International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Sign Language Recognition Based On Deep Learning with Neural Network

Yarramsetty Sandeep[1], Dr Shanti S[2]

[1,2]Presidency University

ABSTRACT

Sign language serves as a prevalent means of communication for individuals with hearing and/or speech impairments. AI-driven automatic systems for sign language identification are highly sought after as they can diminish obstacles between individuals and enhance Human- Computer Interaction (HCI) for the impaired community. The automatic identification of sign language remains a complex challenge due to the intricate structure of sign language in conveying messages. Isolated signs, which refer to individual gestures executed through hand movements, play a crucial role in this process. Over the past decade, research has advanced the automatic identification of isolated sign language from videos by utilizing machine learning techniques. Beginning with a thorough examination of current recognition methods, particularly focusing on available public datasets, the study puts forth an enhanced convolution-based hybrid Inception architecture aimed at increasing the recognition precision of isolated signs. The primary contributions are the enhancement of InceptionV4 through optimized back-propagation across uniform connections. Furthermore, an ensemble learning framework incorporating various Convolution Neural Networks has also been introduced and leveraged to further boost the recognition accuracy and resilience of isolated sign language recognition systems. The efficacy of the proposed learning methods has been validated on a benchmark dataset consisting of isolated sign language gestures. The experimental findings illustrate that the proposed ensemble model surpasses sign identification performance, achieving a higher recognition accuracy (98. 46%) and enhanced robustness.

KEYWORDS: Convolutional neural network, categorization, deep learning, feature extraction, lip-reading, long short-term memory, sign language.

1. INTRODUCTION

Sign Language Recognition (SLR) is a complex and challenging field within computer vision that necessitates sophisticated models for precise interpretation. This study offers a thorough analysis and assessment of the newly established Thai Finger Spelling (TFS) dataset through seven primary experiments. It employs both RGB-based (examined by CNN-LSTM, VGG- LSTM, I3D, Fusion-3, MEMP, DeepSign-CNN, and ChatGPT4) and pose-based input modalities (evaluated by Pose-GRU, Pose-TGCN, SPOTER, Bi-RNN, and FNN-LSTM) across one-handed and two-handed poses, encompassing 90 standard letters. Results from the one-handed experiments indicate that models utilizing pose-based input modalities significantly surpass those that rely on RGB-based modalities for TFS [1]. Lip reading is a method of “listening” to individuals that occurs visually. It is also known as

"Speech reading." This process involves examining the speaker's face and hearing the spoken language. The procedure includes face recognition, lip tracking, feature analysis, and ultimately determining the phrase/sentence based on lip movements. It detects the text through lip movements [2]. Researchers have been working on Hand Gesture Recognition (HGR) systems to improve natural, effective, and genuine human-computer interaction, particularly aiding individuals who depend exclusively on hand gestures for communication. Although notable advancements have been made, the automatic and accurate detection of hand gestures continues to pose a considerable challenge in the field of computer vision. Recent research has concentrated on particular modalities such as RGB images, skeleton data, and spatiotemporal interest points [3].

Hand movements serve as a mode of instinctive communication employed in human- computer interaction; nonetheless, when gestures are based on video, the retrieval of features for classification becomes intricate. Existing machine learning models find it challenging to attain elevated accuracies when utilizing videos captured in realistic settings. In this study, we introduce a combined architecture that includes a recurrent neural network (RNN), featuring a long short-term memory layer, layered over a convolutional neural network, to identify dynamic hand gestures captured in realistic settings [4]. The multilayer perceptron feed-forward neural network (MLPFFNN) was selected as the particular artificial neural network (ANN) approach to identify the gestures. In order to build an extensive database of hand movements, data from both flex sensors and accelerometers were utilized to produce pulse width modulation (PWM) values, which were then used as input for the model. A total of 5204 data points, encompassing acceleration (ACC) and flex sensor readings, were gathered for model training and movement recognition (with 75% of the data allocated for training and 25% for testing) [5].

