An Efficient Model for American Sign Language Recognition using Deep-neural networks

Avtar Chandra
Department of Electronics and Communication
University of Allahabad, Prayagraj,
Uttar Pradesh, India
avi4each@gmail.com

Aditya Ranjan
Department of Electronics and Communication
University of Allahabad , Prayagraj,
Uttar Pradesh, India
adityaranjan953@gmail.com

Dinesh Prasad Sahu
School of Computer Science Engineering &
Technology, Bennett University,Noida, India
dinesh.sahu1230@gmail.com

Shiv Prakash
Department of Electronics and Communication
University of Allahabad, Prayagraj,
Uttar Pradesh, India
shivprakash@allduniv.ac.in

Tiansheng Yang
University of South Wales pontypridd,
United Kingdom
tiansheng.yang1@southwales.ac.uk

Rajkumar Singh Rathore
Cardiff School of Technologies Cardiff Metropolitan
University Cardiff, United Kingdom
rsrathore@cardiffmet.ac.uk

Abhay Vajpayee
Institute of Engineering and Technology, Dr. A.P.J.
Abdul Kalam Technical University, Lucknow, India
2407@ietlucknow.ac.in

Abstract: The Sign Language is a way that is used for communication by people with inability to speak and hear. Thus, improving these languages is recognized as being widely influential across society. Worldwide, about 7,000 sign languages are used for communication, and many studies have been performed using different sign languages. This study considers American Sign Language (ASL) due to its popularity. We have proposed an efficient model using deep learning for 26 alphabets hand gestures in ASL to communicate with people. The proposed model has been assessed with the benchmark dataset and compared with many studies using the same datasets. It has achieved the highest accuracy as compared to the contemporary model. It has been observed that working with complex numbers positively impacted performance by approximately 20% compared to configuring our model to work with real numbers while keeping its structure intact.

Key Words **– American Sign Language (ASL), Feature Extraction, Artificial Intelligence (AI), Deep Learning.**

I. INTRODUCTION

Individual who are not able to speak and hear constitute a significant part of society. As per the World Health Organization (WHO), 5% of the global population which is about 0.5 billion people are experiencing hearing loss, which is expected to double by 2050 [1]. These figures also include people having severe hearing loss.

Sign language is a method used for communication between the general public and people with hearing and speaking disabilities. However, worldwide, about 7,000 sign languages are used for communication, and American Sign Language (ASL) is widely used for research due to its acceptance [1]. These languages are commonly related to gesture, which is motion in any body part, such as the hand or face, and are easily recognized through advanced techniques [1]. Hand gestures, made with either one hand or both, can be used to communicate in sign language [5]. It is a perfectly structured language for these individuals [7].

In the literature, many studies address the issue of recognizing hand gestures, which became a major issue in using sign language among these people to communicate in society. Therefore, we propose to develop an efficient deep learning-

based model to improve and meet the goal of sustainable development. We are creating prediction models that analyze the ASL recognition datasets with intricately correlated features. Here, a fine-based technique is used for classification. We simulated an assistive model for these individuals, which can be efficiently and effectively used in society. The gestures we used in this paper are given in Fig. 1 [5]. We are attempting to communicate words, characters, and phrases in ASL with an understanding of symbols.

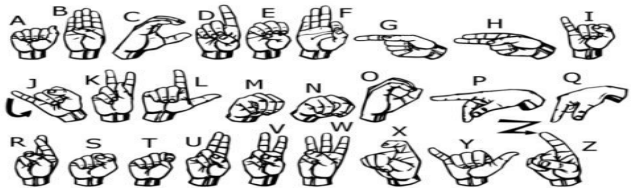

Fig 1- Gestures used for ASL recognition [5]

This paper uses an adaptive deep learning-based method to predict the ASL recognition problem, demonstrating the potential of Artificial Intelligence (AI) in real-time problem-solving. AI's machine learning and deep learning subsets mimic human intelligence for complex tasks and decision-making, respectively.

The organization of this paper is given below-. Introduction is given in Section I, the review of literature is elaborated in Section II, and the problem along with the proposed model are discussed in Section III. The experimental evaluation and findings are discussed in Section IV. Lastly, Section V briefly summarizes the paper in conclusion with future scope.

