model's adaptability and user experience. Expanding the dataset to include different hand orientations, varying lighting conditions, and diverse skin tones will further ensure fairness and inclusivity in recognition accuracy across all users.

Beyond academic and technological significance, this study aligns with the larger goal of inclusivity and accessibility in digital communication. It highlights how AI-driven solutions can bridge the communication gap for marginalized communities, contributing to a more inclusive society. Government and private sector initiatives can benefit from integrating such solutions into public infrastructure, healthcare, and educational services to empower individuals with disabilities.

In conclusion, this research marks a crucial step toward real-time, AI-powered sign language recognition. While our current model successfully demonstrates high-accuracy gesture recognition, its future iterations could revolutionize assistive communication, making digital interactions more inclusive, efficient, and universally accessible. With

continuous advancements in AI, computer vision, and NLP, the dream of a world where technology eradicates communication barriers for differently abled individuals is closer than ever.

FUTURE SCOPES

We plan to boost recognition accuracy by experimenting with different background-subtraction techniques that can cope with visually busy scenes. We are also refining the preprocessing pipeline so the system can identify gestures more reliably in dim lighting.

Deploying the solution as a web or mobile app would make it far more accessible. Although the current work focuses on ASL, the same framework could be extended to other signed languages with adequate data and training.

Because sign languages convey meaning through context—each sign can represent a verb, noun, or other concept—true conversational use will require advanced processing and Natural Language Processing (NLP). Future enhancements could therefore include contextual interpretation alongside improved low-light performance and robust background handling.

REFRENCES

[1] G. O. Young, "Synthetic structure of industrial plastics ," in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
[2] W.-K. Chen, Linear Networks and Systems (Book style), *Belmont, CA: Wadsworth*, 1993, pp. 123–135.
[3] H. Poor, An Introduction to Signal Detection and Estimation, *New York: Springer-Verlag*, 1985, ch. 4.
[4] Au C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
[5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style*)," IEEE Transaction J. Magn. Jp*n., vol. 2, Aug. 1987, pp. 740–741 [Dig. 9th Annu. Conf. Magnetics Japan, 1982, p. 301].
[6] Young, The Techincal Writers Handbook. Mill Valley, *CA: University of Science*, 1989.
[7] Journal Online A. Bhat, V. Yadav, V. Dargan and Yash, "Sign Language to Text Conversion using Deep Learning*," 3rd International Conference for Emerging Technology (INCET),* Belgaum, India,2022, pp.1-7.
[8] Peguda, J., Santosh, V. S., Vijayalata, Y., N, A. D.; Mounish, V. Speech to sign language translation for Indian languages. *8th International Conference on Advanced Computing and Communication Systems (ICACCS),* 2022.
[9] E. P, "Speech To Sign Language Translator For Hearing Impaired," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 10, p. 7, 2021.
[10] O.H. Khalid El-Darymli, "Speech to Sign Language Interpreter System (SSLIS)," *the IEEE International Conference of Computer and Communication Engineering*, 2006.
[11] Lalit Goyal, "Automatic Translation of English Text to Indian Sign Language Synthetic Animations," *13th Intl. Conference on Natural Language Processing*, p. 10, 2016.
[12] Aditya, "ENGLISH TEXT TO INDIAN SIGN LANGUAGE TRANSLATION SYSTEM," *Turkish Journal of Computer and Mathematics Education*, vol. 11, no. 3, 2020
[13] Pankaj Sonawane, "Speech to Indian Sign Language (ISL) Translation System," *International Conference on Computing, Communication, and Intelligent System s(ICCCIS)*, 2021
[14] F. Ali, "Domain Bounded English to Indian Sign Language Translation Model," *International Journal of Computer Science and Informatics*, vol. 1, no. 4
[15] K.M. Suresh Anand, "An Integrated Two Way ISL (Indian Sign Language) Translation System A New Approach," *International Journal of Advance Research in Computer Science*, vol. 4, no. 2, 2013.
[16] V Y. &. J. P. Rokade, "Indian Sign Language Recognition System," International Journal of Engineering and Technology, vol. 9, pp. 189-196, 2017.
[17] D. V. R. Sadhana Bhimrao Bhagat, "Vision based sign language recognition: a survey," *JETIR (ISSN-23495162*), vol. 4, 2017.
[18] C. C. E. Sunitha, "Speaker Recognition using MFCC and Improved Weighted Vector Quantization Algorithm," *International Journal of Engineering and Technology*, vol. 7, 2015.

Journal of Information Systems Engineering and Management
2025, 10(48s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[19] N. S. Ramya, "Implementation of telugu speech synthesis system," *International Conference on Advances in Computing Communications and Informatics (ICACCI),* 2017.
[20] M. R. Reddy, "Transcription of Telugu TV news using ASL," *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2015.
[21] N. K. Sunitha, "A Novel approach to improve rule-based Telugu morphological analyzer", *World Congress on Nature & Biologically Inspired Computing (NaBIC),* 2009.
[22] S. T., "Voice and speech recognition in Tamil language," *2nd International Conference on Computing and Communications Technologies (ICCCT),* 2017, pp. 288-292,
[23] M. S. Nair, A. P. Nimitha and S. M. Idicula, "Conversion of Malayalam text to Indian sign language using synthetic animation," *International Conference on Next Generation Intelligent Systems (ICNGIS*), 2016, pp.1-4.
[24] S. Bansal and S. S. Agrawal, "Development of Text and Speech Corpus for Designing the Multilingual Recognition System," *Oriental COCOSDA - International Conference on Speech Database and Assessments*, 2018, pp. 1-8.
[25] S.C. Sajjan and Vijaya C, "Continuous Speech Recognition of Kannada language using triphone modeling," *International Conference on Communications, Signal Processing and Networking (WiSPNET*), 2016, pp.451-455.