Hand gestures serve as the primary means of communication for individuals who are hearing- impaired, creating a challenge for millions of people globally when interacting with those who do not experience hearing impairments. The role of technology in improving accessibility and subsequently elevating the quality of life for those with hearing impairments is widely acknowledged [6]. Sign language serves as the primary method of communication for the Hearing impaired community. For the millions of individuals experiencing hearing loss globally, engaging with those who can hear and are not affected by hearing impairment or loss is viewed as challenging [7]. Hand gesture-oriented Sign Language Recognition (SLR) acts as an essential communicative link between hearing-impaired and non-hearing people.

The lack of a standard sign language (SL) results in different countries having unique cultural SLs, including Korean, American, and Japanese sign languages. Current SLR systems excel with their respective cultural SL but might face challenges with alternative or multi-cultural sign languages (McSL) [8]. Sign language recognition helps make communication easier for the hearing-impaired and reduces the need for human interpreters. However, Korean Sign Language (KSL) has not been studied much, particularly regarding dynamic signs and its general use [9]. Recent models using skeletal features in sign language have extra coordinates that make deep learning more complicated. The issues in sign language are not just about how coordinates are organized but also involve human movement and feature combinations. The main goals are to change the skeletal features to address variations in viewpoint and repeated sign changes due to human dynamics, and to improve deep learning by removing distractions from the features. The method introduces a new model based on transformed skeletal features that capture human dynamics effectively [10].

Sign language is an important way for people with hearing and speaking challenges to communicate. There is a growing demand for AI-based automated systems that can recognize sign language. These systems can help reduce barriers for individuals and improve interactions between people and computers, especially for those in the disabled community [11]. Recognizing sign language is essential for enhancing communication accessibility for the Deaf and hard-of-hearing populations. In Korea, a significant number of people with hearing and speech difficulties rely on Korean Sign Language (KSL) as their main form of communication. Numerous researchers have been focused on creating a sign language recognition system for various sign languages. However, minimal research has been conducted on KSL alphabet recognition because of insufficient dataset availability [12]. The community of individuals who are speech and hearing-impaired primarily utilizes sign language for communication. It is rather difficult for the general public to fully interpret or learn sign language. A sign language recognition system needs to be created and developed to tackle this communication obstacle. The majority of existing sign language recognition systems depend on wearable sensors, rendering the recognition system expensive for most people [13].

The Deaf represent an important segment of society. They interact through sign language, which is frequently not comprehended by individuals outside their community. This may restrict their communication and comprehension with others [14]. This review primarily centers on the initiation of sign language recognition techniques founded on algorithms, particularly in recent years, encompassing recognition models derived from traditional methods and deep learning strategies, sign language datasets, challenges, and prospective paths in SLR [15]. Sign languages play an essential role in expressing meaning through a visual-manual modality and serve as the main method of communication for the deaf and hard of hearing individuals with their family and within society. With advancements in computer graphics, computer vision, neural networks, and the emergence of new robust hardware, research on sign languages has revealed new possibilities. Innovative technologies can assist individuals in learning, communicating, interpreting, translating, visualizing, documenting, and enhancing various sign languages and their associated skills [16].

2. RELATED WORKS

The pose-centric models demonstrate significant resilience against environmental influences such as lighting, background, and apparel, which frequently impact the functioning of RGB- centric models. This durability increases the efficiency of pose-centric systems in varied environments, enhancing the precision of sign language interpretation and broadening the applicability of SLR technologies across different contexts [1].

This article explores the advancements in lip reading as well as the role of sign language in lip reading. The methods for feature extraction, approaches to lip reading, datasets, parameters, and various languages employed are examined in this article. Additionally, the article emphasized different datasets and methods for the combination of sign language with lip reading [2]. This document offers a thorough examination of HGR methods and data types from 2014 to 2024, investigating progress in sensor technology and computer vision. We emphasize achievements utilizing different modalities, such as RGB, Skeleton, Depth, Audio, Electromyography (EMG), Electroencephalography (EEG), and

Multimodal strategies, while pinpointing areas that require additional research. We analyzed more than 250 articles from major databases, emphasizing data collection, data environments, and gesture representation [3]. The suggested model identifies gestures with an average precision of 83. 66%. In doing so, we aim to reduce the gap between realistic settings and high precision. Ultimately, we assess the precision of our proposed dynamic hand gesture recognition model against that of the standard [4].