II. REVIEW OF LITERATURE

Many research and experiments have been carried out on gesture recognition of hands in the past decade [11], [12], [13]. The steps for identifying hand gestures are acquiring data, Processing data, extracting features, and gesture classification. Electromechanical gadgets are considered to deliver exact hand function (position) and configuration. Various glove-based strategies may be used to abstract records. But it is not user-friendly and steeply-priced [11]. Vision-based approaches know regular contact between computers and humans, and they do not use any additional gadgets because they just need a digital camera [14]. Computer cameras are input devices in vision-based tactics to examine the statistics of hands and fingers. Various machine learning (ML) techniques have been considered for selecting suitable models for detection accuracy. "Hidden Markov Models (HMMs)" [15] are to be used to classify gestures. It focuses on the dynamic aspect of gestures. Tracing skin-colored patches corresponding to hands in both dy and facial space centered on the face of used is used to excerpt gestures from video picture sequences [16]. A Flemish sign language recognition system with a 2.5 percent mistake rate is created using CNN [17], [18]. Construct a recognition model and obtain a defect rate of 10.90% by applying an HMM classifier and a dictionary of 30 words. They accomplished an average accuracy of 86% for 41 static motions in Japanese sign language. Signers who had already been seen obtained an accuracy of 99.99% utilizing depth sensors and 83.61% and 85.39% for original signers.

III. THE PROBLEM AND METHODOLOGY

The main aim of this paper is to simulate a model compared with the contemporary models and predict symbols with the highest accuracy and in the minimum time.

The hand detection method combines color detection by threshold with background removal. Because the face and the hand have the same skin color, we may use an "Adaboost face finder" to tell them apart [19]. We could also apply the Gaussian Blur filter to abstract the images needed for training. We can easily apply filters using Open Computer Vision, also called Open CV, as described in [20].

Moreover, artificial intelligence (AI) mimics human intelligence in performing complex tasks and making decisions. Furthermore, machine learning is a subset of AI used to solve prediction problems and learn patterns from data [10]. However, DL is another evolution of ML, which employs artificial neural networks (ANN) for solving complex real time problems. DL is a computer vision technique that can be recognized and classified in real-time. Therefore, in this paper, the prediction of the ASL recognition problem is solved using an adaptive deep learning (DL) based method. The proposed methodology is given on the flowchart shown below in Fig. 1 [1]:

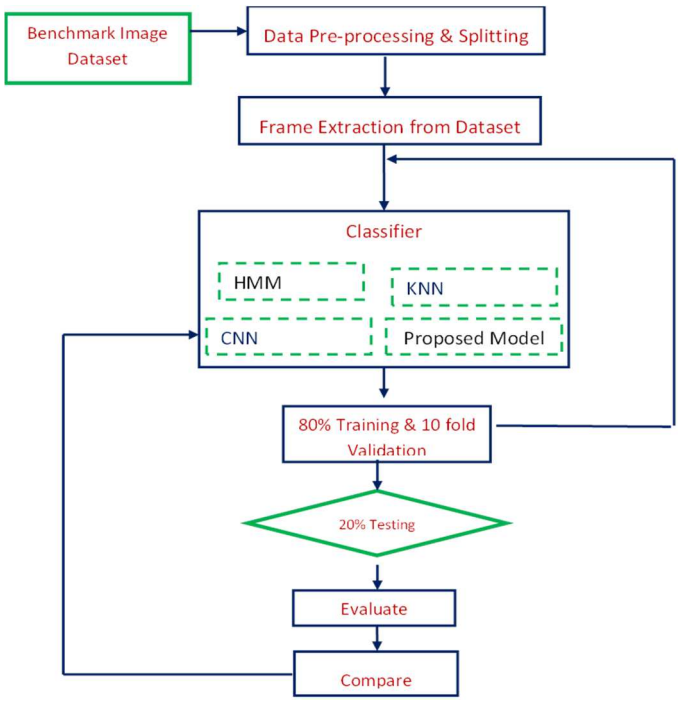

Fig.2- Flowchart of the Proposed Model

All of the convolutional layers employ the ReLU nonlinear activation function [21]:

$$ReLU(x) = max(0, x) \text{ or } ReLU(x) = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x \geq 0 \end{cases} \quad (1)$$

However, in the last layer, the *Soft-max* activation function is used [22]. **T**he values of image pixels are employed along with the DL-based model to extract significant updates. An ANN is a collection of neurons connected in a way that resembles how the human brain is organized. Information is sent from one neuron to another through every connection. Before transmitting the inputs to the buried layer of neurons, they are received and processed through the first layer. The data is transmitted to the final output layer after being processed through multiple hidden levels.