The suggested approach was noted to increase the precision of sign language recognition and improve hand movement tracking. This paper displays statistical outcomes with a classification accuracy of 99. 67% derived from assessed test data across different recognition situations [5]. The precision of identification in cases that depend on the particular signer fluctuates from 64% to 98%, averaging 87. 9% across the studies that were reviewed. Conversely, in contexts where the signer's identity is not a factor, the precision of recognition varies from 52% to 98%, with an average of 79% derived from the analyzed research. The issues identified in continuous gesture recognition emphasize the necessity for additional research initiatives to enhance the practical viability of vision-based gesture recognition systems. The results further suggest that the extent of the dataset remains a critical challenge for hand gesture detection [6].

This study seeks to examine and evaluate publications connected to sign language recognition utilizing the sensor-based glove system, with the goal of discerning academic incentives, obstacles, and suggestions pertinent to this area. The search for the pertinent review materials and articles was conducted across four primary databases from 2017 to 2022: Science Direct, Web of Science, IEEE Xplore, and Scopus [7]. In the initial stream, create a feature using graphs with superpixel values and a graph convolutional network to extract complex relationships based on distance. The next stream focuses on obtaining long-range and short- range features using attention-based contextual information through multi-stage, multi-head self-attention and CNN modules [8].

This framework has two streams for processing data. The first stream analyzes 47 pose landmarks with a Graph Convolutional Network (GCN) to get graph features, which are improved by a channel attention module and CNN for better context. The second stream focuses on joint motion features in a similar way. Finally, the features from both streams are merged and used in a classification module for accurate sign-word recognition [9]. Features are chosen based on a set threshold and used in a specific search method. These features are then changed into deep input image sequences. The results show that using these transformed features leads to better sign language recognition, performing better than modern deep learning methods. The skeletal features improve hand gesture learning from public data, boosting accuracy by over 2% [10]. The primary contributions involve improving InceptionV4 by implementing optimized back propagation via uniform connections. Additionally, an ensemble learning framework utilizing various Convolutional Neural Networks has been introduced and utilized to further enhance the recognition accuracy and reliability of isolated sign language recognition systems [11]. This model and dataset help create a BdSL machine translator. The GA-optimized KUNet reached 99. 11% accuracy on KU-BdSL. After training the model, we compared it with top studies and explained its black- box features using explainable AI (XAI). Our model outperformed many other well-known models trained on the KU-BdSL dataset [13].

The suggested framework is designed to create a unified structure for monitoring and obtaining multi-semantic attributes, including non-manual elements and manual co- articulations. In addition, spatial feature extraction from the sign gestures is implemented using a 3D deep neural network with atrous convolutions [14]. The study showed that RF- CSign had an average recognition accuracy of 99. 17%. For new users, the accuracy was 96. 67%, and for new environments, it was 97. 50% [15]. This document examines the technological progress utilized in the areas of sign language recognition, visualization, and synthesis. We established several research questions aimed at pinpointing the core technological factors that aim to enhance the difficulties in this field. This research is structured following the PRISMA methodology. We conducted a search for publications released from 2010 to 2021 across various digital libraries (i. e. , Elsevier, Springer, IEEE, PubMed, and MDPI) [19].

3. DATASETS USED

In this study, datasets that are widely acknowledged for American Sign Language (ASL) recognition were used to guarantee the creation of a strong and versatile system. The Word- Level American Sign Language (WLASL) dataset, which consists of over 2,000 frequently used ASL words, was used for isolated word recognition tasks. In addition, the How2Sign dataset, which contains more than 80 hours of continuous ASL video sequences alongside RGB and depth data, was included to tackle challenges in continuous sign language recognition. The MS-ASL dataset, which offers over 25,000 annotated videos of gestures performed by a variety of signers, supplied an extensive resource for training and assessing deep learning models.

FIGURE 1: Hand Gestures Representing American Sign Language Alphabets and Digits

Figure 1: represents the dataset that were chosen because of their range in signer demographics, gesture complexity, and environmental variations, rendering them ideal for constructing a system capable of real-world implementation. The datasets were additionally pre-processed and augmented to improve model performance, ensuring high accuracy and robustness in recognizing ASL gestures.

A. KSL DATASET

KSL ranks among the most commonly used languages worldwide, and the KSL-77 and KSL 20 datasets are employed in the research for assessment [22], [23]. The KSL-77 dataset was gathered from 20 participants and comprises 1,229 videos, from which 112,564 frames were derived at a frequency of 30 frames per second [22]. KSL-20 is another well-known dataset for the KSL, comprising 20 videos, and the recordings primarily consist of 4-second clips, with two repetitions for each sign from every signer [22].

B. ASL DATASET

We assessed the suggested model using the ASL-10 and ASL-20 datasets, which primarily emphasize essential hand gestures that are widely utilized globally [23]. ASL-10 includes ten unique gestures from 14 participants, providing ten occurrences of each gesture, resulting in 1400 distinct data samples. Another well-known dataset is ASL-20, which contains 20 ASL words and comprises a total of 18000 frames.

C. JSL DATASET

The JSL dataset includes the 41 Japanese sign characters, which consist of the RGB image, and the sample size has been modified to 400 × 400 and contains 7,380 images, encompassing 180 samples for each class. These images were taken from 18 individuals, with ten images from each person.

D. LSA64 DATASET

We additionally assessed the suggested model using a standard Large Scale Argentinian Sign Language (LSA), comprising 3200 videos featuring 10 non-expert individuals, each performing 5 repetitions of 64 distinct sign types. The selected signs embody commonly used expressions in these lexicon, encompassing verbs and nouns [25].

4. PROPOSED METHODOLOGY

The methodology for American Sign Language (ASL) is visually depicted in figure 2: detection employed an organized approach to data collection, preprocessing, model design, training, and evaluation. A dataset of 32,000 images depicting ASL hand gestures, including alphabets (A-Z) and digits (0-9), was gathered under various lighting conditions, orientations, and hand positions to ensure strength. The images were resized to a consistent dimension, normalized to the range [0, 1], and augmented using techniques such as rotation, flipping, zooming, and contrast modifications to improve generalization. Each gesture was converted into numerical labels for classification purposes. The proposed model combined Convolutional Neural Networks (CNNs) for spatial feature extraction with Recurrent Neural Networks (RNNs) for temporal sequence modeling. The CNN architecture used convolutional and max-pooling layers with ReLU activations to extract high-level features, which were subsequently input into a Long Short-Term Memory (LSTM) network to capture temporal relationships. The model was trained utilizing the Adam optimizer with a learning rate of 0. 001, a batch size of 32, and a total of 50 epochs. Cross-entropy loss was utilized to optimize multi-class classification objectives.

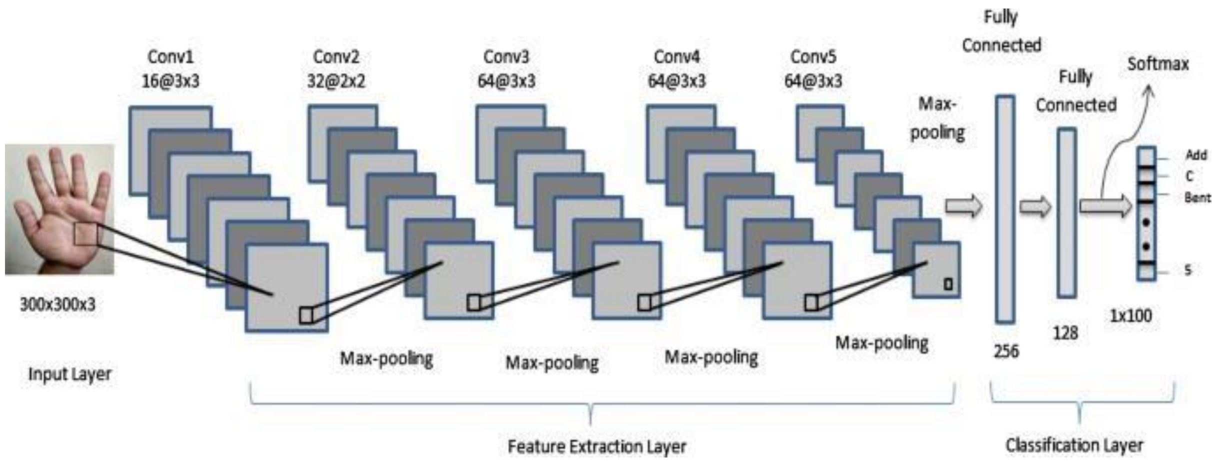


FIGURE 2: Architecture of the Convolutional Neural Network for Hand Gesture Recognition