The ASL-to-text system has been built on the concept of computer vision. Every alphabet sign or gesture is formed using their hands, eliminating the requirement for any artificial device or equipment for engagement. We searched for any previously built datasets for the project, but none had the raw photos we needed. The only datasets that we could discover were in the form of RGB values. Consequently, the following steps are applied to construct the dataset: The Open Computer Vision (OpenCV) library was utilized in our dataset. The first step was to take roughly 200 images for testing and about 800 pictures of each ASL symbol for training. We start by taking a snapshot of each frame the webcam on our computer displays—a blue enclosing square designates each frame's region of interest (ROI). We extracted our RGB, Region of Interest (ROI), from this image and transformed it into a grayscale image, as depicted in the figure below. We employed a Gaussian blur filter to extract several elements from the final image. This is how the image appears following the use of Gaussian blur.

DL is used to train ANNs to learn from datasets. Data gathering and preprocessing, model selection, weight initialization, forward pass, loss computation, back-propagation, and optimizer are some of the steps involved in DL. After the data has been cleaned and normalized, the problem is considered when selecting the model architecture. Every neuron calculates both an activation function and a weighted sum. A function is used to compare the model's output to the accurate labels, and the chain rule is used to determine the gradients. To reduce loss, the optimizer modifies the weights. These actions are repeated over several epochs in the training loop until the model converges. The performance of the model is evaluated on different tests/validation of the benchmark dataset to verify generalization and check for over fitting. The model is then used to conclude fresh, untested data. Our proposed method has two-layer algorithm for predicting the last gesture of the user.

Algorithm Layer "1":

1. Processed images are used by extracting the feature and applying filters and thresholds to Open CV-captured frames.
2. When generating a term, a character is printed and considered if found in more than 60 frames of this processed image, given to the CNN [17],[18] prediction model.
3. Spaces in between words are to be considered by using the blank symbol.

Algorithm Layer "2":

1. A variety of symbol sets are detected which provide similar consequences after detection.
2. Then, we apply a specific classifier to sets to classify between them.

We print and add the character to the current line once the number of detected characters exceeds a particular value and no additional characters are near the threshold (We recorded the value as 50 and the difference threshold as 20 in our code.). Else, clean the present vocabulary with several occurrences of the current character to dodge the possibility of predicting an invalid character. Whenever the total of detected blanks (normal background) surpasses a definite value, no blanks are sensed if the present buffer is empty. Otherwise, it will print one space to predict the finish of a word, and the current word will be added to the sentence below. The user can add a phrase to the current sentence by selecting a list of terms matching the current word. It decreases spelling errors and aids in the prediction of difficult words.

IV. RESULTS AND DISCUSSION

In this paper, we have simulated the proposed method to test the benchmark dataset. In the results, Fig.3 compares accuracies within the benchmark dataset. Ali et al. in 2019 [23] showed an accuracy of 85.2, whereas Tao et al. in 2020 [24] achieved an accuracy of 83.4, respectively. However, Lee et al. in 2021 [25] had an accuracy of 91.2, whereas the proposed model achieved the highest accuracy among all studies with 95.8.

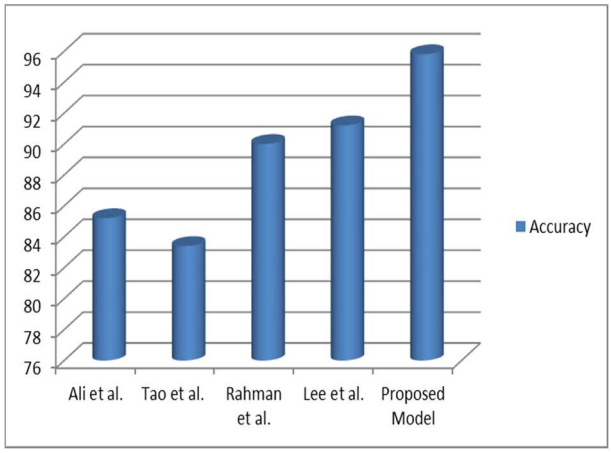

Fig.3- The comparative study of different models with the Proposed Model

In the DL-based model, measuring the model's effectiveness is complex, which requires a confusion matrix (cM) to analyze a model's performance, providing a clear and concise method for evaluation. A cM is a table that compares the true and observed labels of the model, which include True +ves (tP), True -ves (tN), False +ves (fP), and False -ves (fN). The Accuracy is the overall proportion of (tP + tN) divided by (tP + tN + fP + fN). The number of samples was divided by the

number of correct predictions, and the accuracy can be calculated. The accuracy is given by equation1 [21].