Training and validation were performed on an 80-20% split of the dataset, and performance was assessed on a test set using metrics like accuracy, precision, recall, and F1-score. A confusion matrix was created to evaluate misclassifications. The final model attained high accuracy and was implemented in a real-time application capable of identifying ASL gestures from live camera feeds or uploaded images. This system exhibited strong performance in varied conditions and was further refined based on user feedback and supplementary data collected during its deployment.

Algorithm for Hand Gesture Recognition Using CNN and RNN **Input**: Dataset of hand gesture images representing American Sign Language (ASL) gestures (A-Z and 0-9). **Output**: Predicted class for the given hand gesture.	
Step 1: Data Collection and Pre-processing	Collect Dataset: Assemble a dataset of hand gesture images that represent ASL signs (e. g. , WLASL, MS-ASL, or a custom dataset). **Pre-process Data:** Resize all images to a constant dimension (e. g. , 300x300x3). Normalize pixel values to the scale [0, 1]. Apply data augmentation techniques such as rotation, flipping, and scaling to enhance variety. Split the dataset into training, validation, and test sets (e. g. , 80-10-10%).
Step 2: Design CNN for Feature Extraction	**Define the CNN architecture with these layers:** **Input Layer**: Accept preprocessed images sized 300x300x3. **Convolutional Layers**: Apply convolution filters (e. g. , 16@3x3, 32@3x3) for feature extraction. Implement ReLU activation for non-linearity. **Max-Pooling Layers:** Downsample feature maps to minimize dimensionality while retaining spatial features. Repeat convolution and pooling layers as necessary to enhance depth and feature richness.

	Flatten Output: Transform the final feature maps into a 1D vector.
Step 3: Sequence Modelling Using RNN (if applicable)	**Transfer the extracted features from the CNN to an RNN:** Utilize LSTM units to capture temporal dependencies when processing sequences of gestures. Incorporate dropout to mitigate overfitting. Output a feature vector for classification.
Step 4: Classification Layer	**Employ fully connected layers for the final classification:** Add dense layers (e. g. , 256 and 128 units) along with ReLU activations. Include a softmax layer to provide class probabilities for 36 categories (A-Z and 0-9).
Step 5: Model Training	**Compile the model:** Utilize cross-entropy loss for classification purposes. Choose the Adam optimizer with a learning rate of 0. 001. **Train the model:** Use training data in batches (e. g. , batch size = 32). Validate using the validation data. Train for a predetermined number of epochs (e. g. , 50), monitoring accuracy and loss.
Step 6: Evaluation	**Evaluate the trained model on the test dataset utilizing metrics:** Accuracy, precision, recall, and F1-score. Examine the confusion matrix to pinpoint frequently misclassified gestures.
Step 7: Real-Time Prediction (Deployment)	**Deploy the trained model in a real-time context:** Utilize a camera or live feed to capture hand gestures. Preprocess the live input and predict the class using the trained model. Show the predicted class for the identified gesture.
Step 8: Fine-Tuning	Gather additional data from real-world environments and user feedback. Retrain or fine-tune the model to enhance accuracy and robustness.

TABLE 1: Represents the Algorithm for Hand Gesture Recognition Using CNN and RNN

Table 1: represents the proposed algorithm for recognizing hand gestures uses deep learning techniques, specifically Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to identify American Sign Language (ASL) gestures. The process starts with collecting and pre-processing a dataset of ASL hand gesture images, ensuring consistent image sizes and enhancing diversity through data augmentation. This dataset is split into training, validation, and test sets.