$$Acc=[(tP +tN)/(tP + tN + fP + fN)] \quad (2)$$

Using only layer 1 of our algorithm, we received a precision of 91.3 percent in the model, and when we combined both layers' 1' and '2', we could accomplish an accuracy of 95.8 percent. This accuracy surpasses most American Sign Language research articles just released. Most research articles focus on using specified gadgets for possible hand detection. It has benefitted a significant portion of the community, which can benefit from this project for people with disabilities by giving them the means to use sign language that helps them interact with others. This will remove the intermediary, who mainly serves as a translator. The model's simplicity makes it suitable for implementation in mobile applications, which is what we intend to do in the future.

We ran across various issues along the process. A lack of data was the first issue we encountered. Because operating with only square pictures was much more practical in Keras, we intended to deal with raw photographs, primarily square images. We developed our dataset because we could not find an existing one. The second challenge was to decide which filter should be applied to our photographs to extract the required attributes to input into the DL based [17,18] model. After testing with various filters, including canny edge detection, binary threshold, and others, we finally opted for the Gaussian blur filter. Many other disputes have been faced related to the precision of the model we have taught in past phases and have evolved by subsequent increases in the size of the input image and dataset. The accuracy of the proposed model in the real-time environment was 95.8%, which is on par with the contemporary models.

V. CONCLUSION AND FUTURE SCOPE

In this paper, an efficient and CNN-based model for hearing-impaired individuals was developed. We have successfully increased the accuracy of our prediction after building an algorithm with two layers by predicting and confirming symbols. This technique may also be used to translate between text and sign language. A gesture recognition system was released for text conversion. It records the indications and displays them in textual form on the screen. If symbols are adequately presented, noise is avoided, and the lighting is proper, gestures can nearly always be recognized in this way. The entire concept of this study was intended to be applied to smartphones as well. Implementing image processing technologies is challenging when implementing this concept into a mobile phone. The accuracy of the proposed model is better than the contemporary model. This model can be deployed in the natural environment as assistive technology for hearing-impaired.

Reference

1. Devashsih Sethia, P. Singh, and B. Mohapatra, "Gesture Recognition for American Sign Language Using Pytorch and Convolutional Neural Network," Lecture notes in electrical engineering, pp. 307–317, Jan. 2023, doi: https://doi.org/10.1007/978-981-19-6581-4_24.
2. L. Roda-Sanchez, C. Garrido-Hidalgo, A. S. García, T. Olivares, and A. Fernández-Caballero, "Comparison of RGB-D and IMU-based gesture recognition for human-robot interaction in remanufacturing," The International Journal of Advanced Manufacturing Technology, Oct. 2021, doi: https://doi.org/10.1007/s00170-021-08125-9.
3. N. Tran, P. DeVries, M. Seita, R. Kushalnagar, A. Glasser, and C. Vogler, "Assessment of Sign Language-Based versus Touch-Based Input for Deaf Users Interacting with Intelligent Personal Assistants," arXiv (Cornell University), Apr. 2024, doi: https://doi.org/10.1145/3613904.3642094.
4. Kenza Khellas and Rachid Seghir, "Alabib-65: A Realistic Dataset for Algerian Sign Language Recognition," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 22, no. 6, pp. 1–23, Jun. 2023, doi: https://doi.org/10.1145/3596909.
5. E. J. Robert and H. J. Duraisamy, "A review on computational methods based automated sign language recognition system for hearing and speech impaired community," Concurrency and Computation: Practice and Experience, Mar. 2023, doi: https://doi.org/10.1002/cpe.7653.
6. 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, Jan. 2023, doi: https://doi.org/10.1016/j.procs.2022.12.106.
7. S. Sripriya, S. Gnanasambantham, J. Gowtham and N. Logesh, "BeyondWords: A Sign Language Translator," *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India, 2024, pp. 1-5, doi: 10.1109/ACCAI61061.2024.10602433.
8. A. Núñez-Marcos, O. Perez-de-Viñaspre, and G. Labaka, "A survey on Sign Language machine translation," Expert Systems with Applications, vol. 213, p. 118993, Mar. 2023, doi: https://doi.org/10.1016/j.eswa.2022.118993.
9. D. M. Elbourhamy and H. M. Mohammdi, "An intelligent system to help deaf students learn Arabic Sign Language," Interactive Learning Environments, pp. 1–16, Apr. 2021, doi: https://doi.org/10.1080/10494820.2021.1920431.
10. S. Joksimovic, D. Ifenthaler, R. Marrone, M. De Laat, and G. Siemens, "Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review," Computers and Education: Artificial Intelligence, vol. 4, p. 100138, 2023.
11. J. Qi, L. Ma, Z. Cui, and Y. Yu, "Computer vision-based hand gesture recognition for human-robot interaction: a review," Complex & Intelligent Systems, Jul. 2023,.