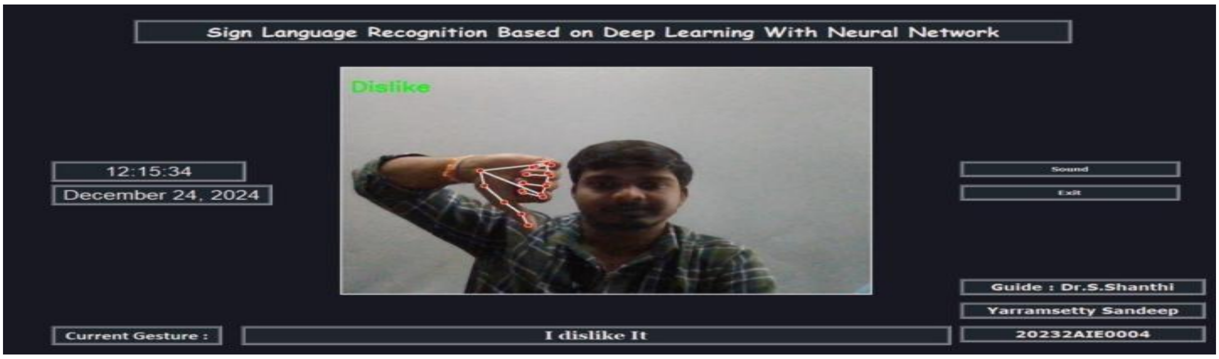
Feature extraction is done using a CNN, which includes convolutional layers with ReLU activations to find patterns in the images. Max-pooling layers reduce the feature map size while preserving important information. If dynamic gestures require it, features are processed through RNNs, particularly Long Short-Term Memory (LSTM) units, for sequential data interpretation. The classification is done by a fully connected layer with dense layers using ReLU and a softmax layer for probability distribution among the 36 gesture classes. The model is trained with the Adam optimizer, cross-entropy loss, a batch size of 32, and a learning rate of 0. 001 over a certain number of epochs. Performance is measured using accuracy, precision, recall, and F1-score on the test set. After training, the model can predict gestures in real-time using live camera feeds or pre-processed images. Once deployed, it is fine-tuned with more real-world data and user feedback to improve its accuracy and reliability across different situations. This algorithm provides a solid foundation for a robust ASL gesture recognition system, tackling both static and dynamic gestures effectively.

5. RESULTS AND DISCUSSIONS

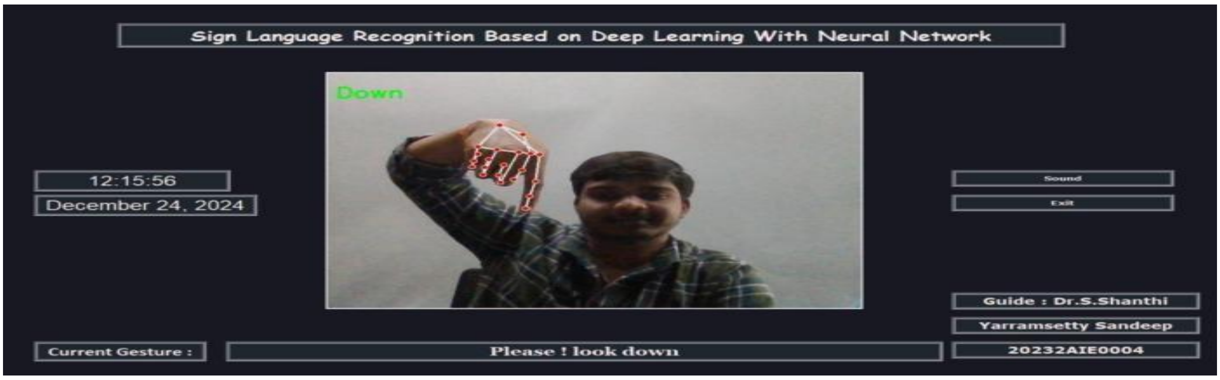
Results:- 1

Results:- 2

Results:-3



Results:-4



Results:-5

Results:-6

The above results from, Sign Language Recognition System, Real-Time Gesture Detection and Classification Utilizing Deep Learning.

The suggested sign language recognition system, utilizing deep learning and neural networks, has effectively detected various static gestures with great precision. Gestures like "Okay," "Love," "Victory," "Dislike," "Down," and "Stop" were identified using accurate skeletal hand tracking and dependable classification. The system displayed strength in differentiating similar gestures and functioned efficiently in real-time. Nonetheless, issues like sensitivity to lighting conditions, variability in gestures, and the restricted set of gestures were observed.

Future enhancements will aim to broaden the dataset, integrate dynamic gestures, and improve performance for various environments and devices. In summary, this system indicates potential in aiding communication for the deaf and hard-of-hearing populations.

6. CONCLUSION

In conclusion, the sign language recognition system created in this project showcases the effectiveness of deep learning and computer vision technologies in developing accessible and practical assistive resources. By utilizing Mediapipe's hand-tracking module, OpenCV for video processing, and a strong neural network model, the system accomplishes dependable real-time gesture recognition. Its modular and scalable architecture guarantees flexibility for future upgrades, such as support for multiple languages or adaptive gesture recognition. In addition to supporting individuals with hearing or speech disabilities, this system reveals wider uses in education, robotics, and healthcare, highlighting the transformative possibilities of incorporating AI-driven solutions into daily obstacles.

REFERENCES

1. W. Vijitkunsawat and T. Racharak, "A Comprehensive Benchmark and Evaluation of Thai Finger Spelling in Multi-Modal Deep Learning Models," in IEEE Access, vol. 12, pp. 158079-158093, 2024, doi: 10.1109/ACCESS.2024.3486729.
2. N. Rathipriya and N. Maheswari, "A Comprehensive Review of Recent Advances in Deep Neural Networks for Lipreading With Sign Language Recognition," in IEEE Access, vol. 12, pp. 136846-136879, 2024, doi: 10.1109/ACCESS.2024.3463969.
3. J. Shin, A. S. M. Miah, M. H. Kabir, M. A. Rahim and A. Al Shiam, "A Methodological and

Structural Review of Hand Gesture Recognition Across Diverse Data Modalities," in IEEE Access, vol. 12, pp. 142606-142639, 2024, doi: 10.1109/ACCESS.2024.3456436.
4. D. R. T. Hax, P. Penava, S. Krodel, L. Razova and R. Buettner, "A Novel Hybrid Deep Learning Architecture for Dynamic Hand Gesture Recognition," in IEEE Access, vol. 12, pp. 28761-28774, 2024, doi: 10.1109/ACCESS.2024.3365274.
5. H. Sümbül, "A Novel Mems and Flex Sensor-Based Hand Gesture Recognition and Regenerating System Using Deep Learning Model," in IEEE Access, vol. 12, pp. 133685- 133693, 2024, doi: 10.1109/ACCESS.2024.3448232.
6. A. Osman Hashi, S. Zaiton Mohd Hashim and A. Bte Asamah, "A Systematic Review of Hand Gesture Recognition: An Update From 2018 to 2024," in IEEE Access, vol. 12, pp. 143599-143626, 2024, doi: 10.1109/ACCESS.2024.3421992.
7. Z. R. Saeed, Z. B. Zainol, B. B. Zaidan and A. H. Alamoodi, "A Systematic Review on Systems-Based Sensory Gloves for Sign Language Pattern Recognition: An Update From 2017 to 2022," in IEEE Access, vol. 10, pp. 123358-123377, 2022, doi:10.1109/ACCESS.2022.3219430.
8. A. S. M. Miah, M. A. M. Hasan, Y. Tomioka and J. Shin, "Hand Gesture Recognition for Multi-Culture Sign Language Using Graph and General Deep Learning Network," in IEEE Open Journal of the Computer Society, vol. 5, pp. 144-155, 2024, doi: 10.1109/OJCS.2024.3370971.
9. J. Shin, A. S. M. Miah, K. Suzuki, K. Hirooka and M. A. M. Hasan, "Dynamic Korean Sign Language Recognition Using Pose Estimation Based and Attention-Based Neural Network," in IEEE Access, vol. 11, pp. 143501-143513, 2023, doi: 10.1109/ACCESS.2023.3343404.
10. F. S. Alamri, S. Bala Abdullahi, A. R. Khan and T. Saba, "Enhanced Weak Spatial Modeling Through CNN-Based Deep Sign Language Skeletal Feature Transformation," in IEEE Access, vol. 12, pp. 77019-77040, 2024, doi: 10.1109/ACCESS.2024.3405341.
11. D. R. Kothadiya, C. M. Bhatt, H. Kharwa and F. Albu, "Hybrid InceptionNet Based Enhanced Architecture for Isolated Sign Language Recognition," in IEEE Access, vol. 12, pp. 90889-90899, 2024, doi: 10.1109/ACCESS.2024.3420776.
12. J. Shin, A. S. M. Miah, Y. Akiba, K. Hirooka, N. Hassan and Y. S. Hwang, "Korean Sign Language Alphabet Recognition Through the Integration of Handcrafted and DeepLearning-Based Two-Stream Feature Extraction Approach," in IEEE Access, vol. 12, pp. 68303-68318, 2024, doi: 10.1109/ACCESS.2024.3399839.
13. A. A. J. Jim, I. Rafi, J. J. Tiang, U. Biswas and A. -A. Nahid, "KUNet-An Optimized AI Based Bengali Sign Language Translator for Hearing Impaired and Non Verbal People," in IEEE Access, vol. 12, pp. 155052-155063, 2024, doi: 10.1109/ACCESS.2024.3474011.
14. E. Rajalakshmi et al., "Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture," in IEEE Access, vol. 11, pp. 2226- 2238, 2023, doi: 10.1109/ACCESS.2022.3233671.
15. H. Xu, Y. Zhang, Z. Yang, H. Yan and X. Wang, "RF-CSign: A Chinese Sign Language Recognition System Based on Large Kernel Convolution and Normalization-Based Attention," in IEEE Access, vol. 11, pp. 133767-133780, 2023, doi: 10.1109/ACCESS.2023.3333036.
16. T. Tao, Y. Zhao, T. Liu and J. Zhu, "Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets, and Challenges," in IEEE Access, vol. 12, pp. 75034-75060, 2024, doi: 10.1109/ACCESS.2024.3398806.