12. Rayane Tchantchane, H. Zhou, Z. Shen, and Gursel Alici, "A Review of Hand Gesture Recognition Systems Based on Noninvasive Wearable Sensors," Advanced intelligent systems, Jul. 2023, doi: https://doi.org/10.1002/aisy.202300207.
13. N. Majdoub Bhiri, S. Ameur, I. Alouani, M. A. Mahjoub, and A. Ben Khalifa, "Hand gesture recognition with focus on leap motion: An overview, real world challenges and future directions," Expert Systems with Applications, vol. 226, p. 120125, Sep. 2023, doi: https://doi.org/10.1016/j.eswa.2023.120125.
14. V. C. Wable, M. Swarna, V. S. Prabhu, N. V. Krishnamoorthy and M. Dinesh, "Experimental Evaluation of Smart Camera based Reading Assistance for Visually Impaired People using Optical Character Recognition Logic," 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2024, pp. 1-6, doi: 10.1109/ACCAI61061.2024.10602441.
15. J. Xi et al., "A Blockchain Dynamic Sharding Scheme Based on Hidden Markov Model in Collaborative IoT," in IEEE Internet of Things Journal, vol. 10, no. 16, pp. 14896-14907, 15 Aug.15, 2023, doi: 10.1109/JIOT.2023.3294234.
16. Nahla Majdoub Bhiri, Safa Ameur, Imen Jegham, Ihsen Alouani, and Anouar Ben Khalifa, "2MLMD: Multi-modal Leap Motion Dataset for Home Automation Hand Gesture Recognition Systems," Arabian Journal for Science and Engineering, Aug. 2024, doi: https://doi.org/10.1007/s13369-024-09396-6.
17. P. Moon et al., "An improved custom convolutional neural network based hand sign recognition using machine learning algorithm," Engineering Reports, Mar. 2024, doi: https://doi.org/10.1002/eng2.12878.
18. P. Moon et al., "An improved custom convolutional neural network based hand sign recognition using machine learning algorithm," Engineering Reports, Mar. 2024, doi: https://doi.org/10.1002/eng2.12878.
19. W. Xu and D. Wang, "A real-time face detection based on skin detection and geometry features," Journal of Optics, Jun. 2024, doi: https://doi.org/10.1007/s12596-024-01949-0.
20. K. R. Kavitha and C. Janani Rakshandha, "Colour correction using open computer vision & python," AIP conference proceedings, Jan. 2024, doi: https://doi.org/10.1063/5.0197667.
21. L. Song, J. Fan, D.-R. Chen, and D.-X. Zhou, "Approximation of Nonlinear Functionals Using Deep ReLU Networks," Journal of Fourier Analysis and Applications, vol. 29, no. 4, Jul. 2023, doi: https://doi.org/10.1007/s00041-023-10027-1.
22. Y. Singh, M. Saini and Savita, "Impact and Performance Analysis of Various Activation Functions for Classification Problems," 2023 IEEE International Conference on Contemporary Computing and Communications (InC4), Bangalore, India, 2023, pp. 1-7, doi: 10.1109/InC457730.2023.10263129.
23. Aly W, Aly S, Almotairi S. User-independent American sign language alphabet recognition based on depth image and PCANet features. IEEE Access. 2019 Sep 2;7:123138-50.
24. Tao W, Leu MC, Yin Z. American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion. Engineering Applications of Artificial Intelligence. 2018 Nov 1;76:202-13.
25. Rahman MM, Islam MS, Rahman MH, Sassi R, Rivolta MW, Aktaruzzaman M. A new benchmark on american sign language recognition using convolutional neural network. In2019 International Conference on Sustainable Technologies for Industry 4.0 (STI) 2019 Dec 24 (pp. 1-6). IEEE.