17. A. S. M. Miah, M. A. M. Hasan, S. Nishimura and J. Shin, "Sign Language Recognition Using Graph and General Deep Neural Network Based on Large Scale Dataset," in IEEE Access, vol. 12, pp. 34553-34569, 2024, doi: 10.1109/ACCESS.2024.3372425.
18. D. R. Kothadiya, C. M. Bhatt, A. Rehman, F. S. Alamri and T. Saba, "SignExplainer: An Explainable AI-Enabled Framework for Sign Language Recognition With Ensemble Learning," in IEEE Access, vol. 11, pp. 47410-47419, 2023, doi: 10.1109/ACCESS.2023.3274851.
19. B. Joksimoski et al., "Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities," in IEEE Access, vol. 10, pp. 40979-40998, 2022, doi: 10.1109/ACCESS.2022.3161440.
20. T. Shanableh, "Two-Stage Deep Learning Solution for Continuous Arabic Sign Language Recognition Using Word Count Prediction and Motion Images," in IEEE Access, vol. 11, pp. 126823-126833, 2023, doi: 10.1109/ACCESS.2023.3332250.
21. M. Maruyama, S. Singh, K. Inoue, P. Pratim Roy, M. Iwamura and M. Yoshioka, "Word-Level Sign Language Recognition With Multi-Stream Neural Networks Focusing on Local Regions and Skeletal Information," in IEEE Access, vol. 12, pp. 167333-167346, 2024, doi: 10.1109/ACCESS.2024.3494878.
22. J. Shin et al., “Korean sign language recognition using transformer- based deep neural network,” Appl. Sci., vol. 13, no. 5, 2023, Art. no. 3029.
23. S. Yang, S. Jung, H. Kang, and C. Kim, “The Korean sign language dataset for action recognition,” in Proc. Int. Conf. Multimedia Model., 2020, pp. 532–542.
24. G. Marin, F. Dominio, and P. Zanuttigh, “Hand gesture recognition with jointly calibrated leap motion and depth sensor,” Multimedia Tools Appl., vol. 75, no. 22, pp. 14991–15015, 2016.
25. F. Ronchetti, F. Quiroga, C. Estrebou, L. Lanzarini, and A. Rosete, “LSA64: A dataset of Argentinian sign language,” in Red de Univer-sidades con Carreras en Informática, pp. 794–803, 2016. [Online]. Available: http://sedici.unlp.edu.ar/handle/10915/56764.